SpatialAgent: An autonomous AI agent for spatial biology

Hanchen Wang[1,2,*,#], Yichun He[3,*,+], Paula P. Coelho[1,*], Matthew Bucci[1,*], Abbas Nazir[1], Bob Chen[1], Linh Trinh[1], Serena Zhang[2], Kexin Huang[2], Vineethkrishna Chandrasekar[1], Douglas C. Chung[1], Minsheng Hao[4,+], Ana Carolina Leote[1], Yongju Lee[1], Bo Li[1], Tianyu Liu[5,+], Jin Liu[1], Romain Lopez[1], Tawaun Lucas[1], Mingyu Ma[6,+], Nikita Makarov[7,8,9], Lisa McGinnis[1], Linna Peng[1], Stephen Ra[1], Gabriele Scalia[1], Avtar Singh[1], Liming Tao[1], Masatoshi Uehara[1], Chenyu Wang[10,+], Runmin Wei[1], Ryan Copping[1], Orit Rozenblatt-Rosen[1], Jure Leskovec[2] and Aviv Regev[1,#]

[1]Genentech, [2]Stanford, [3]Harvard, [4]Tsinghua, [5]Yale, [6]UCLA, [7]Roche, [8]LMU, [9]Helmholtz Munich, [10]MIT

Advances in AI are transforming scientific discovery, yet spatial biology, a field that deciphers the molecular organization within tissues, remains constrained by labor-intensive workflows. Here, we present SpatialAgent, a fully autonomous AI agent dedicated for spatial-biology research. SpatialAgent integrates large language models with dynamic tool execution and adaptive reasoning. SpatialAgent spans the entire research pipeline, from experimental design to multimodal data analysis and hypothesis generation. Tested on multiple datasets comprising two million cells from human brain, heart, and a mouse colon colitis model, SpatialAgent's performance surpassed the best computational methods, matched or outperformed human scientists across key tasks, and scaled across tissues and species. By combining autonomy with human collaboration, SpatialAgent establishes a new paradigm for AI-driven discovery in spatial biology.

1. Introduction

Modern AI is becoming a key tool for scientific discovery (1). Recently, Large language models (LLMs) (2) have shown promises in accelerating this progress by enabling reasoning, planning, and tool integration, through autonomous LLM-equipped agents (3). Autonomous agents are AI systems that iteratively perceive, plan, and act, allowing them to dynamically adapt to tasks with minimal human intervention. This makes them particularly valuable in more open-ended scientific research, where they integrate pre-existing knowledge (via LLMs and retrieval-augmented generation) with analytical actions (in connected tools), and iterative reasoning. In this they are distinct from both prior automated pipelines, which had to be hard-coded and hence could not adapt to the dynamic needs of new data analysis, and human-driven analysis, where a seasoned data scientist must iteratively engage with analysis tools, results, and knowledge, including through prompting of LLMs. Indeed, very recent studies have demonstrated the effectiveness of autonomous agentic systems in scientific workflows. Multi-agent reasoning has been used to iteratively refine hypotheses in biomedical research (4), while domain-specific toolsets have enabled agents to automate complex tasks, such as molecular dynamics simulations (5) and therapeutic reasoning (6). These advances underscore the transformative potential of AI-driven systems in accelerating scientific discovery across disciplines.

Spatial genomics, a rapidly growing field, offers a particularly compelling case in point. Spatial genomics aims to deciphers how biomolecules and cells are spatially organized within tissues to in homeostasis and disease (7). As an emerging area, driven by novel lab technologies that generate complex high dimensional data, spatial genomics relies on computational methods that are quite fragmented, require extensive human intervention and do not yet generalize well across diverse datasets and biological contexts. Moreover, users have to reason over multiple analysis approaches and complex concepts spanning molecules, cells and multicellular interactions, that require deep biological knowledge across many cell types, gene programs and interaction mechanisms to inter-

* Equal contribution, +Work done during internship at Genentech, # Correspondence: {wang.hanchen, regev.aviv}@gene.com

pret. Thus, in practice, most studies demand substantial commitment from expert analysts and time-consuming work. We reasoned that LLM-driven autonomous agents could help drive innovative work in spatial genomics, both accelerating progress and helping make new discoveries.

Here, we introduce SpatialAgent, an LLM agent that autonomously navigates the entire spatial-biology gamut, from experimental design to multimodal analysis and data-driven hypothesis generation. Unlike established computational approaches in spatial genomics, which rely on predefined workflows and rigid models, SpatialAgent employs adaptive reasoning and dynamic tool integration, allowing it to adjust to new datasets, tissue types, and biological questions. It processes multimodal inputs, incorporates external databases, and supports human-in-the-loop interactions, enabling both fully automated and collaborative discovery. We evaluate SpatialAgent on multiple datasets spanning human brain, heart, and mouse colon, with tasks such as gene panel design, cell and tissue annotation, and pattern inference in cell-cell communication and pathway analysis. SpatialAgent outperforms existing computational baselines, matches or surpasses human experts in key biological tasks, and accelerates scientific workflows. Beyond efficiency, SpatialAgent enhances the interpretability of its AI-driven discoveries, bridging the gap between automation and human intuition and legacy knowledge. SpatialAgent sets a new standard for autonomous and collaborative research in spatial biology, with broader implications for biomedical research and precision medicine.

2. Results

Overview of SpatialAgent We designed SpatialAgent, an LLM-based agent specialized for fully autonomous spatial-biology workflows, with the flexibility to optionally incorporate human interactions (**Fig.**1a, **Methods**). It operates through a self-governing loop that integrates LLMs with external tools and databases. SpatialAgent is equipped with both predefined planning templates (akin to those analysts train to use as informal "playbooks") and a curated set of domain-specific tools, which are critical for enhancing LLM performance in scientific problem-solving (8). It also adapts dynamically to human feedback, newly introduced tools, and previously unseen tasks (Supplementary Fig.8).

SpatialAgent consists of three key modules (**Fig.**1b): **memory**, **planning**, and **action**. The **memory** module maintains both semantic memory of high-level objectives and available tools (*e.g.*, *"I want to design a panel of 500 genes for mouse models with prostate cancers, treated with T-cell infiltration."*) and episodic memory of short-term steps and evolving context, ensuring continuity in task execution.

When a user queries, the **planning** module employs chain-of-thought reasoning and self-reflection prompts (Supplementary Information 1.5, Supplementary Fig.7), optionally leveraging predefined templates (Supplementary Information 1.4, Supplementary Fig.1, 2, 3) or generating plans *de novo* (Supplementary Information 1.7), to break down complex tasks into manageable steps. As execution progresses, the memory module iteratively refines stepwise plans, dynamically adapting to new information, such as user modifications or tool execution failures.

The **action** module executes tasks by selecting appropriate tools, such as retrieving relevant scRNA-seq datasets, converting gene names between species, or verifying ligand–receptor interactions via external databases. It also processes data using established libraries (*e.g.*, scikit-learn (9), PyTorch (10), Scanpy (11)) or generates and executes custom code when needed, integrating information from multiple sources. The agent's modular toolbox is easily extensible, enabling customization to suit user needs (Supplementary Information 1.3, 6).

SpatialAgent operates in either **autonomous** or **co-pilot** mode. The autonomous mode executes complex tasks end-to-end, leveraging pre-defined plan templates and generalizing to unseen requests (Supplementary Information 1.7), requiring no human intervention beyond the initial query. The co-

pilot mode enables interactive engagement, allowing users to refine task definitions dynamically (*e.g.*, pose follow-up questions) as the process unfolds (Supplementary Information 1.8), facilitating more adaptable and user-guided task execution.

To assess SpatialAgent's performance we benchmarked it on three core use cases: gene panel design, cell and niche annotation, and cell-cell interaction analysis, spanning three plan templates (one per core use case), 19 specialized tools and five datasets. Specifically, we tested SpatialAgent on gene panel design for spatial profiling of human brain regions, comparing it to four established computational methods and 10 human experts; on multimodal cell and niche annotation of developing human heart tissue compared to seven human scientists; and on discovery of cell-cell interactions in the tissue remodeling in mice, for open ended discovery of key tissue patterns and associated hypotheses. In addition, we assessed SpatialAgent's ability to generalize to unseen user queries (Supplementary Information 1.7), and to engage in two ongoing 'real world' lab experiments.

SpatialAgent out-performed most tools and human experts in gene panel design To evaluate SpatialAgent's ability to autonomously design gene panels, we benchmarked its performance against computational baselines, human scientists, and a hybrid setting where SpatialAgent refined initial designs from human scientists (**Methods**). We used a Visium spatial transcriptomic dataset of 12 sections from the human dorsolateral prefrontal cortex (DLPFC) from three adult human donors (12), capturing spatial gene expression across the six-layered cortical architecture. When a user queries SpatialAgent to design gene panels for the DLPFC, the agent updates its memory with the objective and available tools (all tools were accessible throughout). It then formulates a plan (*i.e.* a sequence of tool calls and decisions), to fulfill the task. This planning is scaffolded by predefined templates, analogous to how human analysts learn from tutorials or recommended workflows: they offer a structured yet adaptable blueprint that guides decision-making while allowing for context-specific variation. These templates help the agent correctly invoke domain-specific tools and reduce failure modes common in self-generated code, such as hallucinations or improper tool usage (mostly self-generated code). In this case, since the gene panel design aligns with a known template, SpatialAgent follows it (**Fig.** 2a), while dynamically adjusting steps based on intermediate results.

Across panel sizes from 50 to 500 genes, SpatialAgent's panels consistently outperformed those from established computational pipelines in terms of cell-type prediction accuracy (**Fig.** 2b,c, 6.0-19.1% improvement) and spatial coordinate prediction (**Fig.** 2d,e), with R^2 gains of up to 47.1% for some methods, and performing on par with others (HVG, -1.4%). Additional results are (**Fig.** 2b-e, Supplementary Fig.17-20). Thus, SpatialAgent autonomously identifies gene sets that capture biological variance and spatial organization better than standard approaches.

We also compared SpatialAgent's performance to that of 10 human experts for the design of gene panels of three sizes (50, 100, 150 genes). SpatialAgent outperformed 90% of human designs in cell-type prediction and 95% in spatial Y-coordinate prediction. Importantly, despite variations in individual human performance and workflow (Supplementary Information 2.2), integrating SpatialAgent with human expertise in the hybrid setting consistently improved outcomes for each human (maximum improvement: 935.1% in predicting Y-coordinates). To support such integration, SpatialAgent was designed to accept human-drafted gene panels as input and initialize its internal plan accordingly. It then evaluated, revised, or extended the design using its reasoning loop and toolset, effectively integrating into diverse human workflows without requiring standardization. Specifically, enhancing human designs with SpatialAgent improved cell-type prediction accuracy in 80% of cases (**Fig.** 2f,g) and spatial coordinate prediction in 90% of cases (**Fig.** 2h,i). Interestingly, incorporating human-designed panels also enhanced the final design compared to the agent alone, with 55% of hybrid designs outperforming SpatialAgent's autonomous runs, highlighting the benefit of human-

machine collaboration (**Fig.** 2f–i).

Reasoning and efficiency of autonomous panel design We next analyzed SpatialAgent's reasoning process, spatial structure preservation, and computational / cost efficiency relative to baseline methods and human scientists.

As part of its output, SpatialAgent's provides detailed reasoning for each gene included in the panel, revealing that the Agent assigned gene importance scores by combining reference datasets, external marker databases, internal knowledge, and human scientist–designed templates (**Fig.** 3a). The agent had appropriately increased its confidence in gene selection by prioritizing genes that are supported by multiple corroborating sources. This integration across knowledge bases increases the likelihood the agent will overcome gaps in individual databases, reduce bias, and ensure robust and generalizable performance. Moreover SpatialAgent produces a transparent aggregation of information, providing interpretable, gene-level rationales in natural language, explicitly connecting each gene in the panel to relevant cell types and functional roles (**Fig.** 3a). This enhances biological interpretability and can expedite human expert validation.

Because the dorsolateral pre-frontal cortex is organized in distinct cortical layers that define functionally distinct cell types and circuits (12), we also evaluated the clustering structure and spatial coherence associated with the gene panels from SpatialAgent , standard pipelines and human experts. Even when reducing the transcriptome to 500 genes, SpatialAgent's panels preserve strong spatial organization and produce well-differentiated clusters (**Fig.** 3b). These are closely aligned with the underlying spatial structure by multiple clustering metrics (Silhouette, Davies-Bouldin, Calinski-Harabasz, and Local Label Agreement), outperforming all baseline approaches (**Fig.** 3c).

Moreover, the genes prioritized by SpatialAgent reflect known functional activities of the DLPFC (*e.g.*, working memory, attention, and executive control) and provide spatially informative signals. For example, the three genes with highest importance scores in the SpatialAgent's panels are key neuronal markers (GAD2 (GABAergic inhibitory neurons), SLC17A7 (glutamatergic excitatory signaling), and GRIN2B (an NMDA receptor subunit)) and display distinct spatial distributions (**Fig.** 3d,e), spanning cortical layers, but absent from adjacent white matter. In contrast, the top three selected genes from Spapros (13) (PCDH15, ATP1A2, NKAIN3) yielded weaker spatial enrichment.

SpatialAgent also provides significant improvements in both runtime and cost, completing its analysis in around 30 minutes, far faster than human-driven panel design and notably quicker than advanced computational methods, such as Spapros (**Fig.** 3f). SpatialAgent offered substantial cost and time savings relative to some (not all) computational pipelines, and compared to human expert work. Thus, SpatialAgent allows users to quickly and efficiently benefit from iteration, seamlessly integrating diverse knowledge bases, with the potential to be combined with expert knowledge in order to design robust panels for spatial-omics experiments.

Enhanced cell type and tissue niche annotation Accurate and scalable annotation of spatially-localized cells and tissue niches remains a bottleneck in spatial biology, particularly when it requires considering multimodal information, such as expression profiles, anatomical annotations or histology. To evaluate SpatialAgent's capabilities, we use a high-resolution molecular and spatial cell atlas of the developing human heart (14), which integrates scRNA-seq and MERFISH spatial measurements across 6 heart samples collected at 9 to 16 post-conception weeks (142,946 single-cell profiles and >1.5 million MERFISH-detected cells). We assessed SpatialAgent's ability to reconstruct cell-type organization with both cellular and spatial fidelity. For cell-type annotation, we compared SpatialAgent to GPTCellType (15), CellTypist (16, 17), our human experts, and author-provided 'ground

truth' using shared UMAP coordinates and ontology terms (**Fig.** 4b,d, **Methods** , Supplementary Information 3.2, 3.3). For tissue niche annotation, we compared SpatialAgent to human experts and Author-provided annotations (**Fig.** 4c), as we are not aware of an existing standard pipelines for this task (Supplementary Information 3.2)).

Overall, SpatialAgent produced consistent, high-quality annotations that overall aligned closely with the author-provided ground truth, outperformed GPTCellType and CellTypist, and was largely on par with the best human experts, in terms of accuracy and precision at both coarse and fine scale (**Fig.** 4b,d). CellTypist struggles reflect its computational design (assigning per-cell annotations using a pre-trained neural network, leading to noisier predictions), training bias toward immune cells, and design for scRNA-seq, rather than MERFISH data with a more limited feature space. GPTCellType and human experts failed to identify immune and neuronal cells, and human experts also miss vascular smooth muscle cells. While SpatialAgent performed well, it did not fully recover epicardial cells, where the author's annotations distinguish between "epicardium-derived progenitor cells" and "epicardial" (as did GPTCellType, see Supplementary Fig.23-26), even though epicardial cells are relatively abundant in the dataset, and were present in both the queries and instructions. Indeed, GPTCellType and SpatialAgent failed to fully capture this tissue-specific feature, even when reference datasets were provided. LLM-based methods, including GPTCellType and SpatialAgent , tended to default to the most common annotations in their internal knowledge rather than incorporating dataset-specific biological context. In contrast, some human experts leveraged this context to provide more accurate epicardial cell annotations (Supplementary Information 3.5).

For tissue niche annotation, we compared SpatialAgent to human experts and Author-provided annotations (**Fig.** 4c), because we are not aware of existing standard pipelines for this task. As possibly the first method for automated tissue niche annotation, SpatialAgent effectively delineated anatomical regions by leveraging multimodal information from anatomical reference images to enhance spatial context. This allowed SpatialAgent to capture fine-grained spatial structures consistently across samples. Its performance benefits from key design choices, such as the removal of small clusters in UTAG (18) (Supplementary Information 3.4) and iterative refinement through sample-wise aggregation (below). SpatialAgent performed on par with or better than human experts in terms of accuracy and precision at coarse and fine scale (**Fig.** 4f), generalizing well across tissue regions. Notably, some human experts introduced creative annotations, such as labeling the outer layer as "Epicardium", which was absent in the original Author-provided annotations. Since this discrepancy lacks a direct reference, we categorize this region as "unmatched" (Supplementary Information 3.2), as some of these could be instances of enhanced performance of a new human annotator compared to the original ones. Overall, SpatialAgent matched the performance of the best human annotators and outperformed the rest (Supplementary Information 3.5). Moreover, SpatialAgent dramatically reduced both annotation costs and time (**Fig.** 4e,g). Thus, SpatialAgent maintains high annotation quality while offering a scalable, cost-effective alternative to manual expert annotation for large-scale spatial transcriptomic studies.

Behavior analysis in multimodal annotation To understand the basis of SpatialAgent 's improved annotation quality we analyzed its behavior and choices. For example, GPTCellType annotated Leiden Cluster 18 as cardiac fibroblasts, whereas SpatialAgent annotated them as neurons, consistent with the Authors' annotation (**Fig.** 5a). While GPTCellType took a gene centric approach that relied on expression of extracellular matrix-related genes (*e.g.*, POSTN, COL2A1, and COL9A2), which are commonly associated with fibroblasts, SpatialAgent combined transferred cell-type composition and recognizes a high neuronal cell proportion in the region (0.54), along with neuronal or neural development markers (NRXN1). SpatialAgent also recognized an overlap with cardiac-related genes with neuronal markers highlighting a potential caveat, which gene-centric methods may overlook.

Similarly, SpatialAgent enhanced tissue niche annotation through a collective intelligence approach, refining spatial assignments across samples by integrating multimodal reasoning (**Fig.** 4b). The agent's initial sample-wise annotations included several inconsistencies when transferring labels from the example anatomical image provided to the spatial transcriptomic data. For example, it mislabeled the 'Right Atrium' of one sample as 'Tricuspid Valve', and the 'Left Ventricle' of another sample as 'Pulmonary Vein'. (**Fig.** 4b, left). These initial discrepancies highlight the challenge of localized annotation variability, partially due to the limitations of visual reasoning in current LLMs (19). However, by aggregating spatial reasoning across all samples, SpatialAgent systematically resolved these ambiguities, leading to a coherent and correct anatomical interpretation of these regions in all samples (**Fig.** 4b, right). Overall, SpatialAgent integrated anatomical reference images, molecular profiles, and cell-type information, to maintain annotation consistency across samples, surpassing traditional per-sample annotation approaches. The use of a standardized naming scheme further ensures comparability across datasets.

Thus, SpatialAgent can effectively automate and standardize cell and tissue niche annotation, providing a biologically meaningful and spatially coherent interpretation of spatial transcriptomics data.

Mining data patterns, summarizing findings, and proposing hypotheses with SpatialAgent

We next evaluated SpatialAgent's capability to extract complex biological insights from spatial transcriptomics data. To avoid any knowledge contamination, we used a recent mouse colitis study (20), which was published after GPT4o's knowledge cutoff date (**Methods**). This dataset consists of MERFISH analysis of 940 genes in 52 tissue sections from 15 mice across four conditions: prior to DSS-induced inflammation, early in colitis, during peak inflammation and after a recovery period. The study includes author-annotated cells and spatial neighborhoods.

Upon receiving a user query (*e.g.*, "I want to infer cell-cell interactions in the tissue and how they change over time. <some additional descriptions on the data file>"), SpatialAgent autonomously executed a multi-step analysis pipeline: it first summarized condition-specific, cell-type compositions and spatiotemporal changes, and then computed cell–cell communication scores using LIANA+ (21), a framework integrating cellChat (22), cellPhoneDB (23), and additional gene set enrichment and factor analysis steps. It culminated by generating a detailed 7,000-word report that uniquely operates in the interaction mode, enabling iterative refinements through user-specified follow-up queries and new tasks via "memory hacking" (**Fig.** 6a, Supplementary Fig. 15).

Comparing SpatialAgent's interpretations with the original study, we observed broad agreement on key findings, particularly regarding inflammation-associated fibroblasts (IAFs) and significant immune infiltration (**Fig.**6b), which are hallmarks of IBD (24) and were particularly highlighted in the original paper (20). Notably, SpatialAgent uncovered additional insights, such as enhanced TGF-β signaling, submucosal remodeling, and fibroblast–pericyte interactions—elements that were not explicitly emphasized in the original study. These findings suggest a pivotal role for TGF-β-induced fibroblast activation in tissue repair and fibrosis (25). In particular, SpatialAgent's analysis proposes fibroblast polarization, mediated through TGF-β and IL-11 signaling, as a critical process orchestrating tissue regeneration (26). SpatialAgent proposes that the observed interplay between IAFs and pericytes appears to drive extracellular matrix remodeling, a mechanism that aligns with prior studies implicating IL-11 in fibrosis and tissue repair (27).

These insights could have significant implications. Identifying key signaling pathways, such as TGF-β and IL-11, offers potential therapeutic targets, and the distinct spatial transcriptomic signatures uncovered by SpatialAgent might be promising biomarker candidates for assessing disease severity and predicting treatment responses.

Adding customized genes to an existing panel for a specialized system Finally, we applied SpatialAgent to optimize experimental design in an ongoing wet lab study. Specifically, we tasked SpatialAgent with selecting 100 additional genes to enhance the Xenium 5k pan-tissue panel for a study on prostate cancer mouse models under different treatment conditions (**Fig.** 7a). This process involved a hybrid collaboration between SpatialAgent , human scientists, and suppliers to ensure the selection was data-driven, biologically informed, and manufacturability-compliant.

To systematically evaluate the impact of these customized gene panels, we first benchmarked their selection against a reference scRNA-seq dataset from a similar prostate cancer system (28). This ensured that the panel captured biologically relevant functions and pathways. The additional 100 SpatialAgent -selected genes improved the resolution of stromal, immune, and epithelial compartments, particularly in regions critical to prostate cancer progression and immune interactions (**Fig.** 7b,c, Supplementary Fig. 75-74).

Specifically, Xenium+SpatialAgent enhanced the definition of basal epithelial cells, fibroblasts, and prostate microvascular endothelial cells, leading to clearer population boundaries (**Fig.** 7b). While the standard Xenium panel performed slightly better in classifying classical monocytes and some T cell populations, Xenium+SpatialAgent provided superior resolution for stromal and epithelial compartments: key components in prostate cancer progression (Supplementary Information 5.2). Moreover, Xenium+SpatialAgent substantially improved clustering quality across multiple metrics (**Fig.** 7c). The selected genes spanned key overlapping processes (**Fig.** 7d) enriched in a reference dataset with notable T-cell infiltration. Analysis using the enhanced Xenium+SpatialAgent set also refined inferred cell–cell interactions within the tumor microenvironment, particularly communication between basal epithelial, fibroblast, and immune compartments (**Fig.** 7e). Interaction strengths, computed via CellPhoneDB, revealed distinct signaling differences, refining our understanding of immune–tumor crosstalk.

We further investigated how SpatialAgent's customized gene selections enhance the detection of cell-cell interactions within the tumor microenvironment, specifically examining communication between basal epithelial cells, fibroblasts, and immune cells (**Fig.** 7e). By calculating interaction strengths using CellPhoneDB, we revealed distinct signaling patterns that provide putative insights into immune-tumor crosstalk. The Xenium-only panel (Supplementary Fig. 76) primarily identified fibroblast-immune interactions involving modulatory ligands such as Jag1-Notch2 and Vcan-Tlr2. In contrast, incorporating SpatialAgent's customized design (Supplementary Fig. 77) highlighted the importance of laminin-integrin signaling networks previously undetected. This enhanced approach revealed cell type-specific communication patterns: fibroblasts engaged in numerous robust interactions across immune subtypes, while basal epithelial cells displayed fewer but functionally important integrin-based signals. Of particular interest, our integrated analysis uncovered distinct regulatory circuits in specific immune populations (*e.g.*, Igf1-Insr in Tregs) and emphasized the role of basal epithelial cells in shaping the tumor microenvironment through targeted signaling pathways. These refined interaction maps demonstrate the critical importance of optimized gene panel selection for accurately dissecting complex cellular communications and may inform future therapeutic strategies targeting these pathways.

3. Discussion

Computational analysis in spatial genomics has been a bottleneck and SpatialAgent introduces a substantial shift that will accelerate progress and empower scientists. By uniting LLM-based reasoning, dynamic tool execution, and multi-modal data analysis within a self-governing system, it merges the interpretability and adaptability typically associated with expert-driven strategies, while offering the

scalability and throughput gains of computational pipelines. In addition to substantially streamlining existing workflows, which were difficult to automate, it enables AI-driven hypothesis generation, guided experimentation, and faster data-to-discovery loops.

One of the most significant advantages of autonomous AI agents in this context is the potential to democratize sophisticated analyses. Current spatial biology pipelines often require specialized computational skills and substantial resources, creating barriers for labs with limited infrastructure or for bench scientists who may not be as well versed in coding or computational analysis. By automatically selecting tools, orchestrating analyses, and making interpretable decisions, SpatialAgent —and AI agents like it—can empower researchers to execute optimal workflows without extensive computational expertise, and combine them with their unique human abilities to formulate interesting questions, and engage with the agent. At the same time it liberates computational scientists from the need to perform relatively rote analyses, and allows them to focus on developing novel analytic strategies, new methods, and deeper analyses. Thus, this broader access has tangible benefits for the research community, from speeding up routine annotations to accelerating the identification of emergent biological patterns that might otherwise be overlooked.

An agent based approach also promises greater synergy between computational predictions and wet-lab validation. Traditionally, refining experimental designs is a lengthy, iterative process: initial hypotheses are tested in the lab, new data are generated, and computational analyses are performed offline. In contrast, AI agents can offer near-real-time feedback and suggest potential avenues for hypothesis refinement, enabling a more "closed loop" where computational insights more directly guide wet-lab experiments. By quickly highlighting the most promising markers or pathways, SpatialAgent can help researchers iterate more efficiently, saving time, resources, and, importantly, allowing investigators to focus on creative exploration rather than manual data wrangling.

Nevertheless, challenges remain. First, building deeper domain-specific knowledge into the system will be crucial, particularly for niche or emerging biological processes underrepresented in current training data. Second, unlike standard computational analyses, agent-driven approaches can have the same limitations as humans, including biases, hallucinations, and other errors. Automated uncertainty quantification would strengthen trust by distinguishing between high- and low-confidence predictions, allowing scientists to concentrate validation efforts where it is most needed. Furthermore, while multi-agent architectures hold promise for distributing specialized tasks among dedicated "expert" agents, their success depends on robust coordination protocols and reasoned consensus-building to avoid compounding errors or generating contradictory outputs (29).

Looking ahead, we envision AI autonomous agents evolving beyond passive analysts into active collaborators capable of context-aware reasoning and counterfactual experimentation. With refined planning algorithms, such agents could propose targeted experimental designs to test competing mechanistic hypotheses, serving as intelligent research partners who generate new questions rather than merely answering old ones. Coupled with advancements in large-scale single-cell multi-omics, clinical diagnostics, and real-time imaging, the next generation of AI-driven systems may redefine the boundaries of biomedical discovery. Ultimately, by reducing drudgery, enhancing interpretability, and unleashing creative potential, SpatialAgent illustrates how autonomous AI agents can save time, augment human ingenuity, and accelerate breakthroughs in spatial biology and beyond.

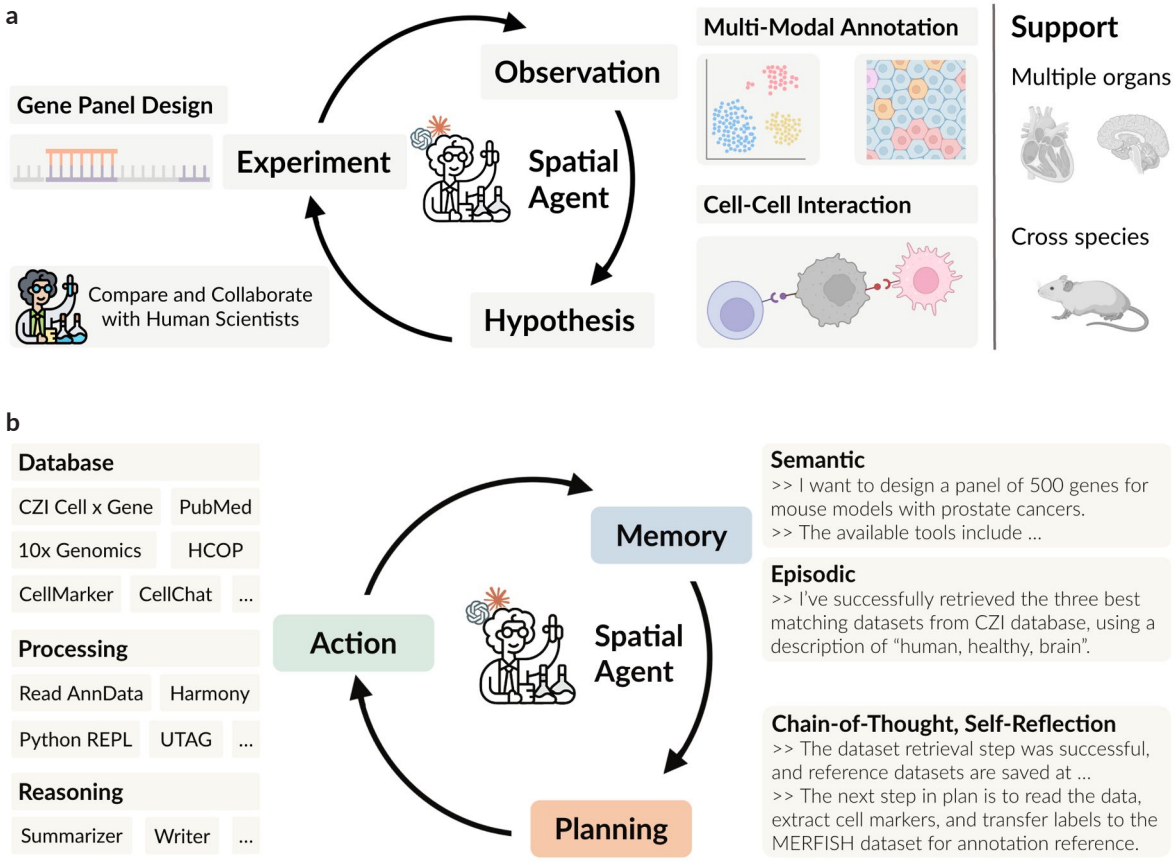


Figure 1: **Overview and modular design of SpatialAgent. (a)** Overview. SpatialAgent is an LLM-equipped autonomous agent that tackles a broad set of tasks in spatial biology, including (1) gene panel design for the experimental stage, (2) cell-type and tissue-niche annotation for the observational analysis stage, and (3) findings summarization and (4) generating hypotheses related to cell–cell interactions. SpatialAgent supports multimodal inputs and cross-species analyses. **(b)** Key modules. The action module (left) executes tasks such as retrieving reference datasets, converting gene names, verifying ligand–receptor interactions using existing databases, processing data with established software packages (*e.g.*, numpy) or generating and executing custom code, while reasoning over and aggregating information from multiple sources. The memory module (top right) maintains both semantic memory (high-level objectives) and episodic memory (short-term steps and context). The planning module (bottom) manages task planning via chain-of-thought reasoning and self-reflection, iteratively refining plans to achieve specific goals.

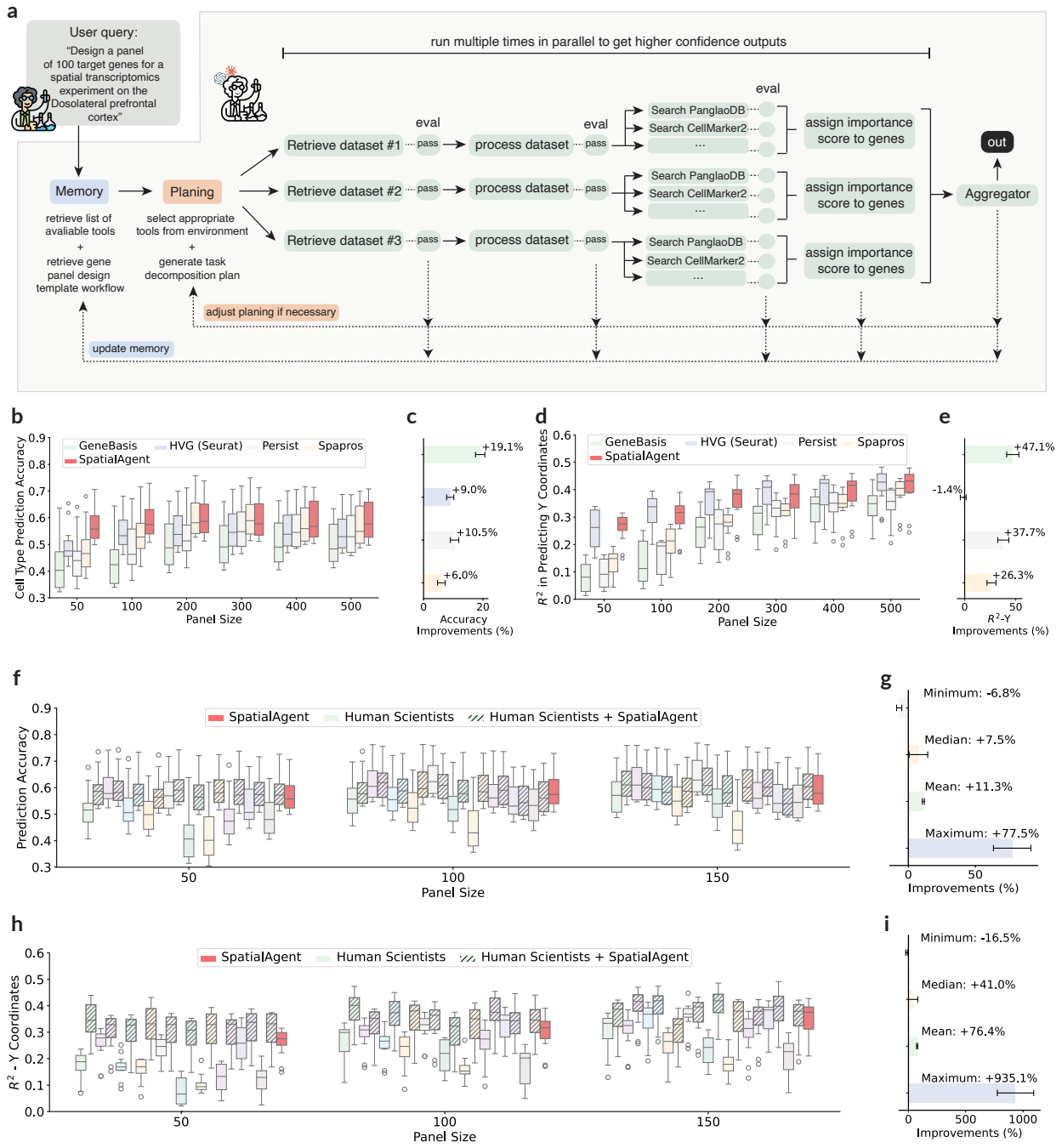


Figure 2: **Gene panel design by SpatialAgent. (a)** Illustrative step wise agent autonomous workflow. Schematic of the first few steps in the SpatialAgent workflow for designing a gene panel in the dorsolateral prefrontal cortex (DLPFC). **(b-f)** SpatialAgent outperforms established computational baselines for cell type and spatial coordinate prediction. **(b,c)** Cell-type prediction accuracy (**b**, y axis) and relative improvements over computational baselines (**c**, x axis) for designing 50-500 gene panels by SpatialAgent or each of several established methods. Box plots show medians (center lines), interquartile ranges (boxes), and 1.5× interquartile ranges (whiskers). Circles denote outliers. Results are averaged over 10 runs across all 12 DLPFC samples. **(d,e)** Spatial coordinate prediction performance (**d**, y axis) and relative improvements (**e**, x axis) by SpatialAgent or each of several established methods. Results are averaged over 10 runs across all 12 DLPFC samples. Box plots as in (b,c). **(f,g)** Cell-type prediction accuracy (**f**, y axis) and relative improvements (**g**, x axis) for SpatialAgent, human scientists, and hybrid approaches in which SpatialAgent incorporates human-designed templates. Solid bars represent individual scientists, and the striped bars to their right show performance gains when combined with SpatialAgent. (we report the minimum, median, mean, and maximum improvements) averaged across all three panel sizes.) **(h,i)** Spatial coordinate prediction performance (**h**, y axis) and relative improvements (**i**, x axis) for SpatialAgent, human scientists, and hybrid approaches. As in **(f, g)**, each bar represents a human scientist, with the paired bar indicating combined performance.

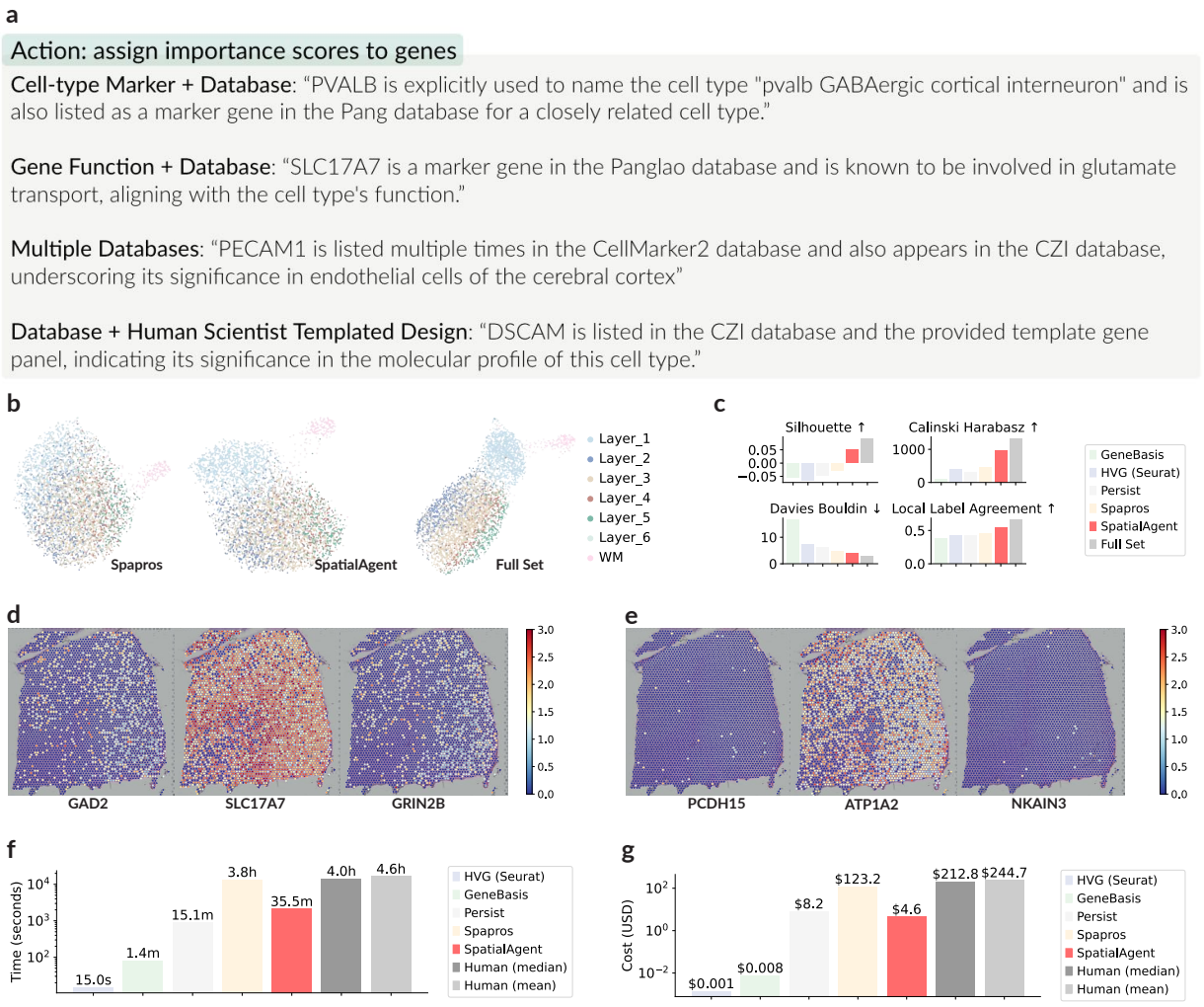
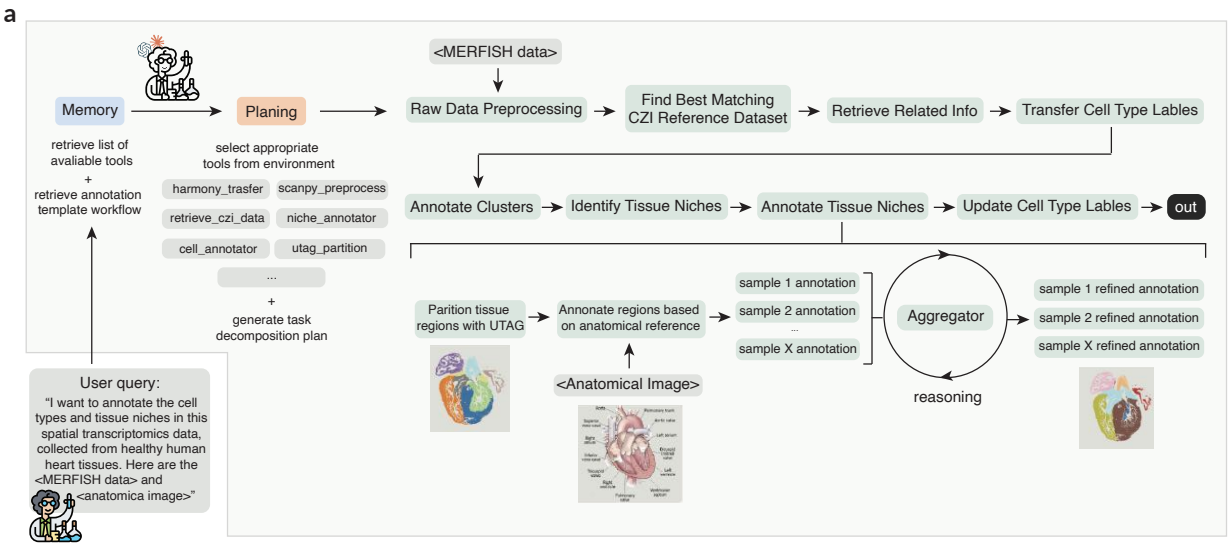
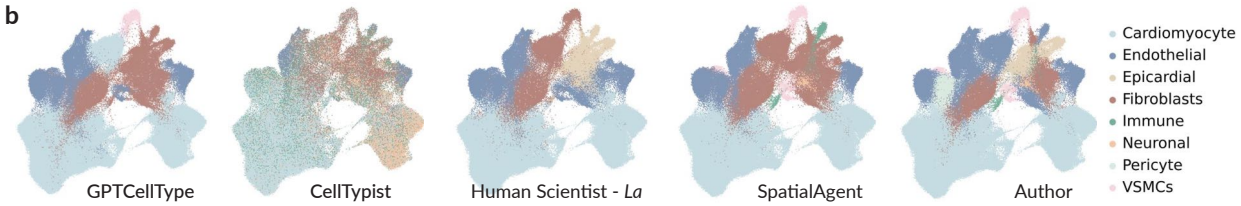
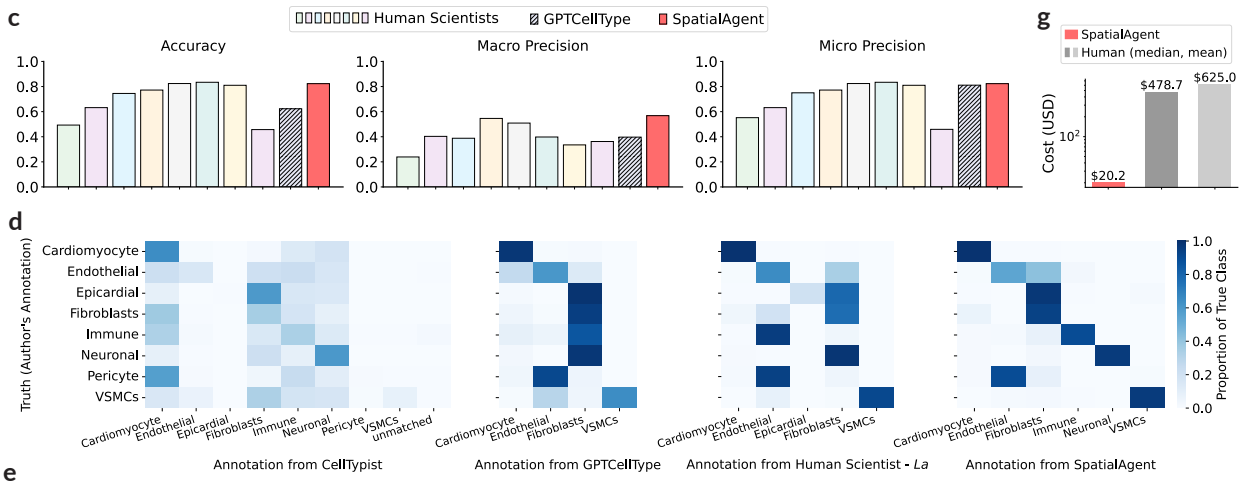
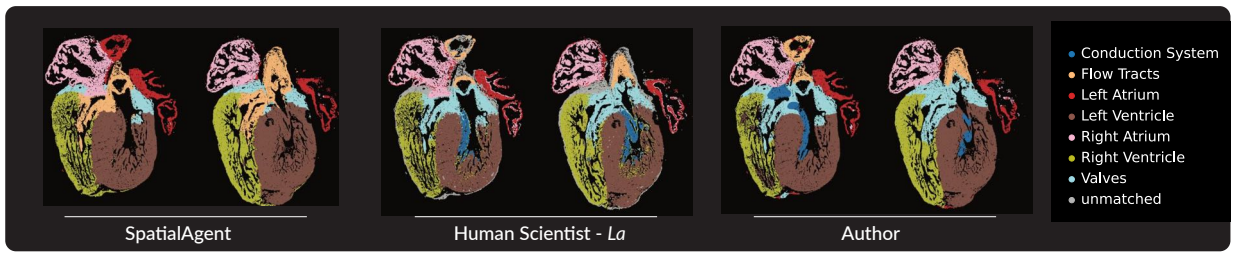


Figure 3: **Reasoning, performance, and efficiency of SpatialAgent in gene panel design.** **(a)** Agent reasoning in annotation. Illustrative examples of multiple sources of information - reference datasets, external marker databases, internal knowledge, and human scientist–designed templates - used by SpatialAgent to assign importance scores in the "Estimate Importance of Each Gene" step for gene panel design in the DLPFC dataset. **(b-e)** Performance on cell type separation by cortical layer. **(b)** UMAP embedding of spatial transcriptomic profiles (dots) consisting of 500 selected genes from Spapros, SpatialAgent , or the full gene set (21,000 genes) and colored by cortical layer annotation. **(c)** Clustering evaluation metrics (y axes, labeled on top) with 500-gene panels selected by each method (x axis) and with all 21,000 genes. Arrows: desired direction of alignment with spatial organization. **(d,e)** Spatial expression patterns of the top three genes selected by SpatialAgent (d) and Spapros (e). **(f,g)** Runtime and cost performance. Runtime (f, y axis, log scale) and cost estimates (g, y axis, log scale) for each method (x axis) across three runs. Human time is averaged over self-reported usage. Cost estimates are based on standard vendor rates (*e.g.*, Amazon Web Service, OpenAI) for computational baselines and SpatialAgent . Human labor costs are projected assuming an annual salary of $100,000 USD (assuming 8 work hours per weekday) (**Methods**).









