OmniCell: Unified Foundation Modeling of Single-Cell and Spatial Transcriptomics for Cellular and Molecular Insights

Jiangshuan Pang[1,2†], Ping Qiu[4†], Youzhe He[5†], Baolong Li[1,2†], Yiting Deng[1,2†], Jun Wang[6], Adi Lin[2], Lei Cao[2], Fei Teng[4], Haoran Wang[2], Shuangsang Fang[2], Shengkang Li[4], Ziqin Deng[2], Yong Zhang[4,*], Yuxiang Li[4,*], Shaoshuai Li[2,*], Xun Xu[3,*]

1. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China.
2. BGI Research, Beijing 102601, China.
3. BGI Research, Shenzhen 518083, China.
4. BGI Research, Wuhan 430074, China.
5. College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.
6. Bioinformatics Centre, Department of Biology, University of Copenhagen, København Ø 2100, Denmark.

*Correspondence: xuxun@genomics.cn (X.X.), lishaoshuai@genomics.cn (L.S.S.), zhangyong2@genomics.cn (Z.Y.), liyuxiang@genomics.cn (L.Y.X.)

† These authors contributed equally as the first authors.

Abstract

Single-cell RNA sequencing (scRNA-seq) enables characterization of cellular heterogeneity but lacks spatial context, while Spatially Transcriptomics maps gene expression in tissues with limited single-cell resolution. Integrating the complementary strengths of these data into a unified framework remains challenging. Here, we present OmniCell, a foundation model for single-cell and spatial transcriptomics, pretrained on a large-scale corpus of 67 million single-cell and spatial transcriptomic profiles, enabling the unified multi-omics representation learning. As the first foundation model to jointly capture intra-cellular gene expression relationships and inter-cellular spatial dependencies within a unified framework, OmniCell explicitly represents tissue spatial topology by serializing spatially adjacent cells during input construction. Leveraging this unified modeling paradigm, OmniCell generates unified representations of genes, cells, and tissue spatial organization. In zero-shot evaluations, it reliably recovers cell-type structure and gene expression patterns, reconstructs co-expression relationships, and outperforms existing methods across all evaluated tasks, including cell-type deconvolution and spatial domain delineation. Applied to real spatial datasets, OmniCell resolves transitional zones at tumor margins and reveals associated inflammatory activation and immune-cell enrichment, demonstrating its capacity for high-resolution spatial profiling.

Introduction

In multicellular organisms, the diversity of cell types and the spatial architecture of tissues jointly shape physiological function and disease progression. The advent of scRNA-seq has enabled the dissection of cellular heterogeneity, regulatory networks, and state transitions at single-cell resolution, providing a detailed view of the molecular organization within tissues[1-3]**.** To achieve this, various scRNA-seq platforms have been developed, including droplet-based methods (10x Genomics[4], Drop-seq[5]), plate-based approaches (Smart-seq2[6], Smart-seq3[7]). However, because scRNA-seq requires physical dissociation of tissues prior to profiling, it inevitably disrupts the native spatial context, leading to the loss of critical information such as a cell's position within the tissue, its local neighborhood, and the microenvironment in which it resides[8, 9].

In recent years, spatial transcriptomics (ST) technologies have advanced rapidly, enabling researchers to measure gene expression directly within tissues while capturing spatial context. Imaging-based methods such as MERFISH[10], seqFISH[11], and STARmap[12] achieve subcellular resolution but typically target predefined gene panels, limiting genome-wide regulatory insights. In situ sequencing approaches, including BaristaSeq[13], enable direct RNA sequencing in tissue sections. Spatial barcoding technologies, such as Visium[14], Slide-seq[15], offer intermediate spatial resolution (10-100 μm) with transcriptome-wide coverage. More recently, genome-scale platforms like Stereo-seq[16] have produced high-resolution spatial expression maps approaching single-cell resolution. These advances enable the study of cell–cell interactions, tissue domains, and disease microenvironments. Nevertheless, current spatial techniques involve trade-offs between resolution and capture efficiency, and they often suffer from limited molecular coverage and substantial technical noise, particularly at near single-cell resolution, posing significant challenges for computational analysis.

As scRNA-seq and ST technologies mature, the amount of transcriptomic data available for analysis has expanded sharply. Large scRNA-seq collections and ST datasets, generated across diverse tissues, species, and experimental conditions, are now accumulating at an unprecedented pace, offering a broader view of gene regulation and cellular organization. This diversity raisean important question: whether such heterogeneous resources can be distilled into representations that capture biological structure shared across platforms and studies. Drawing on developments in language modeling, several groups have begun to explore large pretrained models for single-cell analysis, such as scGPT[17], scFoundation[18] and Geneformer[19]. They have demonstrated that gene expression can be treated as a molecular language, yielding rich cellular representations[20]. Building on this idea, several studies have extended pretrained

single-cell models to spatial data. For example, scGPT-Spatial[21] adapts scGPT via continual pretraining on spot-level ST datasets. While effective for representation transfer, it is limited by spot-level aggregation, which obscures cell-resolved heterogeneity, and separate modality pipelines for scRNA-seq and spatial data, which hinder seamless cross-omics alignment. NicheFormer, despite being described as a unified foundation model for single-cell and spatial data, does not explicitly capture intercellular spatial dependencies [22]. Similarly, CellPLM is constrained by its reliance on cell-level scRNA-seq modeling, which overlooks intracellular gene interactions, and by its modest spatial dataset of only 2 million spots, which is insufficient to fully represent the diversity of spatial organization and cell–cell interactions [23].

These limitations underscore a key challenge: how to integrate complementary but heterogeneous data within a single framework, enabling models to capture gene regulation, cellular states, and tissue spatial organization simultaneously. To address this, we introduce OmniCell, a unified foundation model pretrained on 67 million single-cell and spatial expression profiles to learn shared representations across gene, cellular, and tissue spatial scales. For ST data, we construct spatially ordered sequences by concatenating gene expression profiles from each cell and its neighboring cells, thereby enabling Transformer-style self-attention[24] to capture tissue architecture and spatial context. In the gene space, OmniCell employs a Mixture-of-Experts (MoE) expression embedding module, in which gene identity determines the routing of expression values to gene-specific expert networks, while additional shared experts capture common expression patterns across genes.The learnable Symmetric Bilinear Output Module is a central component of OmniCell, integrating cell-level and gene-level representations through bilinear interactions. By jointly capturing cellular expression patterns and spatial context, it produces stable embeddings that reflect gene interactions and tissue architecture. This design enhances the model's generalization across platforms and multi-omic datasets. Ablation experiments demonstrate that the module is essential for performance, highlighting its role in generating coherent and biologically meaningful representations across modalities.

Building on this design, OmniCell achieves joint representation learning across genes, cells, and tissue spatial organization within a unified framework. Even in zero-shot settings, it accurately recovers cellular structures, captures cell-specific gene programs, and models cross-cell co-expression relationships. Moreover, OmniCell consistently outperforms existing methods in tasks such as cell-type deconvolution and spatial domain identification. In real tumor spatial datasets, it identifies transitional regions at tumor margins and reveals key features such as inflammation pathway activation and immune cell enrichment, demonstrating its remarkable capacity for spatially resolved biological insights.