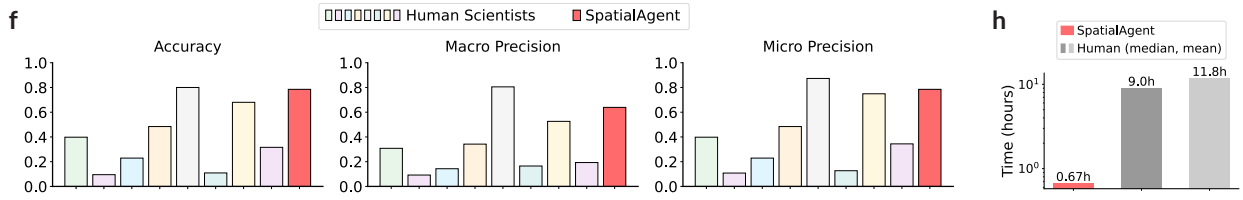


Figure 4: **Cell type and tissue niche annotation by SpatialAgent. (a)** Illustrative workflow. SpatialAgent integrates multimodal information (*i.e.*, anatomical images, MERFISH data) for tissue niche annotation, followed by sample-wise aggregation and refinement via collective intelligence. **(b–d)** Cell type annotations. **(b)** UMAP of cells colored by annotations from GPTCellType, CellTypist, a representative human scientist (*La*, second-best in accuracy), SpatialAgent, and the original study. Colors indicate eight major cell types, where 'VSMCs' denotes 'vascular smooth muscle cells'. Visualizations and scores are aggregated across all samples using the author annotations as ground truth (same rule applies below). **(c)** Annotation performance: accuracy, macro precision, and micro precision (y axis) across methods (x axis). **(d)** Confusion matrices comparing annotations from CellTypist, GPTCellType, human scientist *La*, and SpatialAgent to ground truth, sharing the same coloring scale of 0-1. **(e-f)** Tissue-niche annotations. **(e)** Annotated tissue niches by SpatialAgent, human scientist *La*, and the original author. Colors denote all seven major niches (legend); where 'unmatched' indicates regions with no correspondence to author annotations (**Methods**). **(f)** Accuracy, macro precision, and micro precision (y axis) across methods (x axis). **(g, h)** Cost and time. Estimated cost (g, y axis, USD, log scale) and time (h, y axis, hr) for SpatialAgent and human scientists (x axis), estimated as in **Fig.** 3.

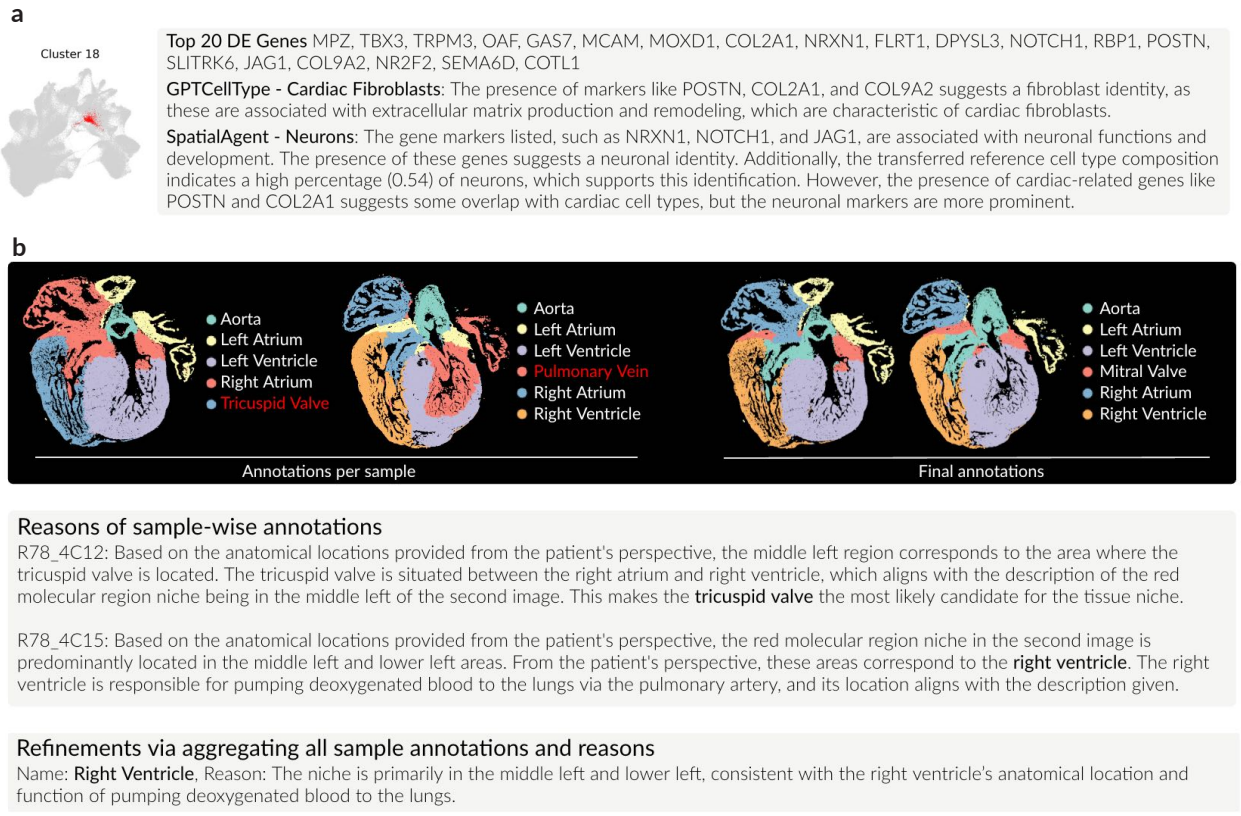


Figure 5: **Multimodal cell and niche annotation with SpatialAgent. (a)** Comparison of GPTCellType and SpatialAgent annotations. Reasoning underlying the annotation of Leiden Cluster 18 (red) from MERFISH data by GPTCelltype (middle) and SpatialAgent (bottom). GPTCellType assigns cell types based on the top 20 differentially expressed genes in the cluster (top), annotating the cluster's cells as cardiac fibroblasts, while SpatialAgent incorporates additional transferred cell-type composition information, suggesting a neuronal identity (with caveats) and matching the authors' annotation (Supplementary Fig.22). **(b)** Collective-intelligence design for tissue-niche annotation in SpatialAgent. Sample-wise annotations (left) are refined by aggregating reasoning across samples (right), improving final niche assignments. Incorrect annotations in the initial sample-wise outputs are marked in red (left). This process optionally integrates molecular and cell-type information to enhance the performance.

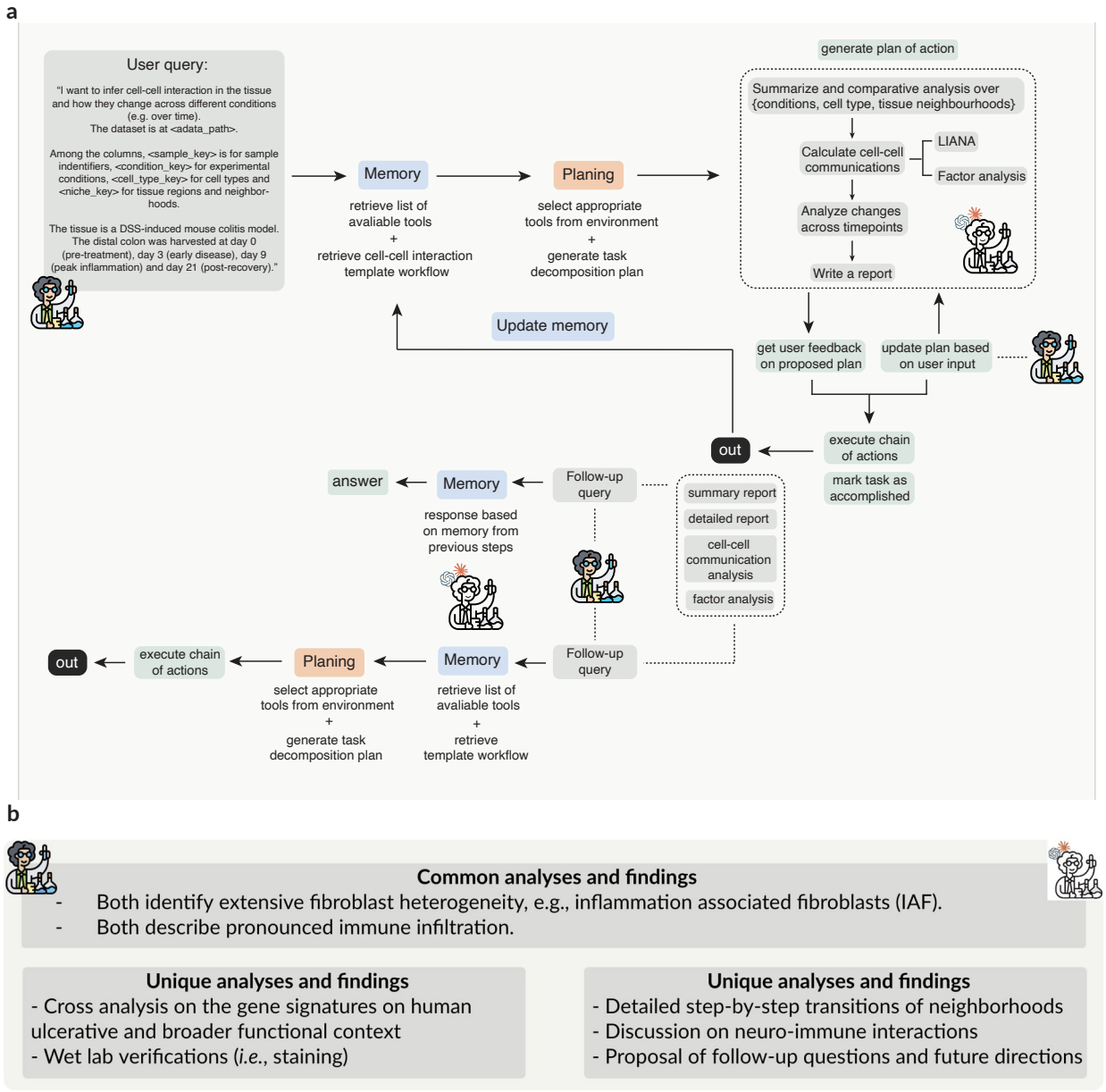
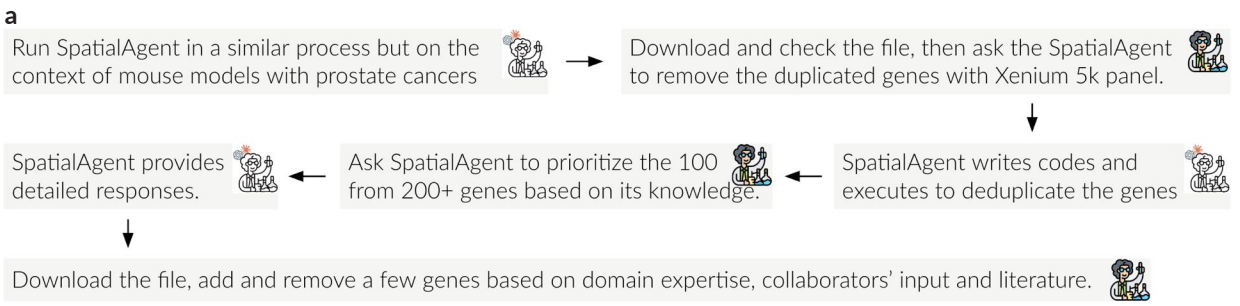
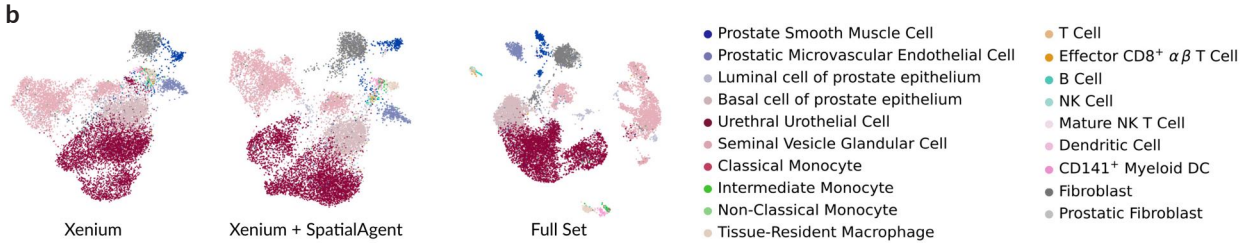
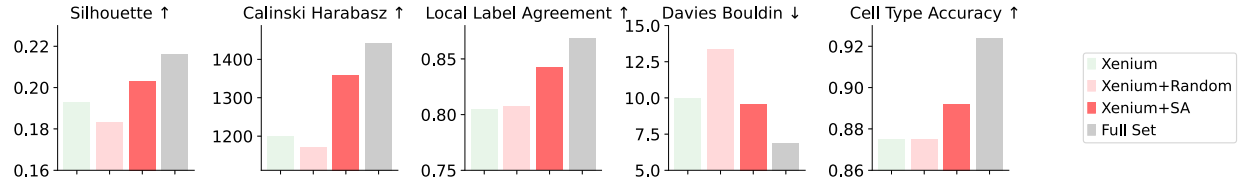
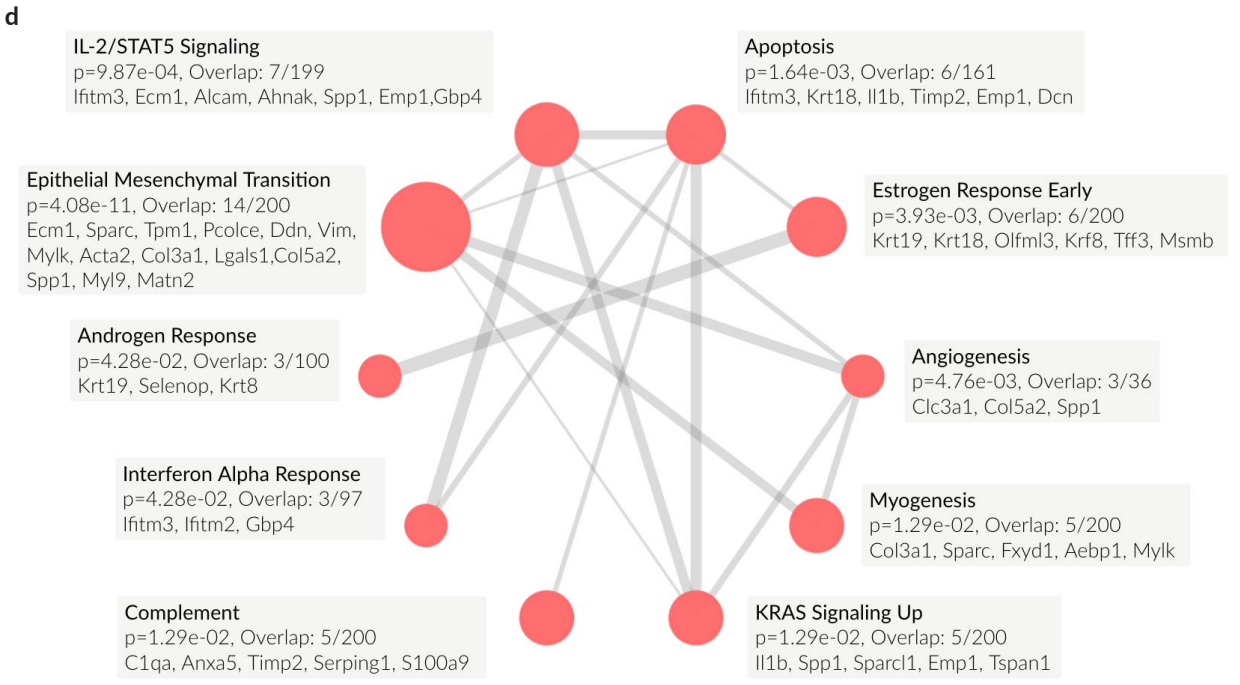


Figure 6: **Mining complex data patterns with SpatialAgent. (a)** Overview of the workflow with the interaction mode of SpatialAgent (**Methods** and Supplementary Information). After SpatialAgent autonomously executes a sequence of actions, the user can ask follow-up questions or initiate a new task. **(b)** Comparison of interpretations between the original authors and SpatialAgent. Shared (top) and distinct (bottom) findings between the authors (left) SpatialAgent (right) analyses.









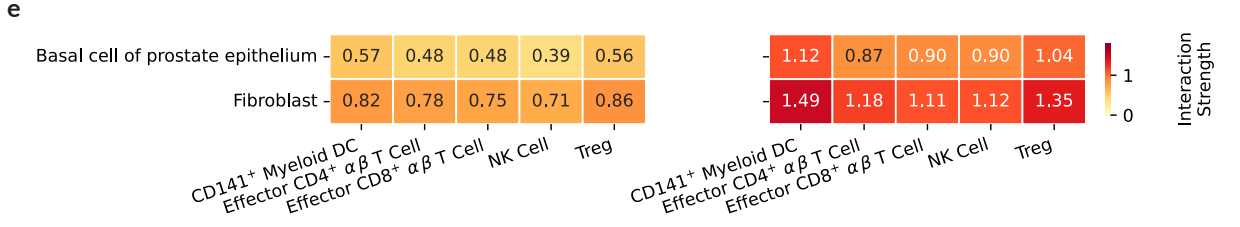


Figure 7: **Designing 100 customized genes alongside the Xenium 5k pan-tissue panel for mouse models of prostate cancer under various treatments.** **(a)** Overview of experimental workflow. **(b,c)** Improved cell type distinctions with Xenium+SpatialAgent panel. **(b)** UMAP embedding of cell profiles using different gene subsets, colored by cell-type annotations. Xenium: standard 5k pan-tissue panel; Xenium + Random: 5k panel combined with 100 randomly selected additional genes; Xenium + SpatialAgent: 5k panel with 100 genes selected by SpatialAgent; Full Set: complete profiles from the reference scRNA-seq dataset. **(c)** Clustering metrics and cell-type accuracy (y axis) for each method (x axis). Xenium + Random results are averaged over three independent random-gene selections. Arrows: direction of improved performance. **(d)** SpatialAgent-selected genes are enriched for key processes (**Methods**). Enrichment (node size, -log(P value)) and overlap (edge width, Jaccard index) of SpatialAgent selected genes with genes from different pathways (text boxes), and the specific overlapping gene names. **(e)** Cell–cell interaction scores are enhanced with Xenium+SpatialAgent panel. Strengths (from CellPhoneDB) of predicted cell-cell interactions from Basal and Fibroblast populations to immune cells, using Xenium (left) or Xenium + SpatialAgent genes (right). (**Methods**).

Methods Summary

SpatialAgent. SpatialAgent's core framework consists of three key components: a memory module that maintains both semantic and episodic information, a planning module that employs chain-of-thought reasoning for task decomposition, and an action module that executes specialized functions through a curated toolbox. The agent operates through a self-governing loop where it first encodes objectives and available tools in memory, formulates a concrete plan with sequential steps, executes these actions while recording outcomes, and iteratively refines its approach based on results and potential execution failures. This architecture enables adaptation to new datasets, tissue types, and biological questions without requiring predefined rigid workflows. Full implementation details and prompt instructions are provided in Supplementary Information.

Co-pilot mode. To support a co-pilot mode for SpatialAgent, memory hacking was employed, where the agent's memory is updated based on the user's new input each time. This allows SpatialAgent to adjust its plans and actions or provide useful content based on the user's queries. Further details are provided in Supplementary Information.

Computational Baselines. SpatialAgent was compared against established computational methods across each of three primary use cases. For gene panel design, it was benchmarked against (1) selection of highly variable genes (HVG) with Seurat (30), (2) GeneBasis (31), (3) Persist (32), and (4) Spapros (13). These range from simple variance-based methods to dedicated methods for spatial transcriptomics. For cell type annotation, SpatialAgent was compared with CellTypist (17) (a traditional approach), and GPTCellType (15) (an LLM based approach). For cell-cell interaction analysis, SpatialAgent was assessed for its ability to integrate tools from the LIANA framework and generate biologically meaningful insights. Detailed implementation parameters for all baseline methods are provided in Supplementary Information.

Human scientist performance. To benchmark SpatialAgent against human expertise, expert scientists were recruited with background in computational method development, experimental spatial transcriptomics, and biological data analysis. For the gene panel design task, 10 scientists were provided with the DLPFC dataset and asked to design gene panels of three different sizes (*i.e.*, 50, 100, and 150 genes). For the cell type and tissue niche annotation task, 7 scientists annotated the developing heart dataset. Participants were explicitly instructed not to use LLMs during their analysis but were otherwise free to employ any tools or resources typically used in their workflow. The approaches, time spent, and final results were documented through a standardized form. Full task descriptions and instructions provided to participants are detailed in Supplementary Information.

Evaluation metrics. For gene panel design, metrics were employed as previously described for Persist and Spapros: cell-type prediction accuracy, spatial coordinate prediction performance, and clustering qualities. These metrics assess both the biological relevance of selected genes and their ability to capture spatial organization within tissues. For cell and niche annotation, performance was evaluated by accuracy, macro precision, and micro precision against author-provided annotations as ground truth. Metrics that are difficult to define or distinguish between baselines (*e.g.*, expression constraints from Spapros) were avoided to ensure fair and interpretable comparisons.

Code availability. The implementation of SpatialAgent and tutorial notebooks for reproducing the results in this manuscript are available at `https://github.com/Genentech/SpatialAgent`.

Acknowledgments

We thank Leslie Gaffney, Yanay Rosen, Xiao Wang, Chuan Xu, Peng He, Russel Littman, Guohao Li, members of Hongyu Zhao's and Mengdi Wang's groups, as well as many colleagues at Genentech and Stanford University, for their invaluable insights. We are also grateful to members of the Leskovec and Regev labs, along with colleagues and friends from various places around the world, for their constructive and insightful discussions. J.L. was supported by NSF under Nos. OAC-1835598 (CINES), CCF-1918940 (Expeditions), DMS-2327709 (IHBEM); Stanford Data Applications Initiative, Wu Tsai Neurosciences Institute, Stanford Institute for Human-Centered AI, Chan Zuckerberg Initiative, Amazon, Genentech, GSK, Hitachi, SAP, and UCB.

References

[1] Hanchen Wang, Tianfan Fu, Yuanqi Du, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[2] Tom Brown et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, 2020.

[3] Shanghua Gao, Ada Fang, Yepeng Huang, et al. Empowering biomedical discovery with ai agents. *Cell*, 187(22):6125–6151, 2024.

[4] Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, et al. Towards an ai co-scientist. *arXiv:2502.18864*, 2025.

[5] Quintina Campbell, Sam Cox, Jorge Medina, et al. Mdcrow: Automating molecular dynamics workflows with large language models. *arXiv:2502.09565*, 2025.

[6] Gao Shanghao et al. Txagent: An ai agent for therapeutic reasoning across a universe of tools. *arXiv:2503.10970*, 2025.

[7] Giovanni Palla, David Fischer, Aviv Regev, et al. Spatial components of molecular tissue biology. *Nature Biotechnology*, 40(3):308–318, 2022.

[8] Botao Yu, Frazier N Baker, Ziru Chen, et al. Chemtoolagent: The impact of tools on language agents for chemistry problem solving. In *NAACL Findings*, 2025.

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[10] Adam Paszke, Sam Gross, Francisco Massa, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 2019.

[11] Alexander Wolf, Philipp Angerer, and Fabian Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19:1–5, 2018.

[12] Kristen R Maynard, Leonardo Collado-Torres, Lukas M Weber, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience*, 24(3):425–436, 2021.

[13] Louis Kuemmerle, Malte Luecken, Alexandra Firsova, et al. Probe set selection for targeted spatial transcriptomics. *Nature methods*, pages 1–11, 2024.

[14] Elie N Farah, Robert K Hu, Colin Kern, et al. Spatially organized cellular communities form the developing human heart. *Nature*, 627(8005):854–864, 2024.

[15] Wenpin Hou and Zhicheng Ji. Assessing gpt-4 for cell type annotation in single-cell rna-seq analysis. *Nature Methods*, pages 1–4, 2024.

[16] Chuan Xu, Martin Prete, Simone Webb, et al. Automatic cell-type harmonization and integration across human cell atlas datasets. *Cell*, 186(26):5876–5891, 2023.

[17] C Domínguez Conde, Chao Xu, Louie B Jarvis, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376(6594):eabl5197, 2022.

[18] Junbum Kim, Samir Rustam, Juan Miguel Mosquera, et al. Unsupervised discovery of tissue architecture in multiplexed imaging. *Nature methods*, 19(12):1653–1661, 2022.

[19] Zane Durante, Qiuyuan Huang, Naoki Wake, et al. Agent ai: Surveying the horizons of multi-

modal interaction. *arXiv:2401.03568*, 2024.

[20] Paolo Cadinu, Kisha N Sivanathan, Aditya Misra, et al. Charting the cellular biogeography in colitis reveals fibroblast trajectories and coordinated spatial remodeling. *Cell*, 187(8):2010–2028, 2024.

[21] Daniel Dimitrov, Philipp Sven Lars Schäfer, Elias Farr, et al. Liana+ provides an all-in-one framework for cell–cell communication inference. *Nature Cell Biology*, 26(9):1613–1622, 2024.

[22] Suoqin Jin, Christian F Guerrero-Juarez, Lihua Zhang, et al. Inference and analysis of cell-cell communication using cellchat. *Nature communications*, 12(1):1088, 2021.

[23] Mirjana Efremova, Miquel Vento-Tormo, Sarah A Teichmann, et al. Cellphonedb: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature protocols*, 15(4):1484–1506, 2020.

[24] Christopher S Smillie, Moshe Biton, Jose Ordovas-Montanes, et al. Intra-and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell*, 178(3):714–730, 2019.

[25] Benjamin Ng, Stuart A Cook, and Sebastian Schafer. Interleukin-11 signaling underlies fibrosis, parenchymal dysfunction, and chronic inflammation of the airway. *Experimental & Molecular Medicine*, 52(12):1871–1878, 2020.

[26] Srinivas Allanki, Boris Strilic, Lilly Scheinberger, et al. Interleukin-11 signaling promotes cellular reprogramming and limits fibrotic scarring during tissue regeneration. *Science advances*, 7(37):eabg6497, 2021.

[27] Takashi Nishina, Yutaka Deguchi, Daisuke Ohshima, et al. Interleukin-11-expressing fibroblasts have a unique gene signature correlated with poor prognosis of colorectal cancer. *Nature communications*, 12(1):2281, 2021.

[28] Diya B Joseph, Gervaise H Henry, Alicia Malewska, et al. Single-cell analysis of mouse and human prostate reveals novel fibroblasts with specialized distribution and microenvironment interactions. *The Journal of pathology*, 255(2):141–154, 2021.

[29] Mert Cemri, Melissa Z Pan, Shuyi Yang, Lakshya A Agrawal, Bhavya Chopra, Rishabh Tiwari, Kurt Keutzer, Aditya Parameswaran, Dan Klein, Kannan Ramchandran, et al. Why do multi-agent llm systems fail? *arXiv:2503.13657*, 2025.

[30] Tim Stuart, Andrew Butler, Paul Hoffman, et al. Comprehensive integration of single-cell data. *cell*, 177(7):1888–1902, 2019.

[31] Alsu Missarova, Jaison Jain, Andrew Butler, et al. genebasis: an iterative approach for unsupervised selection of targeted gene panels from scrna-seq. *Genome biology*, 22:1–22, 2021.

[32] Ian Covert, Rohan Gala, Tim Wang, et al. Predictive and robust gene selection for spatial transcriptomics. *Nature Communications*, 14(1):2091, 2023.

[33] CZI Single-Cell Biology Program, Shibla Abdulla, Brian Aevermann, et al. Cz cell× gene discover: A single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *BioRxiv*, pages 2023–10, 2023.

[34] Oscar Franzén, Li-Ming Gan, and Johan Björkegren. Panglaodb: a web server for exploration of mouse and human single-cell rna sequencing data. *Database*, 2019:baz046, 2019.

[35] Congxue Hu, Tengyue Li, Yingqi Xu, et al. Cellmarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scrna-seq data. *Nucleic acids*

research, 51(D1):D870–D876, 2023.

[36] Michael Schubert, Bertram Klinger, Martina Klünemann, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature communications*, 9(1):20, 2018.

[37] Shunyu Yao, Jeffrey Zhao, Dian Yu, et al. React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023.

Supplementary Information

1. SpatialAgent

1.1. Configurations of base LLMs

SpatialAgent development primarily utilized OpenAI's GPT-4o model (version released on August 6th, 2024, knowledge cutoff **October 2023**). GPT-4o supports a context window of 128,000 tokens and can generate responses up to 16,384 tokens in length, as in OpenAI's documentation: https://platform.openai.com/docs/models.

SpatialAgent was also adjusted for Anthropic's Claude-3.5-Sonnet (released on October 22, 2024, training data cutoff **April 2024**). Claude-3.5-Sonnet supports a context window of 200,000 tokens and can generate up to 8,192 tokens per response, according to its documentation: https://docs.anthropic.com/en/docs/about-claude/models. Without further adaptations, we found that the current agent framework can directly work with Claude-3.7-Sonnet (released on Fed 19, 2025).

1.2. LangChain framework

To enable structured reasoning, dynamic tool invocation, and memory persistence within SpatialAgent , LangChain was used, as a modular framework for building LLM-powered applications. LangChain allows us to compose multi-step workflows that incorporate retrieval-augmented generation (RAG), structured decision-making, and domain-specific tools.

In SpatialAgent , LangChain is employed with two key aspects:

- **Tool curation and orchestration**: LangChain's agent framework is used to orchestrate calls between LLMs and specialized tools for spatial biology. Agents dynamically decide which tools to invoke based on user queries or analysis objectives, ensuring efficiency in various tasks.
- **Memory and context handling**: To maintain consistency across multi-turn interactions and complex workflows, LangChain's memory modules are used, allowing intermediate outputs (*e.g.*, retrieved datasets, computed gene scores) to persist across reasoning steps.

1.3. Tools developments

Three main types of curated tools were used in SpatialAgent . The tool collection below can be readily extended, allowing for customization and integration of additional functionalities to meet specific research needs. This modular design enables researchers to incorporate their own specialized tools, databases, or analytical methods while maintaining compatibility with the existing framework.

1.3.1. Databases

Multiple databases were integrated to support retrieving reference datasets and conduct basic analysis, leveraging their structured metadata and large-scale annotations.

Chan-Zuckerberg Initiative (CZI) CELLxGENE is a scalable single-cell data platform offering access to over 50 million unique cell profiles with standardized metadata, facilitating meta-analyses, cross-dataset integration, and computational modeling (33). CELLxGENE was accessed via its Python API at https://chanzuckerberg.github.io/cellxgene-census/ on November 26, 2024.

The retrieved database, built on November 25, 2024, follows census schema version 2.1.0 and dataset schema version 5.2.0. It comprises 135,560,133 total cell profiles from 1003 datasets, including 71,730,590 unique cell profiles, annotated across 780 distinct cell types, 63 tissues, from 22,009 human donors and 4,752 mouse models.

PanglaoDB is a web-based resource for exploring mouse and human single-cell RNA-seq data, integrating over 1,000 experiments with more than 4 million cell profiles (34). It offers a user-friendly interface, standardized pre-processing pipelines, cell-type inference, and a curated marker gene compendium to enhance data accessibility and biological insights. The database was downloaded in TSV format from `https://panglaodb.se/markers.html` on May 20, 2024 (last update March 27, 2020 (10:44:00 CET)). The database contains 8,286 associations, spanning 178 cell types, 4,679 gene symbols, and 29 tissues.

CellMarker2 provides a manually curated collection of experimentally supported markers for 2,578 cell types across 656 tissues in humans and mice, totaling 83,361 tissue-cell type-marker associations (35), across 26 marker types (*e.g.*, protein-coding genes, lncRNAs) derived from 24,591 studies. The database was downloaded in csv format from `http://www.bio-bigdata.center/CellMarker_download.html` on Oct 20, 2024.

PROGENy provides footprints of pathway activity by focusing on transcriptionally responsive genes rather than pathway components themselves (36). The database contains consensus gene signatures for 14 key signaling pathways (EGFR, Hypoxia, JAK-STAT, MAPK, NFkB, p53, PI3K, TGFb, TNFa, Trail, VEGF, WNT, Androgen, and Estrogen), each represented by 100 responsive genes determined from thousands of perturbation experiments. PROGENy was accessed via its R package (version 1.18.1) from Bioconductor on June 2, 2024, which provides model matrices for both human and mouse organisms and supports weight matrices of varying sizes (100, 300, and 500 genes per pathway) to balance specificity and coverage depending on analytical needs.

LIANA (LIgand-receptor ANalysis frAmework) is a comprehensive resource for cell-cell communication inference that integrates multiple databases and analytical methods into a unified framework (21). It aggregates ligand-receptor pairs from CellPhoneDB, CellChat, NATMI, ConnectomeDB, and other resources, resulting in a comprehensive interaction collection covering more than 3,000 unique interactions. Its curated, non-redundant ligand-receptor interactions are annotated for complex structures (including heteromeric proteins), enabling accurate inference of intercellular signaling networks. LIANA version 0.1.8 was accessed on May 25, 2024, via its Python interface.

1.3.2. Data Processing

CZIInfo is a retrieval tool designed to extract relevant cell-type and tissue information from the CZI reference dataset based on specified criteria. It allows users to query by tissue, organism (*Homo sapiens* or *Mus musculus*), disease state, and dataset identifier. The tool loads census metadata from `czi_census_datasets_v4.csv` (codes on how to curate this are provided in `https://github.com/Genentech/SpatialAgentData`) and leverages an LLM-based tissue name matching pipeline to refine search results. It then filters dataset entries based on tissue and disease annotations, extracting and saving reference single-cell data. Cell types and tissues are further summarized using an LLM, and results are stored as structured text files for downstream annotation. Queries are executed via API-based dataset retrieval and processed using structured filtering and embedding-based

similarity search.

GeneVoting is an integration tool that consolidates multiple rounds of gene importance scoring to generate a refined gene panel. It aggregates importance scores from three iterations, normalizes gene names, and ranks genes based on cumulative importance. To ensure comprehensive coverage, it dynamically adjusts rankings to maintain representation across all cell types. The tool filters and selects the top genes while summarizing their associated reasons. The final gene panel is saved in a structured CSV format for downstream analysis, with automatic cleanup of intermediate files.

PreProcess automates the preprocessing of spatial transcriptomics data using Scanpy (11). It filters low-quality cells and genes, normalizes expression values, and applies log transformation to enhance downstream analyses. The tool performs dimensionality reduction via PCA, computes neighborhood graphs, and generates UMAP embeddings to capture spatial relationships. The processed dataset is saved in `.h5ad` format with standardized gene naming conventions, ensuring compatibility with subsequent computational workflows.

HarmonyTransfer enables the transfer of cell type annotations from CZI reference datasets to spatial transcriptomics data. It preprocesses and integrates single-cell RNA-seq and spatial datasets by identifying shared genes, normalizing expression values, and performing batch correction using Harmony integration. A multi-layer perceptron (MLP) classifier is trained on the reference dataset's principal components and applied to predict cell types in the spatial dataset. The final transferred labels are saved in structured CSV format for downstream analysis, ensuring a robust and scalable approach to cell type annotation in spatial omics studies.

MainUTAG applies the UTAG computational method (18) to identify spatially coherent tissue niches based on transcriptomic similarity. It clusters spatial transcriptomics data using the Leiden algorithm with adaptive resolution selection, optimizing for niche structure while maintaining biological interpretability. Small clusters are reassigned based on nearest-neighbor similarity to ensure robustness. The tool generates and saves structured outputs, including niche annotations in CSV format and visualizations of spatial niche distributions. These results support downstream analyses by providing high-resolution spatial context for cell-type organization.

CellTypeAnnotation similar as GPTCellType (15), is an automated clustering and annotation tool for spatial transcriptomics data, leveraging gene markers and reference-transferred labels to infer main-level cell types. It clusters cells using the Leiden algorithm and identifies marker genes via Wilcoxon rank-sum testing. The tool calculates cell type compositions within each cluster and employs a language model to assign cell types based on marker profiles and ontology constraints. Annotations are further refined through LLM-driven organization. Results, including annotated cell types and justification reasons, are saved in structured formats. The tool also generates UMAP visualizations for both clustering and final cell-type assignments to facilitate interpretation.

TissueNicheAnnotation is a multimodal annotation tool designed to infer tissue niches by integrating spatial transcriptomics data, cell type labels, and anatomical priors. It accepts preprocessed spatial data in `.h5ad` format alongside main-level cell type annotations and niche clustering results. The tool first extracts spatial features and visualizes niche clusters, incorporating anatomical reference images to enhance niche understanding. Using a language model, it annotates tissue niches

by considering spatial location, cell composition, and enriched gene sets. The tool also harmonizes annotations across multiple samples via an LLM-driven consensus mechanism. Final annotations, including tissue niche labels and supporting justifications, are saved in structured formats for downstream analysis. Spatial distributions of annotated niches are visualized to facilitate interpretation.

DynamicsInference performs cross-condition dynamics analysis in spatial transcriptomics data, integrating factor, cell-type, and pathway analyses. It extracts factor loadings, maps sample conditions, and conducts pairwise statistical tests to identify significant changes across experimental groups. The tool ranks cell-type contributions to key factors and detects enriched pathways using structured scoring. Results are saved in a structured `.pickle` format, providing a comprehensive dataset for downstream interpretation and integrative analysis.

CCIContext infers context-specific cell-cell interactions from spatial transcriptomics data using the Tensor-Cell2cell framework. It integrates ligand-receptor analysis, accounts for batch effects, and identifies condition-specific interaction patterns. The tool constructs a four-dimensional interaction tensor across samples, cell types, and signaling pathways, applying tensor decomposition to extract interpretable factors. It further performs pathway enrichment analysis using PROGENy to link ligand-receptor activity with downstream signaling. Results include interaction heatmaps, factor analysis, and pathway scores, saved in structured formats for downstream interpretation.

1.3.3. Aggregation

GeneImportance is an automated pipeline for estimating the importance of marker genes within cell types based on reference datasets. It processes marker gene information from multiple sources, *e.g.*, CZI, PanglaoDB, and CellMarker2, and utilizes an LLM to infer gene importance scores iteratively. The tool first extracts relevant cell-type and gene associations, then queries an LLM to rank genes based on their relevance. The inferred importance scores are reformatted and validated before aggregation into a structured CSV file. The final output provides a ranked list of genes with assigned importance scores, supporting downstream cell-type characterization and feature selection.

ReportGeneration automates the synthesis of spatial transcriptomics analysis results into a structured scientific report. It integrates findings from multiple analyses, including spatial patterns, cell-cell interactions, and condition-specific effects, filtering for statistically significant factors. The tool processes large-scale data using chunking strategies, applies language models for structured summarization, and generates coherent discussions and conclusions. The final report is formatted with background context, dataset descriptions, and analytical insights, ensuring comprehensive documentation for downstream interpretation.

SummarizeConditionTool, SummarizeCellTypeTool, and SummarizeTissueRegionTool facilitate the structured analysis of spatial transcriptomics data by summarizing condition-specific patterns, cell-type distributions, and tissue-region characteristics. The tools leverage an LLM to generate detailed descriptions both per sample and across conditions. They integrate contextual metadata, including tissue type and dataset information, to enhance interpretability. Condition summaries capture key differences across experimental groups, cell-type summaries contextualize distributions across tissue regions, and tissue-region summaries provide insights into spatial organization. The outputs are structured as JSON files for downstream analyses and reporting.