Results

A unified foundation model for single-cell and spatial transcriptomics

OmniCell is a unified foundation model designed to jointly represent scRNA-seq and ST data. The model encodes gene identity, expression level, and spatial coordinates within a shared, technology-agnostic framework, allowing it to seamlessly integrate both dissociated and spatially resolved cells. Building on this architecture, OmniCell was pretrained on approximately 67 million cells drawn from scRNA-seq data across 45 human tissues and ST data spanning 26 tissue types (**Figure 1A**). From this diverse training corpus, the model learns coherent gene- and cell-level representations that capture both transcriptional programs and spatial organization. This enables robust cross-modal reasoning and supports a wide range of downstream analyses.

Model Overview. To jointly model scRNA-seq and ST data, OmniCell encodes cells as sequences of gene tokens, enabling Transformer-style attention to capture intra-cellular gene interactions (**Figure 1B**).

Within this unified tokenization scheme, the two modalities are handled with tailored strategies. For scRNA-seq, we represent each cell by a feature vector prioritizing its non-zero expressed genes. To augment cellular representations in cases of limited non-zero signal, we specifically pad the vector with high-variance zero-expression genes. For ST data, sequences incorporate a target cell alongside its spatial neighbors, thereby capturing both intra- and inter-cellular dependencies.

A central challenge in modeling transcriptomic data lies in handling continuous gene expression values. We address this by first applying a soft-rank transformation, which ranks genes within each cell and rescales them to a fixed interval. This operation preserves the relative ordering of expression levels while mitigating the influence of extreme values, effectively normalizing data for subsequent modeling. Beyond normalization, we further introduce a gene-aware MoE value embedding that combines a unified gene embedding matrix with a MoE-based value embedding module. This design adaptively encodes gene expression in a context-sensitive manner, capturing subtle variations while maintaining biological fidelity and cell-type specificity.

The architectural backbone of OmniCell comprises 10 Transformer layers with RMS normalization, 2D rotary self-attention, and SwiGLU feedforward networks, totaling approximately 74 million parameters. At the output stage, a learnable symmetric matrix integrates gene- and cell-level features: during training, it is optimized under a soft-rank regression objective, while at inference, spectral decomposition projects cell embeddings into a stable low-dimensional space. Together, this architecture produces robust, biologically interpretable embeddings suitable for diverse downstream analyses, including cell-type annotation, spatial

domain delineation, and the construction of gene co-expression networks.

Downstream Tasks. We next evaluated OmniCell across both single-cell and spatial transcriptomic tasks, (**Figure 1C**). At the gene level, OmniCell learns cell-type–specific gene representations that faithfully capture transcriptional programs, enabling marker gene identification and co-expression pattern discovery. At the cell level, the model generates embeddings that integrate transcriptional and spatial context, supporting robust clustering, cell-type annotation, batch correction, spatial domain delineation, and spatial deconvolution. Notably, across these diverse tasks, the model maintains stability under increased single-cell noise and varying tissue contexts. Taken together, these results demonstrate that by jointly learning high-quality gene- and cell-level embeddings, OmniCell effectively bridges dissociated and spatially resolved data, yielding biologically interpretable representations that generalize across modalities.

Ablation Study. Beyond overall performance, we sought to dissect the contribution of key architectural components through systematic ablation studies across clustering and cell-type annotation benchmarks (**Figure 1D**). Removing the gene-aware MoE value embedder modestly reduced performance, reflecting its role in capturing cell-type–specific expression patterns. Excluding spectral subspace projection impaired embedding robustness, and eliminating the symmetric bilinear output module led to the largest performance drop, confirming its central role in modeling cell–gene relationships. We also examined the impact of training data composition: training on scRNA-seq data alone without spatial supervision degraded clustering and annotation accuracy, highlighting the importance of integrating spatial transcriptomics to enrich data diversity and provide structural constraints. Collectively, these ablation results demonstrate that each component of OmniCell contributes synergistically to its robust, biologically interpretable performance.

Cell-Level and Dataset-Level Gene Embeddings Reveal Cell-Type–Specific Signatures and Functional Modules in scRNA-seq and ST data.

We first evaluated the quality of cell-level gene embeddings by extracting contextualized representations from the Transformer model and projecting them into a lower-dimensional space. For this analysis, we employed Zhou's dataset, which enabled us to capture both global gene properties and cell-specific activities. To assess how well these embeddings preserve cell-type information, we performed clustering on the gene embeddings and computed the Adjusted Rand Index (ARI) against known cell-type labels. OmniCell consistently achieved higher ARI scores compared to scGPT, scFoundation, and GeneFormer (**Figure 2A**), indicating that its embeddings retain cell-type information more accurately. To further illustrate these findings, we generated UMAP visualizations of the embeddings for several key genes, including *APOE* (**Figure 2B**), *TREM2* (**Figure 2C**), and *LHFPL3* (**Supplementary Figure 2C**).

These visualizations reveal that OmniCell more clearly distinguishes between cell types, particularly in resolving clusters of high and low expression within microglial cells—patterns with known associations to Alzheimer's Disease that may reflect pathological alterations with diagnostic or therapeutic relevance. We also examined how the model allocates attention across genes by aggregating attention weights into a cell-by-gene attention matrix, which captures the relative importance of each gene across cell types. Multiclass Receiver Operating Characteristic (ROC) analysis (**Figure 2D**) demonstrated that OmniCell outperformed competing models, with all cell types achieving Area Under the Curve (AUC) values above 0.9, underscoring its robust capability in leveraging gene attention to differentiate cell populations.

Having demonstrated that OmniCell embeddings preserve cell-type information, we next investigated whether they could be leveraged to identify marker genes for specific cell types. Using cell-level gene embeddings generated by the foundation model (as described in the Methods), we ranked genes for each annotated cell type in Zhou's scRNA-seq dataset based on their specificity in the embedding space and filtered by transcriptional abundance to retain highly expressed, cell-type–specific markers. Encouragingly, many of the embedding-derived markers corresponded to canonical genes reported in previous studies, such as *SLC17A7* for excitatory neurons[25], *ERBB4* for inhibitory neurons[26], *GFAP* for astrocytes[27], *LRMDA* for microglia[28], *PLP1* for oligodendrocytes[29], and *PTPRZ1* for oligodendrocyte precursor cells (OPCs)[30] (**Figure 2E and Supplementary Figure 2B**). Notably, the majority of the embedding-derived markers for most cell types were also present in the top marker gene lists identified by conventional expression-based ranking methods (**Figure 2F**), supporting the biological validity of the embeddings. Beyond known markers, OmniCell embeddings also identified additional cell-type–enriched genes with strong discriminative power (AUC > 0.7), including *SYT1* for excitatory neurons[31] and *SPARCL1* for astrocytes[32] **(Figure 2G**). Together, these results revealed that OmniCell embeddings not only recapitulate known cell-type–specific gene expression patterns but also enable the systematic identification of novel markers directly from the representation space.

While cell-level embeddings capture gene properties within individual cells, understanding broader co-expression relationships requires aggregating information across the entire dataset. To this end, we generated dataset-level gene embeddings by integrating cell-level embeddings through a generalized PageRank graph neural network [33] (GPR-GNN), which propagates each gene's embedding through a cell similarity graph to reinforce co-expression relationships and context-dependent interactions (**Supplementary Figure 2A**). Final embeddings were obtained by averaging across all cells, capturing both intrinsic expression patterns and dataset-wide functional properties. To validate these embeddings, we

constructed a gene co-embedding network using cell-type–specific gene sets identified previously. Louvain community detection revealed six distinct modules (Modules 0–5) that exhibited highly cell-type–specific activation patterns aligned with known transcriptional programs (**Figure 2H-I**): Module 0 for astrocytes; Module 1 for oligodendrocytes; Module 2 for inhibitory neurons; Module 3 for microglia; Module 4 for both excitatory and inhibitory neurons; and Module 5 for oligodendrocyte precursor cells. Gene Ontology enrichment analysis confirmed functional coherence, with modules enriched for cell-type–appropriate processes including neurotransmitter transport, myelination, synaptic signaling, immune function, and oligodendrocyte differentiation (**Figure 2J**). We further assessed spatial coherence by performing joint module identification on spatial transcriptomic data from Alzheimer's disease and control human hippocampus samples. Five spatial modules emerged with region-specific expression patterns corresponding to known anatomical boundaries: excitatory neuron–enriched modules showed peak activity in the pyramidal cell layers of CA1–CA4 and the granule cell layer of the dentate gyrus, while glial-enriched modules localized to the GC region, consistent with the spatial distribution of astrocytes and oligodendrocyte-lineage cells (**Figure 2K-M and Supplementary Figure 2D, E**)

Collectively, these analyses demonstrate that dataset-level gene embeddings derived from OmniCell capture biologically meaningful gene–gene relationships, organize genes into coherent functional modules aligned with cell-type–specific transcriptional programs and anatomical structures, and provide a principled framework for investigating coordinated gene expression patterns in both single-cell and spatial transcriptomic contexts.

Omnicell Enables Cross-Species Generalization and Robust, Interpretable, Lineage-Aware Cell Embeddings via a MoE Value Embedder.

To evaluate OmniCell's capacity for generalized cell representation learning, we benchmarked its embeddings against those generated by existing pretrained models across human (kidney and blood) and mouse (brain and muscle) datasets. As shown in **Figure 3A**, OmniCell consistently achieved the highest clustering performance in both ARI and NMI, outperforming all baselines. Notably, although OmniCell was pretrained exclusively on human data, it maintained clear cell-type separability in mouse datasets through homology-based gene projection, demonstrating robust cross-species transferability. UMAP visualizations further corroborated these quantitative results (**Figure 3B, 3C and Supplementary Figure 3A, 3B**), showing that OmniCell embeddings form compact and biologically coherent clusters with superior intra-cluster cohesion and inter-cluster separation compared to scGPT and GeneFormer.

Beyond clustering accuracy, a practical challenge in single-cell analysis is mitigating batch effects

while preserving biological variation. We therefore assessed OmniCell's performance on the Zhou dataset, which comprises 11 distinct batches (**Figure 3D, 3E**). OmniCell achieved a favorable balance between integration and biological preservation, effectively mixing technical batches while maintaining underlying structure. Consistent results were observed in a multi-batch human pancreas dataset (**Supplementary Figure 3C, 3D**), where diverse cell types remained well separated across experimental conditions, highlighting OmniCell's robustness to technical variation. Given that real-world single-cell data often suffer from dropout and missing values, we further evaluated embedding stability by applying progressive random dropout to the Zhou dataset (**Figure 3F**). OmniCell retained substantially higher clustering stability than existing models under increasing sparsity, demonstrating its ability to encode noise-resilient transcriptional features. Together, these results establish OmniCell as a robust foundation model capable of reliable generalization across species, platforms, and data quality conditions.

The strong performance observed above prompted us to investigate whether OmniCell's MoE value embedder contributes to its representational capacity. We hypothesized that this modular routing mechanism enables encoding of lineage- and function-specific transcriptional programs by dynamically allocating computational resources. To test this, we computed mean routing weights and standardized activations for each expert across annotated cell types in the human immune cell dataset. The resulting heatmap (**Figure 3G**) revealed lineage-stratified activation patterns, with major populations—including T/NK cells, B cells, myeloid cells, and erythroid progenitors—clearly delineated by distinct expert preferences. Expert Preference Clustering ($Z > 1.0$; **Supplementary Figure 3E**) reinforced these findings, highlighting clear lineage-specific associations across major immune lineages. These activation patterns suggest that the MoE router autonomously partitions computation into biologically coherent modules mirroring immune cell differentiation hierarchies, indicating the model encodes interpretable, lineage-aware functional specialization. In summary, OmniCell excels at learning stable, biologically meaningful cell embeddings, outperforming existing models in clustering, cross-species transfer, and batch-effect mitigation. Its MoE architecture enables dynamic, lineage-specific routing that captures cellular heterogeneity while maintaining interpretable functional organization.

Multi-Scale Spatial Transcriptomic Benchmarking of OmniCell Across Cellular and Tissue Hierarchies.

To systematically evaluate OmniCell's capacity to capture spatially organized biological structure, we conducted a multi-scale assessment spanning single-cell resolution to tissue-level architecture. At the cellular scale, spatial clustering evaluates whether embeddings derived from individual cells faithfully

recover established cell-type identities. At the tissue scale, spatial domain delineation incorporates multi-cell inputs, jointly encoding a central cell with its spatial neighbors to resolve mesoscale anatomical organization. This hierarchical framework enables OmniCell to integrate molecular identity with spatial context within a unified embedding space.

We first assessed spatial clustering using a comprehensive MERFISH-based mouse brain atlas encompassing 31 spatial transcriptomic datasets that capture molecular and cellular dynamics across aging. Single-cell embeddings were subjected to Leiden clustering across multiple resolutions (0.1–1.0), with optimal clustering determined by concordance with reference annotations. OmniCell consistently achieved superior performance relative to scGPT-spatial and Nicheformer, as quantified by both NMI and ARI (**Figure 4A**). Representative visualization of slice 10_0 demonstrated that OmniCell embeddings yield well-resolved clusters in UMAP space (**Figure 4B**) that accurately recapitulate anatomical organization upon projection to spatial coordinates (**Figure 4C**), with comparable results for slice 10_2 (**Supplementary Figure 4A-B**). In contrast, baseline methods exhibited substantial intermixing of spatially adjacent regions.

Moving from single-cell to tissue-scale analysis, we next evaluated spatial domain delineation on a mouse cortex dataset with expert-curated regional annotations encompassing hippocampus, internal capsule, cortical layers 2–6, pia Layer 1, ventricle, and white matter. In this multi-cell configuration, each embedding integrates transcriptomic information from a central cell and its spatial neighborhood. OmniCell achieved optimal reconstruction of cortical architecture (NMI = 0.6623, ARI = 0.6108), significantly outperforming all competing methods (**Supplementary Figure 4C-D**). These results demonstrate that explicit incorporation of spatial context enables OmniCell to resolve mesoscale tissue organization beyond individual cell-type classification.