1.3.4. General purpose

Coding powers SpatialAgent to generate and execute custom functions for *any* analyses. It supports Python and shell scripts with emphasis on data analysis. The tool handles various data input / output formats, maintains execution context, and provides error handling. For spatial transcriptomics, Coding facilitates custom statistical tests, specialized visualizations, and integration with external packages not explicitly covered by other tools. Though it offers great flexibility (ideally all the tools can be generated by the LLM on-the-fly), this tool indeed requires a base LLM with exceptional coding performance (*e.g.*, Claude-3.7-Sonnet far outperforms GPT-4o).

1.4. Plan templates

While SpatialAgent can generate a valid plan for many user queries, which can also be refined by additional user's feedback, using plan templates for high-frequent tasks can largely improve the stability (*i.e.* completion rate) and performance. (Note that in other scientific agentic systems, such as Google's Co-Scientist, most computation / agentic burden is spent on making a concrete research plan). Three plan templates we included for the respective three representative tasks (Supplementary Fig. 1, 2, 3). These are provided to the SpatialAgent during memory initialization stages and can be extended per user's own need and creativity.

Task: Design a comprehensive gene panel through iterative database analysis and marker gene identification.

Input: A description covering the targeted tissue, organ, species, condition.

Process: Execute THREE complete iteration rounds. Label each step with the current iteration number (e.g., "Round X, Step Y").

For each iteration, repeat the five steps:

1. Reference Dataset (CZI Database)
   - Find best matching single-cell reference dataset, ideally different for each round

2. Dataset Analysis
   - Extract cell types and their marker genes from the reference dataset

3 & 4. More Database Search - PanglaoDB, CellMarker2
   - Find additional markers for each cell type, compare with previous findings

5. Gene Assessment
   - Score genes based on: Database frequency, Expression levels, Cell-type specificity

Final integration after all steps of three iterations:
- Combine findings, Create consolidated panel prioritizing: Coverage, Confidence and Non-redundancy.

Supplementary Figure 1: Templated plan for gene panel design.

1.5. Core prompts in the agentic framework

A sophisticated set of core prompts were incorporated in SpatialAgent to orchestrate its autonomous behavior in spatial biology analysis. These prompts form the cognitive architecture of the agent, beginning with a system prompt (Supplementary Fig.4) that establishes the agent's identity as a computational biology analyst and defines its operational parameters with the zero-shot ReAct setting (37). The semantic memory initialization prompt (Supplementary Fig.5) structures the agent's

Task: Annotate cell types and tissue niches in spatial transcriptomics data.

Typical Plan:

1. Data Preprocessing

2. Main Level Cell Type Annotation
   - Find best matching CZI reference dataset, then retrieve related information
   - Transfer cell type labels, then annotate clusters using transferred labels and marker genes

3. Main Level Tissue Niche Annotation
   - Identify spatial clusters based on expression coherence
   - Name clusters using: Anatomical information, Cell type distributions, Marker genes, Spatial patterns

Supplementary Figure 2: Templated plan for cell type and tissue niche annotation.

Task: Analyze cell-cell communications, interaction dynamics in spatial transcriptomics data across conditions.

Typical Plan:

1. Dataset Understanding and Characterization
   - Summarize information w.r.t. condition, then cell type, and finally tissue region.

2. Cell-Cell Communication Analysis
   - Apply computational tools for interaction inference
   - Identify significant ligand-receptor pairs
   - Map communication networks within spatial context

3. Cross-condition Dynamics Inference
   - Compare interaction patterns between conditions
   - Identify condition-specific communication signatures
   - Integrate with spatial and cellular context

4. Comprehensive aggregation, writing a report
   - Synthesize findings across all analysis levels
   - Document key communication changes
   - Highlight biological implications

Supplementary Figure 3: Templated plan for inferring cell-cell communications.

knowledge representation, incorporating task templates and available tools. The planning framework (Supplementary Fig.7) implements a structured thought process with five key components: task status assessment, plan review, action determination, action specification, and success criteria evaluation. This deliberative cycle enables the agent to monitor progress, adapt strategies, and determine appropriate next steps. Finally, the action execution prompt (Supplementary Fig.6) provides a streamlined template for tool calling, ensuring that each operation is precisely formatted with clear input parameters and expected outputs. Together, these prompts create a cohesive cognitive system that guides the agent through complex spatial biology workflows while maintaining contextual awareness and methodical execution.

```
You are an action agent helping with the task.
You must execute the tool exactly once and return its output.
When you receive an Observation from the tool execution:

1. Return the Observation's exact content without any modifications or formatting
2. Do not add any additional text before or after the Observation
3. Do not add any explanations or summaries
4. Do not modify or reformat the Observation in any way
The output should contain only the raw Observation text.
```

Supplementary Figure 4: System prompt on launching the agent.

```
You are a computational biologist called {agent.name}.

Available tools:
{tool_list.values()}

Task templates (create your own plan if no template matches):
{agent.plan_template}

Ask the user what task they would like to complete.
```

Supplementary Figure 5: Prompt on initializing semantic memory.

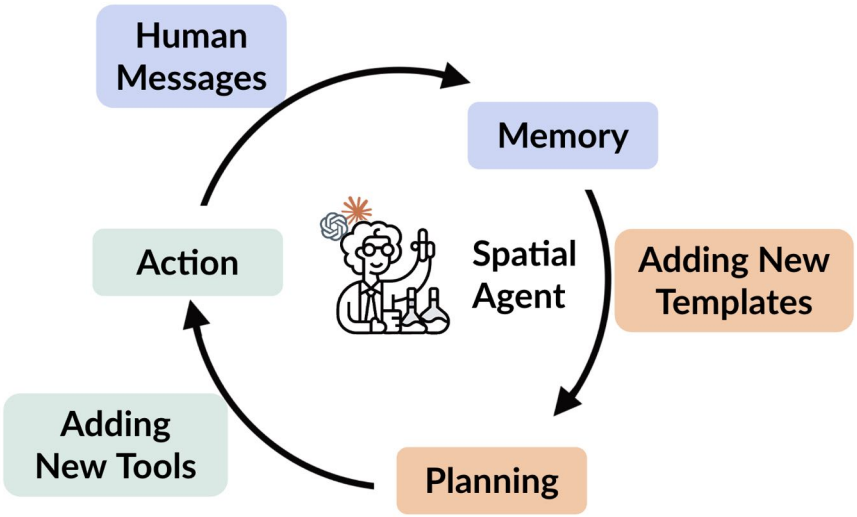
```
The plan is: {plan}

EXTRACT AND FORMAT CURRENT STEP:
[Do not modify information, only reformat according to structure below]

USE_TOOL:
[exact tool name]

INPUT:
[exact input parameters]

EXPECTED_OUTPUT:
[specific expected result]
```

Supplementary Figure 6: Prompt on making actions (tool calling).

```
CONTEXT
Conversation History: {chat_context_memory}

Knowledge Base: {semantic_memory}

Current Task: {task}

Current Plan: {plan}

Recent Actions: {episodic_memory}

Last Action: {most_recent_episodic_memory}

Available Tools: {valid_action_list}


THOUGHT PROCESS

1. Task Status
  Think: Is the task complete based on the current state?
  - If yes -> Return [STOP]
  - If no -> Continue planning

2. Plan Review
  Think: Do we have an existing plan template?
  - If no -> Create new plan using available tools
  - If yes -> Evaluate current progress and needs
  Reasoning: [Your analysis of the situation]

3. Next Action Determination
  Think: What specific step should be taken next?
  Consider:
  - Current progress stage
  - Available tools
  - Required inputs
  - Expected outputs
  Reasoning: [Your step selection logic]

4. Action Specification
  Selected Action:
  - Tool: [exact tool name]
  - Inputs: [precise values/paths]
  - Expected Output: [specific format/location]

5. Success Criteria
  Think: How will we know this step succeeded?
  - Output requirements
  - Validation checks
  Reasoning: [Your success criteria logic]

DECISION
[Provide the exact next action with complete specifications]
```

Supplementary Figure 7: Prompt on proposing and updating plan.

1.6. Extendability



Supplementary Figure 8: **Extendability of SpatialAgent .**

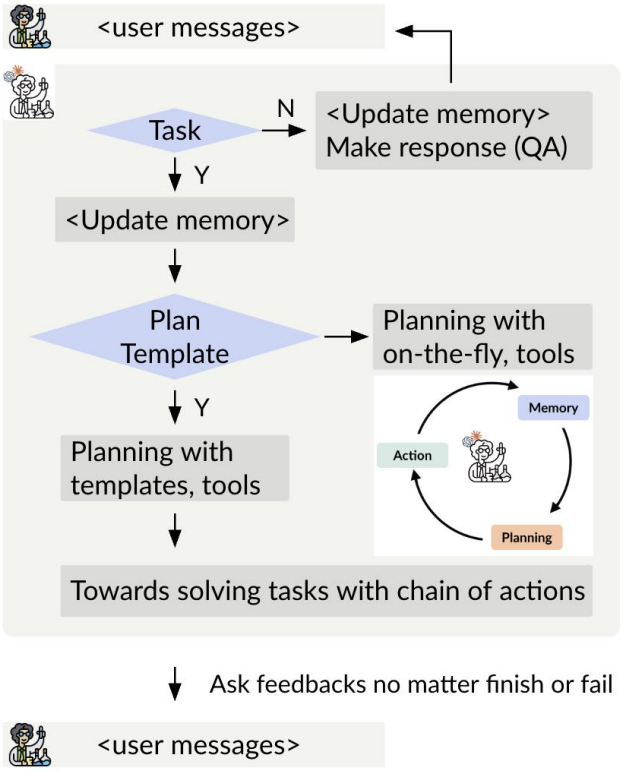
SpatialAgent was designed to be extendable, so it can evolve through three primary mechanisms (Supplementary Fig.8). First, SpatialAgent's **memory continuously updates through human messages**, enabling iterative refinement based on user feedback and new information. Second, SpatialAgent's planning capabilities can be expanded by **adding new plan templates** that encode domain-specific knowledge and problem-solving approaches without modifying the underlying architecture. Third, SpatialAgent's action abilities can be enhanced by **incorporating new tools** that connect to external services, computational resources, or specialized functions through a consistent interface.

This modular design creates a flexible cycle where users can extend the agent's functionality through natural interaction rather than complex reprogramming. For example, researchers can introduce specialized templates for more specific or sophisticated analysis, add tools for spatial computation or visualization, and guide the traces through conversational feedback. The result is a system that adapts to increasingly complex spatial reasoning challenges while maintaining operational coherence across diverse applications.

1.7. Generalization

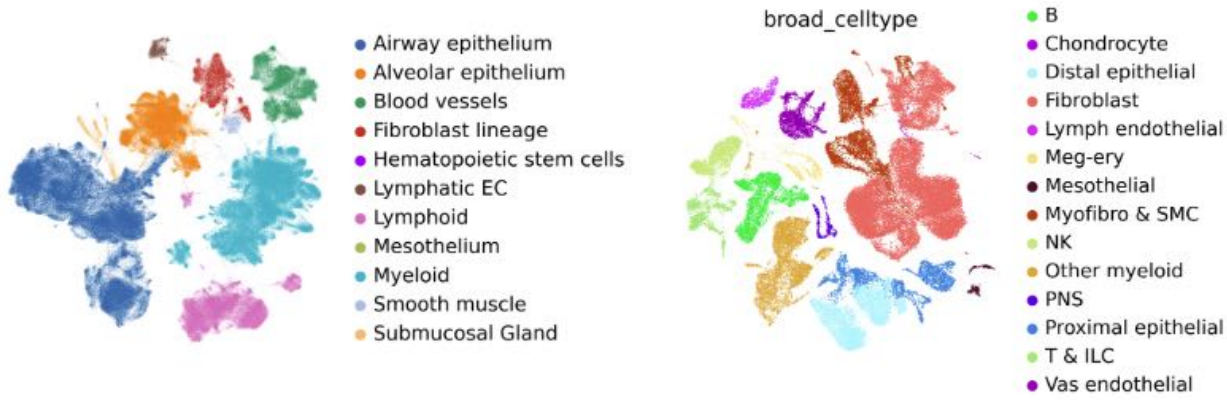
While all results in main-text use autonomous mode with predefined tools and templates, we believe a flexible decision-making architecture is beneficial for robust generalization. It first distinguishes between simple queries and complex tasks. For non-task queries, it updates memory and replies directly without planning. Task-oriented requests activate full capabilities: after memory updates, the agent checks for suitable plan templates. If available, it follows predefined strategies; otherwise, it dynamically composes tools (mostly coding) to form custom plans.

A continuous memory–action–planning loop enables in-execution refinement based on intermediate results. This adaptive loop supports generalization across tasks and incorporates human feedback, often outperforming purely autonomous runs. To test generalization, we next present two unseen tasks requiring novel plans and coding. SpatialAgent succeeds at cell type annotation harmonization (Supplementary Fig.10, 11, 12) but fails at spatial gene regulatory analysis—a task that remains difficult even for Claude-Sonnet-3.7 and GPT-O1 (Supplementary Fig.10, 14).



Supplementary Figure 9: **Generalizability.**

1.7.1. Annotation harmonization



Supplementary Figure 10: **Task overview for cell type annotation harmonization.** The agent is given two human lung scRNA-seq datasets with distinct label sets and asked to harmonize the annotations.

For the two scRNA-seq datasets provided, harmonize the cell type annotations by mapping similar cell populations.

- Dataset 1: /home/wangh256/hanchen/spatialagent_dev_gitlab/data/dataset_harmonization/hlca.h5ad
- Dataset 2: /home/wangh256/hanchen/spatialagent_dev_gitlab/data/dataset_harmonization/fetal.h5ad

- Consider biological markers, transcriptional profiles, and functional characteristics when aligning cell type labels
- Provide a mapping table that shows the correspondence between original annotations and the new unified annotations
- Save the results to /home/wangh256/hanchen/spatialagent_dev_gitlab/experiments/debug/dataset_harmonization/mapping_table.csv

I'll help you harmonize the cell type annotations between the two scRNA-seq datasets. Let me analyze it and create a plan.

Task Analysis
I need to:
1. Load and examine both datasets (HLCA and fetal)
2. Understand the cell type annotations in each dataset
3. Create a mapping between similar cell populations
4. Save the mapping table to the specified output path

Plan
1. Load and explore both datasets
2. Extract cell type annotations from each dataset
3. Analyze cell type similarities based on markers and profiles
4. Create a mapping table between the datasets
5. Save the results to the specified path

Let me execute this plan using Python:

```python
import scanpy as sc
import pandas as pd
import numpy as np
import os

# Load the datasets
hlca_path = "/home/wangh256/hanchen/spatialagent_dev_gitlab/data/dataset_harmonization/hlca.h5ad"
fetal_path = "/home/wangh256/hanchen/spatialagent_dev_gitlab/data/dataset_harmonization/fetal.h5ad"
output_path = "/home/wangh256/hanchen/spatialagent_dev_gitlab/experiments/debug/dataset_harmonization/mapping_table.csv"

# Ensure output directory exists
os.makedirs(os.path.dirname(output_path), exist_ok=True)

# Load datasets
print("Loading HLCA dataset...")
hlca = sc.read_h5ad(hlca_path)
print("Loading fetal dataset...")
fetal = sc.read_h5ad(fetal_path)

# Extract cell type annotations
hlca_celltypes = hlca.obs['cell_type'].unique().tolist()
fetal_celltypes = fetal.obs['cell_type'].unique().tolist()

print(f"HLCA dataset has {len(hlca_celltypes)} cell types")
print(f"Fetal dataset has {len(fetal_celltypes)} cell types")

# Create a mapping table
# This will be a dictionary where keys are fetal cell types and values are HLCA cell types
mapping = {}

# For each fetal cell type, find the most similar HLCA cell type
# This is a simplified approach - in practice, we would use marker genes and expression profiles
for fetal_celltype in fetal_celltypes:
    # Find the most similar HLCA cell type based on name similarity
    # In a real scenario, we would use more sophisticated methods
    best_match = None
    best_score = 0
...
```

Supplementary Figure 11: **Initial user query and agent response.** The agent correctly identifies the task and starts by inspecting the dataset metadata.

<after chains of actions...>

```
mapping = [
        {
          "hlca": "Airway epithelium",
          "fetal": ["Proximal epithelial"],
          "confidence": "high",
          "notes": "Airway epithelium in adults corresponds to proximal epithelial cells in fetal development"
        },
        {
          "hlca": "Alveolar epithelium",
          "fetal": ["Distal epithelial"],
          "confidence": "high",
          "notes": "Alveolar epithelium in adults develops from distal epithelial progenitors in the fetus"
        },
        {
          "hlca": "Blood vessels",
          "fetal": ["Vas endothelial"],
          "confidence": "high",
          "notes": "Adult blood vessel cells correspond to vascular endothelial cells in the fetus"
        },
        {
          "hlca": "Fibroblast lineage",
          "fetal": ["Fibroblast"],
          "confidence": "high",
          "notes": "Direct correspondence between adult and fetal fibroblasts"
        },
...
```

Supplementary Figure 12: **Final result: a successful annotation mapping.** After a few iterations using the coding tool, the agent produces a reasonable mapping across datasets.

1.7.2. Gene Regulatory Network Inference with the Spatial Context

Spatial Gene Regulatory Network Inference

Analyze the provided 10x Visium spatial transcriptomics dataset from mouse brain cortex to infer gene regulatory networks with spatial context.

- Input dataset: ./data/spatial_grn/mouse_brain_visium.h5ad
- Gene list: ./data/spatial_grn/transcription_factors.txt

Tasks:
1. Identify the top 50 spatially variable genes across cortical layers
2. For each transcription factor in the provided list, predict its target genes based on spatial co-expression patterns
3. Build a directed network where edges represent likely regulatory relationships
4. Identify network modules specific to cortical layers

Output files (save to log/spatial_grn/):
- spatial_variable_genes.csv: List of spatially variable genes with statistics
- tf_target_predictions.csv: Table of TF-target gene pairs with confidence scores
- spatial_network.graphml: Network file compatible with Cytoscape
- cortical_layer_modules.pdf: Visualization of layer-specific regulatory modules

Supplementary Figure 13: **Task overview for spatial GRN inference.** The agent is asked to infer gene regulatory relationships within spatial transcriptomics data.

I'll help you build a spatial gene regulatory network inference pipeline for the mouse brain cortex dataset.
Let me understand the task and create a plan.

Task Analysis

You need to analyze a 10x Visium spatial transcriptomics dataset from mouse brain cortex to infer gene regulatory networks with spatial context. The goal is to identify regulatory relationships between transcription factors and their target genes, with special attention to spatial patterns across cortical layers.

Plan for Spatial Gene Regulatory Network Inference

1. **Data Loading and Preprocessing**
   - Load the Visium dataset from the provided path
   - Perform quality control and normalization
   - Explore dataset structure and spatial organization

2. **Spatial Variable Gene Identification**
   - Identify the top 50 spatially variable genes across cortical layers
   - Calculate spatial statistics for these genes
   - Save results to spatial_variable_genes.csv

3. **Transcription Factor Analysis**
   - Load the provided transcription factor list
   - For each TF, identify potential target genes based on spatial co-expression
   - Calculate confidence scores for TF-target relationships
   - Save results to tf_target_predictions.csv

4. **Network Construction**
   - Build a directed network with TFs as source nodes and targets as destination nodes
   - Weight edges based on confidence scores
   - Export network in graphml format for Cytoscape compatibility

5. **Cortical Layer Module Identification**
   - Identify network modules specific to different cortical layers
   - Visualize layer-specific regulatory modules
   - Save visualization to cortical_layer_modules.pdf

Let me implement this plan using Python code.

USE_TOOL:
python

INPUT:
```python
# Step 1: Data Loading and Preprocessing
import scanpy as sc
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import os

# Set random seed for reproducibility
np.random.seed(42)

# Load the Visium dataset
print("Loading Visium dataset...")
adata = sc.read_visium("./mouse_brain_visium")
print(f"Dataset loaded with {adata.shape[0]} spots and {adata.shape[1]} genes")

# Basic dataset exploration
print("\nBasic dataset information:")
print(adata)

# Quality control metrics
sc.pp.calculate_qc_metrics(adata, inplace=True)
print("\nQuality control metrics calculated")

...
```

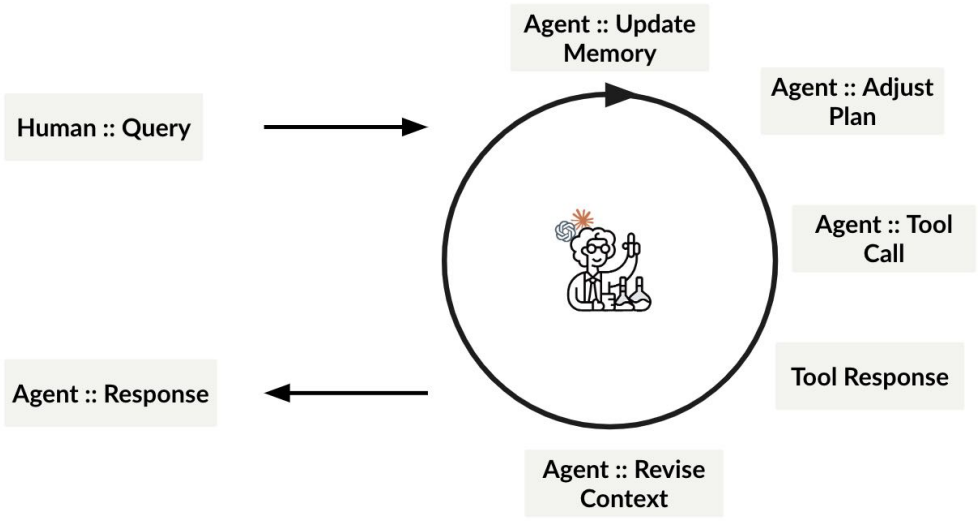
Supplementary Figure 14: **Initial query and failed agent response.** The agent attempts to reason about the task and identify relevant steps, but ultimately fails due to the task's complexity.

1.7.3. Summary

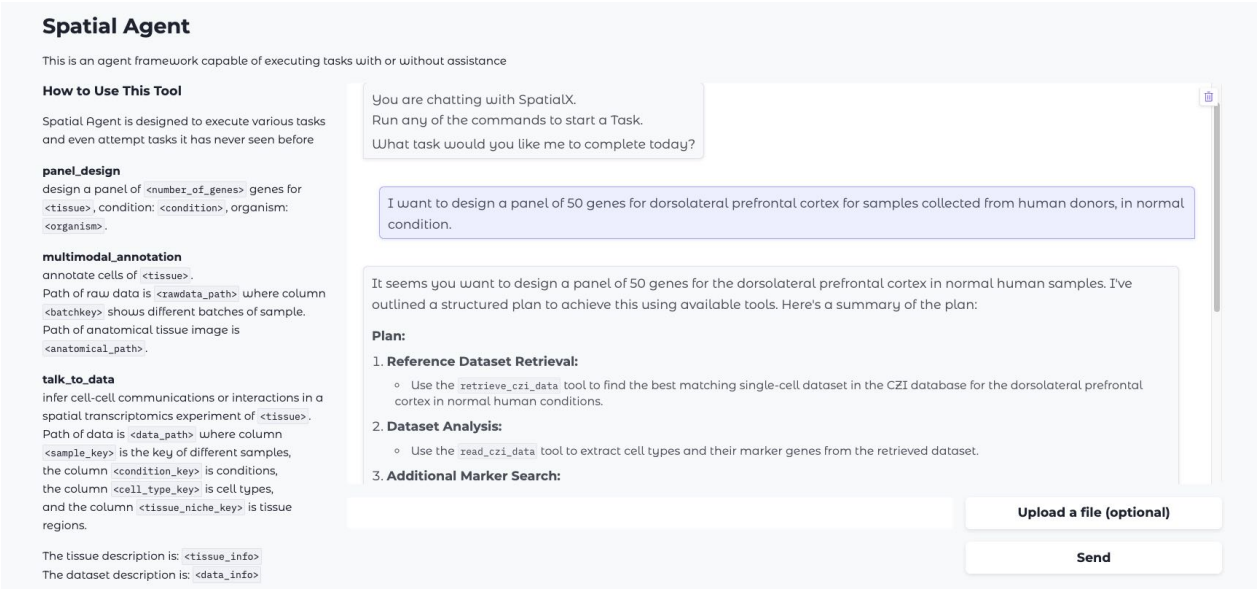
These two tasks demonstrate both the strengths and current limitations of SpatialAgent's generalization. In the annotation harmonization task, the agent adapts effectively through iterative planning and code generation, leveraging its flexible architecture. In contrast, the spatial GRN task reveals difficulties in multi-step reasoning and specific tool orchestration - challenges that remain unsolved even for the most advanced LLMs.

Addressing these limitations will require stronger coding capabilities, improved reasoning for dynamic plan construction, integration of domain-specific tools (*e.g.*, PySCENIC+), and clearer user prompts that reflect first-hand know-hows. These improvements are critical for boosting the agent's completion rates and performance on complex, open-ended scientific tasks.

1.8. Interaction design of SpatialAgent



Supplementary Figure 15: **Design on the interaction mode of SpatialAgent.**



Supplementary Figure 16: **User interface of SpatialAgent, developed using Gradio package.**

SpatialAgent enables dynamic interaction between users and the agentic system (Supplementary Fig. 15, 16). A user query triggers the agent to update its memory and context. It then revises its execution plan, refining templates or generating new ones, and invokes external tools or databases as needed. Tool outputs are integrated with the agent's knowledge to revise its context. The agent then responds with insights, visualizations, or recommendations based on the analysis.

This cyclical design enables continuous refinement through interaction, as each human query triggers a complete processing cycle that builds upon previous exchanges. The structured workflow maintains consistency while allowing for adaptive responses to diverse inquiries, creating an intuitive collaborative experience between researchers and the automated assistant.

2. Gene panel design

2.1. Computational baselines

Several computational baseline methods were used. Each relies on scRNA-seq data as a reference from which to select a set of genes predicted to be most informative in spatial contexts.

HVG(Seurat) identifies genes with high expression variance across cells, typically after normalizing for the relationship between mean expression and variance. This approach prioritizes genes with biological heterogeneity over technical noise.

GeneBasis greedily selects genes that maximize mutual information with principal components of the full expression matrix. It iteratively builds a gene panel that preserves the overall structure of the dataset while minimizing redundancy among selected genes.

Persist uses a persistence-based approach to identify genes that contribute significantly to topological features in the data. It quantifies how each gene affects persistent homology metrics, selecting genes that maintain important structural information across different scales.

Spapros applies a sparse regression framework to identify a minimal set of genes that can reconstruct the full expression matrix with high fidelity. It leverages L1-regularization to enforce sparsity while optimizing for genes that capture maximum information about cell states and transitions.

2.2. Human scientists evaluation

2.2.1. Study design

Human expert performance in panel design was systematically evaluated using a two-stage approach. The human experts included both computational and experimental scientists, who were either pursuing a PhD or held an MD/PhD in a related field (*e.g.*, cell biology, biostatistics, computer science) and had published work.

Stage 1: Domain experts selected 50, 100, or 150 genes for spatial transcriptomics experiments focusing on the DLPFC. They were provided with the same scRNA-seq dataset used by baseline methods and SpatialAgent , along with exploratory data analysis (EDA) notebooks. Experts documented their selections with gene symbols, rankings, and detailed reasoning. They also provided their research backgrounds and were allowed to consult literature (excluding original dataset papers) and online resources, except for AI tools like chatGPT. We then evaluated these panels using the same set of metrics as we benchmarked the computational baselines and SpatialAgent .

Stage 2: We implemented a comparative evaluation framework where experts assessed outputs from various sources (SpatialAgent , GPT-4o, human experts, and SpatialAgent + human combinations) using a structured 5-point rubric. The rubric evaluated accuracy, reasoning quality, completeness, and conciseness of gene selections. This approach enabled a quantitative comparison of human and AI performance while exploring the potential of human-AI collaboration in gene panel design for spatial transcriptomics.

	Panel Design Workflow
D	Used literature-based approach with integrated human PFC atlas. Selected cell type markers from literature and added markers for neuronal activity and neuropathologies (Alzheimer's, Schizophrenia).
Ro	Used reference dataset only. Applied Persist algorithm to select top 50, 100, and 150 genes with algorithmic approach without manual curation.
RB	Submitted previously designed panel (incorrect tissue).
T	Combined reference dataset with literature. Used gene enrichment ranked with Cohen's mean and supplemented with known marker genes for key cell types.
C	Used reference dataset only. Applied iterative greedy algorithm for kNN graph reconstruction with focus on graph-based representation of data.
M	Used alternative dataset (not the provided one). Selected genes to differentiate predicted cell types by integrating scRNA-seq with Visium data and using spot-level deconvolution.
P	Used multiple datasets (provided + cell type-specific studies). Selected cell type markers from multiple references and DE genes. Included genes targeting intracellular pathways and cross-referenced markers across studies.
TY	Used reference dataset only. Selected highly variable genes with optimal normalization techniques.
V	Used literature only (did not use reference). Selected cell type markers from broad and targeted studies, including presynaptic markers, transcription factors, and Schizophrenia risk genes in GRNs.
HC	Used reference dataset only. Selected top DE genes for each cell type with focus on differential expression.

Supplementary Table 1: Comparison of panel design workflows for spatial transcriptomics

All materials, including instructions, EDA notebooks, and anonymized human scientists' solutions, are provided in the supplementary files.

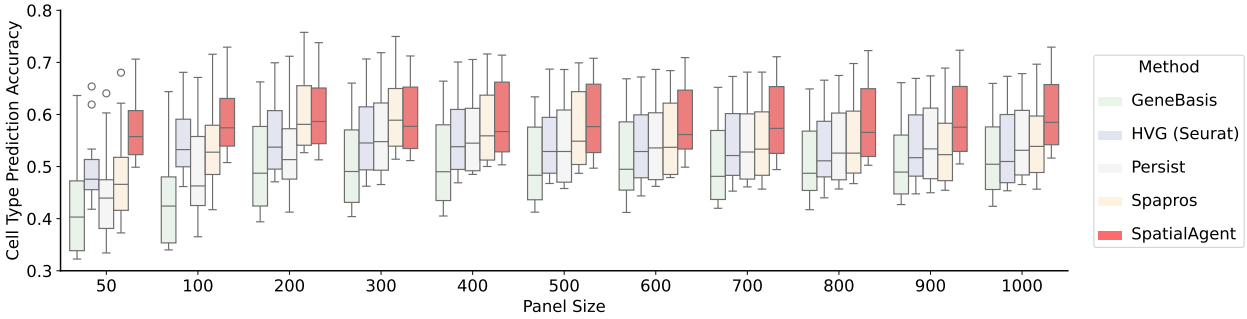
2.2.2. Overview of human solutions

Researchers employed several distinct strategies when selecting genes for spatial transcriptomics panels (Supplementary Table. 1). Most researchers relied on either the provided reference dataset or literature sources, with only a few integrating multiple data sources. Their selection methods typically fell into three categories: (1) algorithmic approaches (Persist algorithm, iterative greedy methods, or highly variable gene selection), (2) differential expression analysis to identify cell type-specific markers, or (3) knowledge-driven selection of known markers from literature. The more comprehensive approaches, such as those from P and V, combined multiple strategies by cross-referencing markers from different sources and including genes related to specific biological processes or pathways of interest. Some researchers focused exclusively on cell type identification, while others extended their panels to capture pathway activities or disease-relevant genes.

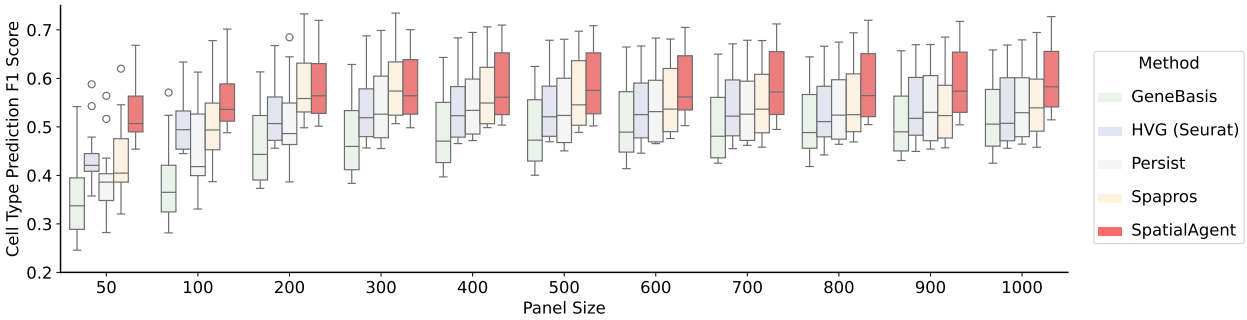
We provide anonymized human expert designs in the supplement to support future refinement and improvement by other practitioners.

2.3. Pairwise comparisons on cell type prediction

We also conducted a pairwise statistical comparison of cell-type prediction performance across different methods on the DLPFC dataset, using gene panels of 50 and 100 genes. Each entry in the tables below reports the t-statistic and corresponding p-value (in parentheses) from independent two-sample t-tests. A positive t-statistic indicates that the method in the row outperforms the method in the column. Statistically significant differences ($p < 0.05$) are highlighted in bold. As panel size increases, performance gaps between SpatialAgent and other baselines become less pronounced.



Supplementary Figure 17: Accuracy in predicting cell type.



Supplementary Figure 18: F1 score in predicting cell type.

Supplementary Table 2: Pairwise statistical comparison of cell-type prediction using 50-gene panels. Values represent test statistics from independent t-tests, with p-values in parentheses. Statistically significant differences are bolded ($^*p < 0.05$), same notations below.

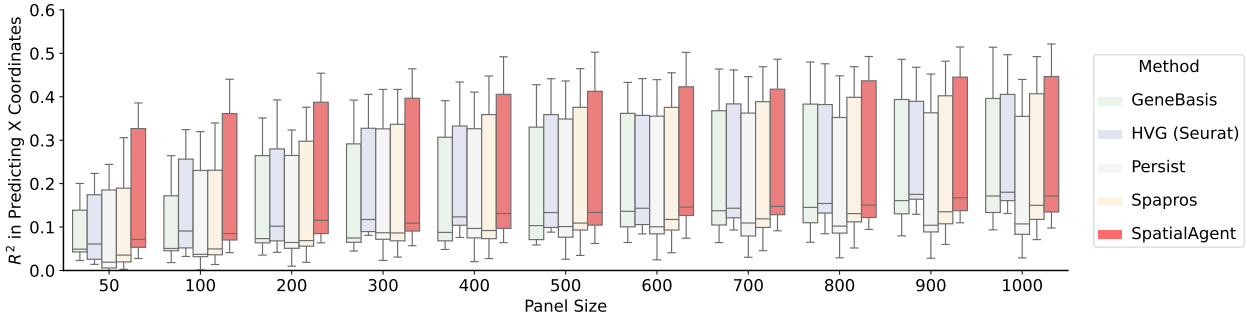
Method	SpatialAgent	GeneBasis	Persist	HVG (Seurat)	Spapros
SpatialAgent	-	**2.668** (0.014)	**3.629** (0.001)	**2.625** (0.015)	**2.668** (0.014)
GeneBasis	**-2.668** (0.014)	-	0.774 (0.447)	-0.440 (0.664)	0.000 (1.000)
Persist	**-3.629** (0.001)	-0.774 (0.447)	-	-1.329 (0.198)	-0.774 (0.447)
HVG (Seurat)	**-2.625** (0.015)	0.440 (0.664)	1.329 (0.198)	-	0.440 (0.664)
Spapros	**-2.668** (0.014)	0.000 (1.000)	0.774 (0.447)	-0.440 (0.664)	-

2.3.1. More metrics on location restoration

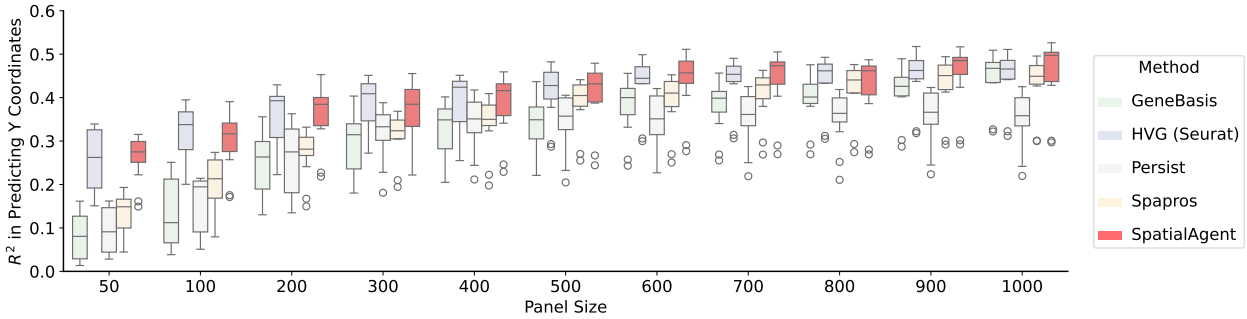
Essentially most of other metrics (clustering-based) as well as the spatially expressed variations are highly correlated with the cell type and location prediction performance, the reader can easily calculate on their ends as we release the data.

Supplementary Table 3: Pairwise statistical comparison of cell-type prediction using 100-gene panels.

Method	SpatialAgent	GeneBasis	Persist	HVG (Seurat)	Spapros
SpatialAgent	-	1.533 (0.139)	**2.903** (0.008)	1.518 (0.143)	1.533 (0.139)
GeneBasis	-1.533 (0.139)	-	1.362 (0.187)	-0.203 (0.841)	0.000 (1.000)
Persist	**-2.903** (0.008)	-1.362 (0.187)	-	-1.687 (0.106)	-1.362 (0.187)
HVG (Seurat)	-1.518 (0.143)	0.203 (0.841)	1.687 (0.106)	-	0.203 (0.841)
Spapros	-1.533 (0.139)	0.000 (1.000)	1.362 (0.187)	-0.203 (0.841)	-



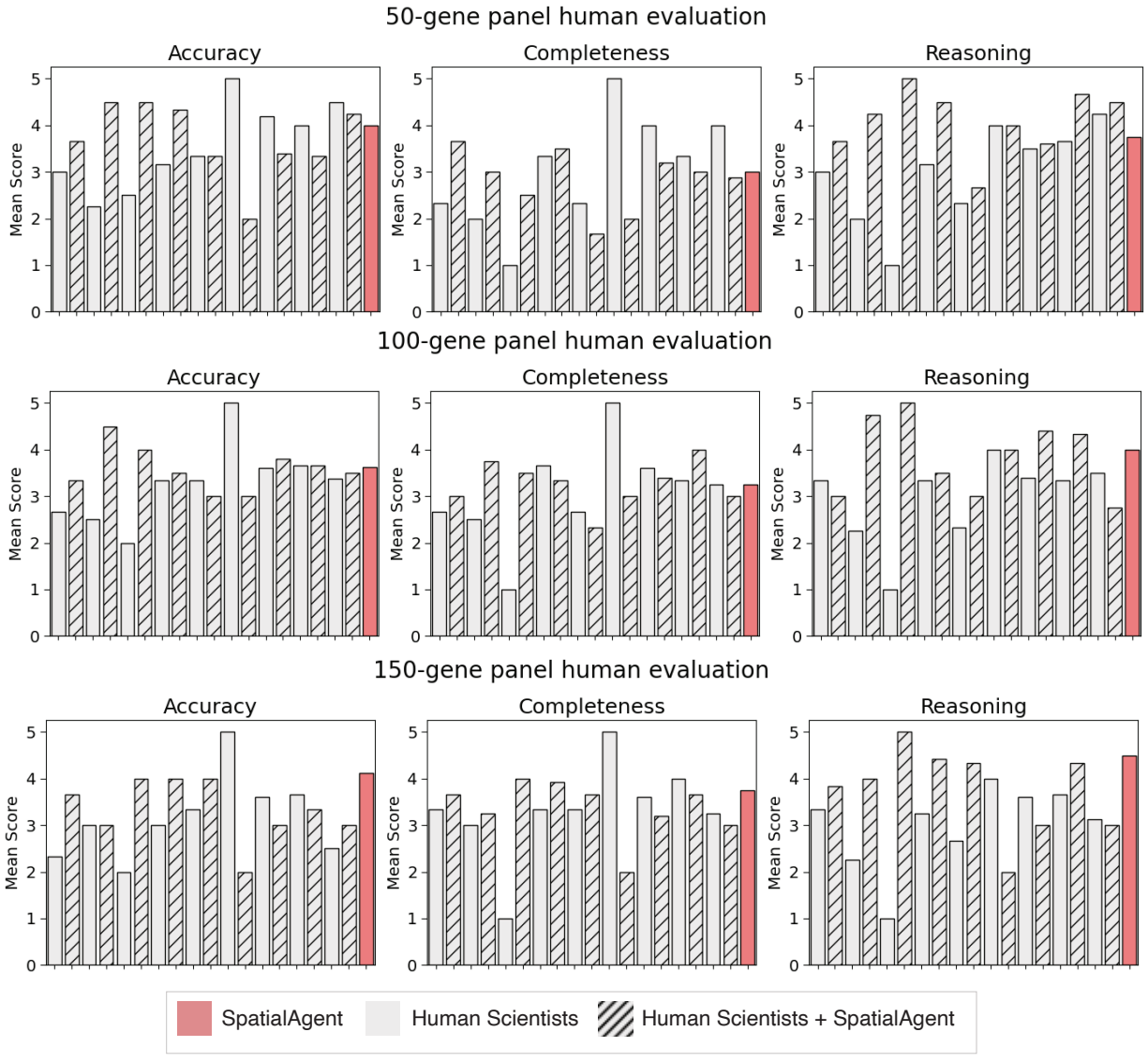
Supplementary Figure 19: Correlation in predicting x coordinates.



Supplementary Figure 20: Correlation in predicting y coordinates.

2.3.2. Performance via human expert rating

Human scientists were asked to rate anonymized gene panels that included a mixture of those designed by other scientists, the SpatialAgent and hybrid designs of SpatialAgent incorporating human-panels. Participants were asked to rate each panel across three categories: accuracy (can the given panel accurately map the DLPFC?), completeness (is the designed panel proposed complete?) and reasoning (is the reasoning given for the gene panel selection logical and insightful?). Scientist scored panels from 1 (Poor) to 5 (Excellent). Mean score for each panel are summarized in Supplementary Fig.21 below.



Supplementary Figure 21: Human evaluation of panel design

3. Cell type and tissue niche annotation

3.1. Overview

Cell type and tissue niche annotation is a crucial step in spatial transcriptomics, enabling the biological interpretation of complex spatial gene expression patterns. Unlike scRNA-seq, spatial transcriptomics preserves tissue context, allowing researchers to identify both cell types and their spatial organization within micro-environments. This integration is essential for understanding cellular interactions, developmental processes, and disease mechanisms in their native spatial contexts.

The annotation process typically follows a hierarchical structure, from broad cell type categories (Tier 1) to finer subtypes (Tiers 2–3), alongside spatial domain identification. Annotations are derived from marker gene expression, spatial coordinates, and morphological features. The integration of these multimodal data presents computational and biological challenges, especially in developing or pathological tissues where cell states exist along a continuum rather than discrete categories.

3.2. Baselines

We compare SpatialAgent with two automated cell type annotation methods:

CellTypist is a supervised machine learning approach using multinomial logistic regression trained on curated reference datasets. It supports hierarchical classification with probabilistic outputs but is limited by the quality and scope of its training data.