Having established OmniCell's performance on well-annotated datasets, we sought to challenge the model with a more complex scenario: the Stereo-seq LC5-M hepatocellular carcinoma dataset, which encompasses tumor core, peri-tumoral transition zone, and adjacent non-malignant parenchyma without predefined annotations [34]. Comparative analysis across six methods revealed that only OmniCell successfully delineated a discrete transitional domain (Cluster 3) at the tumor-parenchyma interface (**Figure 4D**). Gene Ontology enrichment analysis of this transition zone revealed significant activation of pathways including acute-phase response, inflammatory signaling, complement activation, coagulation regulation, and copper ion detoxification (**Figure 4E**). Consistent with these inflammatory signatures, spatial mapping of cell type distributions (**Figure 4F**) and quantitative composition analysis (**Figure 4G)**

demonstrated marked enrichment of immune populations—particularly macrophages and T/NK cells—within the transition zone relative to tumor and paratumor tissues[34]. The transition zone also exhibited pronounced activation of the copper detoxification pathway (Figure 4H) and elevated expression of metallothionein family genes (*MT2A, MT1E, MT1X, MT1M, MT1F*)[35](**Figure 4I**). Given that metallothioneins modulate immune cell function and oxidative stress responses within the tumor microenvironment, their spatial co-localization with immune infiltrates reinforces the biological coherence of this OmniCell-defined niche. Concordant spatial gradients observed for acute inflammatory response, complement activation, and coagulation regulation pathways (**Supplementary Figure 4E**) further delineate a spatially coordinated inflammatory-metabolic program characteristic of the invasive tumor margin. These findings establish the transition zone as a molecularly and immunologically distinct microenvironment uniquely resolved by OmniCell. More broadly, our multi-scale analyses demonstrate that OmniCell embeddings effectively bridge molecular and spatial hierarchies—from accurate cell-type recovery at single-cell resolution, through faithful anatomical domain reconstruction at tissue scale, to de novo identification of functionally distinct niches within complex pathological tissues.

OmniCell achieves state-of-the-art performance in cell type annotation and spatial deconvolution.

OmniCell was designed to jointly leverage scRNA-seq and ST data, capturing spatially resolved contextual information while learning fine-grained cellular features. By integrating these complementary modalities, OmniCell enhances representation capacity, more faithfully characterizes cellular states, and resolves cellular heterogeneity within complex biological environments. To quantify these advantages, we systematically benchmarked OmniCell against leading foundation models for single-cell analysis, including scGPT, scFoundation, and Geneformer. Across five datasets, OmniCell consistently achieved the highest annotation performance as reflected by F1 score and accuracy (**Figure 5A**), demonstrating superior generalizability across heterogeneous biological contexts. A particularly stringent test of annotation quality is the ability to identify rare cell populations. On the Zheng68k dataset (**Figure 5B and 5C**), OmniCell predicted CD34+ cells with 94% accuracy despite their comprising only 0.29% of the dataset—a performance 13 percentage points higher than the next-best models, scGPT and Geneformer. For extremely rare populations, OmniCell achieved 20% accuracy for CD4+/CD45RA+/CD25- Naive T cells (2.8% of dataset) and 6% for CD4+ T Helper2 cells (0.14% of dataset), whereas all competing models failed entirely. Summary metrics across all cell types (**Figure 5E**) and full confusion matrices (**Supplementary Figure 5A–B**) further underscore OmniCell's capacity to capture biologically meaningful low-frequency states.

Robust annotation must also generalize across technical batches. Using the two-batch hPBMC dataset with train-test splits across batches, UMAP projections showed that OmniCell maintained close alignment between predicted and true labels (**Figure 5D**), outperformed competing models across all metrics (**Figure 5E**), and achieved the most accurate predictions in full-cell-type confusion matrices (**Supplementary Figure 5C**). This demonstrates its ability to disentangle technical confounders for scalable cross-study annotation. Consistent superiority was observed on the lung cancer dataset (**Figure 5E; Supplementary Figure 5D**), confirming robust performance across complex, heterogeneous contexts. These advantages extend naturally to spatial transcriptomics, where cell-type annotation must additionally respect tissue architecture. Across 30 benchmark spatial datasets, OmniCell outperformed leading methods including scGPT-spatial, Nicheformer, and cell2location (**Figure 5G**), with boxplots demonstrating the most stable F1 and accuracy distributions—indicative of stronger generalization. OmniCell further reconstructed spatial maps that closely matched ground-truth cellular organization (**Figure 5F**), achieving higher fidelity than competing methods (**Supplementary Figure 5E**). Confusion matrices from MERFISH-based mouse brain atlas slices 9_0 and 11_2 (**Supplementary Figure 5F and 5G**) corroborated its superior accuracy in recovering fine-grained tissue architecture.

In summary, OmniCell robustly integrates scRNA-seq and ST data, achieving superior performance in cell-type annotation, rare cell detection, and spatial reconstruction. Its consistent results across diverse datasets highlight both its robustness and utility for dissecting complex tissue architectures.

Discussion

In this study, we present OmniCell, a unified foundation model that integrates scRNA-seq and ST data within a single architecture. By learning shared representations across genes, cells, and tissue spatial organization, and by incorporating spatial neighborhood serialization, gene-specific MoE value embedder, and a learnable symmetric bilinear output module, OmniCell achieves robust performance across diverse tissues, platforms, and species. Importantly, it captures both cell-intrinsic gene programs and intercellular spatial dependencies, enabling accurate zero-shot recovery of cellular structures, co-expression relationships, and spatial domains.

We demonstrate OmniCell's capability across multiple downstream tasks at cellular, gene, and spatial levels. At the cellular level, it enables integrative clustering, resolving heterogeneity across multi-slide and multi-platform datasets while mitigating technical batch effects. At the gene level, OmniCell generates

cell-specific gene embeddings that capture expression programs of individual cell types, alongside dataset-level gene embeddings that reveal co-expression networks and regulatory relationships across cells. At the spatial level, the model produces spatially aware embeddings that enhance cell-type deconvolution and spatial domain delineation. In real tumor spatial datasets, OmniCell identifies transitional regions at tumor margins and highlights biologically relevant features such as immune cell enrichment and inflammation pathway activation, illustrating its capacity to provide meaningful insights across molecular, cellular, and tissue scales.

Despite its advances, OmniCell faces several limitations. Representation quality may decrease in extremely sparse spatial datasets, and rare cell types or lowly expressed genes may be less reliably captured. Training on tens of millions of cells and spots also requires substantial computational resources. Moreover, while OmniCell provides a unified multi-omic framework, modeling complex scenarios—such as dynamic spatial processes, additional omics layers, or cellular responses to perturbations—remains challenging, as models trained on control cells alone often fail to predict transcriptional shifts induced by chemical or genetic perturbations[36-38]. Future work could address these limitations by incorporating additional modalities, perturbation-aware objectives, and temporal or multi-modal spatial information, enhancing OmniCell′s applicability and biological insight.

Overall, OmniCell establishes a unified framework for multi-omic representation learning, bridging scRNA-seq and ST data. By capturing gene regulation, cellular states, and tissue spatial organization within a single architecture, it provides a versatile tool for exploring biological complexity across diverse contexts and lays the groundwork for future integrative studies in single-cell and spatial omics.

Methods

Data collection

Single-cell data for pretraining. scRNA-seq datasets were obtained from publicly available CELL×GENE repositories. For datasets originating from CELL×GENE, we used the CZ CELL×GENE Discover Census v.2023-07-15, retrieved from the project's publicly accessible Amazon S3 bucket. All datasets associated with this census release were downloaded directly from S3 to ensure full consistency with the published resource. The collected datasets span a broad range of human tissues, including brain, blood, lung, liver, digestive and immune systems, skin, and kidney, among others. The final corpus comprises 38 million human single cells across 45 tissues, covering multiple sequencing platforms, developmental stages, and disease contexts curated from the CZ CELL×GENE database. In total, single-cell profiles account for approximately 57.8% of all cells analyzed in this study.

Spatial transcriptomic data for pretraining. Spatially resolved transcriptomic data were generated primarily using the STOmics Stereo-seq whole-transcriptome platform. These datasets originate from large-scale internal resources produced at the BGI Research Institute. Stereo-seq data from multiple spatial resolutions—including Bin20 and Cellbin—account for approximately 27.2% and 14.8% of the total cell count, respectively. Portions of these datasets are publicly accessible through STOmics data portals, whereas the remaining fraction is part of ongoing internal releases. All Stereo-seq data were processed following the standard analysis pipeline, including spatial binning, UMI quantification, and spatial quality assessment.

Datasets for downstream tasks and evaluations. To rigorously benchmark our framework across diverse biological contexts, we curated a comprehensive collection of datasets spanning single-cell analysis, batch integration, cell-type annotation, spatial domains, spatial deconvolution, and model interpretability. This collection enables extensive evaluation across both single-cell and spatial transcriptomic modalities.

For gene module analysis, we used the human single-nucleus transcriptomic dataset from Zhou[39] and the human hippocampus dataset from both Alzheimer's disease (AD) and control samples profiled by Stereo-seq[40]. Cell-level tasks, including clustering and annotation, were assessed using a diverse panel of datasets primarily sourced from the CELL×GENE repository, covering multiple human and mouse tissues (e.g., blood[41], kidney[42], brain[43], muscle[44]) and sequencing platforms to ensure heterogeneity and cross-platform generalizability. Batch integration performance was further evaluated on the hPBMC[45] and Zhou[39] datasets, while cell-type annotation benchmarks included five CELL×GENE atlas-level datasets(e.g., Brain[46], Hippocampus[46], Cortex[47], Immune[48], Great_Apes[49] in addition to the Zheng 68k[4] and hPBMC[45] datasets.

For spatial analyses—including clustering, domain delineation, and deconvolution—we employed high-resolution spatial transcriptomic datasets, comprising the MERFISH mouse brain aging dataset (31 samples)[50] from the Allen Institute, the LC5-M Stereo-seq dataset[34] of invasive regions in human liver cancer, and the osmFISH mouse cortex dataset[51] from the Linnarsson Lab. All datasets are publicly accessible at https://modelscope.cn/datasets/PJSucas/OmniCell-test-data .

OmniCell architecture and pretraining

A unified foundation model was developed to generalize across scRNA-seq and ST data by encoding gene identity, expression levels, and spatial information in an assay-agnostic format. Gene identities are represented using Ensemble IDs (e.g., ENSG00000139618) and mapped to integer indices through a

standardized vocabulary to ensure consistency across platforms. Each cell is denoted as $i \in 1,2,\ldots,N$, where N is the total number of cells across all the datasets For each cell i, the model covers: (1)g_{ij}, the integer-mapped identifier for the gene j; (2) v_{ij}, the raw expression value of the gene j in the cell i; and (3) (x_i, y_i), the spatial coordinates of the cell i. While ST cells retain their native spatial positions, scRNA-seq cells are assigned default coordinates (0, 0), preserving structural consistency while reflecting the absence of spatial resolution in standard single-cell protocols.

Rank-Based Normalization of Gene Expression via Soft Rank Scaling

To harmonize expression values across sequencing technologies and mitigate platform-specific variability[52], we implemented a rank-based normalization strategy. For each cell i, genes are ranked by their expression levels in descending order, producing a rank ρ_{ij} for each gene j, where $\rho_{ij} \in \{1,2,\ldots,G\}$, and G represents the total number of unique expression values in cell i. To preserve expression ordering while smoothing magnitude difference, we define a soft rank r_{ij} by linearly rescaling the ranks into a predefined interval $[a,b]$ as follows:

$$r_{ij} = a + (b-a)\cdot\frac{G-\rho_{ij}}{G-1} \quad (1)$$

In our implementation, we set $a = 0$ and $b = 5$, such that $r_{ij} \in [0,5]$. This transformation retains the relative expression order of genes within each cell while yielding a bounded, continuous representation suitable for downstream modeling. The soft rank reduces the influence of extreme values and enables effective integration across both scRNA-seq and ST datasets, without reliance on raw expression magnitudes.

Construction of Input Sequences for Transformer Modeling

To integrate scRNA-seq and ST data into a unified Transformer-based modeling framework, we construct gene-centric token sequences for each sample. The embedding of each token is derived from the sum of two components: gene identity and expression level. This dual-component embedding design enables the model to effectively combine discrete genetic signatures with context-dependent expression profiles, facilitating a comprehensive analysis of transcriptional heterogeneity across diverse data modalities.

Unified Sequence Construction Across scRNA-seq and Spatial Transcriptomics. For scRNA-seq data, each cell is treated as an independent sequence, with genes prioritized for inclusion based on

biological relevance: non-zero expression genes are selected first, followed by high-variance zero-expression genes to maintain signal diversity. In the case of ST data, such as Stereo-seq, we extend this approach by incorporating local tissue context. Each spatial token sequence is centered on a given cell and includes gene-level profiles from its k nearest spatial neighbors, determined by Euclidean distance in tissue coordinates. Using the same gene prioritization strategy as in scRNA-seq, n genes are selected from each of the 1+k cells. The gene sequences from each cell are enclosed with <RNA_START> and <RNA_END> tokens, resulting in a structured, multi-cellular input sequence that jointly encodes both spatial relationships and transcriptional states. These special tokens serve as markers to denote the boundaries of cellular contexts and guide position-aware encoding within the spatial neighborhood.

Gene Identity Embedding. Each gene is associated with a unique Ensemble identifier and mapped to an integer index through a unified vocabulary. A shared embedding matrix $E_{\text{gene}} \in \mathbb{R}^{G\times d}$ then transforms this index into a dense vector representation:

$$e_{ij} = E_{\text{gene}}(j) \tag{2}$$

where G is the total number of unique genes, d is the embedding dimension, and (i,j) refers to the j-th gene in the i-th cell or spatial spot. This embedding captures global identity information and is consistent across scRNA-seq and spatial transcriptomics.

Gene-Aware Value Embedder via MoE. Inspired by DeepSeek-V1[53], we introduce a gene-aware MoE architecture to quantitatively embed each gene's expression level $\boldsymbol{v_{ij}}$, capturing both the magnitude and the biological context of transcriptional activity. This module is conditioned on the gene identity embedding, enabling the model to adapt its encoding strategy based on gene-specific context. The architecture integrates $\boldsymbol{S}$ shared experts that model general expression-magnitude patterns across all genes, together with $\boldsymbol{R}$ routing experts that are dynamically selected through a learnable routing network, allowing the expression magnitude of each gene to be encoded in a gene-sensitive and context-adaptive manner.