GPTCellType is LLM-based method that interprets gene expression profiles and generates cell type annotations with explanatory rationales. It integrates diverse information sources and adapts to novel cell types but may hallucinate annotations when faced with ambiguous expression patterns. To enable side-by-side comparison, we run the GPTCellType directly on the leiden cluster calculated from SpatialAgent .

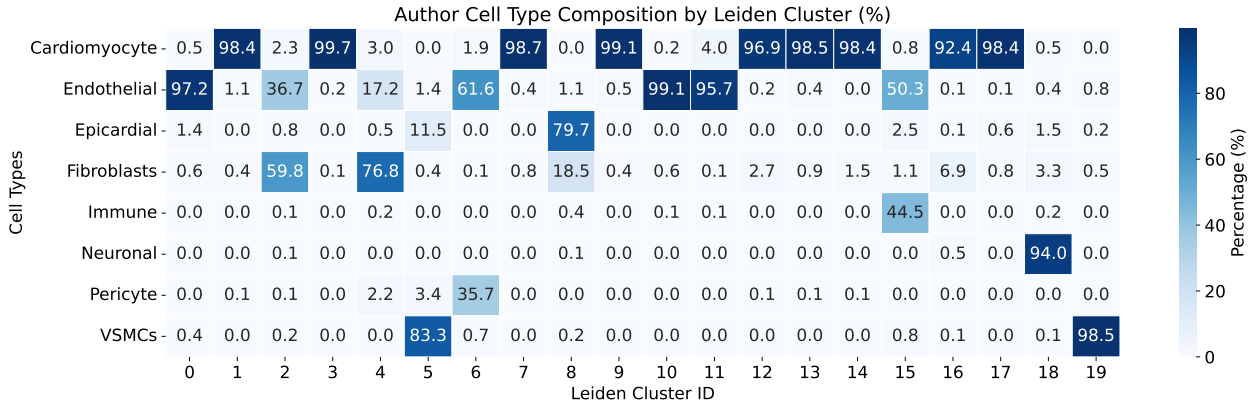
To ensure fair comparisons, we harmonize all annotations, including those from computational methods, human scientists, and SpatialAgent , to a common tier-1 reference set. We use the original author-provided annotations as ground truth, acknowledging the inherent bias in this setting.

3.3. Per cluster annotation between GPTCellType and SpatialAgent

We next compare SpatialAgent with GPTCellType on the same set of leiden clustering, noting that the authors did not use the same set cluster for annotation, but we notice that SpatialAgent can provide a descent annotation in Tier-1 annotations.

3.4. Removing small clusters in UTAG

Building on the original UTAG algorithm, we employ a nearest-neighbor approach to reassign cells from small clusters to larger, more established ones. Specifically, we identify clusters with fewer than a specified threshold (default: 100 cells) for removal. For each batch, we reassign cells in small clusters by finding their five nearest spatial neighbors among reference cells (*i.e.*, those in larger clusters). The most common cluster label among these neighbors is then assigned to the cell, ensuring spatial coherence by favoring proximity-based reassignments over arbitrary ones.



Supplementary Figure 22: Composition of author annotations across Leiden clusters used by SpatialAgent and GPTCellType.

3.5. Analysis of human expert annotations

Researchers employed consistent approaches for cell and niche annotation in spatial transcriptomics data (Supplementary Table. 4). For cell annotation, most relied on computational clustering (primarily Leiden) followed by either manual annotation with marker genes or automated label transfer methods like CellTypist or TACCO. Many provided hierarchical (multi-tier) annotations to capture both major cell types and subtypes. Only one researcher (TL) used a unique co-expression pattern approach instead of clustering.

For niche annotation, UTAG spatial clustering was the predominant method, showing remarkable consistency across researchers. The differences emerged primarily in how researchers assigned biological meaning to these spatial clusters - most used a combination of spatial positioning and reference to anatomical images, with some incorporating prior anatomical knowledge (particularly of heart structures). Overall, researchers balanced computational methods with biological knowledge, using spatial context to refine their annotations.

We provide the original human annotations in Supplementary Fig. 27-32. For cell annotations, we visualize them within the same UMAP clustering structure, where each scientist's annotations represent their individual interpretations. For tissue niche annotations, we map them onto the spatial coordinates of each MERFISH spot.

3.6. Confusion matrix

We plot the confusion matrices for cell type annotations (Supplementary Fig.33–41) and tissue niche annotations (Supplementary Fig.42–50).

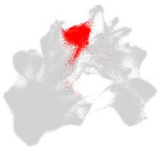
3.7. Summary

Accurate cell type and tissue niche annotation is key to interpreting spatial transcriptomics. With detailed comparisons, SpatialAgent outperforms existing methods in Tier-1 cell type annotation and aligns well with authors. We also improved UTAG clustering by removing small clusters to boost spatial coherence in niche annotation.

	Cell Annotation Workflow	**Niche Annotation Workflow**
HC	Combined annotation on Leiden clusters with CellTypist' transferred labels as reference	Used UTAG for spatial clustering
L	Performed Leiden clustering with majority voting for consensus-based cell type annotation. Provided 3 tiers of annotation	Labeled niches based on cell annotations with clear distributions (e.g., Atrium, Ventricular) for first tier; considered spatial left/right position for tier 2
Lh	Performed Leiden clustering with manual annotation using marker gene sets and DEG (per Scanpy tutorial). Projected cell types spatially. Used spatial position and key marker expression for final annotation. Provided 3-tier annotation	Used UTAG for spatial clustering. Labeled structures based on position and anatomical knowledge of heart (*e.g.*, "chamber wall is thicker on the left ventricle")
La	Performed Leiden clustering and analyzed expression of predefined marker genes in clusters. Used multiple genes per cell type. Mapped both major cell type and subtypes	Used UTAG for spatial clustering. Labeled structures based on position, provided anatomical image, and additional sources
MH	Performed Leiden clustering and analyzed predefined marker genes in clusters. Used (typically) single marker gene to differentiate cell types. Projected cell types spatially and used position for final annotation	Used UTAG for spatial clustering. Labeled structures based on position and provided anatomical image
TL	Annotated based on gene co-expression patterns	NA
Lu	Unknown	Unknown, likely misordered annotations
P	Performed label transfer using TACCO with scRNA-seq reference dataset of human heart. Projected cell types spatially. Performed Leiden clustering and DEG for marker genes, using key markers for second tier annotation	Used UTAG for spatial clustering. Labeled structures based on position and provided anatomical image

Supplementary Table 4: Comparison of cell and niche annotation methodologies

Cluster 0

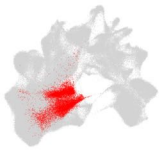


Top 20 DE Genes: POSTN, COL9A2, FBLN2, VCAN, FMOD, DPYSL3, INHBA, PENK, FN1, PDLIM3, SOX9, COL15A1, DKK3, CGNL1, NELL2, TPBG, PRRX1, MECOM, CRABP2, PDGFRA

GPTCellType: Cardiac Fibroblasts,0.9,The markers POSTN, COL9A2, FBLN2, VCAN, and PDGFRA are strongly associated with cardiac fibroblasts, which are involved in extracellular matrix production and tissue repair in the heart.

SpatialAgent: fibroblast of cardiac tissue; 0.9; The gene markers provided, such as POSTN, COL9A2, FBLN2, VCAN, and FN1, are strongly associated with fibroblasts, particularly those involved in the extracellular matrix and structural support, which are characteristic of cardiac tissue fibroblasts. The transferred reference cell type composition also supports this identification, with a 0.50 proportion of fibroblasts of cardiac tissue, which is the highest among the listed cell types. The presence of genes like SOX9 and PDGFRA further supports the fibroblast identity, as they are involved in mesenchymal and fibroblast cell functions.

Cluster 1

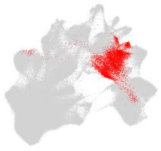


Top 20 DE Genes: CNN1, MYH7, TBX3, IRX4, TTN, CRABP2, XPO4, NKX2-5, BAMBI, LBH, IGFBP5, TNFRSF12A, MYH6, FZD1, OSR1, VCAN, DPYSL3, PPM1K, KCNH2, DES

GPTCellType: Cardiomyocytes,0.9,The presence of markers such as MYH7, MYH6, TTN, NKX2-5, and DES strongly indicates cardiomyocytes as these are key markers associated with cardiac muscle cells.

SpatialAgent: regular atrial cardiac myocyte; 0.9; The gene markers provided, such as MYH7, TTN, NKX2-5, and MYH6, are highly indicative of cardiac muscle cells, particularly those involved in contractile function. The presence of TBX3 and IRX4 further supports the identification of cardiac myocytes, as these genes are involved in cardiac development and function. The transferred reference cell type composition shows a high percentage (0.67) for regular atrial cardiac myocytes, which aligns with the gene expression profile. The presence of ventricular cardiac muscle cell markers (0.31) suggests a possible mixture, but the dominant expression of atrial markers and the high confidence from the reference data lead to a strong identification as regular atrial cardiac myocytes.

Cluster 2

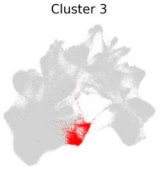


Top 20 DE Genes: FN1, NR2F2, NOTCH1, PDGFRB, MYH6, TCF21, FBLN2, COL15A1, DCN, MSC, APOE, FNDC1, CXCL12, IGFBP4, COL14A1, PIEZO2, JAG1, DPYSL3, TNC, MCAM

GPTCellType: Cardiomyocytes,0.6,The presence of MYH6 suggests cardiomyocytes, but markers like PDGFRB and TCF21 indicate potential involvement of fibroblasts or pericytes, suggesting a mixed cell type.

SpatialAgent: fibroblast of cardiac tissue; 0.9; The gene markers FN1, NR2F2, PDGFRB, TCF21, FBLN2, COL15A1, DCN, MSC, APOE, FNDC1, IGFBP4, COL14A1, and TNC are strongly associated with fibroblasts, particularly in cardiac tissue. The presence of MYH6, a cardiac muscle marker, suggests a cardiac-specific fibroblast. The transferred reference cell type composition also supports this, with a high percentage (0.43) of fibroblast of cardiac tissue. The confidence score is high due to the strong alignment of gene markers and reference data.

Cluster 3

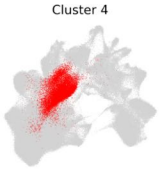


Top 20 DE Genes: IRX1, IRX2, IRX3, HCN4, IGFBP5, BMP2, BRINP3, RCAN1, RRAD, SCN5A, TBX3, DKK3, NKX2-5, DES, PRSS35, IRX4, GJA5, VCAN, TBX5, MMP11

GPTCellType: Cardiomyocytes,0.9,The presence of markers such as NKX2-5, SCN5A, and GJA5 strongly indicates cardiomyocytes, as these are key markers for cardiac muscle cells. Additionally, the presence of TBX3, TBX5, and DES further supports this identification, as they are associated with cardiac development and structure.

SpatialAgent: regular atrial cardiac myocyte; 0.9; The gene markers provided are highly indicative of cardiac myocytes, particularly those found in the atria. Key genes such as IRX1, IRX2, IRX3, HCN4, SCN5A, TBX3, NKX2-5, and TBX5 are known to be involved in cardiac development and function, specifically in the conduction system and atrial identity. The transferred reference cell type composition also supports this identification, with a high percentage (0.76) of regular atrial cardiac myocytes, which aligns with the gene expression profile. The presence of ventricular cardiac muscle cell markers is less pronounced, suggesting a primary identity as atrial myocytes.

Cluster 4

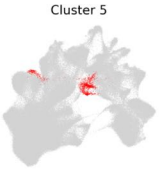


Top 20 DE Genes: TCF21, DCN, SCN7A, COL15A1, FN1, CD34, ALDH1A2, SPON2, HHIP, PDGFRA, FBLN2, GAS7, SOX9, OAF, COLEC11, SULF1, NKD2, PDGFRB, DPYSL3, POSTN

GPTCellType: Fibroblasts,0.9,The presence of markers such as TCF21, DCN, COL15A1, FN1, PDGFRA, and PDGFRB strongly suggests fibroblast identity, as these are commonly associated with cardiac fibroblasts involved in extracellular matrix production and tissue repair.

SpatialAgent: fibroblast of cardiac tissue; 0.9; The gene markers TCF21, DCN, COL15A1, FN1, CD34, ALDH1A2, PDGFRA, FBLN2, SOX9, and POSTN are strongly associated with fibroblasts, particularly in cardiac tissue. The presence of PDGFRB and GAS7 also supports this identification, as these genes are involved in fibroblast function and development. The transferred reference cell type composition indicates a 0.53 probability for fibroblast of cardiac tissue, which is significantly higher than the threshold of 0.5, further supporting this identification.

Cluster 5



Top 20 DE Genes: JAG1, MCAM, CXCL12, NOTCH1, MYH11, PDGFRB, FN1, FBLN5, NR2F2, IGFBP4, MECOM, GJA5, CPE, MSC, SULF1, CD34, CAV1, HEY2, KLF2, RGS5

GPTCellType: Pericytes,0.7,The markers JAG1, PDGFRB, MCAM, and CD34 are commonly associated with pericytes, which are cells that wrap around the endothelial cells of capillaries and venules. The presence of MYH11 and RGS5 also supports this identification, as they are often expressed in vascular smooth muscle cells and pericytes. However, the presence of other markers like NOTCH1 and FN1, which can be found in various cell types, slightly lowers the confidence.

SpatialAgent: smooth muscle cell; 0.9; The gene markers listed, such as MYH11, PDGFRB, and CAV1, are strongly associated with smooth muscle cells. MYH11 is a marker for smooth muscle myosin, PDGFRB is a receptor involved in the development of vascular smooth muscle cells, and CAV1 is involved in caveolae formation in smooth muscle cells. The transferred reference cell type composition also supports this identification, with a high percentage (0.58) of smooth muscle cells. Other markers like JAG1 and NOTCH1 are involved in signaling pathways that can be active in smooth muscle cells. The presence of genes like FN1 and NR2F2, which are involved in extracellular matrix interactions and vascular development, further supports the identification of this cluster as smooth muscle cells.

Supplementary Figure 23: Annotation of GPTCellType and SpatialAgent on Leiden clusters (1/4)

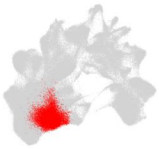
Cluster 6

Top 20 DE Genes: COL15A1, CD34, MCAM, NOTCH1, IGFBP4, CAV1, PDGFRB, ARHGAP18, PGF, RGS5, RBP1, KLF2, PLK2, PECAM1, ECE1, CLDN5, ARHGAP29, NOS3, NRP2, NR2F2

GPTCellType: Endothelial Cells,0.9,The markers such as PECAM1, NOS3, and CLDN5 are strongly associated with endothelial cells, indicating a high confidence in this identification.

SpatialAgent: endothelial cell; 0.9; The gene markers provided, such as PECAM1, NOS3, and CLDN5, are strongly associated with endothelial cells. These genes are involved in endothelial cell functions like cell adhesion, nitric oxide synthesis, and tight junction formation. The transferred reference cell type composition also supports this identification, with a significant percentage (0.59) of cardiac endothelial cells. The presence of other markers like CD34 and MCAM, which are also expressed in endothelial cells, further supports this classification. The confidence score is high due to the strong alignment of gene markers with endothelial cell characteristics and the supporting reference data.

Cluster 7



Top 20 DE Genes: IRX3, SCN5A, MYH7, IRX4, GJA1, RRAD, LBH, PPP1R12B, CASQ2, KCNH2, NKX2-5, TECRL, NAV1, HAND1, SLC1A3, GJA5, TTN, BRINP3, FZD1, DES

GPTCellType: Cardiomyocytes,0.9,The markers IRX3, SCN5A, MYH7, GJA1, NKX2-5, and TTN are strongly associated with cardiomyocytes, indicating a high confidence in identifying this cell type.

SpatialAgent: regular atrial cardiac myocyte; 0.9; The gene markers provided, such as IRX3, SCN5A, MYH7, GJA1, NKX2-5, and TTN, are strongly associated with cardiac myocytes, particularly those found in the atria. The transferred reference cell type composition indicates a high percentage (0.70) of regular atrial cardiac myocytes, which supports this identification. The presence of genes like GJA5 and HAND1 further supports the association with atrial myocytes. The confidence score is high due to the strong alignment of gene markers with known atrial cardiac myocyte markers and the high percentage in the reference composition.

Cluster 8

Top 20 DE Genes: DCN, TCF21, COL14A1, OSR1, FBLN2, FN1, FBLN5, POSTN, TBX18, IGFBP4, FLRT2, FNDC1, PDGFRA, DPYSL3, DLK1, MAF, HHIP, SPON2, PDGFRB, SOX9

GPTCellType: Cardiac Fibroblasts,0.9,The markers DCN, TCF21, COL14A1, and PDGFRA are strongly associated with cardiac fibroblasts, which are crucial for extracellular matrix production and structural support in the heart.

SpatialAgent: fibroblast of cardiac tissue; 0.9; The gene markers provided, such as DCN, TCF21, COL14A1, and POSTN, are strongly associated with fibroblasts, particularly those in cardiac tissue. The presence of PDGFRA and PDGFRB further supports this identification, as these genes are known to be involved in fibroblast signaling pathways. The transferred reference cell type composition also indicates a high percentage (0.74) of fibroblasts of cardiac tissue, which aligns with the gene expression profile, providing additional support for this classification.

Cluster 9

Top 20 DE Genes: PAM, MYH6, TBX5, COL2A1, NR2F1, DKK3, TECRL, HCN4, PRSS35, GJA5, RABGAP1L, ADAMTS8, ANGPT1, NR2F2, TTN, BTG1, MAF, ADM, C1QTNF3, JAG1

GPTCellType: Cardiomyocytes,0.7,The presence of markers such as MYH6, TTN, and GJA5 suggests a cardiomyocyte identity, as these are commonly associated with cardiac muscle cells. However, the presence of other markers like NR2F1 and NR2F2, which are not specific to cardiomyocytes, indicates potential involvement of other cell types or a mixed cell population.

SpatialAgent: regular atrial cardiac myocyte; 0.9; The gene markers MYH6, TTN, and GJA5 are strongly associated with cardiac myocytes, particularly atrial myocytes. The presence of TBX5 and NR2F2 further supports this identification, as these genes are involved in cardiac development and function. The transferred reference cell type composition also indicates a high percentage (0.70) of regular atrial cardiac myocytes, which aligns with the gene expression profile. The confidence score is high due to the strong alignment of gene markers with known atrial cardiac myocyte markers and the supporting reference data.

Cluster 10

Top 20 DE Genes: ECE1, FN1, POSTN, FOXC1, CGNL1, NPR3, SMOC1, NOS3, COLEC11, NOTCH1, LEPR, NR2F2, FMOD, IGFBP4, CD34, CD9, NRP2, MAF, MCAM, PECAM1

GPTCellType: Endothelial Cells,0.9,The markers such as NOS3, PECAM1, and CD34 are strongly associated with endothelial cells, which line the blood vessels in the heart.

SpatialAgent: cardiac endothelial cell; 0.95; The gene markers provided, such as PECAM1, NOS3, and NOTCH1, are strongly associated with endothelial cells, particularly in the cardiac context. The transferred reference cell type composition also supports this identification, with a high percentage (0.95) indicating cardiac endothelial cells. The presence of other markers like CD34 and NRP2 further corroborates the endothelial nature of the cells in this cluster.

Cluster 11

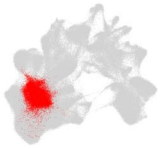
Top 20 DE Genes: ECE1, POSTN, CGNL1, NOTCH1, IGFBP4, FOXC1, NOS3, FN1, SMOC1, COLEC11, KLF2, NSG1, CD34, FMOD, CD9, FBLN2, APOE, NPR3, MECOM, NRP2

GPTCellType: Endothelial Cells,0.85,The presence of markers such as NOS3, CD34, and KLF2 strongly suggests endothelial cells, as these are commonly associated with vascular endothelium.

SpatialAgent: cardiac endothelial cell; 0.9; The gene markers ECE1, NOS3, and CD34 are strongly associated with endothelial cells, particularly in the cardiac context. The presence of these markers, along with the high transferred reference cell type composition score of 0.76 for cardiac endothelial cells, supports this identification. Other markers like KLF2 and FN1 also align with endothelial cell functions, further reinforcing this classification.

Supplementary Figure 24: Annotation of GPTCellType and SpatialAgent on Leiden clusters (2/4)

Cluster 12



Top 20 DE Genes: PPP1R12B, LBH, NAV1, DES, CASQ2, PLN, CKMT2, RYR2, PCDH7, TTN, MYH7, XPO4, KCNH2, DHRS3, FZD1, RRAD, LMOD3, GJA1, FGF12, IGFBP4

GPTCellType: Cardiomyocytes,0.9,The markers provided such as DES (desmin), CASQ2 (calsequestrin 2), PLN (phospholamban), RYR2 (ryanodine receptor 2), TTN (titin), MYH7 (myosin heavy chain 7), and GJA1 (connexin 43) are strongly associated with cardiomyocytes, which are the muscle cells of the heart responsible for contraction. These markers are indicative of the structural and functional properties of cardiomyocytes, supporting a high confidence in this identification.

SpatialAgent: regular atrial cardiac myocyte; 0.9; The gene markers provided, such as MYH7, TTN, RYR2, and PLN, are strongly associated with cardiac muscle cells, particularly myocytes. The transferred reference cell type composition indicates a high percentage (0.57) of regular atrial cardiac myocytes, which aligns with the gene expression profile. The presence of genes like GJA1 (connexin 43) and CASQ2, which are involved in cardiac muscle function, further supports this identification. The confidence score is high due to the strong alignment of gene markers with known cardiac myocyte markers and the significant percentage in the reference composition.

Cluster 13

Top 20 DE Genes: MKI67, TOP2A, KIF20A, KPNA2, PLK1, PTTG1, MYH7, ARL6IP1, MAZ, CKMT2, IRX4, FZD1, HEY2, RRM2, MYBL2, LBH, NKX2-5, CASQ2, GJA1, HAND1

GPTCellType: Cardiomyocytes,0.7,The presence of markers like MYH7, NKX2-5, CASQ2, GJA1, and HAND1 suggests cardiomyocytes, but the expression of proliferation markers such as MKI67, TOP2A, and PLK1 indicates a proliferative state, which is less typical for mature cardiomyocytes, suggesting a possible mixture or a less differentiated state.

SpatialAgent: regular atrial cardiac myocyte; 0.9; The gene markers listed, such as MYH7, NKX2-5, CASQ2, and GJA1, are strongly associated with cardiac muscle cells, particularly atrial myocytes. The presence of IRX4 and HEY2 further supports this identification, as they are involved in cardiac development and function. The transferred reference cell type composition indicates a high percentage (0.87) of regular atrial cardiac myocytes, which aligns with the gene expression profile. The presence of cell cycle-related genes like MKI67 and TOP2A suggests active proliferation, which can occur in cardiac myocytes under certain conditions. The confidence score is high due to the strong alignment of gene markers with atrial cardiac myocytes and the supporting reference data.

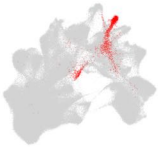
Cluster 14

Top 20 DE Genes: HEY2, MYH7, NAV1, RRAD, FZD1, GJA1, IRX4, CKMT2, LBH, HAND2, RYR2, DHRS3, TTN, PCDH7, SLC1A3, FGF12, LMOD3, KCNH2, CASQ2, PLN

GPTCellType: Cardiomyocytes,0.9,The markers MYH7, TTN, RYR2, and CASQ2 are strongly associated with cardiomyocytes, indicating a high confidence in identifying this cell type.

SpatialAgent: regular atrial cardiac myocyte; 0.9; The gene markers provided, such as MYH7, GJA1, TTN, and PLN, are strongly associated with cardiac muscle cells, particularly myocytes. The presence of genes like HEY2, HAND2, and IRX4 further supports the identification of cardiac myocytes, as these are involved in cardiac development and function. The transferred reference cell type composition indicates a high percentage (0.67) of regular atrial cardiac myocytes, which aligns with the gene expression profile. The presence of ventricular cardiac muscle cell markers is also noted, but the higher percentage of atrial myocytes in the reference data and the specific gene markers suggest a stronger association with atrial myocytes.

Cluster 15



Top 20 DE Genes: MAF, LYVE1, COTL1, IGFBP4, NRP2, NOTCH1, CD36, F13A1, NR2F1, KLF2, C1QB, OAF, RBP1, DCN, FLRT2, NR2F2, DLK1, PIEZO2, KPNA2, SOX9

GPTCellType: Endothelial Cells,0.7,The presence of markers like LYVE1, NRP2, and NOTCH1 suggests endothelial cell characteristics, although the presence of other markers indicates potential overlap with other cell types.

SpatialAgent: macrophage; 0.8; The gene markers such as LYVE1, COTL1, CD36, and C1QB are commonly associated with macrophages. The presence of NR2F2 and SOX9, which are involved in macrophage differentiation and function, further supports this identification. The transferred reference cell type composition also indicates a high percentage (0.43) of macrophages, which aligns with the gene expression profile. Although cardiac endothelial cells are also present in the reference composition, the gene markers are more indicative of macrophage identity.

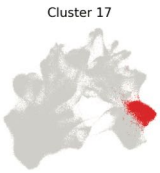
Cluster 16

Top 20 DE Genes: SHOX2, HCN4, MYH6, TBX5, TENM2, BMP2, VSNL1, IGFBP5, JAG1, DKK3, ISL1, TBX18, TRPM3, NR2F2, MYH11, CXCL12, TBX3, NR2F1, CPE, SFRP1

GPTCellType: Pacemaker cells,0.9,The presence of markers such as SHOX2, HCN4, and TBX3, which are associated with pacemaker cells, suggests that this cluster is likely composed of pacemaker cells. These markers are crucial for the development and function of the sinoatrial node, the heart's natural pacemaker.

SpatialAgent: cardiac muscle cell; 0.9; The gene markers provided, such as SHOX2, HCN4, MYH6, TBX5, and ISL1, are strongly associated with cardiac muscle cells, particularly those involved in the conduction system of the heart, like pacemaker cells. The presence of MYH6 and TBX5 further supports the identification as cardiac muscle cells, as these are key markers for cardiac myocytes. The transferred reference cell type composition also indicates a high percentage of regular atrial cardiac myocytes and ventricular cardiac muscle cells, which aligns with the gene expression profile. The confidence score is high due to the strong correlation between the gene markers and known cardiac muscle cell markers, despite the presence of some genes that could be associated with other cell types.

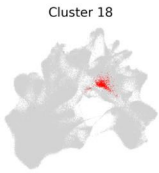
Supplementary Figure 25: Annotation of GPTCellType and SpatialAgent on Leiden clusters (3/4)



Top 20 DE Genes: COL2A1, MYH6, ADAMTS8, PAM, GJA5, TBX5, NR2F1, TTN, DAPK2, TRPM3, NR2F2, SFRP1, PPM1K, BTG1, HCN4, JAG1, OSR1, TECRL, NRXN1, HAND2

GPTCellType: Cardiomyocytes,0.7,The presence of MYH6 and TTN suggests cardiomyocytes as these are key markers for cardiac muscle cells; however, the presence of other markers like COL2A1 and GJA5 indicates potential involvement of other cell types or developmental stages.

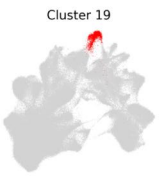
SpatialAgent: regular atrial cardiac myocyte; 0.9; The gene markers provided, such as MYH6, TTN, and HCN4, are strongly associated with cardiac muscle cells, particularly atrial myocytes. The presence of TBX5 and NR2F2 further supports this, as these genes are involved in cardiac development and function. The transferred reference cell type composition also indicates a high percentage (0.53) of regular atrial cardiac myocytes, which aligns with the gene expression profile. Although the transferred information should be cautiously considered, the high percentage and the gene markers strongly suggest this cell type.



Top 20 DE Genes: MPZ, TBX3, TRPM3, OAF, GAS7, MCAM, MOXD1, COL2A1, NRXN1, FLRT1, DPYSL3, NOTCH1, RBP1, POSTN, SLITRK6, JAG1, COL9A2, NR2F2, SEMA6D, COTL1

GPTCellType: Cardiac Fibroblasts,0.7,The presence of markers such as POSTN, COL2A1, and COL9A2 suggests a fibroblast identity, as these are associated with extracellular matrix production and remodeling, which are characteristic of cardiac fibroblasts.

SpatialAgent: neuron; 0.8; The gene markers listed, such as NRXN1, NOTCH1, and JAG1, are associated with neuronal functions and development. The presence of these genes suggests a neuronal identity. Additionally, the transferred reference cell type composition indicates a high percentage (0.54) of neurons, which supports this identification. However, the presence of cardiac-related genes like POSTN and COL2A1 suggests some overlap with cardiac cell types, but the neuronal markers are more prominent.

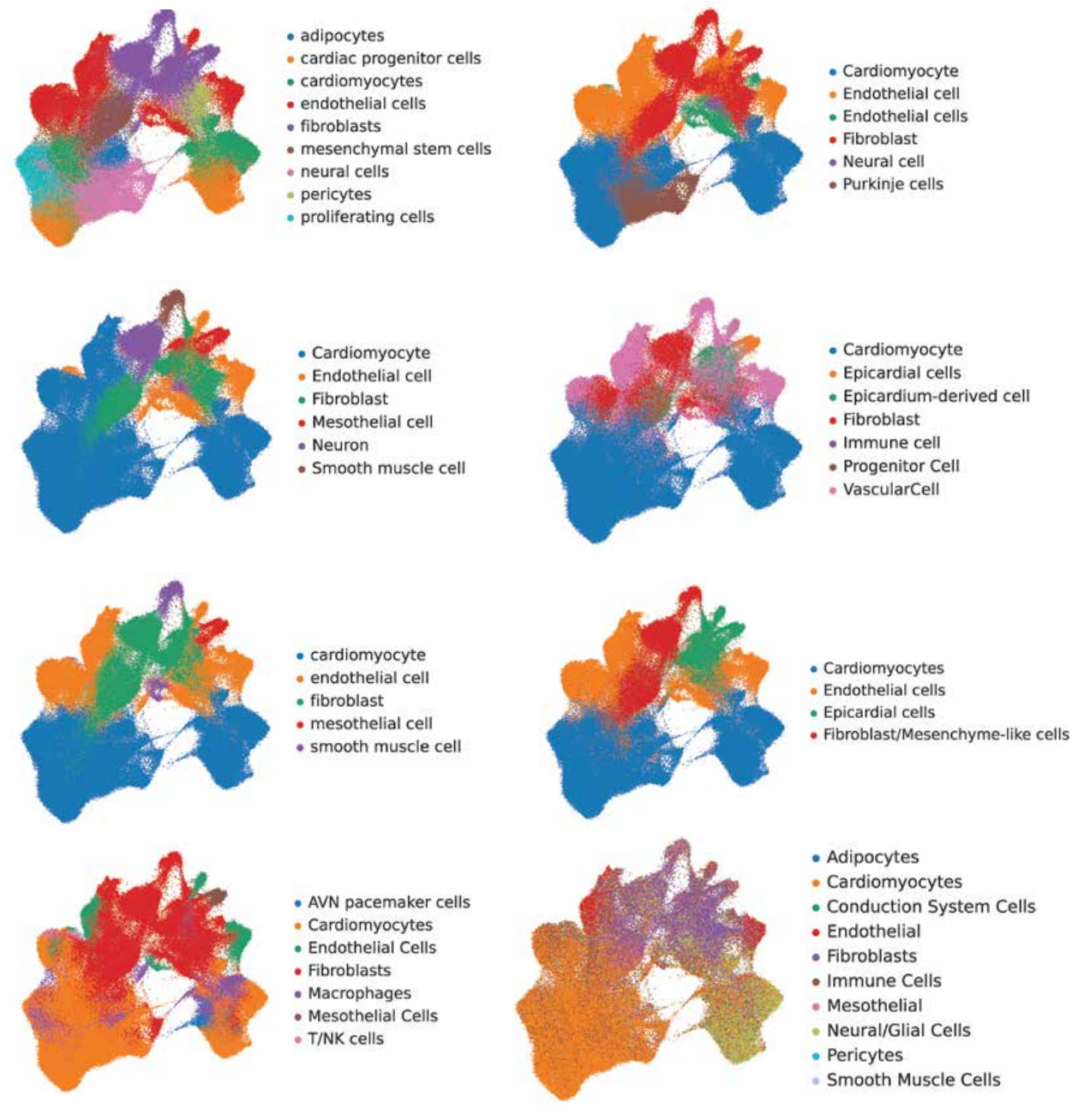


Top 20 DE Genes: FBLN5, MYH11, CRABP2, OSR1, CXCL12, FN1, PRSS35, JAG1, PLK2, COL14A1, FOXC1, COL9A2, INHBA, CGNL1, TRPM3, TNNT1, FNDC1, SFRP1, TBX18, PDGFRB

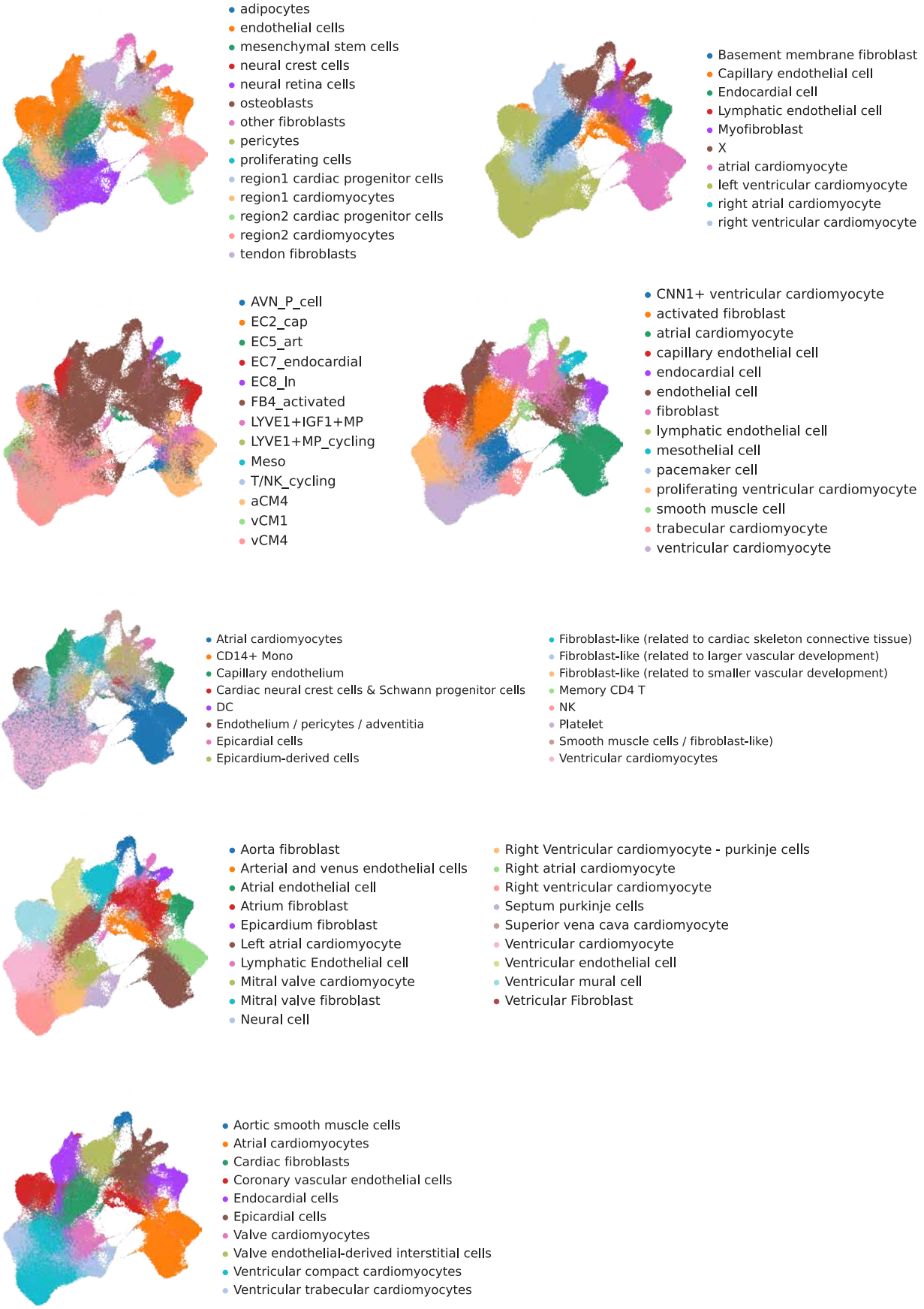
GPTCellType: Fibroblast-like cells,0.7,The presence of markers such as FBLN5, FN1, and COL14A1 suggests a fibroblast-like phenotype, while the presence of MYH11 and PDGFRB indicates some smooth muscle cell characteristics. The combination of these markers suggests a mixed cell type with fibroblast-like properties.

SpatialAgent: smooth muscle cell; 0.9; The gene markers MYH11, FBLN5, and PDGFRB are strongly associated with smooth muscle cells, which are known for their contractile function and presence in the heart. The transferred reference cell type composition also supports this identification, with a significant percentage (0.37) of smooth muscle cells. Other markers like FN1 and JAG1 are also expressed in smooth muscle cells, further supporting this classification. The confidence score is high due to the strong alignment of gene markers with the smooth muscle cell profile and the supporting reference data.

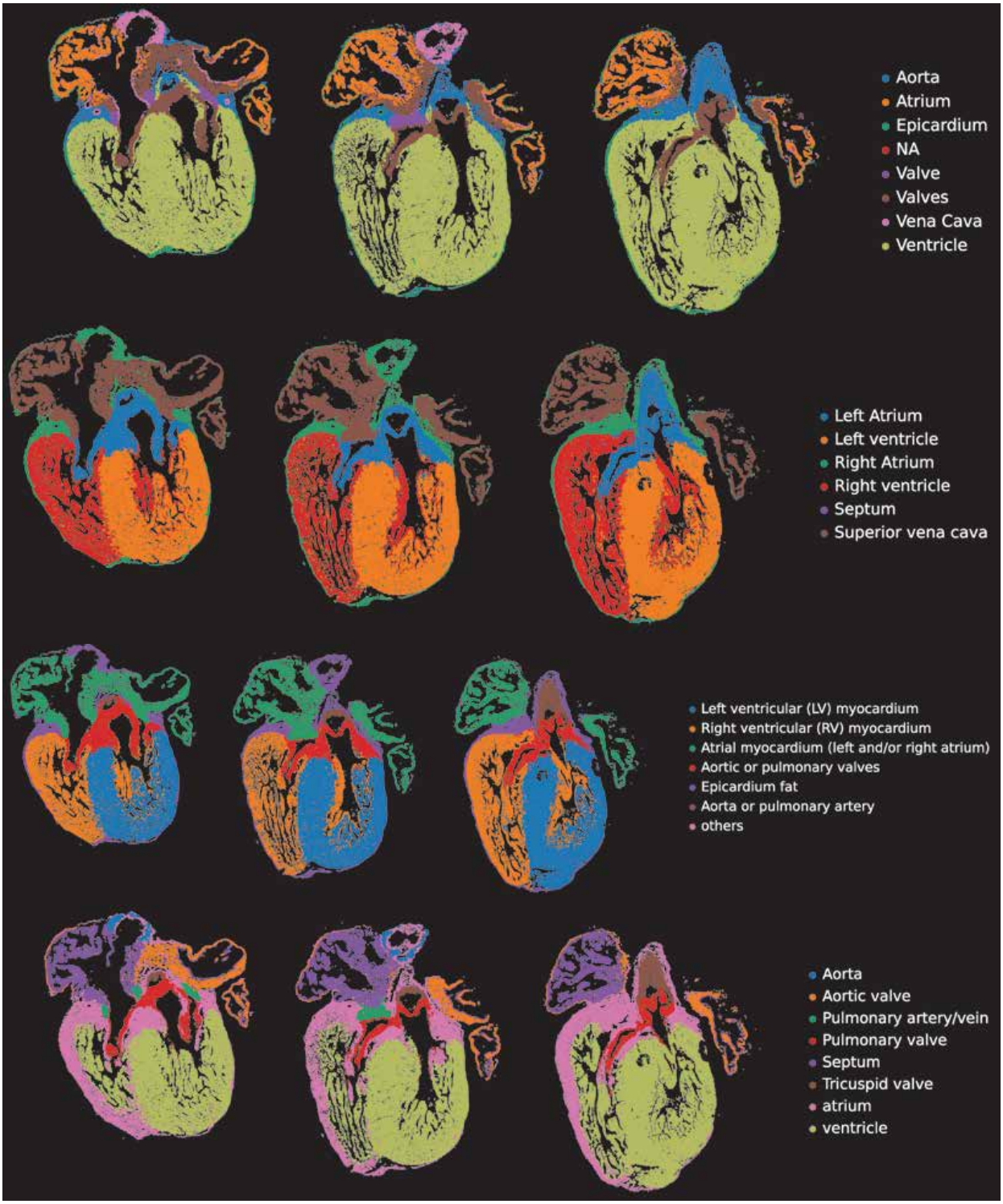
Supplementary Figure 26: Annotation of GPTCellType and SpatialAgent on Leiden clusters (4/4)



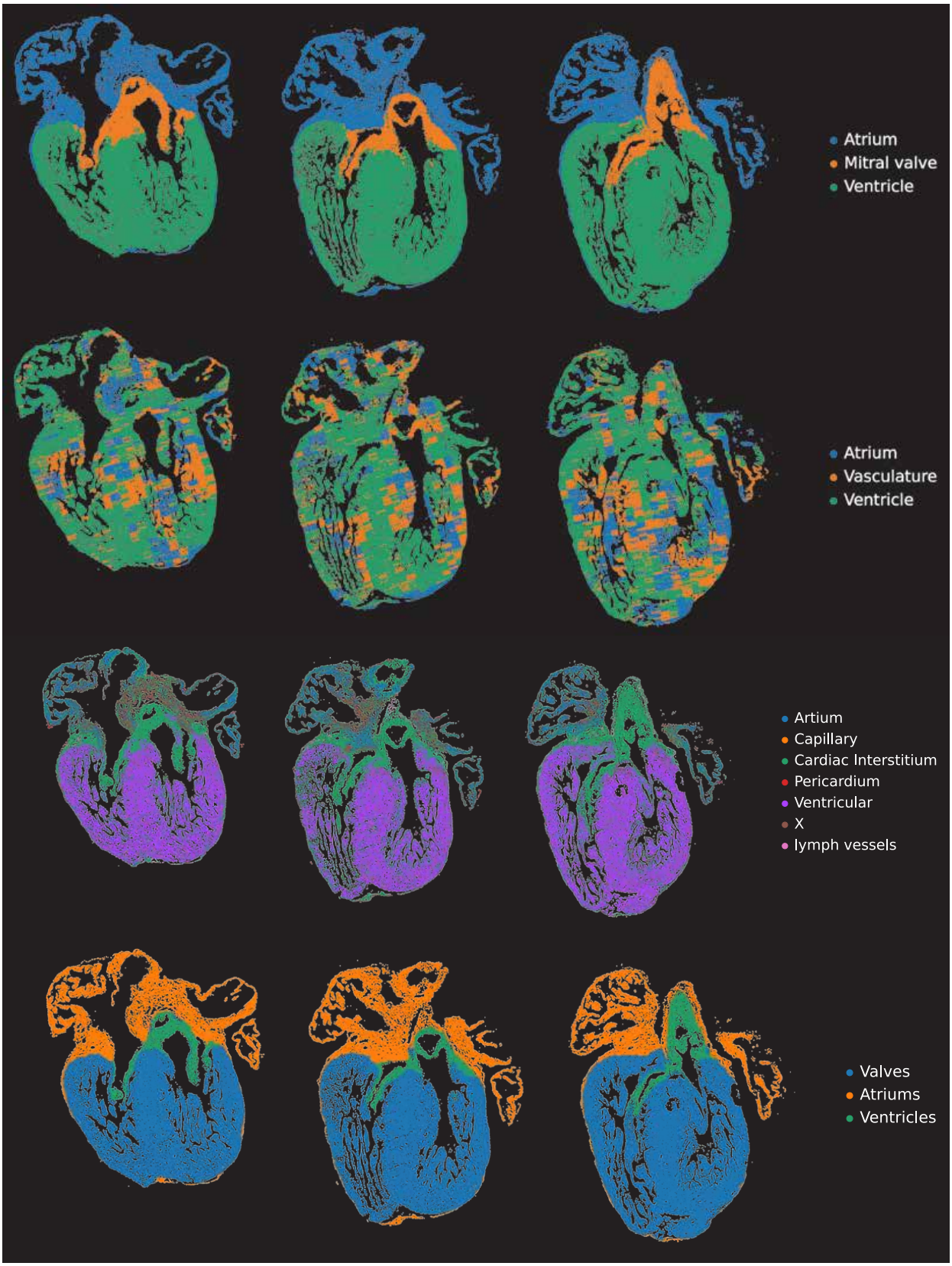
Supplementary Figure 27: Original Tier 1 cell type annotations.



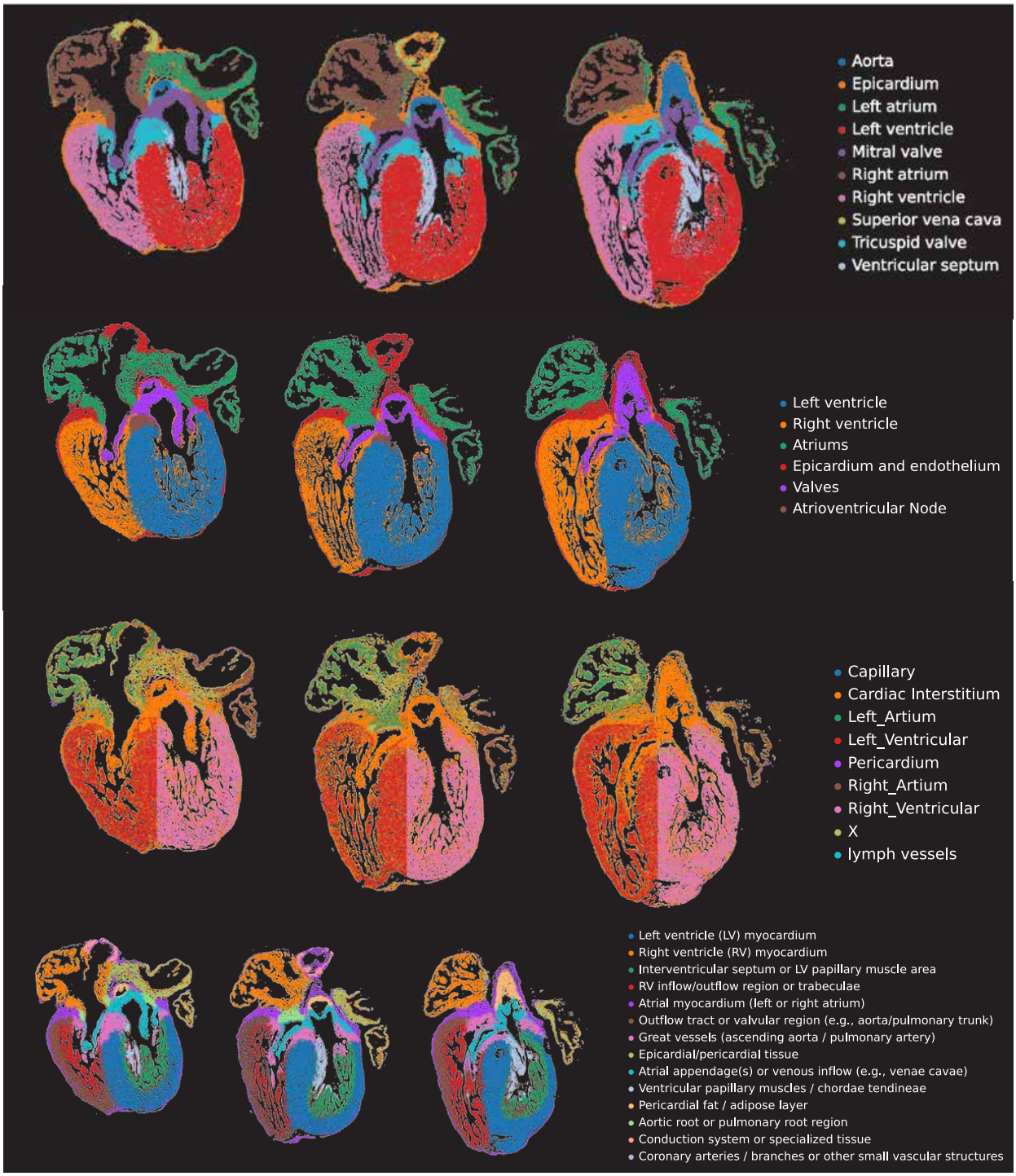
Supplementary Figure 28: Original Tier 2 cell type annotations.



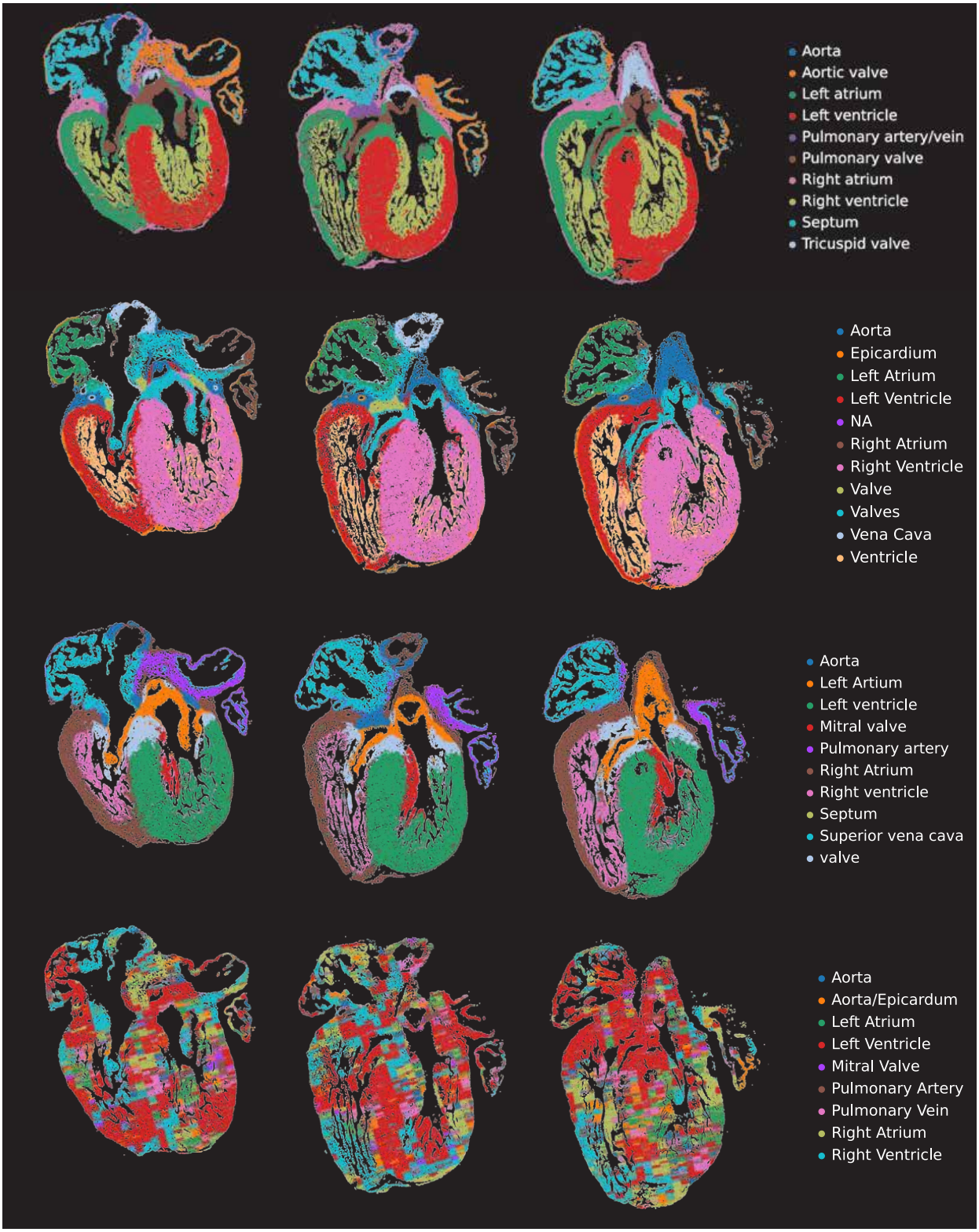
Supplementary Figure 29: Original Tier 1 tissue niche annotations (1/2).



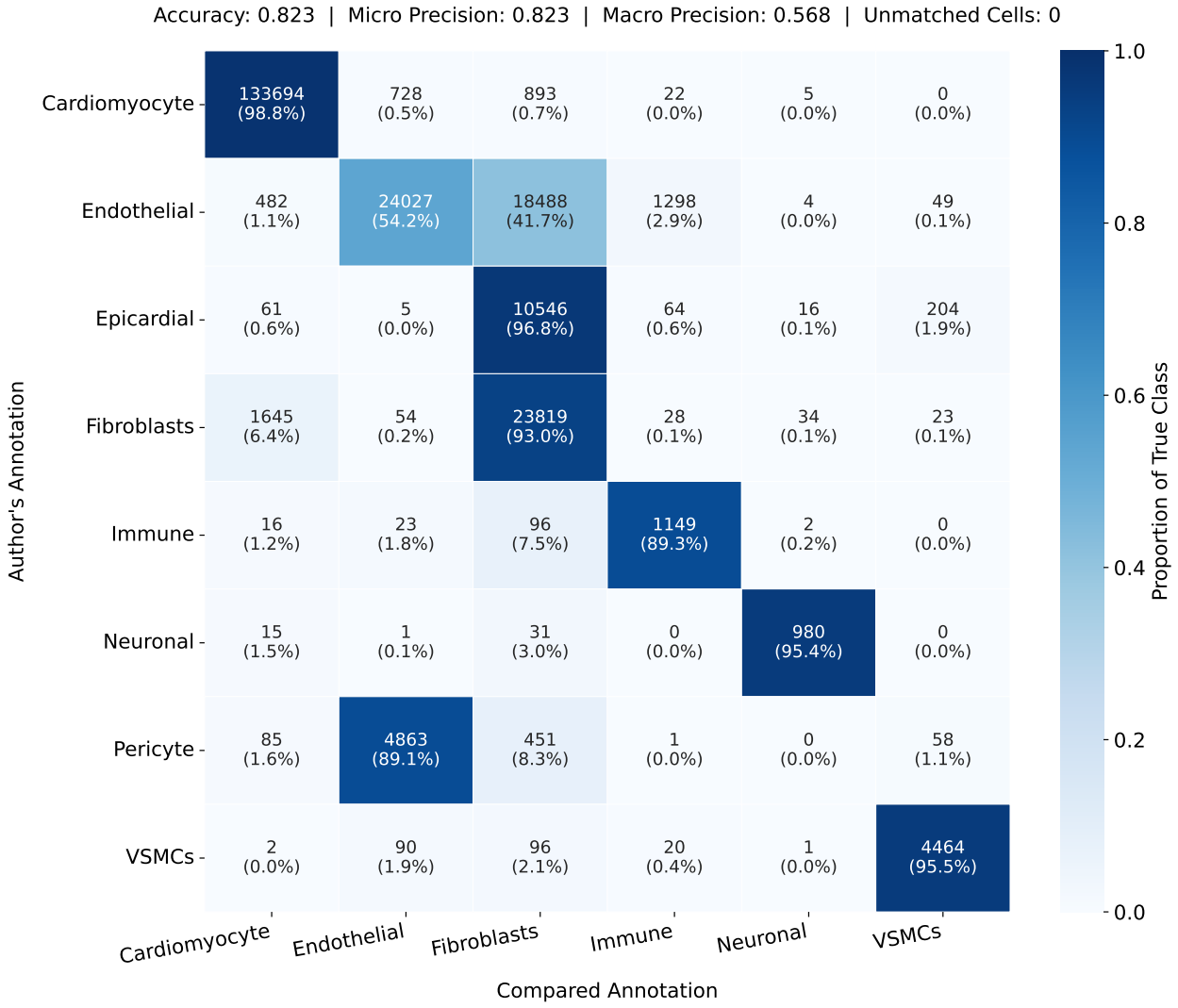
Supplementary Figure 30: Original Tier 1 tissue niche annotations (2/2).



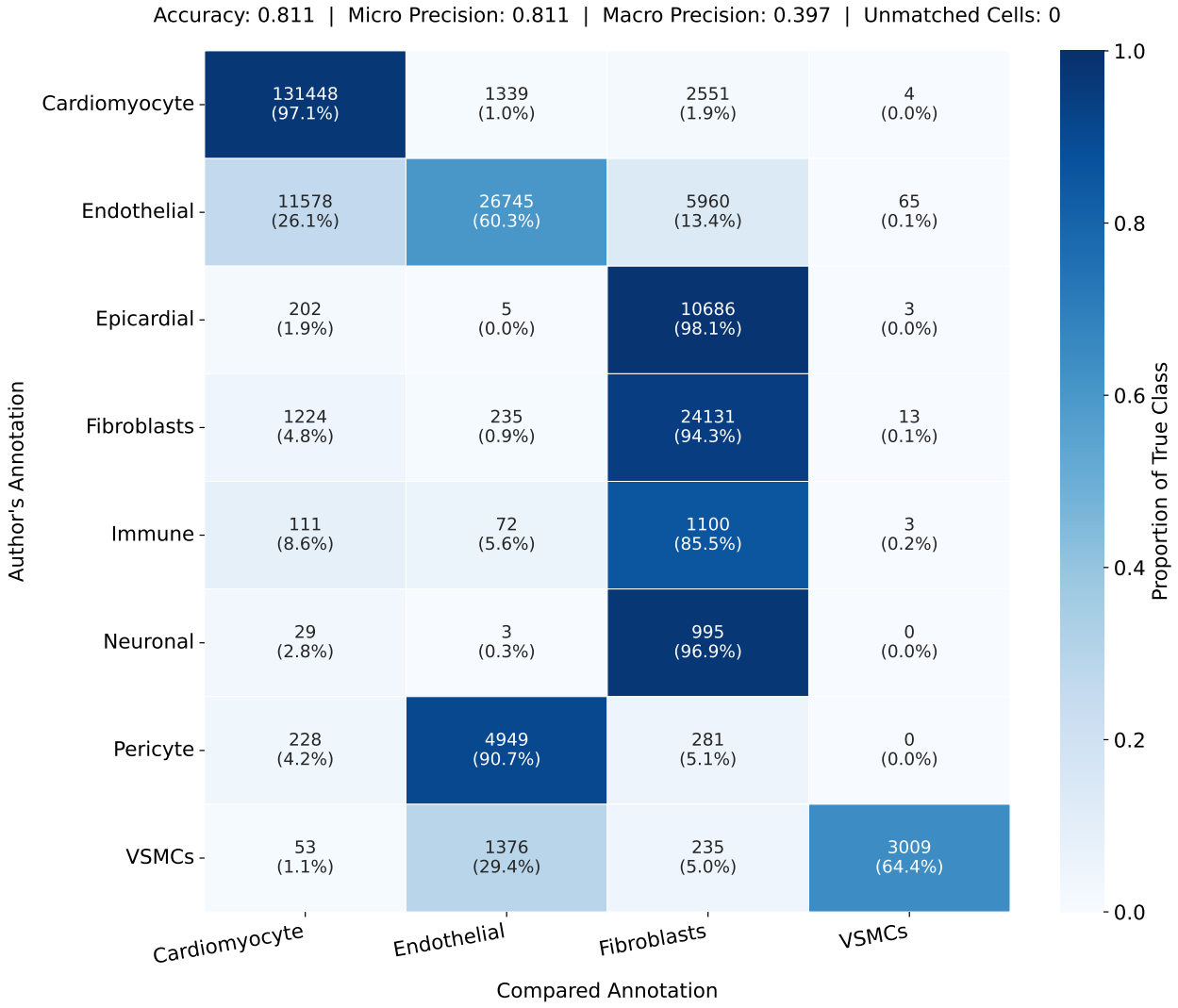
Supplementary Figure 31: Original Tier 2 tissue niche annotations (1/2).



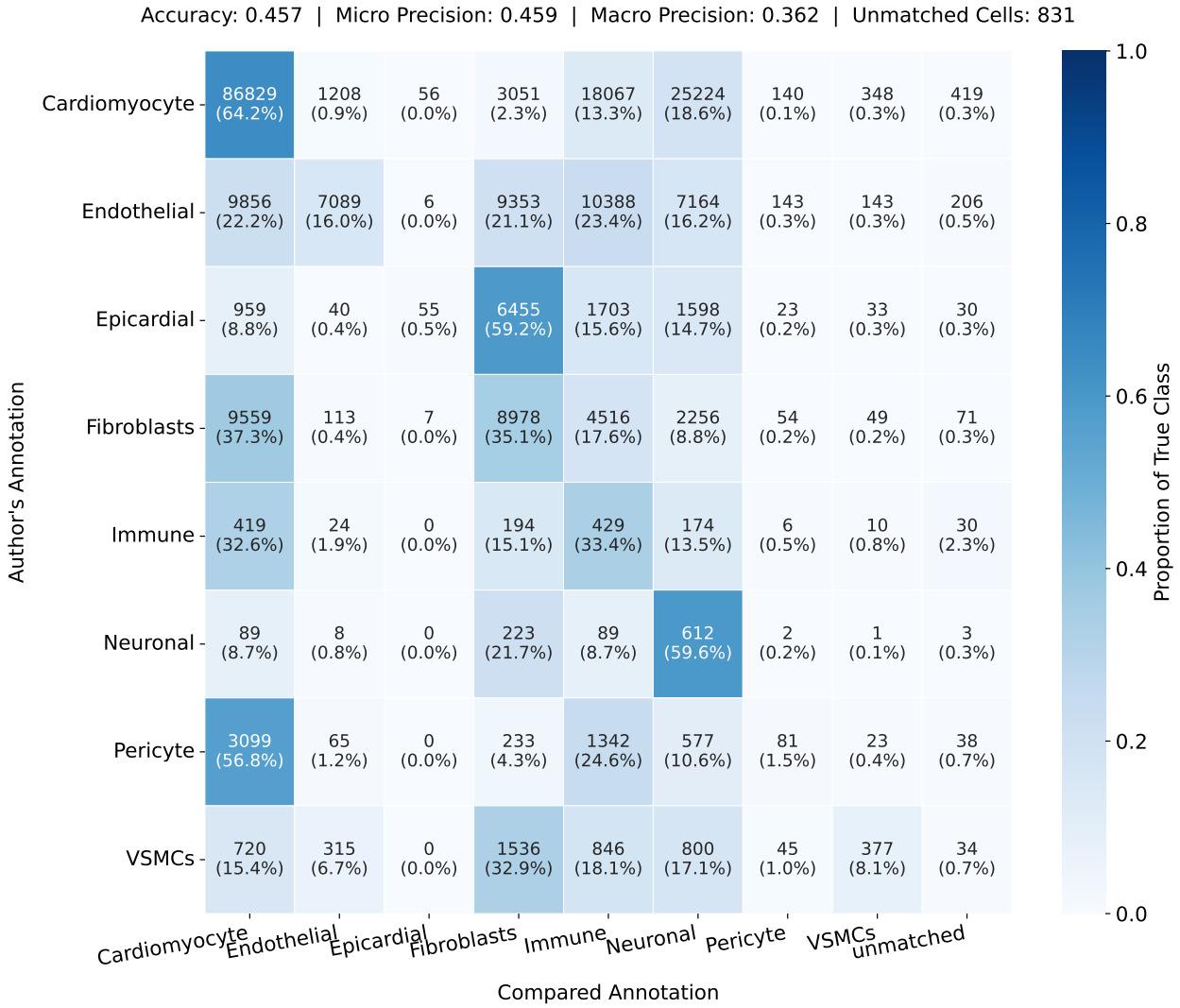
Supplementary Figure 32: Original Tier 2 tissue niche annotations (2/2).



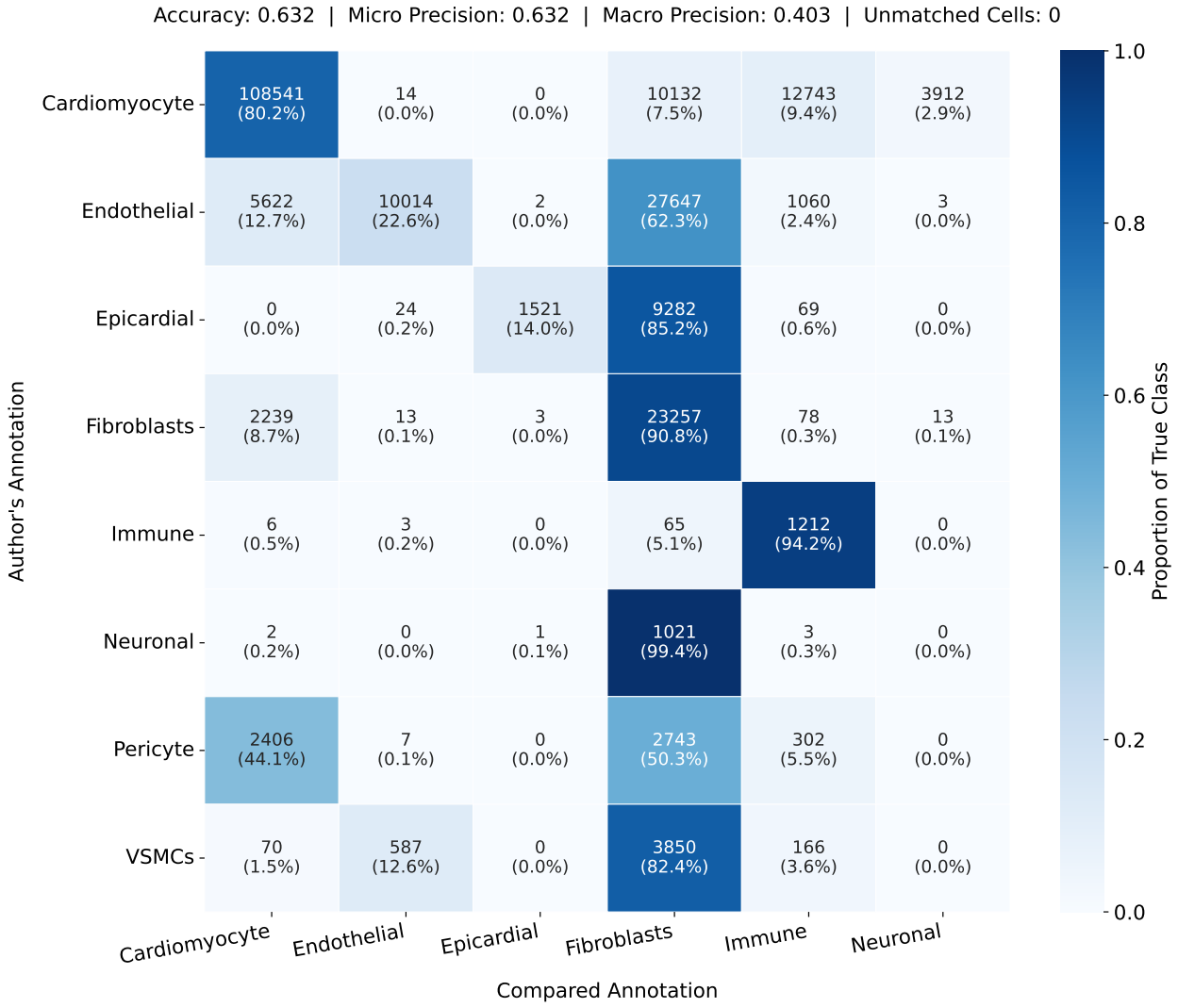
Supplementary Figure 33: Confusion matrix of SpatialAgent cell type annotation.



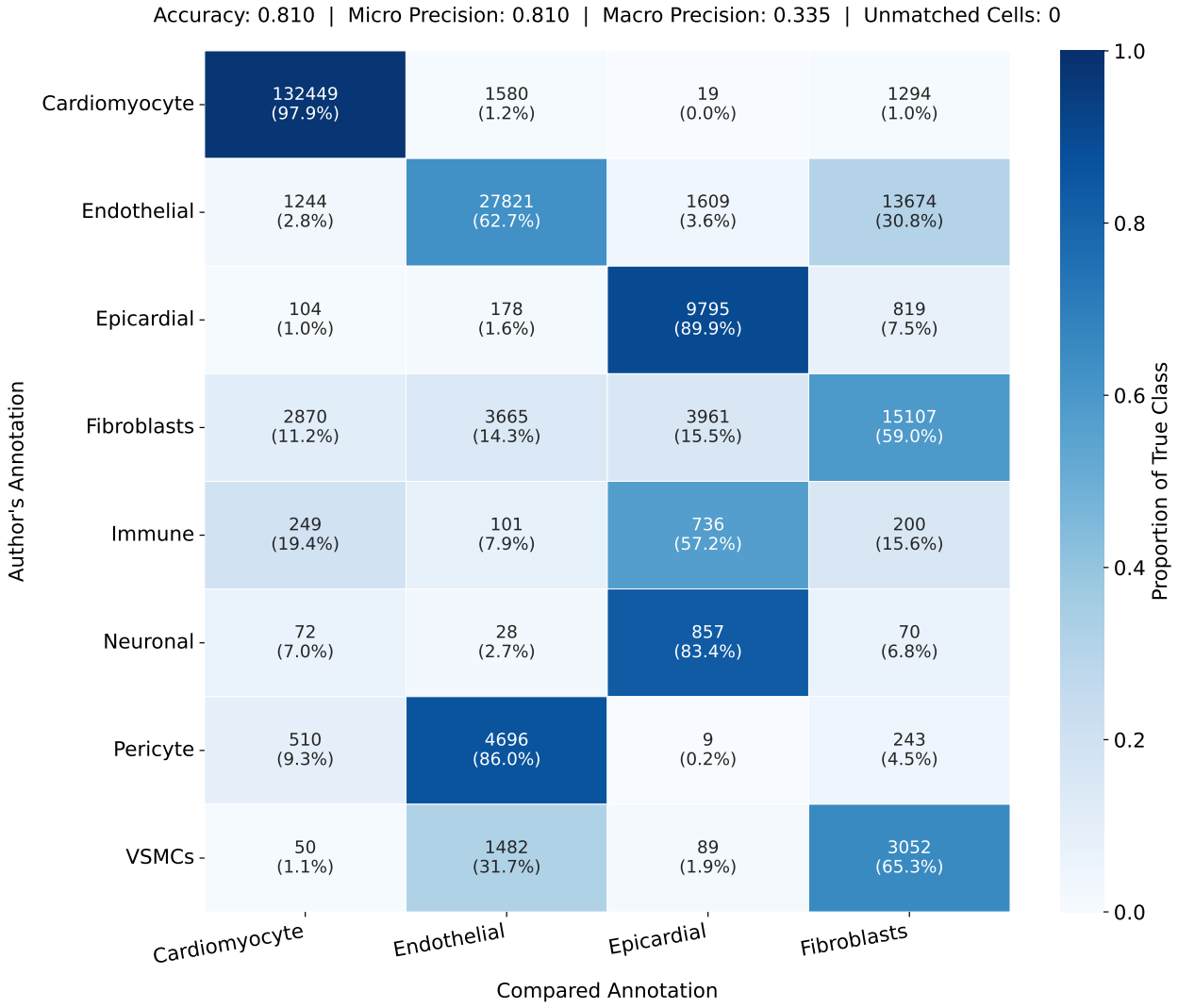
Supplementary Figure 34: Confusion matrix of GPTCellType cell type annotation.



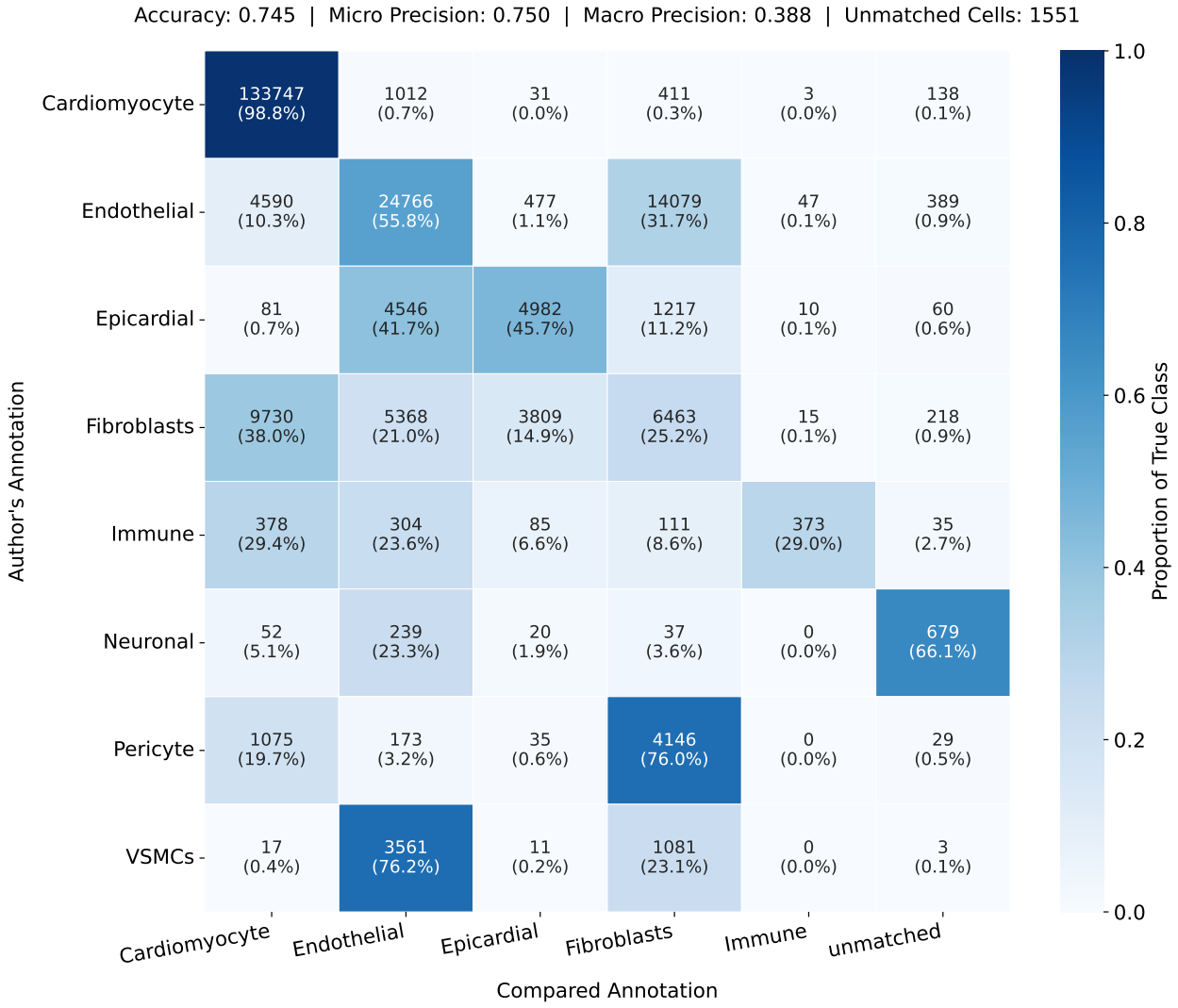
Supplementary Figure 35: Confusion matrix of scientist *HC* (also CellTypist) cell type annotation.



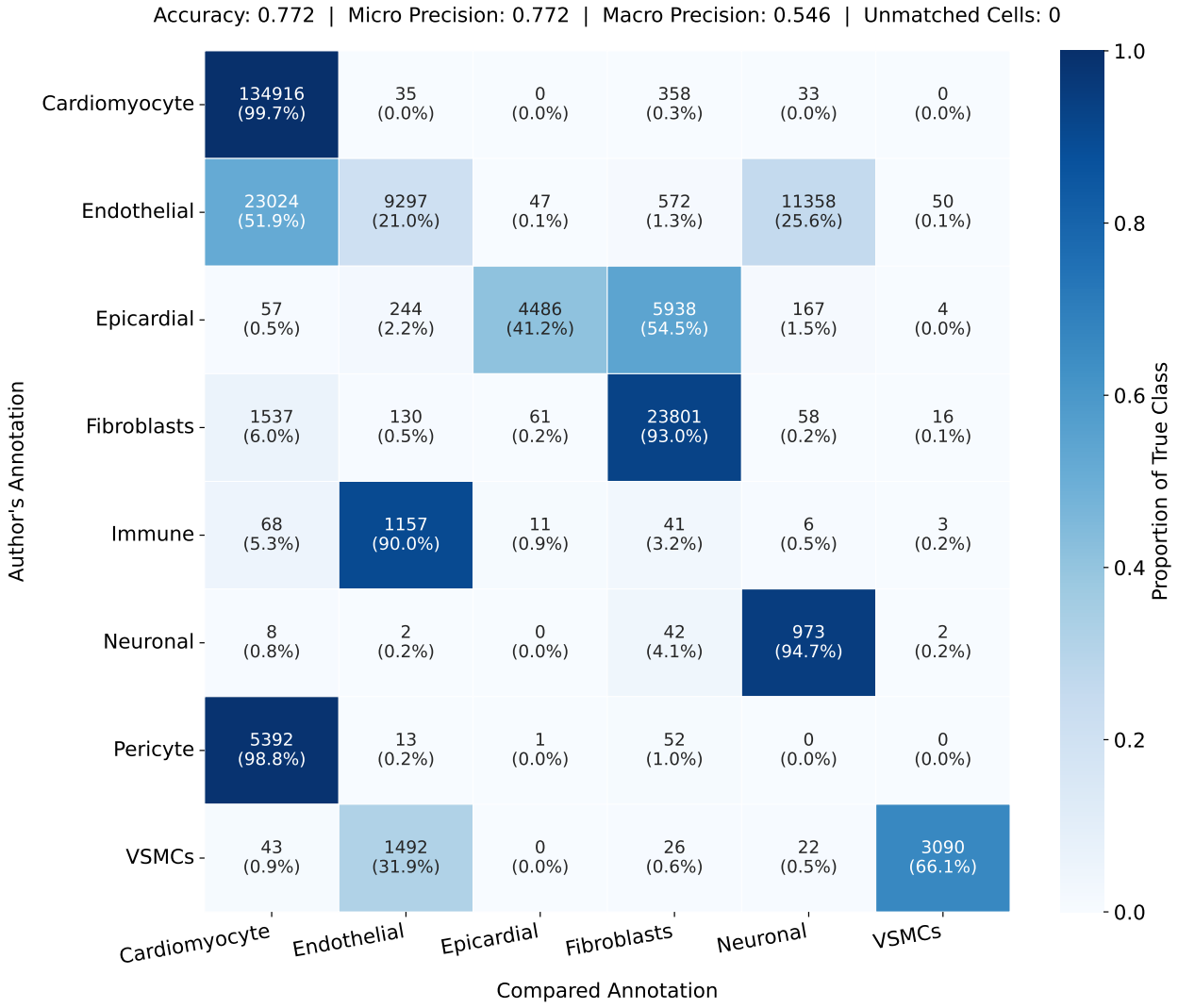
Supplementary Figure 36: Confusion matrix of human scientist *L* cell type annotation.



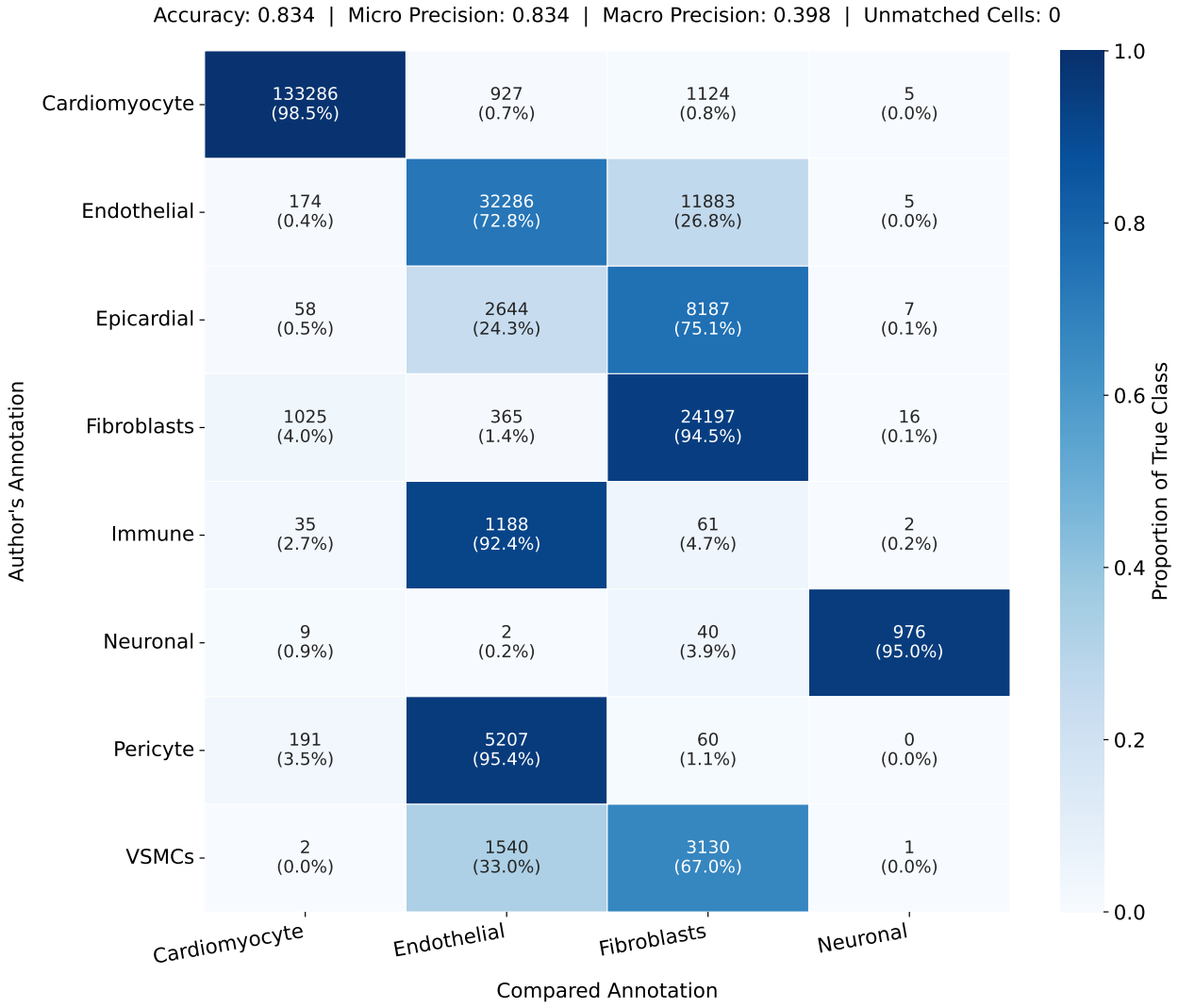
Supplementary Figure 37: Confusion matrix of human scientist *Lh* cell type annotation.



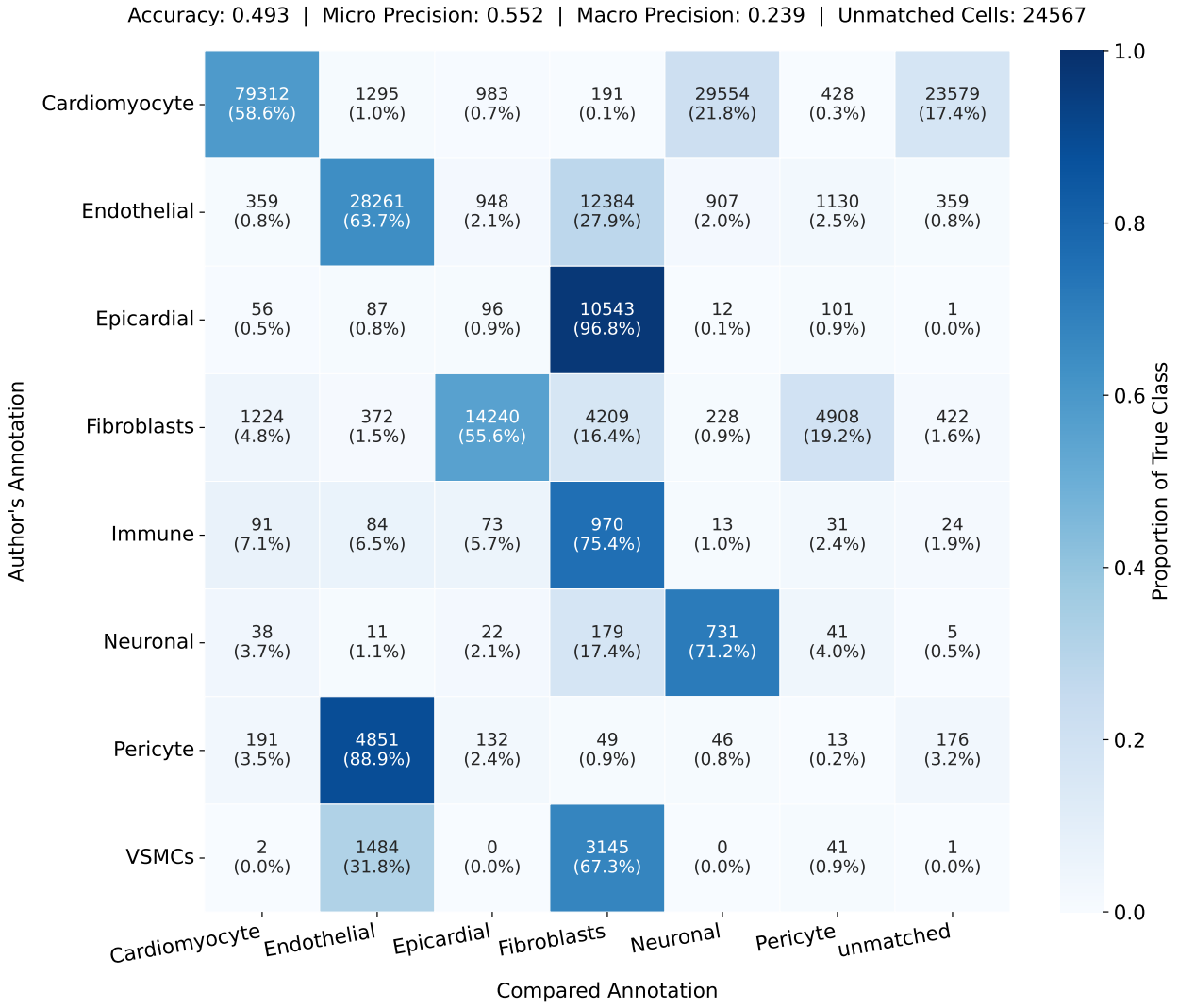
Supplementary Figure 38: Confusion matrix of human scientist *Lu* cell type annotation.