The routing mechanism takes the gene embedding $\mathbf{e}_{ij}$ as input and produces a softmax distribution over the routing experts:

$$p_{ij} = \text{softmax}(W_2 \cdot \text{ReLU}(W_1 \cdot e_{ij})) \in R^{r\times 1} \tag{3}$$

where r denotes the number of routing experts.

To enhance the representational capacity and generalization of the model, only the top k experts with

the highest probabilities (denoted as $\mathcal{T}_{ij}$) are activated for each gene expression embedding. A sparse weight vector $\boldsymbol{\alpha}_{ij} \in \mathbb{R}^r$ is constructed, retaining the original routing probabilities for the selected experts while assigning zero to the remainder:

$$\alpha_{ijk} = \begin{cases} p_{ijk}, & k \in \mathcal{T}_{ij} \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

Importantly, these sparse weights are not renormalized post-selection, preserving the relative magnitude of expert contributions and maintaining the inherent sparsity.

Concretely, each expert (comprising both shared experts and routing experts) transforms the expression value v_{ij} into an embedding vector. The shared experts, consisting of S linear transformations, capture general patterns of expression magnitude:

$$\mathbf{h}_{ij}^{\text{shared}} = \sum_{s=1}^{S} \text{SharedExpert}_s(v_{ij}), \tag{5}$$

while the routing experts provide gene-specific transformations aggregated according to the sparse routing weights:

$$\mathbf{h}_{ij}^{\text{routing}} = \sum_{k=1}^{R} \alpha_{ijk} \cdot \text{RoutingExpert}_k(v_{ij}). \tag{6}$$

The final embedding of the gene expression level is the sum of these two components:

$$h_{ij} = h_{ij}^{\text{shared}} + h_{ij}^{\text{routing}}. \tag{7}$$

This architecture enables the model to flexibly and selectively capture both global expression trends and gene-specific nuances, with the sparse routing mechanism improving the model's expressive power and generalization.

Load Balancing for Expert Utilization. To prevent expert collapse and encourage balanced expert usage, we employ DeepSeek-V1's[53] load balancing loss. Let T be the total number of tokens in a batch and K the number of active experts per token. The expected routing frequency for expert k is approximated as:

$$f_k = \frac{1}{K \cdot T} \sum_{i,j} \mathbb{1}(k \in T_{ij}) \cdot R, \tag{8}$$

and the average routing probability as:

$$P_k = \frac{1}{T}\sum_{i,j} p_{ijk}. \tag{9}$$

The load balancing loss is computed as:

$$\mathcal{L}_{\text{balance}} = \sum_{k=1}^{R} f_k \cdot P_k, \tag{10}$$

which encourages both the frequency and the confidence of expert routing to be uniformly distributed across experts. This regularization promotes diversity among expert participation and improves generalization.

Transformer with Two-Dimensional Rotary Positional Encoding

We developed a transformer-based encoder architecture incorporating two-dimensional rotary positional encoding (2D-RoPE), inspired by the Qwen2-VL[54] model. The architecture follows the standard Transformer encoder block structure, consisting of multi-head self-attention, feedforward networks, RMS normalization, and residual connections. Spatial inductive bias is introduced via block-diagonal rotations acting on the query and key vectors, parameterized by the 2D coordinates associated with each token. Let the input to the Transformer be

$$\mathbf{E} \in \mathbb{R}^{B\times T\times d} \tag{11}$$

where B is the batch size, T is the sequence length and d is the model dimension. The model is composed of L identical Transformer layers, each applying the following operations.

Root Mean Square Normalization. Each layer begins with RMS normalization applied to the input $\mathbf{E}$. For a token embedding $\mathbf{e}_t \in \mathbb{R}^d$, the normalized vector is defined as:

$$\text{RMSNorm}(\mathbf{e}_t) = \frac{\mathbf{e}_t}{\sqrt{\frac{1}{d}\sum_{j=1}^{d} e_{t,j}^2 + \varepsilon}} \odot \boldsymbol{\gamma} \tag{12}$$

where ε is a small constant for numerical stability, and $\boldsymbol{\gamma} \in \mathbb{R}^d$ is a learned scaling vector.

Multi-Head Self-Attention with Two-Dimensional Rotary Encoding. We introduce a multi-head self-attention mechanism enhanced with two-dimensional rotary encoding. Initially, token embeddings are

projected to form queries, keys, and values, which are then reshaped for multi-head attention. Each attention head is divided into $\boldsymbol{m}$ four-dimensional subspaces, with each subspace further split into two 2D planes corresponding to the spatial x- and y-dimensions. These planes are rotated independently using log-spaced angular frequencies, allowing the model to capture positional relationships in both spatial dimensions. Attention scores are computed for each head by performing dot-product interactions between the rotated queries and keys, and the outputs from all heads are concatenated and projected to produce the final attention output. This method ensures efficient integration of spatial information within the attention mechanism.

Feedforward Network with SwiGLU. Following attention and residual connection, the output is passed through a feedforward network:

$$\mathbf{E}_{\text{res}} = \mathbf{E} + \mathbf{E}_{\text{attn}}, \quad \mathbf{E}' = \mathbf{E}_{\text{res}} + \text{FFN}(\text{RMSNorm}(\mathbf{E}_{\text{res}})) \tag{13}$$

The feedforward network consists of a Swish-Gated Linear Unit:

$$\text{FFN}(\mathbf{z}) = \mathbf{W}_3\left[\text{Swish}(\mathbf{W}_1\mathbf{z}) \odot (\mathbf{W}_2\mathbf{z})\right], \quad \text{Swish}(x) = x \cdot \sigma(x) \tag{14}$$

where $\sigma(x)$ denotes the sigmoid function，$\odot$: element-wise (Hadamard) product.

Layer Stacking. The above operations are repeated across L layers. Each layer processes the output from the previous one, maintaining the dimensionality $\mathbf{E} \in \mathbb{R}^{B\times T\times d}$ throughout.

Symmetric Bilinear Output Module and Pretraining Objective

To more faithfully capture the subtle dynamics of expression levels, we reformulate the pretraining objective as a soft-rank regression task. In contrast to conventional approaches that discretize expression values into predefined bins for classification, our method preserves the intrinsic continuity and relative ordering of expression by predicting their soft ranks within a local semantic context. This formulation avoids the loss of relational information between expression levels introduced by discretization, enabling finer-grained modeling of expression variation.

Let $\mathbf{E} \in \mathbb{R}^{B\times T\times d}$ denote the final Transformer representations, where B is the batch size, T the sequence length, and d the hidden dimension. The sequence is partitioned into C semantic cells, each of length $t + 2$ and comprising t content tokens flanked by two boundary markers. In our setup, we adopt $C = 1$ for single-cell transcriptomes and $C = 10$ for spatial transcriptomics, where each cell corresponds to a spatial region encompassing a focal cell and its neighbors.

For each cell, the t interior tokens are extracted and reshaped into $\mathbb{R}^{B\times C\times t\times d}$. A pooled cell-level representation $\mathbf{h}_c \in \mathbb{R}^d$ is obtained via mean aggregation:

$$\mathbf{h}_c = \frac{1}{t}\sum_{i=1}^{t}\mathbf{e}_{c,i}. \tag{15}$$

We then apply a trainable symmetric matrix transformation:

$$\mathbf{W}_{\text{sym}} = \frac{1}{2}(\mathbf{W} + \mathbf{W}^{\top}), \tag{16}$$

and compute token-level scores via dot product:

$$s_{c,i} = \left(\mathbf{W}_{\text{sym}}\mathbf{h}_c\right)^{\top}\mathbf{e}_{c,i}. \tag{17}$$

These scores $s_{c,i} \in \mathbb{R}$ represent predicted soft ranks within each semantic cell. To supervise this regression, ground-truth soft ranks $\hat{s}_{c,i}$ are derived from observed expression magnitudes. We adopt the Huber loss (Robust estimation of a location parameter) as the training objective. The loss is computed over masked positions as follows:

$$\mathcal{L}_{\text{MLM}} = \frac{1}{|\mathcal{M}|}\sum_{(c,i)\in\mathcal{M}}\mathcal{L}_{\delta}(s_{c,i} - \hat{s}_{c,i}), \tag{18}$$

where $\mathcal{M}$ denotes the set of masked positions, and $\mathcal{L}_{\delta}(\cdot)$ is the Huber loss function defined as:

$$\mathcal{L}_{\delta}(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \le \delta, \\ \delta(|x| - \frac{1}{2}\delta) & \text{otherwise.} \end{cases} \tag{19}$$

We set the threshold parameter to $\delta = 1$ in all experiments. This formulation encourages stability during optimization while preserving sensitivity to moderate errors.

To encourage uniform usage of semantic cells and avoid representational collapse, we incorporate a load balancing regularization term $\mathcal{L}_{\text{balance}}$, previously introduced. The full pretraining loss is given by:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MLM}} + \alpha \cdot \mathcal{L}_{\text{balance}}, \quad \alpha = \begin{cases} \alpha_0 & \text{if } \mathcal{L}_{\text{MLM}} < 10, \\ 10 & \text{otherwise,} \end{cases} \tag{20}$$

where α_0 is a small constant (typically 0.05). This adaptive weighting ensures that cell utilization remains balanced even during early or unstable phases of regression training.

Construction the Cell and Gene Embeddings

Cell Embedding. We compute cell embeddings in two steps. First, final-layer hidden states are aggregated using a weighted averaging strategy in which each gene contributes based on whether it is expressed (non-zero count), ensuring that only actively expressed genes influence the representation. Second, an optional refinement step projects the aggregated embedding into a low-dimensional subspace using a symmetric projection matrix learned during pretraining. This projection reduces technical noise—including dropout artifacts and batch effects—while preserving biologically meaningful structure in the resulting embedding space.

Specifically, the matrix $\mathbf{W}_{\text{sym}} \in \mathbb{R}^{d \times d}$ is trained directly during pretraining as a symmetric transformation. We perform spectral decomposition:

$$\mathbf{W}_{\text{sym}} = \mathbf{Q} \Lambda \mathbf{Q}^{\top} \tag{21}$$

where Λ contains eigenvalues and $\mathbf{Q}$ contains orthonormal eigenvectors. We select the top-m eigenvectors that collectively explain at least τ of the total spectral energy, computed using the absolute values of eigenvalues. This forms an orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times m}$ used to denoise each cell embedding $\mathbf{h}_c$:

$$\mathbf{h}_c = \mathbf{V}^{\top} \mathbf{h}_c \tag{22}$$

The projected vector $\mathbf{h}_c$ retains task-relevant variation while suppressing noise directions that dominate in raw representations. In practice, this spectral projection improves the separation of cell types and reduces technical confounding across spatial and single-cell modalities.

Gene Embedding. Gene embeddings were derived at two complementary levels, capturing both local and global gene properties. At the cellular level, the final-layer hidden states from a Transformer encoder were extracted, producing a contextualized embedding for every gene within each cell. To reduce technical noise, these embeddings were projected onto a low-dimensional subspace as described previously, yielding representations that reflect cell-state- or cell-type–specific gene characteristics.

Dataset-level gene embeddings were then obtained by integrating these cell-level gene embeddings across all cells, with the explicit objective of bringing genes with similar expression patterns closer in the embedding space. **Supplementary Figure 2A** illustrates this integration process, showing how cell-level embeddings are combined to form dataset-level representations, which can then be utilized for further analysis. A cell similarity graph was constructed based on the previously described cell embeddings, connecting each cell to its k nearest neighbors, with edge weights determined using a decaying kernel[55, 56]. Each gene's cell-level embeddings were propagated through this graph using a generalized PageRank graph neural network (GPR-GNN)[33], which aggregates information from neighboring cells while optimizing reconstruction of observed gene expression. This procedure reinforces co-expression patterns and context-dependent relationships, ensuring that genes exhibiting similar transcriptional behavior across the dataset acquire similar dataset-level embeddings. The final embedding for each gene was obtained by averaging its propagated embeddings across all cells, capturing both intrinsic expression patterns and global, dataset-wide functional properties.

Pretraining of the OmniCell Model

All single-cell and spatial transcriptomic datasets were first converted from the AnnData (h5ad) format into LMDB files to enable high-throughput sequential access and memory-efficient data streaming during large-scale optimization. Pretraining was conducted in two sequential stages to hierarchically integrate spatial context and single-cell precision within the same model.

In the first stage, spatial transcriptomic data were used to expose the model to tissue-level organization and neighborhood-dependent gene co-variation, allowing it to capture local transcriptional continuity across adjacent spots. Because spatial data are typically noisier and have lower molecular capture efficiency, a moderate masking ratio was adopted to balance information removal and reconstruction difficulty. Specifically, the non-zero expression values of central nodes were randomly masked at a rate of 0.3, and those of neighboring nodes at 0.05, while zero entries were largely preserved. This design guided the model to infer missing gene expression patterns from partially masked spatial contexts, enhancing its ability to recover structured co-expression signals while maintaining training stability.

In the second stage, the pretrained model was further optimized on scRNA-seq datasets, which provide higher data quality and reduced technical noise. To fully exploit the intrinsic information redundancy of transcriptomic data—analogous to pixel redundancy in natural images—a high masking ratio was employed to strengthen contextual inference. The non-zero expression values of each cell were masked at a rate of 0.3, while zero expression values were masked at a rate of 0.1, forcing the model to predict large portions of the expression vector based on the remaining signals. This large-mask regime encourages the model to learn robust gene–gene dependencies and to generalize effectively across diverse cellular states.

Training was performed using a large batch size (3,200 cells per step) evenly distributed across 32 NVIDIA A100 GPUs (40 GB memory each), enabling efficient optimization at scale. Gradient clipping (maximum norm = 1.0) was applied to stabilize optimization. The AdamW optimizer—a weight-decoupled variant of Adam—was used with a learning rate of 1×10^{-6}, weight decay of 0.01, and momentum coefficients $\beta_1 = 0.9$ and $\beta_2 = 0.99$. A cosine annealing scheduler with a minimum learning rate of 5×10^{-7} ensured smooth and stable convergence.