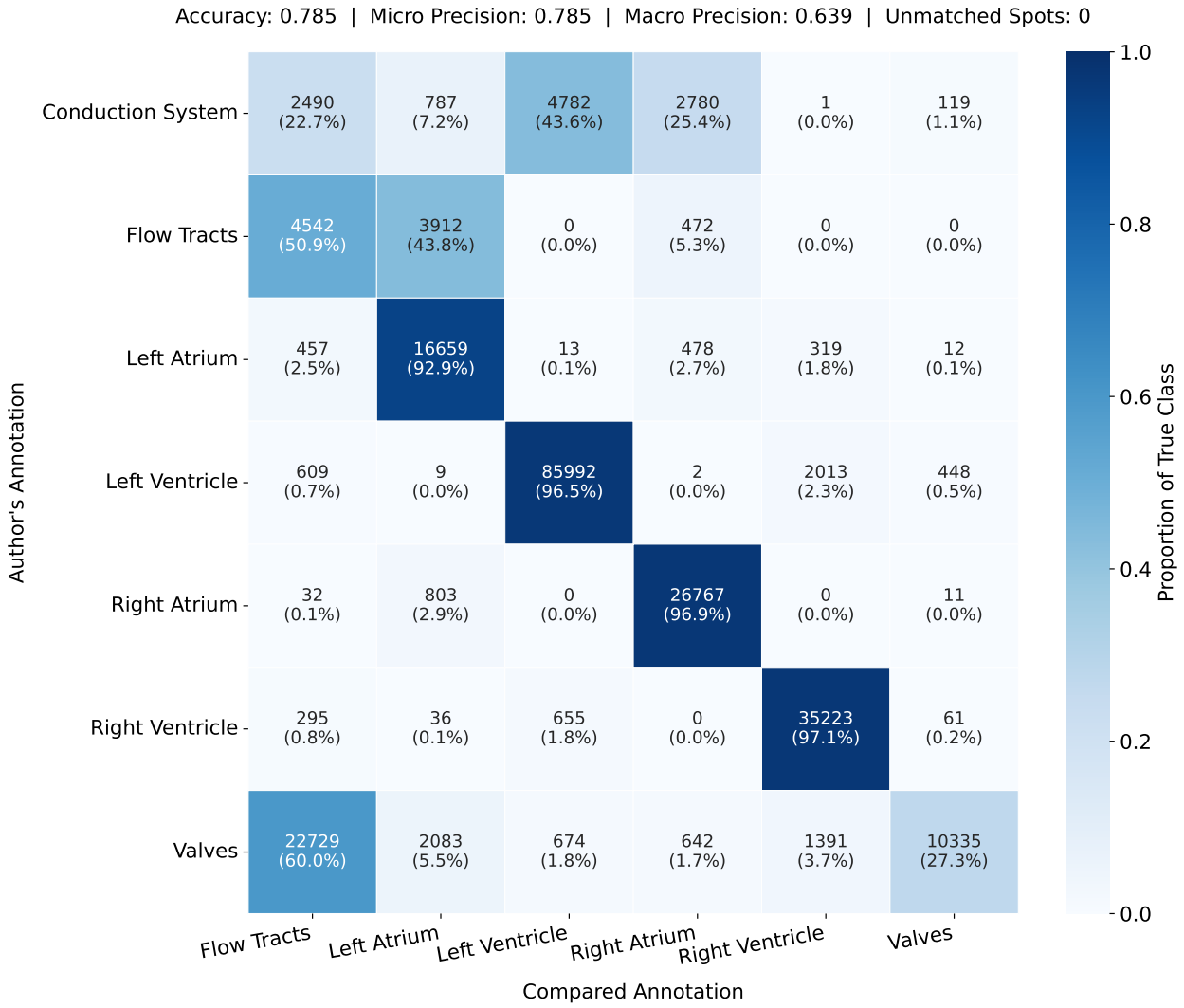
Supplementary Figure 39: Confusion matrix of human scientist *MH* cell type annotation.



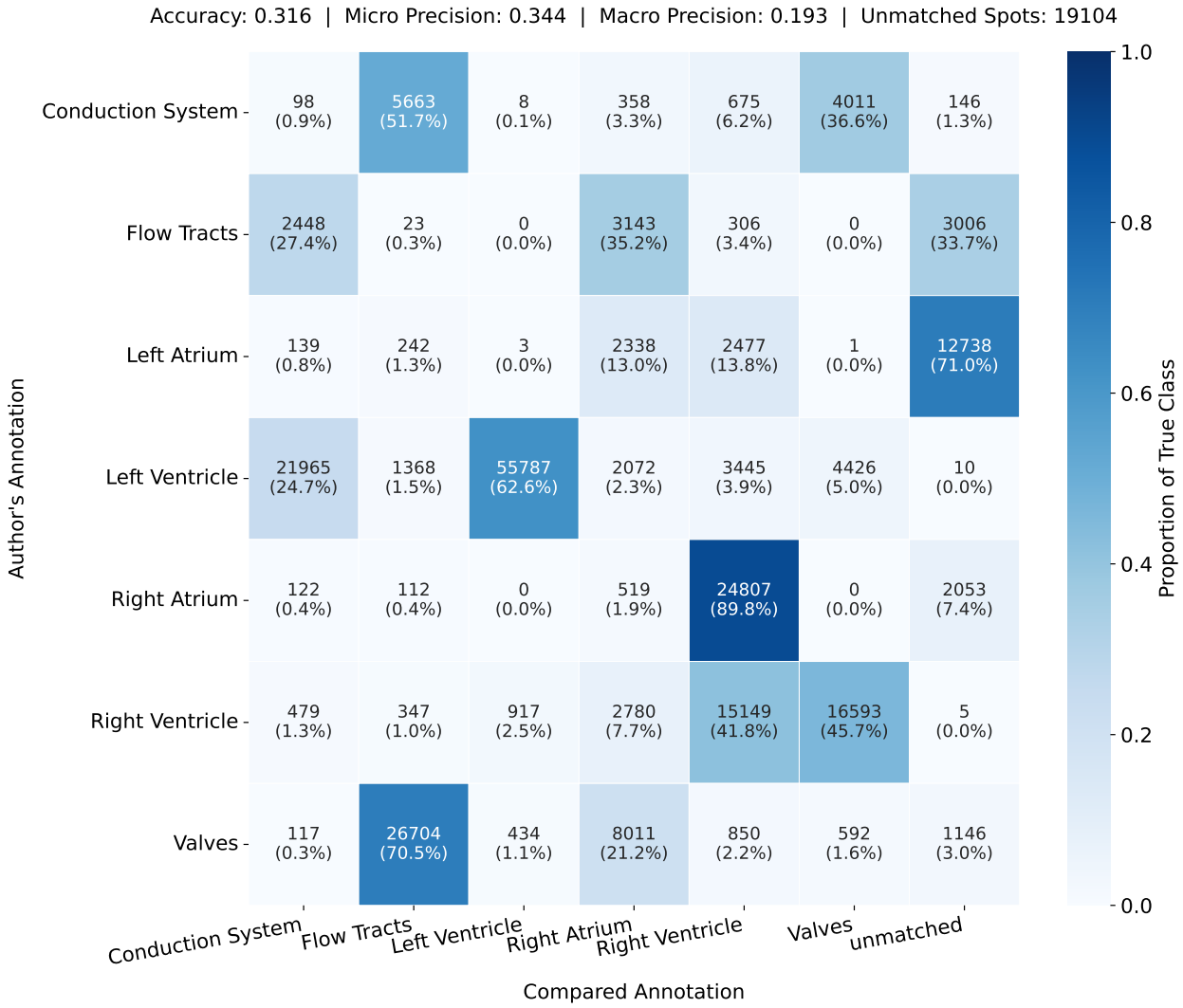
Supplementary Figure 40: Confusion matrix of human scientist *P* cell type annotation.



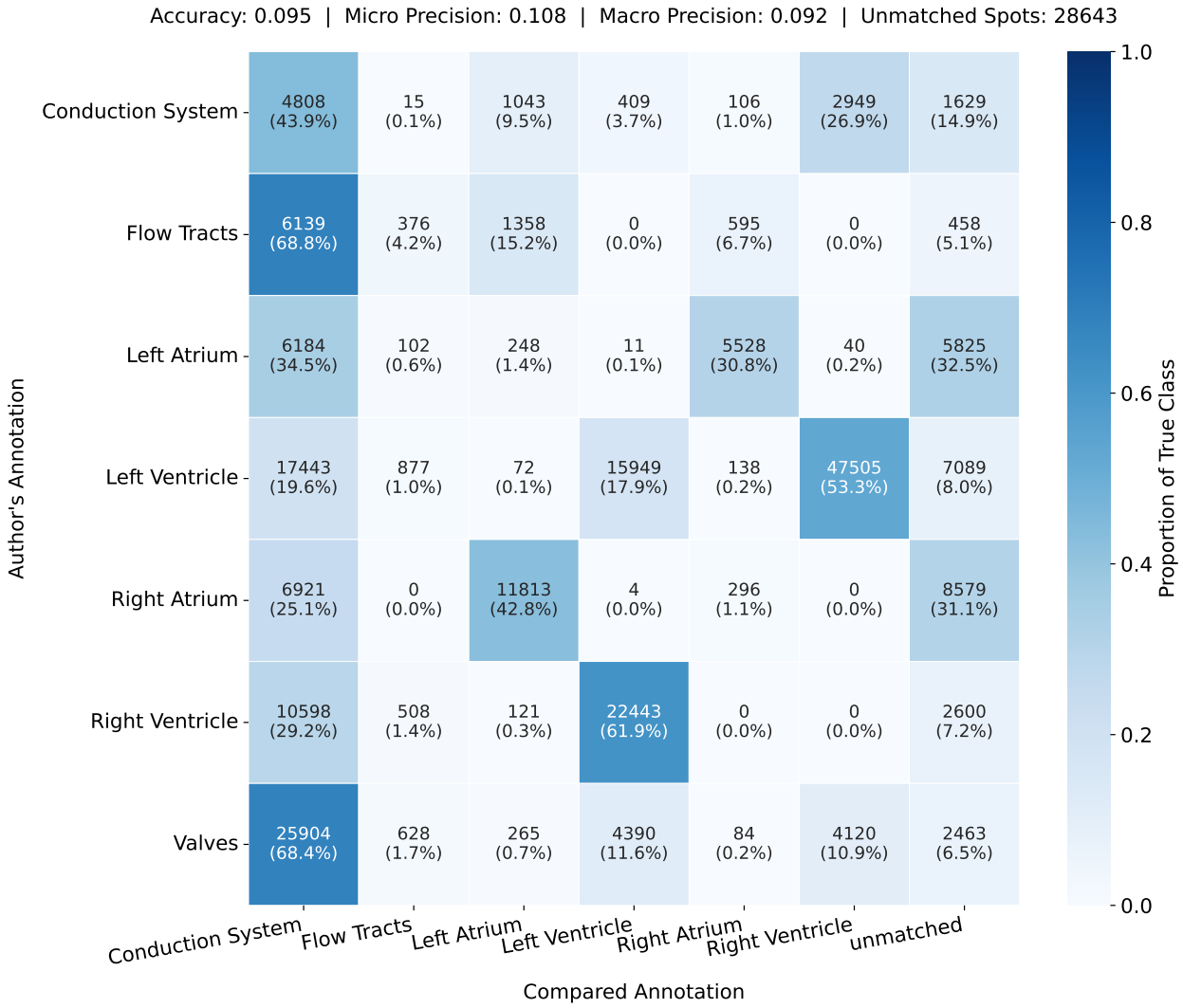
Supplementary Figure 41: Confusion matrix of human scientist *TL* tissue niche annotation.



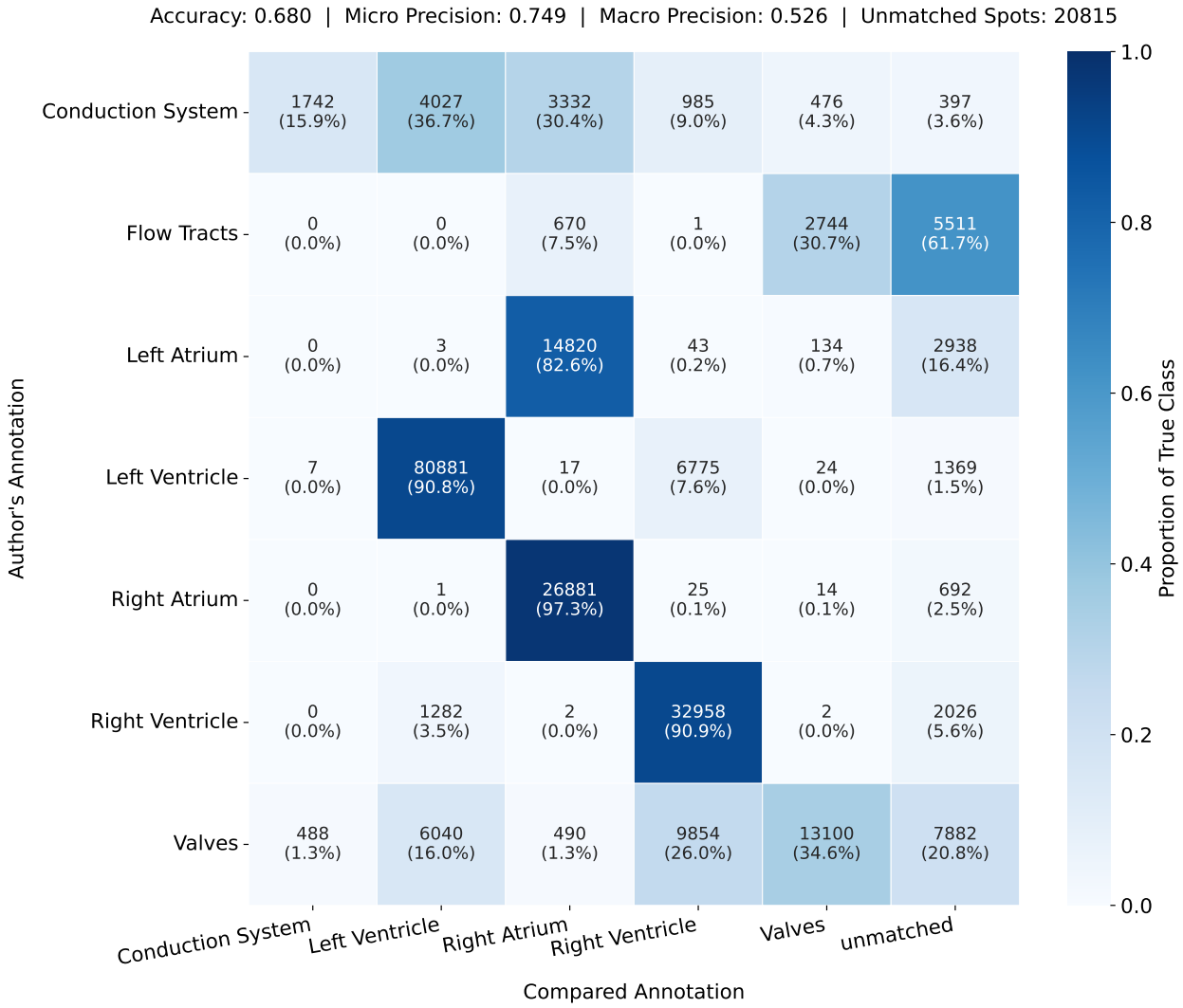
Supplementary Figure 42: Confusion matrix of SpatialAgent tissue niche annotation.



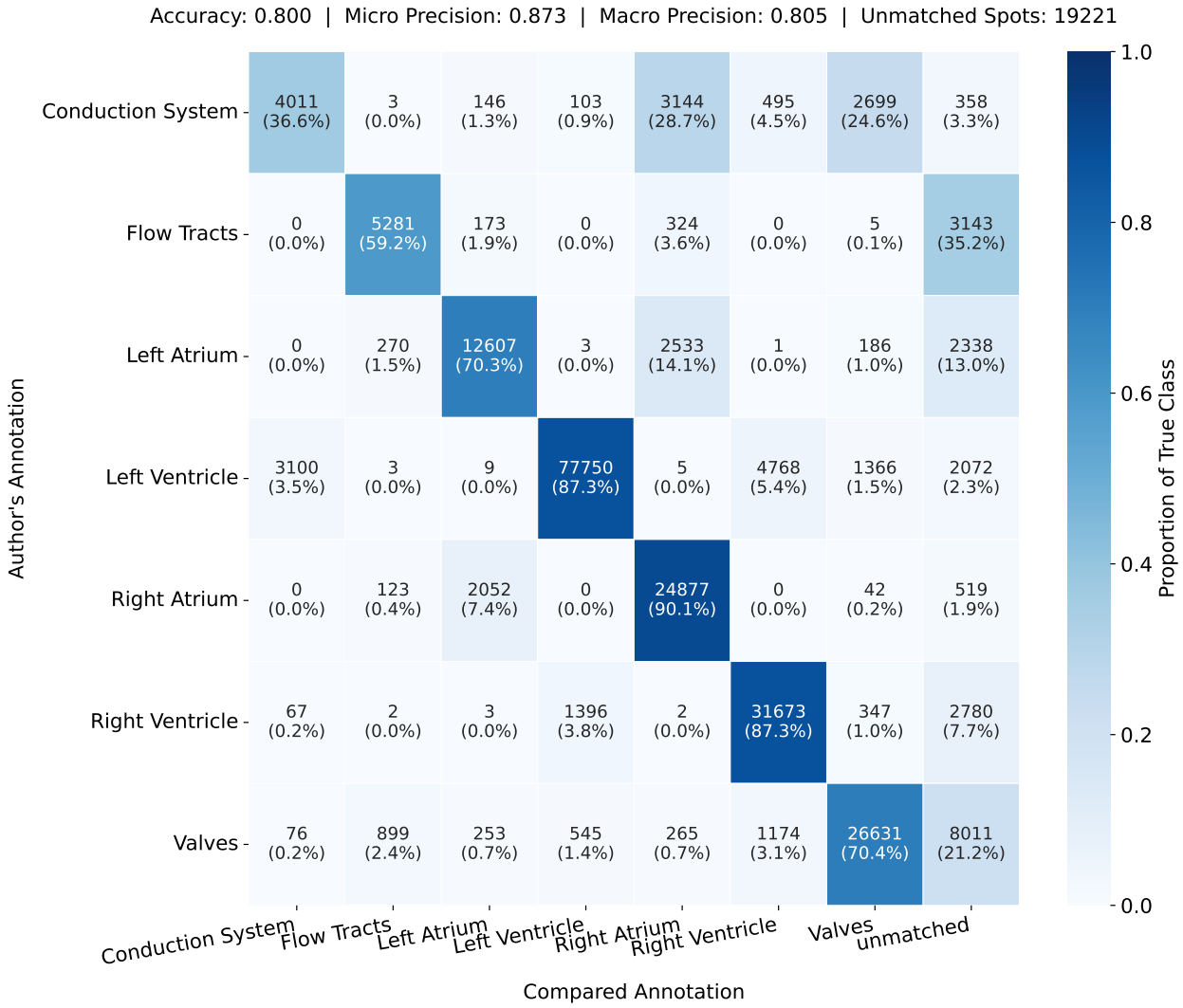
Supplementary Figure 43: Confusion matrix of human scientist *HC* tissue niche annotation.



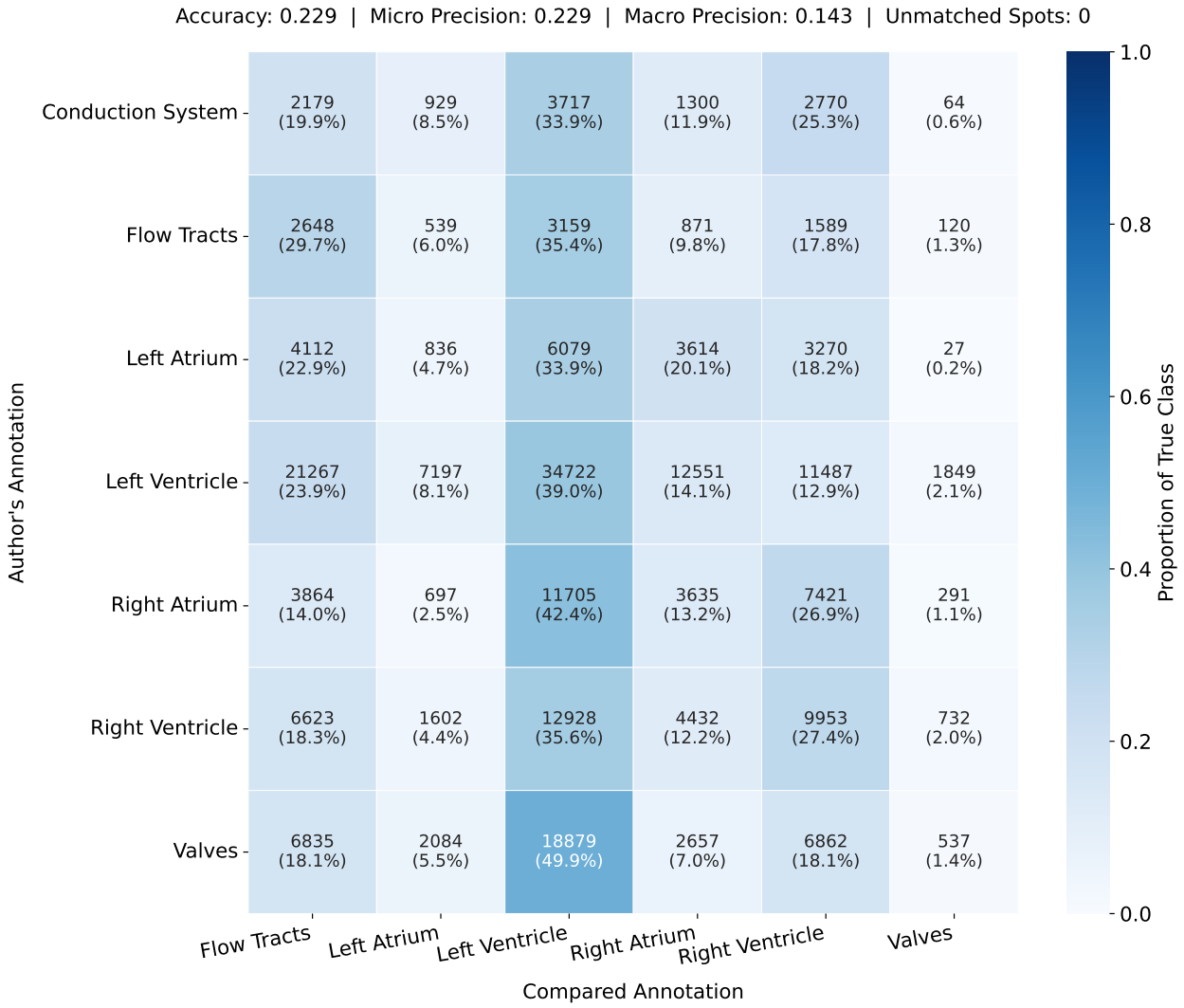
Supplementary Figure 44: Confusion matrix of human scientist L tissue niche annotation.



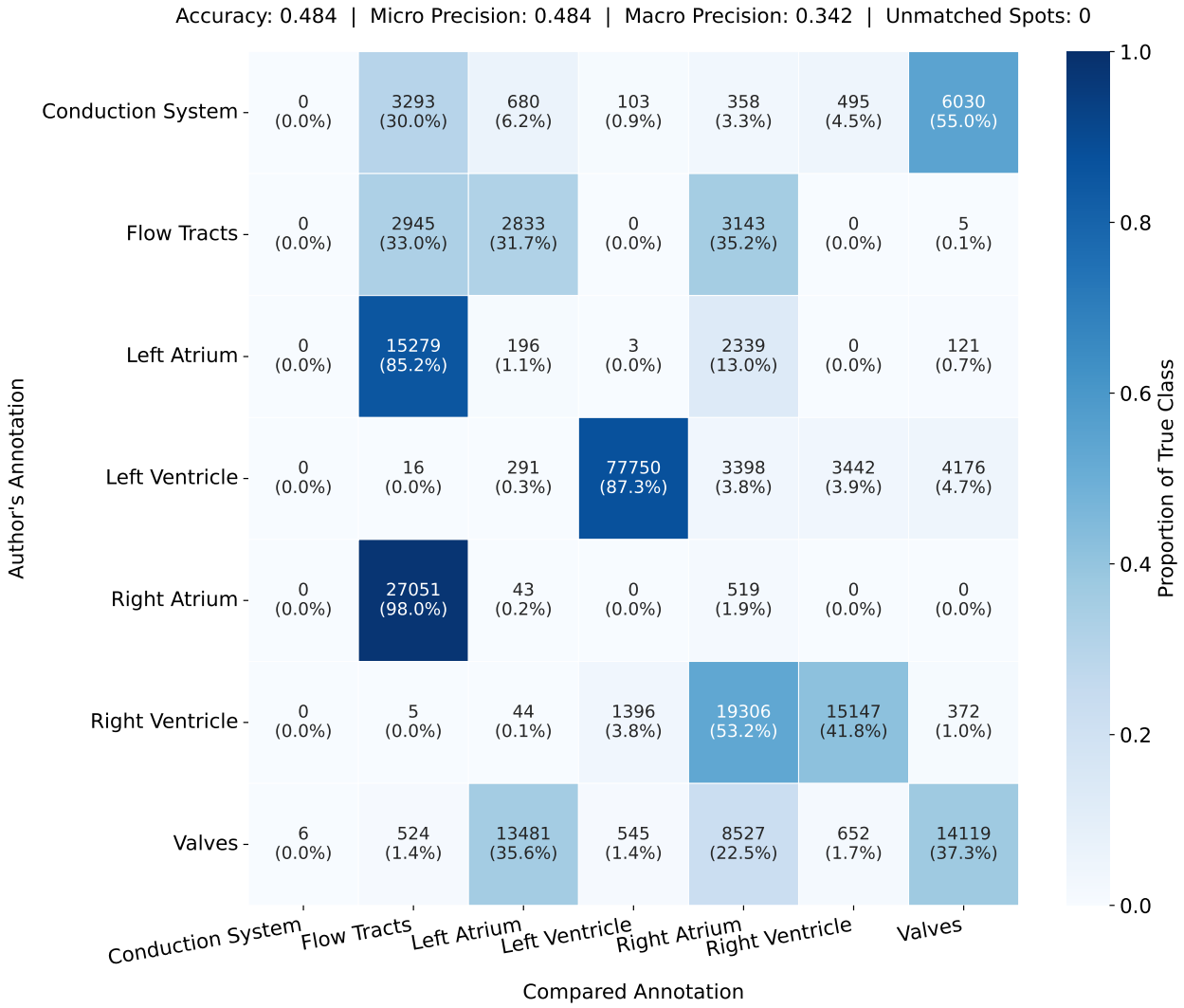
Supplementary Figure 45: Confusion matrix of human scientist *Lh* tissue niche annotation.



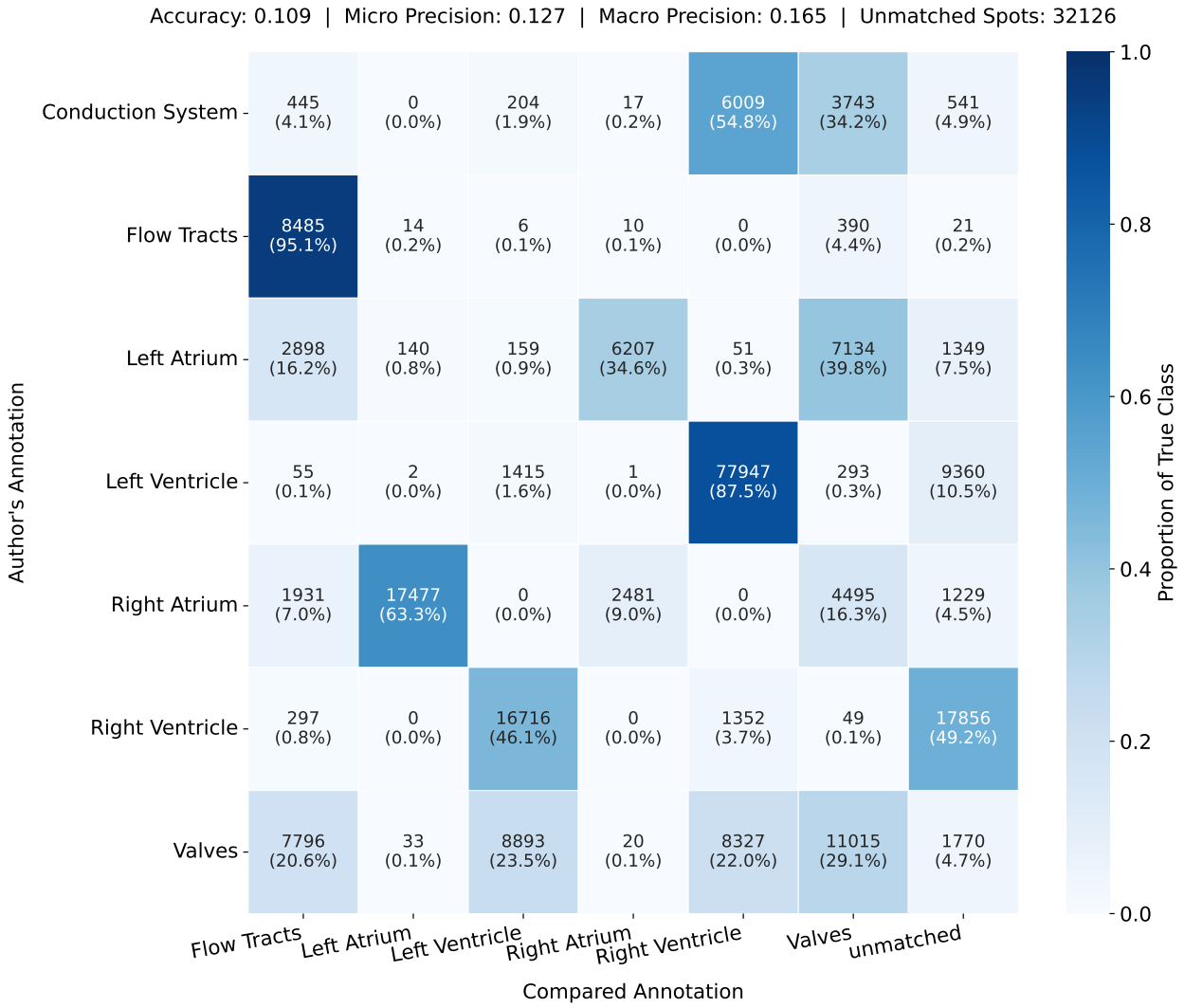
Supplementary Figure 46: Confusion matrix of human scientist *La* tissue niche annotation.



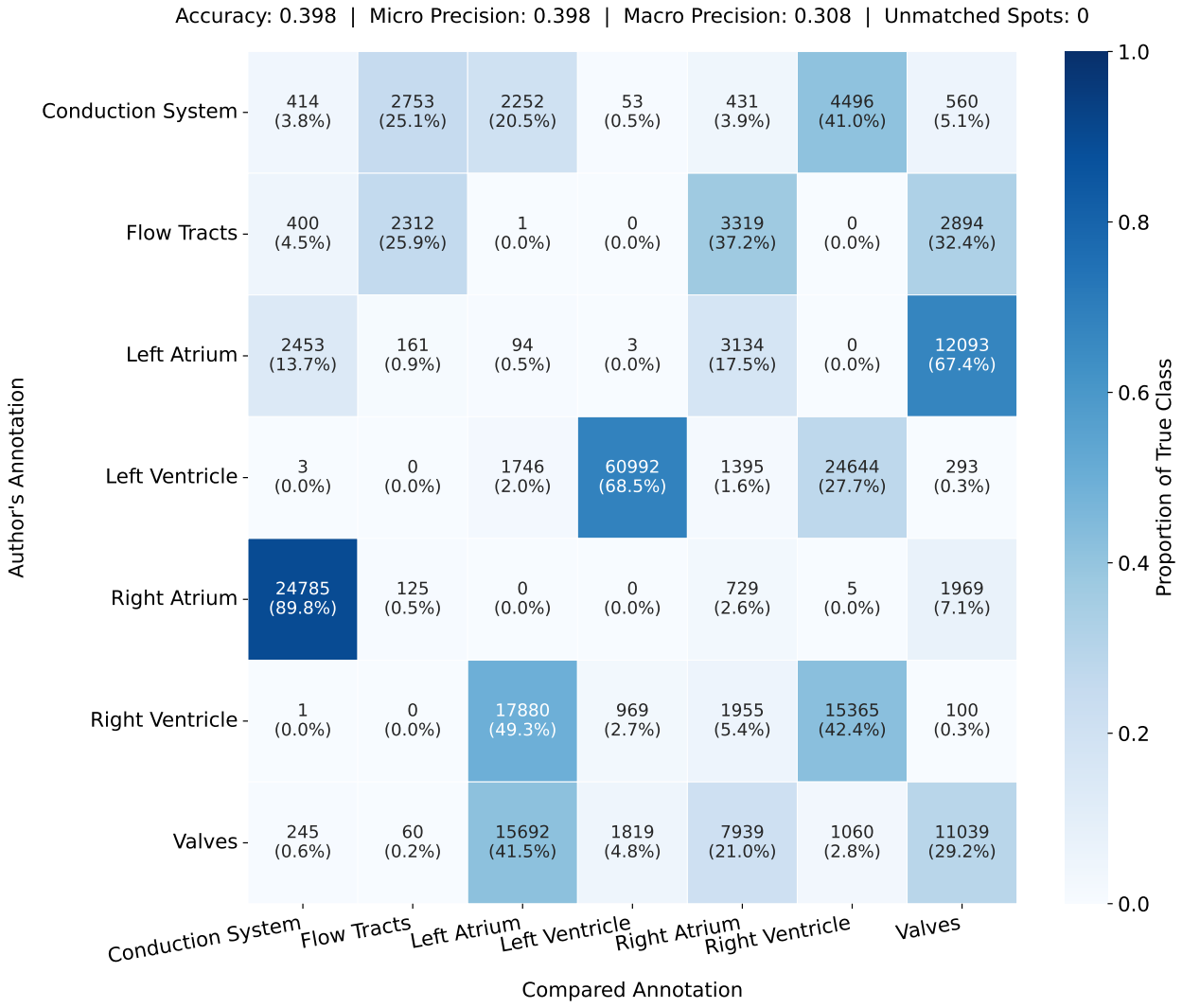
Supplementary Figure 47: Confusion matrix of human scientist *Lu* tissue niche annotation.



Supplementary Figure 48: Confusion matrix of human scientist *L* tissue niche annotation.



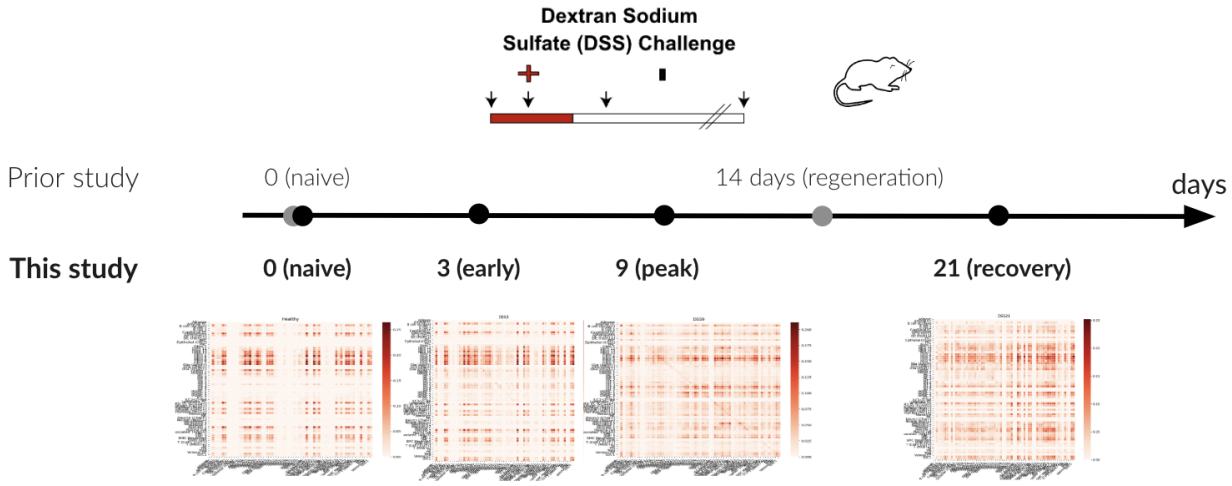
Supplementary Figure 49: Confusion matrix of human scientist *P* tissue niche annotation.



Supplementary Figure 50: Confusion matrix of human scientist *TL* tissue niche annotation.

4. Mining cell-cell interaction and report generation

4.1. Overview



Supplementary Figure 51: Overview of spatial transcriptomics data used for cell–cell interaction analysis. The dataset captures spatial and molecular profiles across disease stages—onset, peak inflammation, and recovery—enabling investigation of dynamic tissue remodeling and intercellular communication.

Advances in spatial transcriptomics (*e.g.*, MERFISH) enable high-resolution mapping of cellular identity, location, and molecular output within tissues, providing critical insights into disease progression and tissue remodeling (Supplementary Fig.51). By capturing spatial and molecular changes across disease onset, peak inflammation, and recovery, these techniques facilitate a deeper understanding of inflammatory conditions such as colitis.

Comparative analysis plays a central role in extracting biological insights. Systematic comparisons between healthy and diseased states, or across disease stages, reveal key shifts in gene expression, cell populations, and spatial organization. These analyses highlight tissue remodeling events, such as immune infiltration or the emergence of inflammation-associated fibroblasts (IAFs), pinpointing mechanisms of damage and potential therapeutic windows.

Equally important is understanding cell-cell interaction (CCI), which governs tissue responses. In inflamed environments, fibroblasts and immune cells exchange molecular signals (*e.g.*, cytokines, chemokines) that orchestrate inflammation and repair. Mapping these interactions uncovers regulatory networks that maintain or disrupt tissue balance, explaining immune escalation, tissue remodeling, and healing processes.

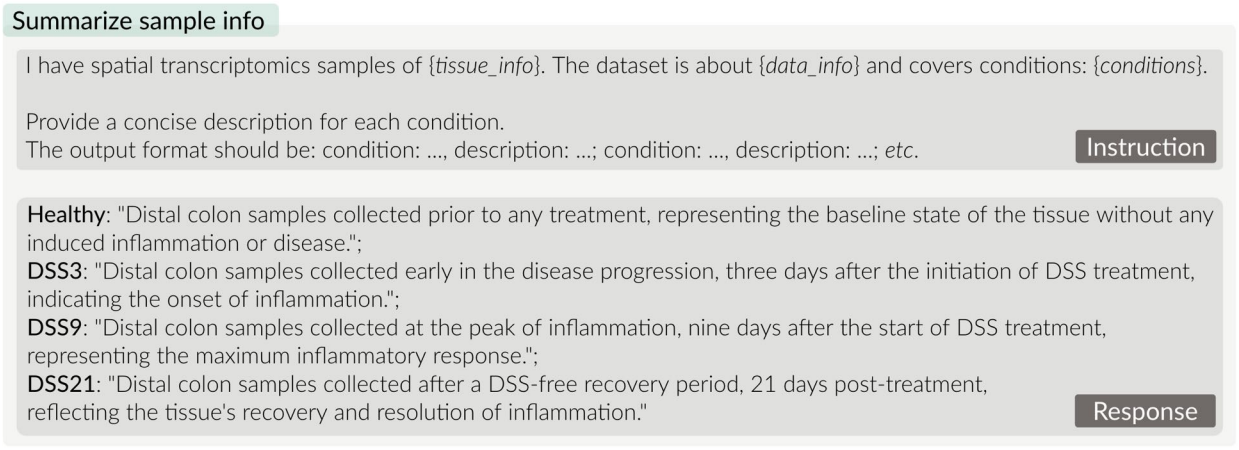
SpatialAgent leverages multimodal spatial transcriptomic datasets to systematically extract CCI insights. It integrates ligand-receptor analysis, neighborhood-based correlations, and factor-driven networks through a multi-step pipeline: preprocessing, interaction inference using LIANA+ (aggregating multiple ligand-receptor tools), and factor analysis to determine key drivers of intercellular communication. The findings are compiled into structured reports highlighting biological patterns, key signaling pathways, and potential therapeutic targets.

4.2. Summarization

This section presents a comparative summary of CCI across different conditions, focusing on interaction network shifts, cell type distributions, and tissue structures.

4.2.1. Condition specific analysis

SpatialAgent examines CCI changes across disease stages, such as increased pro-inflammatory signaling (*e.g.*, TNF, IL-6) during peak inflammation and fibroblast-driven repair signals in recovery (Supplementary Fig. 52).



Supplementary Figure 52: Summary of sample information.

4.2.2. Cell type specific analysis

Comparative analysis identifies dynamically interacting cell populations. For example, inflammatory fibroblasts in colitis show increased signaling with T cells and monocytes via TGF-β and IL-11, while epithelial-stromal interactions remain stable, underscoring immune-driven inflammation (Supplementary Fig. 53–59).

4.2.3. Tissue region specific analysis

SpatialAgent assesses spatial variation in CCI across tissue regions. In colitis, immune infiltration and fibroblast expansion localize to specific regions, refining insights into tissue heterogeneity and microenvironment-specific signaling. Summaries for representative tissue regions are shown in Supplementary Fig. 60–63.

4.3. Factor analysis in CCI

To interpret complex intercellular signaling patterns, SpatialAgent performs factor analysis to uncover underlying regulatory programs and communication modules. This approach identifies key drivers of cell–cell interactions by combining tensor decomposition, pathway enrichment, and regulatory network inference.

LIANA and Cell2Cell integration SpatialAgent integrates LIANA for robust ligand-receptor interaction quantification and employs Cell2Cell to identify global communication patterns through

Summarize cell type info per condition

Analysis of Inflammatory Associated Fibroblasts (IAF 2) Spatial Distribution in DSS3 Sample

1. Spatial Specificity:
   - Specific in Two or Three Tissues: The IAF 2 cells are primarily distributed in the [MU8] Mucosa (subtype 8) at 50%, [MU7] Mucosa (subtype 7) at 15%, and [ME3] Muscularis externa (subtype 3) at 12%. These three regions together account for 77% of the distribution, indicating a specificity towards these tissues.

2. Potential Reasons for Spatial Specificity:
   - Mucosa (subtypes 8 and 7): The mucosa is the innermost layer of the colon and is involved in absorption and secretion. Fibroblasts in this region may play a role in maintaining the structural integrity and repair of the mucosal layer, especially during inflammation.
   - Muscularis Externa (subtype 3): This layer is responsible for peristalsis and movement of the colon. Fibroblasts here might contribute to the structural support and repair of the muscular layer during inflammatory responses.

3. Interesting or Special Findings:
   - The high concentration of IAF 2 cells in the mucosal subtypes suggests a significant role in the inflammatory response and tissue repair processes during peak inflammation. The presence in the muscularis externa also indicates a potential involvement in maintaining motility and structural integrity under inflammatory stress.
   - The absence of IAF 2 cells in several other regions, such as the lumen and certain mucosal subtypes, highlights their specific roles and potential functional specialization in the regions where they are present.

Supplementary Figure 53: Summary of IAF2 cells within DSS3 samples.

tensor-based decomposition.

Pathway and network analysis SpatialAgent contextualizes communication modules within biological pathways using PROGENy, linking interactions to transcriptional responses. This enables:

- Pathway enrichment analysis (*e.g.*, TGF-β in fibroblast activation, IFN-γ in immune response).
- Cross-species gene mapping for translational insights.
- Regulatory network inference, predicting transcription factors (*e.g.*, NF-κB, SMAD3) mediating key interactions.

4.4. Generated report

SpatialAgent produces structured reports from spatial transcriptomic analyses by integrating statistical significance filtering ($p<0.01$) with contextual scientific narratives. Reports summarize spatial pattern information, cellular interactions, and condition-specific effects, transforming complex computational results into accessible scientific insights (Supplementary Fig. 64–71).

Summarize cell type info per condition

Analysis of Macrophage (Itgax+) Spatial Distribution in DSS3 Sample (Early Inflammation)

1. Spatial Specificity:
- Specific in three tissues: MU4 (32%), MU3 (23%), and MU2 (19%), collectively accounting for 74% of distribution
Clear shift to different mucosal regions compared to healthy baseline

2. Potential Reasons for Spatial Specificity:
- These regions likely represent initial sites of inflammatory response or epithelial damage
Macrophages are recruited to these areas to initiate repair processes and respond to epithelial damage

3. Interesting Findings:
- Absence in several regions suggests targeted immune response rather than generalized inflammation
Shift from MU3-dominant in healthy tissue to MU4-dominant during early inflammation indicates specific target regions for initial inflammatory response

Supplementary Figure 54: Summary of Macrophage cells within DSS3 samples.

Summarize cell type info per condition

Analysis of Treg Spatial Distribution in DSS21 Sample (Recovery Phase)

1. Spatial Specificity:
- Distributed across SM3 (27%), MU3 (17%), and MU10 (16%), collectively accounting for 60% of distribution
More evenly distributed than inflammatory cells during acute phases

2. Potential Reasons for Spatial Specificity:
- Regulatory T cells (Tregs) modulate immune responses and promote resolution of inflammation
Distribution in both submucosal (SM3) and mucosal regions suggests active immune regulation during recovery

3. Interesting Findings:
- Broader distribution compared to inflammatory cells indicates widespread immunomodulatory activity
Presence in multiple regions reflects the complex nature of inflammation resolution and tissue repair

Supplementary Figure 55: Summary of Treg cells within DSS21 samples.

Summarize cell type info across all conditions

Analysis of IAF2 Spatial Distribution across all samples / stages

To analyze the spatial distribution of Inflammatory Associated Fibroblasts (IAF 2) across different stages of DSS-induced colitis, let's break down the observations step by step:

1. Pattern of Spatial Distribution Change:
   - Healthy: IAF 2 cells are moderately distributed across the submucosa and mucosa, indicating a baseline role in maintaining tissue homeostasis.
   - DSS3 (Early Inflammation): There's an increased concentration in the submucosa and mucosa, suggesting an early response to inflammation, likely involved in structural support and initial tissue repair.
   - DSS9 (Peak Inflammation): A significant shift towards the mucosa, particularly subtypes 8 and 7, and some presence in the muscularis externa, indicating a role in intense tissue repair and possibly maintaining motility during peak inflammation.
   - DSS21 (Recovery): Distribution shifts back towards the submucosa and mucosa, reflecting involvement in tissue repair and resolution of inflammation.

2. Possible Reasons Behind the Changes:
   - Submucosa and Mucosa Association: Throughout the stages, the presence of IAF 2 cells in these regions suggests their role in maintaining structural integrity, supporting epithelial barriers, and modulating immune responses. During inflammation, their increased presence in the mucosa indicates active participation in repair processes.
   - Muscularis Externa Involvement (DSS9): The presence here during peak inflammation suggests a role in maintaining structural support and possibly aiding in motility under stress.

3. Interesting or Special Findings:
   - Dynamic Role in Inflammation and Recovery: IAF 2 cells exhibit a dynamic spatial distribution, aligning with their roles in structural support, immune modulation, and tissue repair across different stages of colitis.
   - Specificity and Versatility: While not highly specific to a single region, their moderate distribution across key regions highlights their versatile role in maintaining tissue homeostasis and responding to environmental changes.
   - Potential Therapeutic Target: The focused presence in the mucosal and submucosal layers during recovery suggests potential for targeted therapeutic strategies in inflammatory bowel diseases.

Supplementary Figure 56: Summary of IAF2 cells across all conditions.

Summarize cell type info across all conditions

Analysis of Macrophage (Itgax+) Spatial Distribution across all samples/stages

To analyze the spatial distribution of macrophages (Itgax+) across different stages of DSS-induced colitis, let's break down the observations step by step:

Pattern of Spatial Distribution Change:
- Healthy (Day 0): Macrophages are primarily concentrated in MU3 (55%), with significant presence in MU2 (17%) and MU1 (13%). This suggests a baseline distribution focused on immune surveillance in specific mucosal regions.
- DSS3 (Day 3): The distribution shifts to MU4 (32%), MU3 (23%), and MU2 (19%), indicating an early inflammatory response. The increase in MU4 suggests an expansion of macrophage activity to new areas as inflammation begins.
- DSS9 (Day 9): At peak inflammation, macrophages are concentrated in MU8 (25%), MU7 (23%), and MU5 (15%). This shift to different regions suggests these areas are experiencing intense inflammatory activity.
- DSS21 (Day 21): Post-recovery, macrophages are more evenly distributed, with the highest presence in SM3 (26%) and MU3 (18%). This indicates a return to tissue homeostasis and ongoing tissue repair.

Possible Reasons Behind the Changes:
- Healthy: Macrophages are positioned in regions likely involved in routine immune surveillance and microbial interaction.
- DSS3: The shift to MU4, MU3, and MU2 suggests these regions are early sites of inflammation, possibly due to epithelial damage and the need for repair.
- DSS9: Concentration in MU8, MU7, and MU5 indicates these regions are experiencing peak inflammatory responses, likely due to severe tissue damage and immune cell infiltration.
- DSS21: The distribution in SM3 and MU3 suggests these areas are involved in tissue remodeling and resolution of inflammation, reflecting recovery processes.

Interesting or Special Findings:
- Healthy: The absence of macrophages in certain regions (e.g., ME1, SM3) suggests these areas have different functional roles or microenvironments.
- DSS3: The early concentration in specific mucosal regions highlights macrophages' role as indicators of inflammation onset.
- DSS9: The specific regions with high macrophage presence could be potential targets for therapeutic interventions to modulate inflammation.
- DSS21: The high presence in SM3 suggests significant prior inflammation, with macrophages playing a role in tissue repair. The absence in other regions indicates a return to homeostasis.

Supplementary Figure 57: Summary of Macrophage cells across all conditions.

Summarize cell type info across all conditions

Analysis of IAE 1 Distribution across all samples/stages
To analyze the spatial distribution changes of cell type IAE 1 (Inflammatory Associated Epithelial) across different stages of DSS-induced colitis, let's break down the observations step by step:

1. Pattern of Spatial Distribution Change:
- Healthy (Day 0): IAE 1 is predominantly found in MU3, MU2, and MU4, accounting for 85% of its distribution. This suggests a baseline presence in mucosal regions.
- DSS3 (Day 3): IAE 1 shows a strong preference for MU4, with 82% of its presence concentrated there. This indicates a shift towards a specific mucosal region during early inflammation.
- DSS9 (Day 9): The distribution shifts to MU7 (39%), MU6 (24%), and MU4 (10%), indicating a broader spread across different mucosal regions during peak inflammation.
- DSS21 (Day 21): IAE 1 is primarily found in MU3 (39%), MU4 (15%), and MU9 (12%), suggesting a return to some of the initial regions with an additional presence in MU9 during recovery.

2. Possible Reasons Behind the Changes:
- Early Inflammation (DSS3): The concentration in MU4 suggests a role in initial immune response or mucosal barrier maintenance, critical during the onset of inflammation.
- Peak Inflammation (DSS9): The spread to MU7 and MU6 may reflect active inflammation or tissue remodeling, where IAE 1 cells contribute to immune response or repair.
- Recovery (DSS21): The presence in MU3, MU4, and MU9 suggests involvement in inflammation resolution and tissue repair, with these regions possibly being key sites for recovery processes.

3. Interesting or Special Findings:
- Specificity and Functionality: The dynamic changes in spatial distribution highlight the potential functional roles of IAE 1 in different phases of colitis, from immune response to tissue repair.
- Potential Biomarker: The specific localization of IAE 1, especially during early inflammation (DSS3), could serve as a biomarker for disease progression.
- Absence in Certain Regions: The consistent absence of IAE 1 in regions like LUM, ME3, and others across conditions suggests these areas are less involved in the inflammatory response or recovery processes.

Supplementary Figure 58: Summary of IAE1 cells across all conditions.

Summarize cell type info across all conditions

Analysis of Colonocytes Spatial Distribution across all samples/stages

To analyze the spatial distribution of colonocytes across different stages of DSS-induced colitis, let's break down the observations step by step:

1. Pattern of Spatial Distribution Change:
- Healthy (Day 0): Colonocytes are highly concentrated in the MU3 region (80%), with complete absence in regions like FOL1, MU6, SM1, ME2, ME1, and MU7.
- DSS3 (Day 3): Colonocytes are distributed across MU4 (50%), MU3 (32%), and MU2 (17%). This indicates a shift from a single-region concentration to a broader distribution across multiple regions.
- DSS9 (Day 9): Colonocytes are primarily localized in MU6 (53%), with lesser concentrations in MU3 (21%) and MU2 (9%). This suggests a further shift in distribution, possibly due to peak inflammation.
- DSS21 (Day 21): Colonocytes show a strong preference for MU3 (65%), with MU9 (16%) and MU2 (7%) also having notable concentrations. This indicates a partial return to the original distribution pattern seen in healthy tissue.

2. Possible Reasons Behind the Changes:
- Healthy: The high concentration in MU3 suggests this region is structurally or functionally optimal for colonocyte activity, focusing on absorption and secretion.
- DSS3: The broader distribution across MU4, MU3, and MU2 may reflect an adaptive response to early inflammation, with colonocytes possibly migrating to areas that require enhanced barrier and absorptive functions.
- DSS9: The concentration in MU6 during peak inflammation suggests this region may be more actively engaged in inflammatory response or has structural adaptations that support colonocyte function under stress.
- DSS21: The return to MU3 dominance during recovery suggests this region plays a crucial role in tissue repair and resolution of inflammation, possibly due to its regenerative capacity.

3. Interesting or Special Findings:
- Absence in Specific Regions: The consistent absence of colonocytes in certain regions (e.g., FOL1, SM1) across conditions suggests these areas may have distinct roles or cellular compositions that do not support colonocyte presence.
- Inflammation Impact: The shifts in colonocyte distribution during DSS3 and DSS9 highlight the dynamic nature of tissue response to inflammation, with specific regions becoming more prominent in supporting colonocyte functions.
- Recovery Phase: The re-concentration of colonocytes in MU3 during DSS21 underscores the importance of this region in recovery and regeneration, suggesting it may be a key site for therapeutic targeting in colitis treatment.

Supplementary Figure 59: Summary of Colonocyte cells across all conditions.

Summarize tissue regions across all conditions

MU10 Region: Critical for Early Repair Response

Early Inflammation (DSS3)

1. Cell Type Composition:
   - The tissue region MU10 is highly specific in one cell type, with the "Repair associated (Arg1+)" cells comprising 85% of the cell population. This indicates a strong dominance of this cell type in the region.

2. Potential Reasons for Cell Type Specificity:
   - The high presence of "Repair associated (Arg1+)" cells suggests that this region is actively involved in tissue repair and regeneration processes. Given that the sample is from the early stage of DSS-induced colitis (day 3), the onset of inflammation likely triggers a repair response. Arg1+ cells are often associated with anti-inflammatory and tissue repair functions, which aligns with the body's attempt to counteract the initial inflammatory damage caused by DSS treatment.

3. Interesting or Special Findings:
   - The near absence of other cell types, particularly immune cells like macrophages and neutrophils, is notable. This could indicate a localized or specific repair response that is not yet heavily infiltrated by other immune cells, which might be more prevalent at later stages of inflammation. Additionally, the presence of a small percentage of "Fibro 5" and "Capillarial EC" cells suggests some level of vascular and structural support, which may be necessary for the repair process. The low diversity of cell types at this stage might reflect an early, focused repair mechanism before broader immune involvement.

Peak Inflammation (DSS9)

No data provided for MU10 at peak inflammation (DSS9).

Recovery Phase (DSS21)

1. Cell Type Composition:
   - The highest percentage of a single cell type is 51% for "Repair associated (Arg1+)" cells. This indicates a specificity in one cell type, though reduced from early inflammation.

2. Potential Reasons for Cell Type Specificity:
   - The high presence of repair-associated cells suggests active tissue repair and resolution of inflammation. This is consistent with the sample being collected after a DSS-free recovery period, indicating the tissue is in a healing phase.
   - The presence of fibroblasts (Fibro 1, 4, 5) and immune cells (e.g., Macrophage, Neutrophil) supports tissue remodeling and immune surveillance, which are crucial during recovery from inflammation.

3. Interesting or Special Findings:
   - The absence of colonocytes and goblet cells might indicate that epithelial regeneration is still ongoing or that these cells are underrepresented in this specific region.
   - The presence of IAF (intramuscular adipose fibroblasts) and IASMC (intestinal smooth muscle cells) suggests involvement in structural support and motility restoration.
   - The diverse immune cell types, including macrophages and neutrophils, highlights ongoing immune regulation and potential resolution of any residual inflammation.

Supplementary Figure 60: Summary of MU10 tissue region across all conditions.

Summarize tissue regions across all conditions

LUM Region: Persistent Neutrophil Activity

Early Inflammation (DSS3)
No data provided for LUM at early inflammation (DSS3).

Peak Inflammation (DSS9)
1. Tissue Region's Cell Type Composition:
- The tissue region LUM is highly specific in one cell type, with Neutrophil 2 comprising 76% of the cell population. This indicates a strong dominance of this cell type.
- Other cell types present include Neutrophil 1 (7%), Macrophage (Cxcl10+) (5%), and IASMC 1 (2%). These, along with minor contributions from other cell types, do not significantly alter the dominance of Neutrophil 2.

2. Potential Reasons for Cell Type Specificity:
- The high presence of Neutrophil 2 suggests an acute inflammatory response, as neutrophils are key players in inflammation and innate immunity. This aligns with the sample being collected at the peak of inflammation in a DSS-induced colitis model.
- The presence of Macrophage (Cxcl10+) further supports an inflammatory environment, as these cells are involved in cytokine production and immune response modulation.
- The presence of IASMC 1 (intestinal smooth muscle cells) might indicate tissue remodeling or repair processes, which are common during inflammation.

3. Interesting or Special Findings:
- The overwhelming dominance of Neutrophil 2 is notable and suggests a robust neutrophilic response, which is characteristic of acute inflammation.
- The absence or near absence of other immune cell types like T cells and B cells is interesting, as it highlights the acute nature of the inflammatory response, potentially overshadowing adaptive immune responses at this stage.
- The presence of a small percentage of Treg cells (0.01%) might indicate an attempt to regulate or resolve inflammation, although their low numbers suggest limited regulatory activity at this peak inflammation stage.

Recovery Phase (DSS21)
1. Tissue Region's Cell Type Composition:
- The tissue region LUM is highly specific in one cell type, with Neutrophil 2 comprising 55% of the cell population. This indicates a strong predominance of this cell type.
- Neutrophil 1 also has a significant presence at 14%, making the combined neutrophil population 69%.
- Other cell types are present in much lower percentages, with none exceeding 10%.

2. Potential Reasons for Cell Type Specificity:
- The high presence of neutrophils, particularly Neutrophil 2, suggests ongoing or recent resolution of inflammation. Neutrophils are key players in the acute inflammatory response and are often involved in the initial stages of tissue repair.
- The presence of macrophages (Cxcl10+) at 7% and other immune cells like monocytes and dendritic cells (DC) indicates an active immune environment, possibly involved in clearing residual inflammation and promoting tissue repair.
- The presence of epithelial cells (Clu+) and fibroblasts, although low, suggests ongoing tissue remodeling and repair processes.

3. Interesting or Special Findings:
- The dominance of Neutrophil 2 suggests a specific role or state of these cells in the recovery phase post-DSS treatment. This could indicate a specialized function or adaptation of neutrophils in the resolution of inflammation in the distal colon.
- The absence of certain cell types like stem cells and T (Mki67+) cells might suggest a reduced proliferative activity in this recovery phase, focusing more on resolution and repair rather than regeneration.
- The low presence of regulatory T cells (Treg) might indicate a shift from active immune regulation to a more resolved state of inflammation.

Supplementary Figure 61: Summary of LUM tissue region across all conditions.

Summarize tissue regions across all conditions

MU1 Region: Dynamic Stem Cell Response

Healthy (Baseline)

1. Cell Type Composition:
   - The tissue region MU1 is specific in two or three cell types. The stem cells (42%) and transit-amplifying (TA) cells (11%) together account for 53% of the cell population, indicating a focus on these cell types. The remaining cell types are present in much smaller proportions, with none exceeding 10%.

2. Potential Reasons for Cell Type Specificity:
   - Stem Cells (42%): The high percentage of stem cells suggests that this region is likely involved in maintaining the regenerative capacity of the distal colon. Stem cells are crucial for tissue homeostasis and repair, especially in the gastrointestinal tract where cell turnover is high.
   - Transit-Amplifying (TA) Cells (11%): These cells are progeny of stem cells and are involved in rapid proliferation to replenish the epithelial lining. Their presence supports the regenerative function of the tissue.
   - The presence of fibroblasts and smooth muscle cells, although in smaller proportions, indicates roles in structural support and motility of the colon.

3. Interesting or Special Findings:
   - The absence or very low presence of immune cells (e.g., T cells, B cells, macrophages) is notable, reflecting the baseline state of the tissue without inflammation or disease. This suggests a healthy, non-inflamed tissue environment.
   - The presence of various fibroblast subtypes, albeit in small percentages, may indicate a diverse stromal environment that could play roles in tissue architecture and signaling.
   - The low presence of endothelial cells (arterial, capillarial, lymphatic, and venous) suggests a well-maintained vascular network, essential for nutrient supply and waste removal in the tissue.

Early Inflammation (DSS3)

1. Tissue Region's Cell Type Composition:
   - The tissue region MU1 does not have specificity in one cell type, as no single cell type exceeds 80%.
   - It is specific in two or three cell types, with stem cells (16%), transit-amplifying (TA) cells (15%), and fibroblast 6 (14%) collectively making up 45% of the composition.
   - The remaining cell types are more evenly distributed, with none exceeding 16%.

2. Potential Reasons for Cell Type Specificity:
   - Stem Cells and TA Cells: The presence of a high percentage of stem cells and TA cells suggests active tissue regeneration and repair processes, which are typical during the early stages of inflammation as the tissue attempts to recover from damage.
   - Fibroblast 6: The significant presence of fibroblast 6 may indicate a role in extracellular matrix remodeling and tissue repair, which are crucial during the onset of inflammation.
   - Goblet Cells (13%): Goblet cells are responsible for mucus production, which is essential for protecting the epithelial lining of the colon. Their presence suggests a response to maintain the mucosal barrier during inflammation.

3. Interesting or Special Findings:
   - The presence of macrophages (Mrc1+) at 9% suggests an active immune response, as these cells are involved in phagocytosis and cytokine production during inflammation.
   - The low presence of immune cells like T cells and neutrophils might indicate that the immune response is still in the early stages, focusing more on tissue repair and barrier maintenance rather than full-scale immune activation.
   - The absence of certain cell types, such as B cells and dendritic cells, could suggest a localized response focused on immediate repair rather than adaptive immune activation at this stage.

....

Supplementary Figure 62: Summary of MU1 tissue region across all conditions (1/2).

Summarize tissue regions across all conditions

MU1 Region: Dynamic Stem Cell Response

...

Peak Inflammation (DSS9)

1. Cell Type Composition:
   - The highest percentage of any cell type is 46% (Stem cells), which does not reach the 80% threshold for specificity in one cell type. The sum of the top three cell types (Stem cells: 46%, Goblet 1: 7%, TA: 7%) is 60%, which does not reach the 90% threshold for specificity in two or three cell types. Therefore, the cell type distribution is mixed or evenly distributed.

2. Potential Reasons for Cell Type Specificity:
   - Stem Cells (46%): The high presence of stem cells suggests active tissue regeneration and repair, which is expected at the peak of inflammation as the tissue attempts to recover from damage.
   - Goblet Cells (7%): Goblet cells are responsible for mucus production, which is crucial for protecting the epithelial lining of the colon, especially during inflammation.
   - Transit-Amplifying (TA) Cells (7%): These cells are indicative of active cell proliferation, supporting the regeneration process in response to inflammation.

3. Interesting or Special Findings:
   - The presence of a significant proportion of stem cells and TA cells highlights the tissue's response to inflammation, focusing on repair and regeneration.
   - The low presence of immune cells (e.g., T cells, B cells, macrophages) is notable, suggesting that the peak inflammatory response might be more focused on epithelial and stromal cell activity rather than immune cell infiltration in this specific region.
   - The presence of various fibroblast subtypes, although in low percentages, indicates a complex stromal environment that may contribute to tissue remodeling and fibrosis during inflammation.

Recovery Phase (DSS21)

1. Cell Type Composition:
   - The highest percentage of any cell type is 37% (Stem cells), which is significant but not over 80%. The next highest is Goblet 1 at 12%, followed by TA (Transit Amplifying) cells at 8%. This indicates a mixed distribution rather than specificity in one or two cell types.

2. Potential Reasons for Cell Type Specificity:
   - Stem Cells (37%): The high presence of stem cells suggests active tissue regeneration and repair, which is consistent with the recovery phase post-DSS treatment. Stem cells are crucial for replenishing damaged cells and restoring normal tissue architecture.
   - Goblet Cells (12%): Goblet cells are responsible for mucus production, which is essential for protecting the epithelial lining of the colon. Their presence indicates a restoration of the mucosal barrier, which is vital after inflammation.
   - TA Cells (8%): These cells are indicative of active cell proliferation, supporting the regeneration process in the colon.
   - Fibroblasts and SMCs (Smooth Muscle Cells): Various fibroblast subtypes and SMCs are present, suggesting ongoing tissue remodeling and structural support.

3. Interesting or Special Findings:
   - Low Immune Cell Presence: There is a notably low presence of immune cells such as T cells, B cells, and macrophages, which suggests a resolution of inflammation and a return to homeostasis.
   - Presence of Repair-Associated Cells: The presence of repair-associated cells (Arg1+) and macrophages (Itgax+, Mrc1+) at low levels indicates ongoing tissue repair and resolution of inflammation.
   - Neuronal and Glial Cells: Minimal presence of neuronal and glial cells suggests limited neural involvement in this recovery phase, focusing more on epithelial and structural cell recovery.

Supplementary Figure 63: Summary of MU1 tissue region across all conditions (2/2).

Spatial Transcriptomics Analysis Report (1/8)

Abstract
This study investigates spatial transcriptomics data from Dextran-sodium-sulfate (DSS)-induced mouse colitis model, focusing on We harvested distal colon prior to treatment (day 0), early in disease (day 3), at peak inflammation (day 9), and after a DSS-free recovery period (day 21)..

Introduction
Spatial transcriptomics is a powerful technique that allows researchers to study gene expression within the context of tissue architecture. This approach is particularly useful in understanding complex diseases like colitis, where spatial heterogeneity plays a crucial role in disease progression and tissue response.

Background
1. Dextran-Sodium-Sulfate (DSS)-Induced Colitis Model:
- DSS is a chemical compound used to induce colitis in mice, serving as a model for human inflammatory bowel disease (IBD). It disrupts the epithelial barrier in the colon, leading to inflammation and ulceration.
- The DSS-induced colitis model is widely used due to its reproducibility and similarity to human ulcerative colitis in terms of symptoms and histopathological features.

2. Temporal Sampling:
- **Day 0 (Prior to Treatment)**: This baseline measurement provides insights into the normal gene expression profile and tissue architecture of the distal colon before any inflammatory insult. It serves as a control to compare changes induced by DSS treatment.
- **Day 3 (Early in Disease)**: At this stage, initial inflammatory responses are observed. Spatial transcriptomics can reveal early changes in gene expression, including upregulation of pro-inflammatory cytokines and chemokines, and alterations in epithelial cell function.
- **Day 9 (Peak Inflammation)**: This is typically when the most severe inflammation and tissue damage occur. Spatial transcriptomics can identify regions of high immune cell infiltration, epithelial cell death, and changes in stromal cell populations. It can also highlight the spatial distribution of immune responses and tissue remodeling.
- **Day 21 (Recovery Period)**: After DSS withdrawal, the colon undergoes repair and regeneration. Spatial transcriptomics can elucidate the processes involved in tissue healing, such as stem cell activation, re-establishment of epithelial integrity, and resolution of inflammation.

3. Spatial Heterogeneity in Colitis:
- Colitis involves complex interactions between various cell types, including epithelial cells, immune cells, fibroblasts, and endothelial cells. Spatial transcriptomics allows for the mapping of these interactions and the identification of specific cellular niches that contribute to disease progression and recovery.
- Understanding spatial heterogeneity is crucial for identifying potential therapeutic targets and for developing strategies to modulate specific cellular responses.

4. Applications of Spatial Transcriptomics:
- **Cellular Interactions**: By preserving spatial context, researchers can study how different cell types communicate and influence each other during colitis progression and recovery.
- **Gene Expression Patterns**: Spatial transcriptomics can reveal localized gene expression changes that are not apparent in bulk RNA sequencing, providing insights into region-specific responses.
- **Biomarker Discovery**: Identifying spatially distinct gene expression signatures can lead to the discovery of biomarkers for disease diagnosis, prognosis, and treatment response.

5. Challenges and Considerations:
- **Technical Limitations**: Spatial resolution and sensitivity can vary depending on the technology used, which may affect the ability to detect subtle changes in gene expression.
- **Data Complexity**: The integration and interpretation of spatial transcriptomics data require sophisticated computational tools and expertise in bioinformatics.

Overall, spatial transcriptomics offers a comprehensive approach to studying DSS-induced colitis, providing valuable insights into the spatial dynamics of gene expression and cellular interactions during disease progression and recovery.

Dataset Description
We harvested distal colon prior to treatment (day 0), early in disease (day 3), at peak inflammation (day 9), and after a DSS-free recovery period (day 21).

Methods
Analysis pipeline included spatial pattern analysis, cell-cell interaction mapping, and condition-specific effect evaluation.

Supplementary Figure 64: SpatialAgent generated report (1/8)

Spatial Transcriptomics Analysis Report (2/8)

Results - Spatial Organization Patterns

To provide a detailed analysis of the spatial distribution patterns in the DSS-induced mouse colitis model, we will follow a structured approach, focusing on cell type distributions, factor associations, and region-specific patterns.

1. Examining Cell Type Spatial Distributions and Changes Across Conditions

Colonocytes:
- **Healthy (Day 0)**: Colonocytes are predominantly located in MU3 (80%), indicating a stable environment conducive to their function in absorption and secretion.
- **DSS3 (Day 3)**: Colonocytes show a broader distribution, with 50% in MU4, suggesting an adaptive response to early inflammation.
- **DSS9 (Day 9)**: Colonocytes shift to MU6 (53%), reflecting their involvement in the inflammatory response, possibly due to structural changes or increased inflammatory signaling.
- **DSS21 (Day 21)**: Colonocytes return to MU3 (65%), indicating recovery and tissue repair.

Other Cell Types:
- **TA Cells**: Concentrated in MU2 and MU1 during healthy conditions, they spread to MU6 during DSS9, reflecting their role in the peak inflammatory response and subsequent recovery.
- **Smooth Muscle Cells (SMCs):** Consistently present in regions like ME1 and SM1, indicating their role in maintaining structural integrity and motility throughout the inflammatory process.
- **Fibroblasts**: Increase during DSS3 and DSS9, particularly in regions like MU1 and MES1, indicating tissue remodeling and repair.
- **Immune Cells:** Their presence increases during DSS3 and peaks at DSS9, especially in regions like MU7 and MU9, reflecting active inflammation.

2. Identifying Factors with Strong Cell Type Associations

Factor 2:
- **Strong Association with DSS9**: This factor shows significant changes between DSS9 and DSS3, indicating a strong association with immune cells and fibroblasts, which are prominent during peak inflammation.

Factor 4:
- **Strong Association with DSS9 and Healthy**: This factor highlights significant differences between DSS9 and both Healthy and DSS3 conditions, suggesting a major shift in epithelial and immune cell types during peak inflammation.

Factor 7:
- **Strong Association with DSS9 and Healthy**: This factor reflects changes in immune and epithelial cells, emphasizing the dynamic interplay between inflammation and epithelial integrity.

Factor 8 and Factor 10:
- **Association with Peak Inflammation**: These factors show significant differences between DSS9 and Healthy, indicating their role in peak inflammation and association with immune cell activity.

3. Finding Region-Specific Patterns by Connecting Factor Loadings with Tissue Regions

MU3:
- **Healthy and Recovery**: Dominated by colonocytes, with minimal immune presence during healthy conditions. During DSS9, there's a shift towards fibroblasts and immune cells, indicating tissue remodeling and inflammation. Factors 2 and 4 are likely linked to these changes.

MU6
- **DSS9**: High presence of colonocytes and immune activity, indicating a focus on barrier maintenance during peak inflammation. Factors 2 and 7 may be relevant, highlighting epithelial responses to inflammation.

MU9
- **DSS9**: Diverse cell types, including macrophages and fibroblasts, reflecting active inflammation and repair. Factors 4 and 7 likely play a role, emphasizing immune and stromal interactions.

Supplementary Figure 65: SpatialAgent generated report (2/8)

Spatial Transcriptomics Analysis Report (3/8)

Results - Spatial Organization Patterns

4. Highlighting the Most Significant Spatial Relationships

- Colonocytes and MU3: The strong preference for MU3 in both healthy and recovery phases suggests a critical role in maintaining and restoring mucosal integrity. This aligns with Factor 2, indicating a strong association with epithelial repair processes.
- Immune Cells and MU9: The peak presence of immune cells during DSS9, especially in regions like MU9, correlates with Factors 4 and 7, highlighting their role in inflammation.
- Fibroblasts and MU1/MES1: The increase in fibroblasts during DSS3 and DSS9 aligns with Factor 2, reflecting tissue remodeling and repair.

Conclusion

The analysis reveals a complex interplay between cell types, factors, and spatial distribution patterns in the DSS-induced colitis model. Factors 2, 4, 7, 8, and 10 show strong associations with specific cell types and regions, highlighting the dynamic changes in epithelial, immune, and stromal interactions during inflammation and recovery. These insights provide a deeper understanding of the spatial and cellular dynamics in colitis, emphasizing the importance of specific regions and cell types in disease progression and resolution.

Results - Cell-Cell Interaction Analysis

To analyze the cell-cell interaction patterns in the distal colon under different conditions, we need to delve into the key interacting cell pairs, dominant signaling pathways, condition-specific interactions, biological implications, and potential regulatory mechanisms. This analysis provides insights into the complex interplay of cells and pathways involved in inflammation, tissue repair, and immune regulation, which are crucial for developing therapeutic strategies for inflammatory diseases.

1. Key Interacting Cell Pairs

Factor 1

- **Macrophage (Itgax+) and Monocyte**: These cells are pivotal in the innate immune response, particularly in inflammation. Their interaction suggests a coordinated effort in pathogen response and tissue damage repair, highlighting their role in initiating and sustaining inflammatory responses.
- **cDC1 and B cell 1**: This interaction is indicative of antigen presentation and B cell activation, essential for adaptive immunity. It underscores the role of dendritic cells in priming B cells for antibody production.

Factor 2

- **Fibro 7 and Fibro 15**: These fibroblast interactions suggest a focus on tissue structure and repair, likely involving extracellular matrix production and remodeling, crucial for maintaining tissue integrity during and after inflammation.
- **Glia (Gfap+) and Glia (Gfra3+)**: These interactions may support neural signaling and integrity within the colon, emphasizing the role of glial cells in maintaining the nervous system's function in the gut.

Factor 3:

- **Pericyte 1 and Pericyte 2**: These cells are essential for vascular stability and blood-brain barrier maintenance, indicating their role in maintaining vascular integrity and possibly influencing blood flow and nutrient delivery during inflammation.
- **IASMC 1 and Fibro 2**: This interaction likely contributes to smooth muscle function and extracellular matrix remodeling, important for maintaining tissue structure and function.

Factor 4:

- **cDC1 and Macrophage (Itgax+)**: This interaction is crucial for coordinating immune responses, particularly in antigen presentation and activation of other immune cells, highlighting the interplay between innate and adaptive immunity.
- **Plasma cell and Fibro 2**: This interaction may involve antibody production and tissue repair, indicating a dual role in immune response and tissue maintenance.

Factor 5:

- **Tfh and Treg**: These interactions are vital for maintaining immune homeostasis and preventing autoimmunity, balancing immune activation and suppression.
- **T (Mki67+) and T (Cd4+ Ccr7+)**: This suggests active T cell proliferation and migration, important for adaptive immune responses and ensuring effective immune surveillance.

Supplementary Figure 66: SpatialAgent generated report (3/8)

Spatial Transcriptomics Analysis Report (4/8)

Results - Cell-Cell Interaction Analysis

2. Dominant Signaling Pathways

- **TGFb**: This pathway is central across multiple factors, regulating immune responses, fibrosis, and tissue repair. TGFb is a key cytokine involved in immune regulation and tissue homeostasis.
- **VEGF**: Involved in angiogenesis, particularly relevant in Factors 1 and 6, suggesting roles in vascular development and repair, crucial during inflammation and recovery.
- **EGFR and PI3K**: These pathways are common in several factors, indicating roles in cell proliferation and survival, important for both immune responses and tissue maintenance.
- **NFkB and TNFa**: Key pathways in inflammation, particularly in Factors 1 and 5, indicating roles in immune activation and inflammatory responses.

3. Condition-Specific Interactions

- **Healthy**: Baseline interactions likely involve maintenance of tissue homeostasis and immune surveillance, with balanced signaling pathways ensuring tissue integrity.
- **DSS3 (Early Inflammation)**: Increased interactions among immune cells (e.g., Macrophages, Monocytes) and activation of inflammatory pathways (e.g., NFkB, TNFa), indicating the onset of an immune response.
- **DSS9 (Peak Inflammation):** Heightened immune cell interactions and signaling through pathways like TGFb and VEGF, indicating tissue remodeling and angiogenesis as the tissue responds to peak inflammation.
- **DSS21 (Recovery)**: Shift towards interactions involving fibroblasts and pathways like TGFb, reflecting tissue repair and resolution of inflammation, as the tissue returns to a homeostatic state.

4. Biological Implications of Interactions

- **Immune Regulation**: Interactions among T cells, B cells, and macrophages are crucial for orchestrating immune responses and maintaining tolerance, preventing excessive inflammation or autoimmunity.
- **Tissue Repair and Fibrosis:** Fibroblast interactions and TGFb signaling are indicative of tissue repair processes, which can lead to fibrosis if dysregulated, highlighting the balance needed in tissue remodeling.
- **Angiogenesis and Vascular Stability:** VEGF and pericyte interactions suggest roles in restoring vascular integrity post-inflammation, crucial for tissue recovery and function.

5. Potential Regulatory Mechanisms

- **Cytokine Networks**: TNFa, TGFb, and other cytokines likely regulate the balance between inflammation and resolution, modulating immune responses and tissue repair.
- **Cell-Cell Contact**: Direct interactions between immune cells (e.g., Tfh and Treg) are essential for modulating immune responses, ensuring appropriate activation and suppression.
- **Extracellular Matrix Remodeling**: Fibroblast activity and signaling pathways like PI3K and EGFR contribute to changes in the extracellular matrix, affecting cell migration and tissue structure, crucial for both repair and maintaining tissue integrity.

In summary, understanding these interactions provides insights into the mechanisms of inflammation, tissue repair, and immune regulation, which are crucial for developing therapeutic strategies for inflammatory diseases. The analysis highlights the importance of balancing immune responses and tissue repair processes to prevent chronic inflammation and fibrosis.

Supplementary Figure 67: SpatialAgent generated report (4/8)

Spatial Transcriptomics Analysis Report (5/8)

Results - Condition-Specific Effects

The analysis of the DSS-induced mouse colitis model provides a comprehensive understanding of the dynamic changes occurring at various stages of inflammation and recovery. Here's a detailed breakdown of each aspect:

1. Condition-Specific Cellular Changes

- **Healthy State (Day 0)**: In the baseline condition, immune cells such as B cells, T cells, and macrophages are strategically positioned in regions like FOL1 and MU1, maintaining immune readiness and homeostasis. Epithelial cells ensure the integrity of the mucosal barrier, crucial for normal tissue function.
- **Early Inflammation (DSS3, Day 3)**: There is a notable redistribution of immune cells, including Treg and Cd8+ T cells, to regions like MU3 and SM1, indicating the onset of an immune response. Neutrophils are recruited to follicular regions, marking early immune activation. Colonocytes show a broader distribution, suggesting an adaptive response to maintain barrier and absorptive functions.
- **Peak Inflammation (DSS9, Day 9)**: Immune cell concentration intensifies, particularly in regions like MU8, with macrophages and neutrophils indicating heightened immune activity. The presence of fibroblasts suggests active tissue remodeling. Colonocytes shift to regions like MU7 and MU6, reflecting structural changes and increased inflammatory activity.
- **Recovery Phase (DSS21, Day 21)**: The distribution of many cell types returns to a baseline-like state, such as B cells in FOL1, signifying inflammation resolution. Treg cells in SM3 indicate ongoing immune regulation and tissue repair. Colonocytes concentrate in MU9, focusing on recovery and repair.

2. Differential Pathway Activation

- **Early and Peak Inflammation (DSS3 and DSS9)**: Inflammatory pathways, including NF-κB and cytokine signaling, are activated, driving immune cell recruitment and activation. The presence of fibroblasts and epithelial cells suggests activation of repair and regeneration pathways, such as TGF-β signaling and EMT.
- **Recovery Phase (DSS21)**: Immune regulatory pathways, including IL-10 and TGF-β, are activated, facilitating inflammation resolution and homeostasis restoration. Factor analysis indicates significant differences in pathway activation between DSS9 and DSS3, highlighting dynamic changes in response to inflammation.

3. Altered Spatial Organization

- Early Inflammation (DSS3): Redistribution of immune cells to regions like MU3 and SM1 indicates an initial response to tissue damage. Colonocytes show a broader distribution, suggesting adaptive migration to support barrier function.
- Peak Inflammation (DSS9): Immune cells concentrate in regions like MU8, highlighting focal points of inflammation and tissue remodeling. Colonocytes shift to MU7 and MU6, indicating critical roles in managing peak inflammatory responses.
- Recovery Phase (DSS21):** The spatial distribution of cells reverts to a baseline-like pattern, reflecting successful inflammation resolution and tissue architecture restoration. Colonocytes concentrate in MU9, indicating a localized healing process.

4. Key Molecular Signatures

- **Inflammatory Markers**: Increased expression of pro-inflammatory cytokines (e.g., TNF-α, IL-6) and chemokines (e.g., CCL2, CXCL10) during DSS3 and DSS9 facilitates immune cell recruitment and activation. Upregulation of MMPs during peak inflammation indicates active tissue remodeling.
- **Recovery Markers**: Elevated levels of regulatory molecules (e.g., IL-10, TGF-β) during recovery suggest mechanisms for resolving inflammation and promoting tissue repair. The return to a healthy-like distribution pattern in DSS21 indicates activation of pathways involved in tissue repair and resolution.

Supplementary Figure 68: SpatialAgent generated report (5/8)

Spatial Transcriptomics Analysis Report (6/8)

Results - Condition-Specific Effects

5. Potential Biological Mechanisms

- **Inflammation Response**: Cytokine and chemokine signaling drive immune activation and recruitment during early and peak inflammation. The shift in colonocyte distribution and significant factor differences indicate an active inflammatory response.

- **Tissue Remodeling and Repair**: The presence of fibroblasts and epithelial cells, along with upregulated MMPs, suggests mechanisms of tissue remodeling and repair. These processes involve growth factors and are crucial for restoring tissue integrity.

- **Immune Regulation and Resolution:** Treg cells and regulatory cytokines facilitate immune regulation and resolution during recovery, promoting tissue homeostasis. The return to a healthy-like state in DSS21 suggests mechanisms of tissue repair and regeneration.

In summary, the DSS-induced mouse colitis model provides valuable insights into the complex interactions between cellular changes, pathway activations, and molecular signatures throughout the stages of inflammation and recovery. This understanding is crucial for identifying potential therapeutic targets for colitis management, emphasizing the importance of spatial context in understanding cell type distribution and function.

Discussion

1. Integration of Key Findings
The DSS-induced mouse colitis model reveals intricate spatial and cellular dynamics across different stages of inflammation and recovery. Key findings include:

- **Spatial Patterns**: Colonocytes exhibit dynamic shifts in spatial distribution, reflecting their role in barrier maintenance and recovery. Immune cells, particularly macrophages and neutrophils, concentrate in specific regions during peak inflammation, indicating focal points of immune activity. Fibroblasts increase during inflammation, suggesting tissue remodeling.

- **Cell Interactions**: Interactions between immune cells (e.g., macrophages and monocytes) and fibroblasts highlight coordinated efforts in inflammation and tissue repair. Signaling pathways such as TGF-β and NF-κB are central to these processes, driving immune responses and tissue remodeling.

- **Condition Effects**: Cellular changes are marked by redistribution of immune cells during early inflammation, peak immune activity during DSS9, and return to baseline-like states during recovery. Differential pathway activation underscores the dynamic nature of inflammation and resolution.

2. Biological Significance
The spatial and cellular dynamics observed in the DSS-induced colitis model underscore the complexity of inflammatory processes and tissue repair:

- **Immune Response:** The concentration of immune cells in specific regions during peak inflammation highlights the body's targeted response to tissue damage, emphasizing the role of spatial organization in effective immune function.

- **Tissue Remodeling:** The increase in fibroblasts and activation of pathways like TGF-β during inflammation suggest mechanisms of tissue repair and remodeling, crucial for restoring tissue integrity and function.

- **Resolution and Recovery:** The return to baseline-like spatial patterns and activation of regulatory pathways during recovery indicate successful resolution of inflammation, highlighting the body's ability to restore homeostasis.

Supplementary Figure 69: SpatialAgent generated report (6/8)

Spatial Transcriptomics Analysis Report (7/8)

Discussion

3. Potential Mechanisms

- **Cytokine Signaling**: Pro-inflammatory cytokines (e.g., TNF-α, IL-6) drive immune cell recruitment and activation during inflammation, while regulatory cytokines (e.g., IL-10, TGF-β) facilitate resolution and tissue repair.
- **Cell-Cell Interactions**: Direct interactions between immune cells and fibroblasts are crucial for coordinating immune responses and tissue remodeling, ensuring effective inflammation management and repair.
- **Extracellular Matrix Remodeling**: Fibroblast activity and signaling pathways like PI3K and EGFR contribute to changes in the extracellular matrix, affecting cell migration and tissue structure, essential for repair and maintaining tissue integrity.

4. Clinical Implications
Understanding the spatial and cellular dynamics in colitis provides valuable insights for developing therapeutic strategies:

- **Targeted Therapies**: Identifying key regions and cell types involved in inflammation and repair can inform targeted therapies, potentially improving treatment efficacy and reducing side effects.
- **Biomarker Development**: Molecular signatures associated with different stages of inflammation and recovery can serve as biomarkers for disease progression and treatment response, aiding in personalized medicine approaches.
- **Preventing Chronic Inflammation**: Insights into immune regulation and tissue repair mechanisms can inform strategies to prevent chronic inflammation and fibrosis, improving long-term outcomes for patients with colitis.

5. Future Research Directions
- **Spatial Transcriptomics**: Employing advanced techniques like spatial transcriptomics can provide deeper insights into gene expression patterns across different tissue regions, enhancing understanding of spatial dynamics in colitis.
- **Longitudinal Studies**: Conducting longitudinal studies to track cellular and molecular changes over time can provide a comprehensive view of disease progression and recovery, informing therapeutic strategies.
- **Mechanistic Studies**: Investigating the specific roles of signaling pathways and cell-cell interactions in inflammation and repair can uncover novel therapeutic targets, improving colitis management.
- **Comparative Models**: Comparing findings from the DSS-induced colitis model with other models of inflammatory bowel disease can enhance understanding of disease mechanisms and inform cross-disease therapeutic approaches.

In summary, the DSS-induced mouse colitis model offers valuable insights into the spatial and cellular dynamics of inflammation and recovery, highlighting potential therapeutic targets and informing future research directions. Understanding these complex interactions is crucial for developing effective strategies to manage colitis and improve patient outcomes.

Supplementary Figure 70: SpatialAgent generated report (7/8)

Spatial Transcriptomics Analysis Report (8/8)

Conclusions

The analysis of the DSS-induced mouse colitis model provides a comprehensive understanding of spatial patterns, cell interactions, condition effects, and biological implications across different stages of inflammation and recovery. Key conclusions include:

1. **Spatial Patterns**: Colonocytes exhibit dynamic shifts in spatial distribution, reflecting their role in barrier maintenance and recovery. Immune cells, particularly macrophages and neutrophils, concentrate in specific regions during peak inflammation, indicating focal points of immune activity. Fibroblasts increase during inflammation, suggesting tissue remodeling.

2. **Cell Interactions**: Interactions between immune cells (e.g., macrophages and monocytes) and fibroblasts highlight coordinated efforts in inflammation and tissue repair. Signaling pathways such as TGF-β and NF-κB are central to these processes, driving immune responses and tissue remodeling.

3. **Condition Effects**: Cellular changes are marked by redistribution of immune cells during early inflammation, peak immune activity during DSS9, and return to baseline-like states during recovery. Differential pathway activation underscores the dynamic nature of inflammation and resolution.

4. **Biological Significance**: The spatial and cellular dynamics observed underscore the complexity of inflammatory processes and tissue repair. The concentration of immune cells in specific regions during peak inflammation highlights targeted responses to tissue damage. The increase in fibroblasts and activation of pathways like TGF-β during inflammation suggest mechanisms of tissue repair and remodeling.

5. **Clinical Implications**: Understanding these dynamics provides insights for developing therapeutic strategies, such as targeted therapies and biomarker development. Insights into immune regulation and tissue repair mechanisms can inform strategies to prevent chronic inflammation and fibrosis.

6. **Future Research Directions**: Employing techniques like spatial transcriptomics, conducting longitudinal studies, investigating specific roles of signaling pathways, and comparing findings with other models can enhance understanding and inform therapeutic approaches.

Overall, the DSS-induced mouse colitis model offers valuable insights into the spatial and cellular dynamics of inflammation and recovery, highlighting potential therapeutic targets and informing future research directions.

Supplementary Figure 71: SpatialAgent generated report (8/8)

5. Designing additional customized genes upon Xenium 5k panel

5.1. Overview

The standard Xenium 5k panel provides broad coverage but may lack key genes necessary for specialized biological contexts. Here, we designed a customized gene set to enhance spatial transcriptomic resolution in prostate cancer mouse models under different treatments. By selecting 100 additional genes using SpatialAgent , we improved the characterization of stromal, immune, and epithelial compartments: critical for understanding tumor progression and immune interactions. This customization refined cell-type resolution, enhanced clustering quality, and uncovered key ligand-receptor interactions within the tumor microenvironment.

5.2. Customized gene panel improves cell-type resolution and clustering

To assess the impact of the additional genes, we compared clustering performance across different gene selection strategies on the reference scRNA-seq data. Although improvements are subtle in visualizations (Supplementary Fig. 72-75), quantitative metrics indicate enhanced clustering resolution. Specifically, incorporating 100 genes selected via SpatialAgent led to improved clustering scores compared to both the Xenium 5k panel alone and a random selection of 100 genes. The random selection did not contribute meaningfully to the resolution, slightly degrading performance compared to the Xenium 5k panel alone.

5.3. Customized gene panel enhances the capture of cell-cell interactions

Beyond improving clustering, adding 100 genes via SpatialAgent enhanced detection of ligand–receptor interactions in the tumor microenvironment, based on a scRNA-seq reference dataset. Expanded gene coverage revealed richer cell–cell communication, especially in stromal and immune compartments, offering deeper insight into tumor–immune interactions and treatment responses. We compared interaction detection across gene selection strategies and found that SpatialAgent -selected genes yielded more significant ligand–receptor pairs than random panels, highlighting its value.

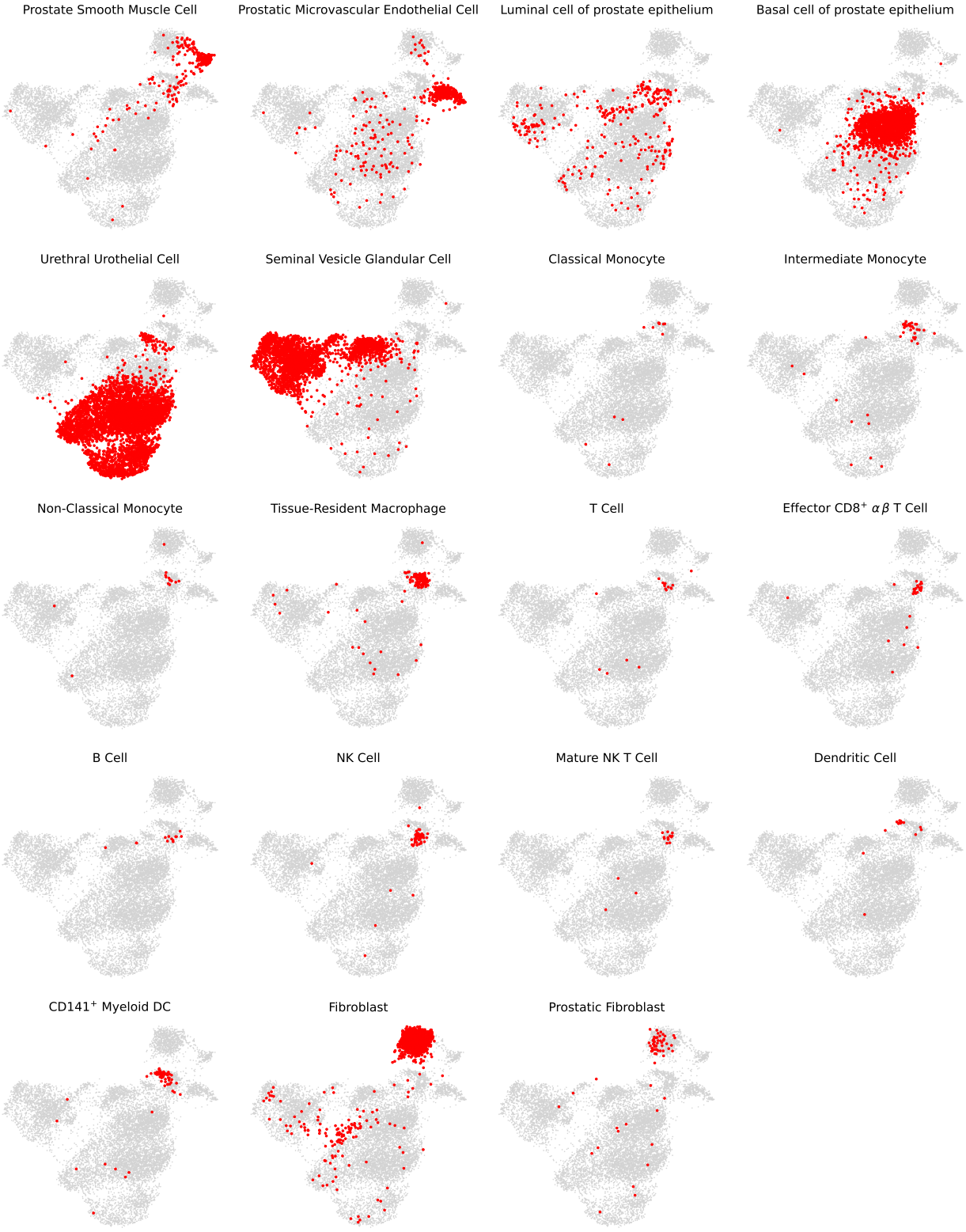
In the Xenium-only analysis (Supplementary Fig. 76), the strongest signals centered on immune-modulatory ligands such as Jag1–Notch2 and Vcan–Tlr2, which dominate fibroblast–immune crosstalk. In contrast, incorporating SpatialAgent (Supplementary Fig. 77) shifted the focus toward laminin - integrin pairs (*e.g.*, Lamb1–Itga3_Itgb1), highlighting enhanced extracellular matrix remodeling and cell-adhesion pathways.

Distinct interaction patterns also emerged across immune subpopulations. Larger or more intensely colored circles for certain ligand-receptor pairs (*e.g.*, Igf1–Insr in Tregs vs. myeloid DCs) suggest diverse regulatory circuits at play. Notably, fibroblasts engaged in multiple robust interactions, reinforcing their role in shaping immune infiltration and the local microenvironment.

Though basal epithelial cells had fewer strong signals, they mediated key integrin-based interactions potentially relevant for tumor maintenance and immune evasion. Overall, integrating Xenium with spatially informed gene panels sharpens our view of tumor–immune dynamics, emphasizing both immune-modulatory and ECM-related signaling.

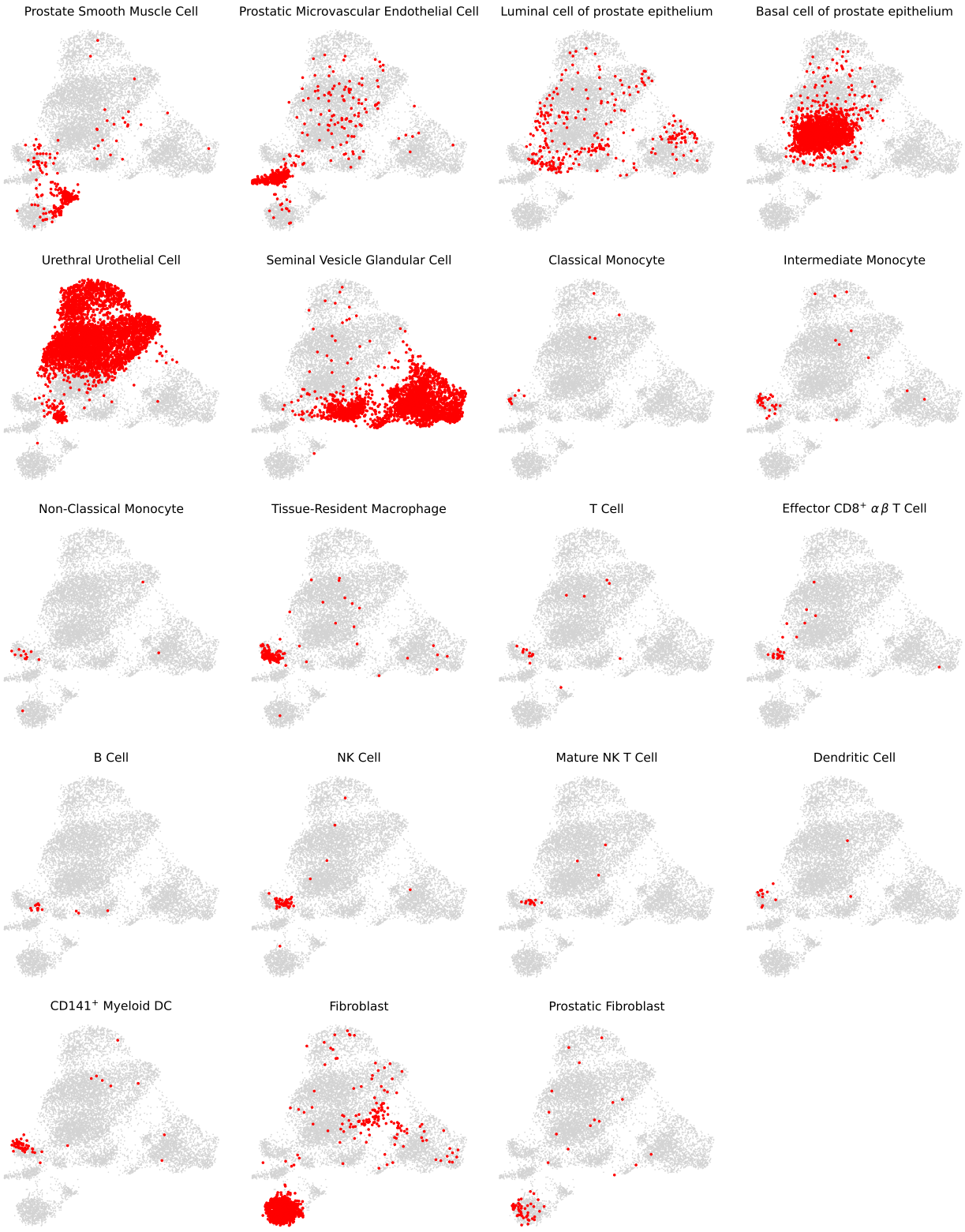
5.4. Summary

Our results demonstrate that a customized gene panel can significantly improve spatial transcriptomic resolution, refining clustering and cell-cell interaction analyses. We will update the results as

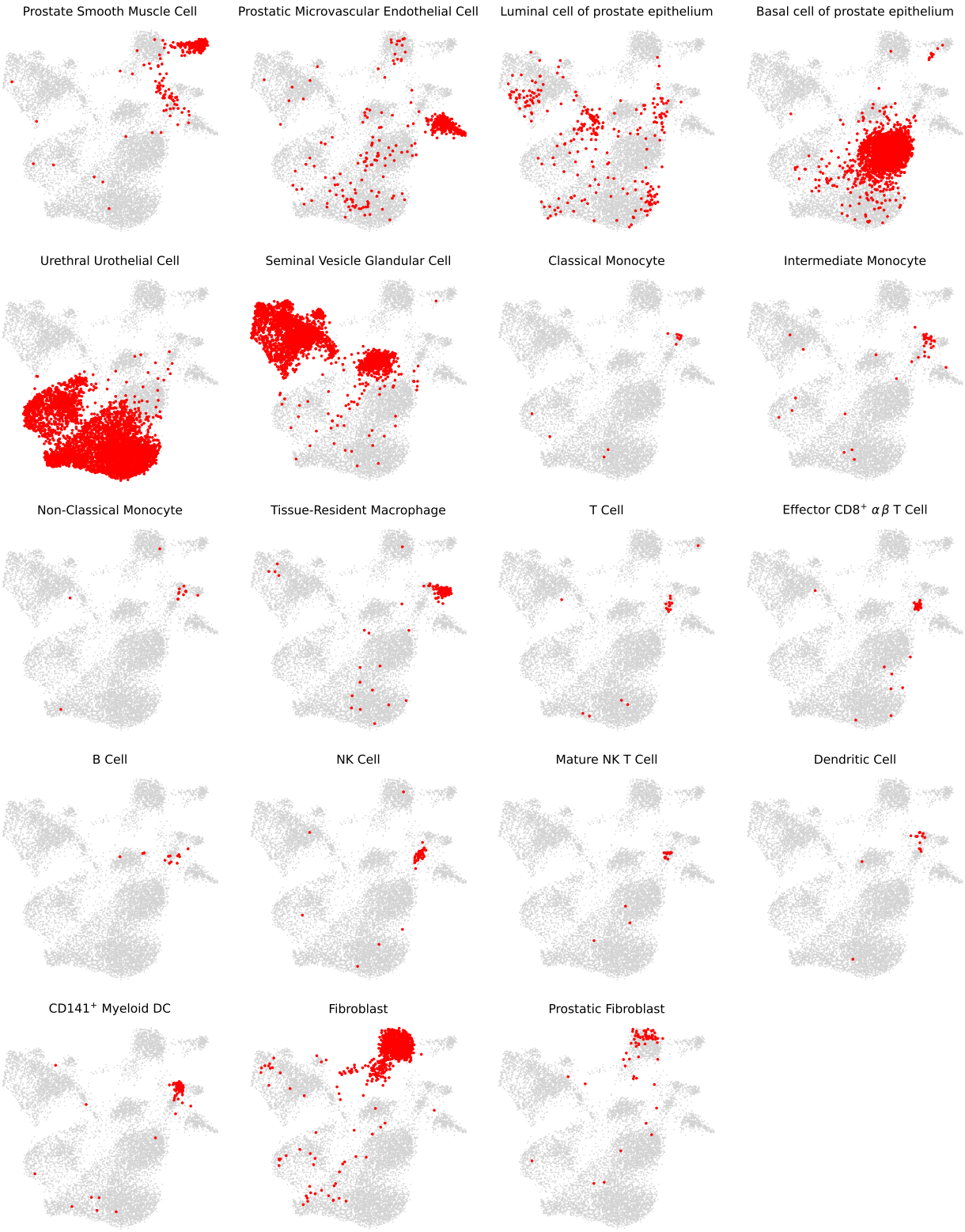


Supplementary Figure 72: UMAPs using Xenium 5k Pan Tissue panel

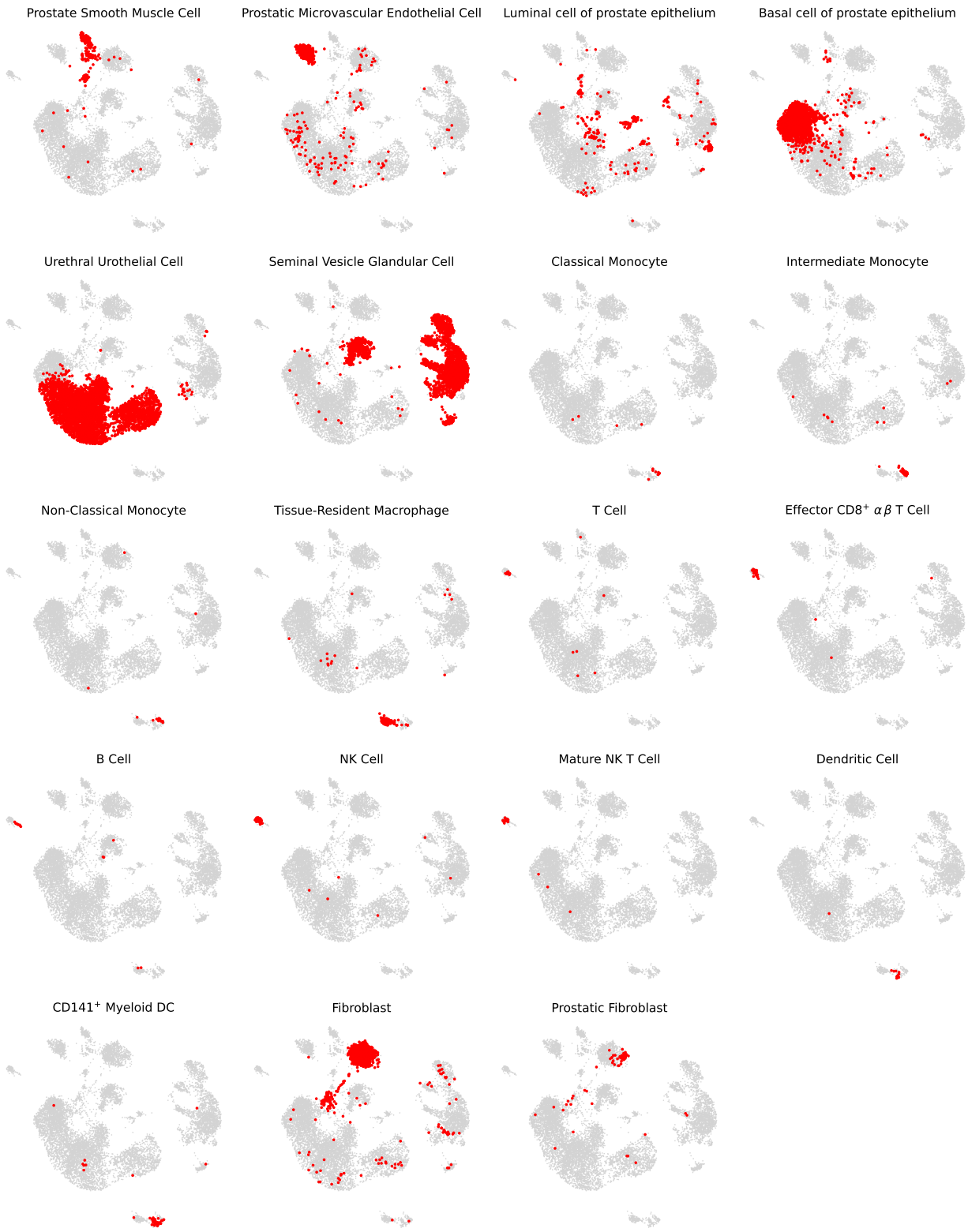
wet-lab data finishes the collection progresses.



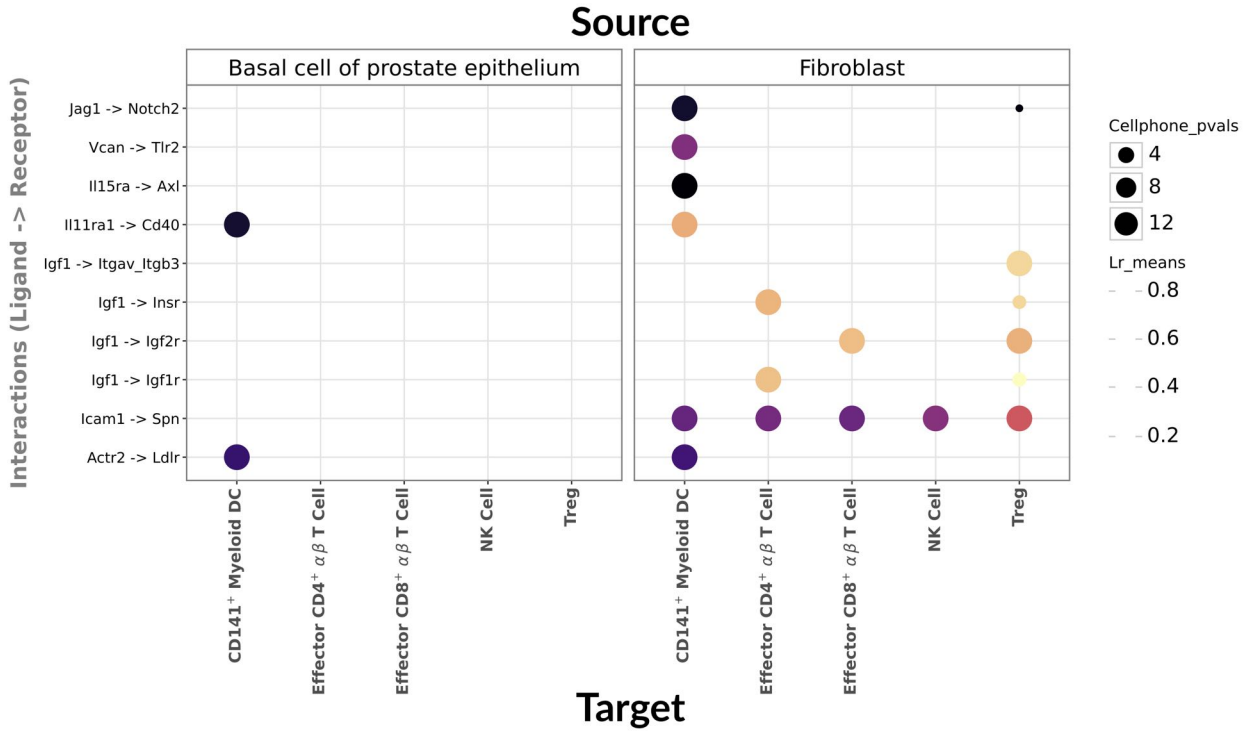
Supplementary Figure 73: UMAPs using Xenium 5k panel and 100 random selected genes



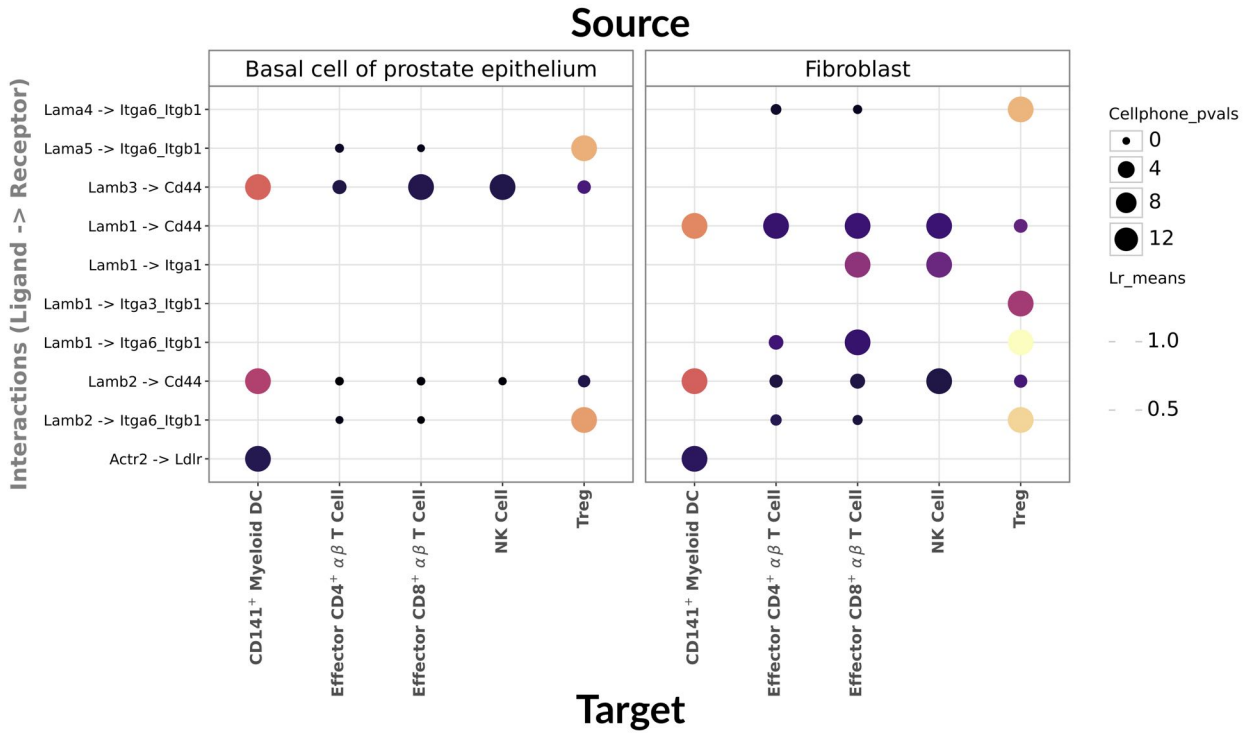
Supplementary Figure 74: UMAPs using Xenium 5k panel and 100 genes from SpatialAgent



Supplementary Figure 75: UMAPs using full 23k genes



Supplementary Figure 76: Ligand receptor analysis using Xenium



Supplementary Figure 77: Ligand receptor analysis using Xenium + SpatialAgent

6. Tools

We attached the description (doc strings) of each tool (Supplementary Fig.78-92) and prompts (Supplementary Fig.93-108.)

retrieve_czi_data
Set up a retriever tool for CZI data with configurable number of datasets to retrieve.

Args:
llm_embed: Embedding model to use
chroma_path: Path to save/load Chroma database
n_datasets: Number of datasets to retrieve (default: 1)

Returns:
create_retriever_tool: Configured retriever tool

Supplementary Figure 78: Tool description (1/15)

read_czi_data
Reads and processes CZI database reference datasets to extract cell types and marker genes, for panel design.

Input:
- dataset_id: a string or a list of strings of the retrieved czi dataset ids
- organism: must be "Homo sapiens" or "Mus musculus"
- iter_round: iteration number for file naming

Output:
A string like "Successfully read reference data, ... and saved the result in csv at the folder ..."

Supplementary Figure 79: Tool description (2/15)

get_czi_info
Retrieves relevant information from the CZI reference dataset based on input criteria, for annotation.

To use this tool, format the input as follows:

Input:
- tissue: the tissue name (e.g., "brain", "lung")
- organism: the species name, must be "Homo sapiens" or "Mus musculus"
- disease: the disease name (e.g., "normal" or a specific disease like "lung adenocarcinoma")
- dataset_id: the dataset ID, must be a valid identifier (e.g., "1234-5678-90ab-cdef")

Output:
A confirmation message indicating where the extracted cell types and tissue information has been saved.

Supplementary Figure 80: Tool description (3/15)

search_panglao
This tool searches for marker genes of each cell type in the Panglao database.

Input:
- save_path: the path of the folder where the reference data info is saved
- organism: must be Hs or Mm (Hs is human, homo sapiens; Mm is mouse, mus musculus)
- tissue: the tissue of target cell types, e.g., brain, liver, pancreas
- iter_round: iteration number for file naming

Output:
A string like "Successfully saved the result at ..."

Supplementary Figure 81: Tool description (4/15)

search_cellmarker2
For each cell type identified in a study of a tissue or organ in human or mouse, this tool searches for marker genes of each cell type in the CellMarker2 database.

Input:
- save_path: path to the folder where the reference data is saved.
- organism: the organism of target cell types: Human or Mouse.
- tissue: the tissue of target cell types: e.g. brain, liver, pancreas, vasculature.
- condition: the condition of target cell types: e.g. normal, disease, treatment.
- iter_round: current iteration round.

Output:
A string like "Successfully saved ..."

Supplementary Figure 82: Tool description (5/15)

gene_importance
For collected information about specific cell types and marker genes, this tool estimates the importance of each marker.

Input:
- save_path: Path where previous information about cell types and marker genes is saved.
- iter_round: Current iteration round.
- num_gene: Number of target genes requested by the user.

Output:
A message like "Successfully saved the importance score at ..."

Supplementary Figure 83: Tool description (6/15)

gene_final_integration
Tool for integrating multiple iteration rounds of gene scoring to produce final gene panel.

Input:
- save_path: Path where previous importance scores of genes are saved.
- num_gene: Number of target genes requested by the user.

Output:
Returns status message "Successfully saved the final gene panel at ..."

Supplementary Figure 84: Tool description (7/15)

scanpy_preprocess
Preprocess spatial transcriptomics data using Scanpy.

Input:
- adata_path: the file path to the raw spatial transcriptomics dataset.

Output:
A confirmation message, e.g., "Successfully saved [results] to [output_path] "

Supplementary Figure 85: Tool description (8/15)

harmony_transfer
Transfers cell type annotations from CZI reference datasets to query spatial transcriptomics data.

Input:
- adata_path: the file path to the preprocessed spatial transcriptomics dataset.

Output:
A confirmation message, e.g., "Successfully saved [results] to [output_path] "

Supplementary Figure 86: Tool description (9/15)

cell_annotator
Clusters preprocessed spatial transcriptomics data and annotates clusters with main-level cell types based on gene markers and transferred labels from reference datasets.

To use this tool, provide the following input:

Input:
- adata_path: the file path to the preprocessed spatial transcriptomics dataset.
- transferred_celltype: the file path to the saved transferred cell type labels.
- data_info: metadata about the dataset, including tissue type, disease and species (e.g., "mouse brain").

Output:
A confirmation message, e.g., "Successfully saved [results] to [output_path] "

Supplementary Figure 87: Tool description (10/15)

main_utag
Identifies main-level spatial neighborhood expression coherent clusters using the UTAG computational method.

Input:
- adata_path: file path to the preprocessed spatial transcriptomics data

Output:
A confirmation message, e.g., "Successfully saved [results] to [output_path].

Supplementary Figure 88: Tool description (11/15)

niche_annotator
Annotates main-level tissue niches by combining multimodal information from
spatial transcriptomics data and existing cell type labels.

To use this tool, provide the following input:

Input:
- adata_path: file path to the preprocessed spatial transcriptomics data
- celltype_csv: file path to the annotated main-level cell type labels
- utag_csv: file path to the main-level utag results of tissue niche clusters
- data_info: metadata about the dataset, including tissue and species (e.g., "mouse brain")
- tissue_txt: text file path to saved possible tissues

Output:
A confirmation message, e.g., "Successfully saved [results] to [output_path] "

Supplementary Figure 89: Tool description (12/15)

summarize_condition
Summarizes condition information from spatial transcriptomics data by analyzing condition-specific patterns and characteristics.

Input:
- adata_path: file path to the spatial transcriptomics data
- condition_key: column name containing condition labels
- data_info: description of the dataset
- tissue_info: description of the tissue

Output:
A confirmation message, e.g., "Successfully saved [results] to [output_path]"

Supplementary Figure 90: Tool description (13/15)

summarize_celltype
Analyzes and summarizes cell type distributions and patterns from spatial transcriptomics data, incorporating condition and tissue region context.

Input:
- adata_path: file path to the spatial transcriptomics data
- summary_condition: path to saved condition information summary
- data_info: description of the dataset
- tissue_info: tissue type information
- condition_key: column name containing condition labels
- cell_type_key: column name containing cell type annotations
- tissue_region_key: column name containing tissue region labels

Output:
A confirmation message, e.g., "Successfully saved [results] to [output_path]"

Supplementary Figure 91: Tool description (14/15)

summarize_tissue_region
Analyzes and summarizes tissue region information from spatial transcriptomics data, incorporating condition and cell type context.

Input:
- adata_path: file path to the spatial transcriptomics data
- summary_condition: path to saved condition information summary
- data_info: description of the dataset
- tissue_info: tissue type information
- condition_key: column name containing condition labels
- cell_type_key: column name containing cell type annotations
- tissue_region_key: column name containing tissue region labels

Output:
A confirmation message, e.g., "Successfully saved [results] to [output_path]"

Supplementary Figure 92: Tool description (15/15)

Prompt used in "retrieve_czi_data" tool for retrieving dataset in CZI cell x gene database
Find the top {n_datasets} matching single-cell dataset{'s' if n_datasets > 1 else ''} in CZI database based on the query.

Input is a string containing: tissue, condition, and organism.
Output must be information of retrieved reference scRNA-seq data.

Each dataset at least contains: dataset_id ands dataset_title.
Additional available attributes can be also included.

Supplementary Figure 93: Tool prompt (1/16)

Prompt used in "get_czi_info" tool for finding the matched tissue name in CZI from user's query
Which is the best matching tissue name for {tissue} in all categories: {all_tissue}?
Output format must be only the best matching tissue name.

Supplementary Figure 94: Tool prompt (2/16)

Prompt used in "get_czi_info" tool for summarizing all possible tissue names for the context
Please organize all possible tissues considering:

- Tissue: {tissue}
- Organism: {organism}
- Disease status: {disease}
- Current tissue list: {all_tissue}

Instructions:
1. Combine similar tissues, remove duplicates but keep details such as spatial info (left, right)
2. Suggest new relevant tissues
3. Format as: tissue1; tissue2; tissue3

Supplementary Figure 95: Tool prompt (3/16)

Prompt used in "get_czi_info" tool for summarizing all possible cell type names for the context

```
Please organize all possible cell types considering:

- Tissue: {tissue}
- Organism: {organism}
- Disease status: {disease}
- Current cell type list: {all_cell_type}

Instructions:
1. Combine similar cell types, remove duplicates
2. Suggest new relevant cell types
3. Format as: celltype1; celltype2; celltype3
```

Supplementary Figure 96: Tool prompt (4/16)

Prompt used in "gene_importance" to aggregate information for each gene

```
Here is the information I collected:

The cell type in reference dataset is {cell_info.get('cell_type', 'Not Available')}
Description of this cell type is {cell_info.get('description', 'Not Available')}

Marker Genes from CZI Database:
- Computational markers: {cell_info.get('czi_marker_genes', 'Not Available')}
- Canonical markers: {cell_info.get('czi_cano_marker_genes', 'Not Available')}

Panglao Database Match:
- Cell type: {cell_info.get('pangdb_cell_type', 'Not Available')}
- Marker genes: {cell_info.get('pangdb_marker_genes', 'Not Available')}

CellMarker2 Database Match:
- Cell type: {cell_info.get('cellmarker2_cell_type', 'Not Available')}
- Marker genes: {cell_info.get('cellmarker2_marker_genes', 'Not Available')}
```

Supplementary Figure 97: Tool prompt (5/16)

Prompt used in "gene_importance" to estimate importance score of each gene based on all previous information

```
Understand the information, your task is to output a list of most important {num_each} genes for this cell type with the highest importance score.

Scoring Rules (0 to 1, where 0 is least important and 1 is most important):
1. each cell type has unique marker genes which are important.
2. When a gene is used to name this cell type and marked as its unique marker, such marker gene is very important. You must include this gene in any case. For example, gene 'Pvalb' is unique marker for cell type 'pvalb GABAergic cortical interneuron'.
3. When a gene appears in multiple database resources like all CZI, Panglao, and CellMarker2 databases, it has a higher confidence to be a valid and important marker.
4. When a gene appears as marker in multiple cell types, it is important. However, it may less useful for differentiate the two cell types so in this case you need additional unique markers.
5. choose the genes that are non repeatable.
6. You must provide the correct number of genes as requested.

Required Output Format:
1: [gene name]
Importance Score: [score]
Reason: [specific justification based on above rules]

[directly repeat for all {num_each} genes, don't ask]
```

Supplementary Figure 98: Tool prompt (6/16)

Prompt used in "gene_importance" to write python code to reformat the previous output into a csv

```
Generate executable Python code to parse a text file into a pandas DataFrame.

Input file {txt_path} contains entries in this consistent format:
   1: GENE1
   Importance Score: 0.9
   Reason: Example reason text

   2: GENE2
   Importance Score: 0.8
   Reason: Another reason text
   ...

Requirements:
1. Create a pandas DataFrame with columns: 'Gene', 'Importance Score', and 'Reason'.
2. Print out any error messages, do not print out success messages
3. Save complete DataFrame to {csv_path}
4. DO NOT include if __name__ == "__main__" or main() function
5. Put execution code directly at module level

Previous error (if any):
{error}

Return only the executable Python code.
```

Supplementary Figure 99: Tool prompt (7/16)

Prompt used in "cell_annotator" to annotate each leiden cluster

```
Please think carefully, and identify the cell type in {info["data_info"]} based on the gene markers.
Optionally refer to the transferred cell type information but do not trust it when the percentage is lower than 0.5.

{{cluster_gene_info}}

The cell type names should come from cell ontology: {info["ontology"]}.
Only provide the cell type name, confidence score (0-1), and detailed reason.
Output format: "name; score; reason". No numbers before name or spaces before number.
Some can be a mixture of multiple cell types.
```

Where the cluster_gene_info is:

```
The enriched genes in this cluster are: {', '.join(marker_genes)}.
For a starting point, the transferred reference cell type composition {cluster_id} is: ...
```

Supplementary Figure 100: Tool prompt (8/16)

Prompt used in "cell_annotator" to reformat the annotation of each cluster

```
Task: Organize and standardize cell type annotations in a scRNA dataset of {data_info}.

Input Cell Types:
{cell_type}

Instructions:
1. Analyze the provided cell type names
2. Combine similar cell type namings
3. Remove redundant naming variations
4. Preserve biological distinctions and details
5. Standardize naming conventions

Required Output Format:
- Return and only return dictionary, whose keys are original cell type names and values are standardized cell type names.
- Example format: {{"original_name": "standardized_name"}}

Note: Ensure consistent naming patterns while maintaining biological accuracy.
```

Supplementary Figure 101: Tool prompt (9/16)

Prompt used in "niche_annotator" to understand the tissue niche information in the reference anatomical images

```
    {"role": "system", "content": """
You are a expert in {data_info}.
Remember that in anatomical images, 'left' and 'right' are defined from the patient's perspective,
which is opposite to the viewer's perspective.
"""},
     {"role": "user", "content": [{"type": "text", "text": """
You need to describe the anatomical regions in the first anatomical image and the red molecular region niche in the second patient sample image based on the following information.

The first image show the locations of anatomical tissue regions.
First, describe the location of anatomical regions in the first anatomical image.
For example, Right ventricle: lower left; Left ventricle: lower right etc.

**Please note that in the anatomical image, left and right are defined from the patient's perspective, which is opposite to the viewer's perspective (second image).**
For example, 'Left atrium' may appear on the right side of the anatomical image, but it should still be described as 'Left atrium' in the patient's image.

Second, describe the spatial location of the majority of the red molecular region niche in the second image in detail, using terms like upper left, upper middle, etc.

Do not include other locations where only a small part is present.
"""},
{"type": "image_url", "image_url": {"url": f"{anatomical_image_b64}"}},
{"type": "image_url", "image_url": {"url": f"{zone_cluster_b64}"}}
]}
```

Supplementary Figure 102: Tool prompt (10/16)

Prompt used in "niche_annotator" to annotate each tissue niche

```
      {"role": "system", "content": (
f"You are a expert in {data_info}. Remember that in anatomical images, 'left' and 'right' are defined from the patient's perspective, which is opposite to the viewer's perspective."
)},
       {"role": "user", "content": [
          {"type": "text", "text": f"""
You need to decide the name of the red molecular region niche in the second image based on following information.

{loc_info}

Based on previous location description and additional information, reason the most possible tissue niche name of the red region. You must output the tissue niche name of the red molecular region niche in the second image in the format:
    'Name:...; Score:...; Reason:...'
""")}]}
```

where the data_info is the metadata about the dataset, including tissue and species (e.g., "mouse brain"),
and the loc_info is LLM's output from the understand the tissue niche information in the reference anatomical images

Supplementary Figure 103: Tool prompt (11/16)

Prompt used in "summarize_condition" to describe each condition

```
I have spatial transcriptomics samples of {tissue_info}.
The dataset description is {data_info}.
The conditions are: {conditions}

Provide a concise description for each condition.
The output format should be: condition: ..., description: ...; condition: ..., description: ...; etc.
```

Supplementary Figure 104: Tool prompt (12/16)

Prompt used in "summarize_condition" to summarize cell type per sample

```
I have spatial transcriptomics samples in "{tissue_info}"
The dataset description is "{data_info}"

Give me analysis of the cell type, especially in terms of their spatial distributions.
Now I'm analyzing cells from a sample whose condition is "{condition_info}"
Percentage of one cell type in each tissue region is: "{celltype_comp}"

For each cell type, think step by step:
  1. what is the cell type's spatial specificity:
     (1) have specificity in one tissue (with more than 80%),
     (2) specific in two or three tissues (sum up to 90%),
     (3) evenly distributed in each tissue (highest percentage is lower than 10%)?
  2. from the annotations, give potential reasons of the cell type's spatial specificity,
     e.g. the cell types functions related to any tissue structures.
  3. is there any interesting or special findings?

Be concise.
```

Supplementary Figure 105: Tool prompt (13/16)

Prompt used in "summarize_condition" to summarize cell type across sample

```
I have spatial transcriptomics samples in "{tissue_info}"
The dataset description is "{data_info}"

Give me analysis of the cell type, especially in terms of their spatial distributions.
All condition information is: "{batch_info}"
The cell type information is: "{cell_type_across_sample}"

Think step by step:
   1.what is pattern of each cell type's spatial distribution change?
   2.what is possible region behind the changes?
   3.Is there any interesting or special findings?

Be concise.
```

Supplementary Figure 106: Tool prompt (14/16)

Prompt used in "summarize_condition" to summarize tissue region per sample

```
I have spatial transcriptomics samples in {tissue_info}.
The dataset description is {data_info}.
Give me analysis of the tissue region, especially in terms of their features and functions.
Now I'm analyzing cells in condition {condition_id} sample.
Percentage of cell types in this tissue region is: {celltype_comp}

For each tissue region, think step by step:
   1. what is the tissue region's cell type composition:
      (1) have specificity in one cell type (with more than 80%),
      (2) specific in two or three cell type (sum up to 90%),
      (3) mixed or evenly distributed of cell type (where the highest percentage is lower than 10%)?
   2. from the annotations, give potential reasons of the tissue region's cell type specificity,
      e.g. the tissue region's functions related to in this sample.
   3. is there any interesting or special findings?

Be concise.
```

Supplementary Figure 107: Tool prompt (15/16)

Prompt used in "summarize_condition" to summarize tissue region across sample

```
I have spatial transcriptomics samples in {tissue_info}.
The dataset description is {data_info}.
Give me analysis of the tissue region features and patterns.

The condition information is {condition_info}.
The tissue region information is {tissue_region_across_sample}.

Think step by step:
   1.what is pattern of each tissue region's cell type composition change?
   2.what is possible region behind the changes?
   3.Is there any interesting or special findings?

Be concise.
```

Supplementary Figure 108: Tool prompt (16/16)