Downstream tasks

Evaluation of Cell-Type Specificity of Gene Embeddings at the Cell Level

To assess the cell-type specificity captured by the learned gene embeddings, we analyzed the gene-level representations derived from the final layer of each foundation model, including OmniCell, scGPT, Geneformer, and scFoundation. This evaluation reflects how well each model learns the context-dependent relationships for each gene, providing insight into how gene embeddings are organized across different cell

types. For every gene, we collected its embeddings across all cells and performed unsupervised clustering in the corresponding embedding subspace. The resulting gene-specific clustering was then quantitatively compared with the known cell-type annotations using the Adjusted Rand Index (ARI). This procedure yields a cell-type specificity score for each gene, which reflects the degree to which its embedding structure recapitulates known cellular identities. Higher ARI values indicate that the model captures more distinct and biologically meaningful gene representations across diverse cell populations.Gene–cell importance scores were obtained by pooling gene–gene attention weights across all layers of the transformer. To further evaluate the performance of our model, we performed a multiclass ROC analysis based on the gene–cell importance scores derived from the attention weights. For each gene, we calculated a gene–cell importance matrix, which reflects how each gene's importance varies across different cells based on the attention scores. This matrix was then used to assess the model's ability to discriminate between cell types. Specifically, for each cell type, we computed a prototype importance profile by averaging the importance scores of all cells in that type. The similarity between each cell's gene importance vector and the prototype for each cell type was calculated using cosine similarity, yielding predicted scores that reflect the likelihood of each cell belonging to a specific cell type. The resulting multiclass ROC analysis, including AUC scores, allowed us to assess how well the model captures and distinguishes the context-dependent gene relationships across diverse cell populations. This analysis was conducted using the Zhou dataset, with 2000 highly variable genes selected for evaluation, providing a comprehensive assessment of the model's ability to learn cell-type specificity in cell-level gene embeddings.

Identification of Cell-Type–Specific Genes from OmniCell Cell-Level Gene Embeddings

To identify genes that characterize specific cell types or cellular states, we developed a method leveraging cell-level gene embeddings generated by the foundation model. For each gene, its embedding distributions across cells were compared between a target cell type and all other cells using the mean one-dimensional Wasserstein distance across embedding dimensions, thereby quantifying how distinctively the gene is represented in the target population. Genes were ranked by these distance scores, and the top candidates were further filtered based on their transcriptional abundance in the corresponding single-cell data, retaining only those with higher mean expression in the target cell type relative to others. This two-step procedure—embedding-based discrepancy analysis followed by expression validation—yields interpretable gene sets that reflect both representation-level and expression-level specificity, providing a principled approach for discovering cell-type–specific or state-associated marker genes from foundation

model embeddings. To evaluate the consistency between marker genes identified from OmniCell embeddings and those from conventional methods, we compared the embedding-derived markers with cell-type markers generated by the Seurat[57] *FindAllMarkers* function. Results showed that the majority of embedding-derived markers for most cell types overlapped with the top marker gene lists from Seurat[57] *FindAllMarkers*, which strongly validated the biological reliability of the markers identified via OmniCell embeddings. Beyond the overlapping known markers, OmniCell embeddings also identified additional cell-type-enriched novel markers that were not detected by the conventional method. To assess the discriminative power of these novel markers, we performed ROCanalysis and calculated the AUCvalues. Further confirming the validity of both the overlapping and novel markers derived from OmniCell embeddings.

Validation of Co-expression Structure in OmniCell Dataset-Level Gene Embeddings

To determine whether dataset-level gene embeddings learned by OmniCell capture biologically meaningful co-expression structure, we constructed a gene–gene similarity network using cosine similarity or a Gaussian-transformed Euclidean metric and applied soft thresholding to enhance high-confidence relationships before identifying modules via Louvain community detection. We evaluated the biological relevance of these embedding-derived modules using two independent datasets: the Zhou scRNA-seq dataset and a Stereo-seq dataset comprising both Alzheimer's disease and control samples. Cell-type marker genes identified from OmniCell cell-level embeddings and region-specific marker genes curated from prior literature[40] served as biological priors for the scRNA-seq and ST datasets, respectively. We then quantified whether these biological priors aligned with the embedding-derived network structure by calculating expression score profiles for each module. Higher scores indicate that genes grouped together in the embedding space also exhibit coordinated expression patterns at the transcriptomic level, a defining property of biologically meaningful co-expression modules. Consistently higher scores in modules enriched with known biological marker genes indicate that OmniCell's dataset-level gene embeddings capture biologically coherent and context-dependent co-expression patterns rather than arbitrary latent structure. These findings support that the learned embeddings provide a faithful representation of regulatory organization across both scRNA-seq and ST datasets.

Single-Cell and Spatial Cell Clustering

To evaluate clustering performance, we assessed OmniCell alongside both single-cell and spatial foundation models. The single-cell models included scGPT, Geneformer, and scFoundation, whereas scGPT-spatial and Nicheformer represented spatial models. For models not pretrained on mouse datasets,

684 homologous gene mapping was applied to enable cross-species evaluation. To ensure fairness across
685 modalities, a unified gene selection strategy was applied: the top 2,000 highly variable genes were used for
686 all scRNA-seq datasets, including human kidney and blood, as well as mouse brain and muscle datasets,
687 while the full gene set was retained for the MERFISH mouse brain spatial transcriptomics dataset.
688 Embeddings generated by each model were clustered using the Leiden algorithm, with the resolution
689 systematically varied from 0.1 to 1.0; for each model, the resolution yielding the highest agreement with
690 reference annotations was selected as the final result. Clustering performance was quantified using
691 Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI), providing complementary
692 measures of concordance between predicted clusters and biologically defined cell identities.

693 **Batch correction evaluation**

694 For batch correction tasks, model-derived cell embeddings were evaluated using the scIB[58]
695 benchmarking framework. Four baselines—PCA, scGPT, Geneformer, and scFoundation—were included
696 for comparison against OmniCell. Two human scRNA-seq datasets, hPBMC and the Zhou dataset, were
697 used to assess performance. To ensure comparability across models, the top 2,000 highly variable genes
698 were selected as input features for all evaluations. The scIB framework reports two composite metrics: a
699 biological conservation score, reflecting how well intrinsic cellular structure and annotations are preserved
700 across batches, and a batch correction score, measuring the effectiveness of batch integration based on
701 silhouette and graph connectivity–based metrics. Higher values in both scores indicate a more optimal
702 balance between maintaining biological fidelity and mitigating batch-driven technical variation.

703 **Noise robustness evaluation**

704 Noise robustness was assessed under increasing dropout perturbations using the Zhou dataset. Starting
705 from the original count matrix, random dropout noise was incrementally introduced from 0.0 to 1.0 in steps
706 of 0.05. To ensure fairness across methods, the top 2,000 highly variable genes were used as input features
707 for all models. Three foundation models—scGPT, Geneformer, and scFoundation—were evaluated
708 alongside OmniCell. For each noise level, cell embeddings were regenerated using the pretrained model
709 and subsequently subjected to Leiden clustering at a fixed resolution of 0.1, which was empirically
710 identified as optimal across models. Clustering consistency was quantified using Normalized Mutual
711 Information (NMI), selected as the primary metric due to its stability under varying cluster numbers and its
712 sensitivity to biological structure despite increasing data sparsity. All experiments were performed with a
713 fixed random seed to ensure reproducibility, and the resulting NMI–dropout curves capture each model's
714 robustness to transcriptional noise.

Spatial Domain Identification

To evaluate the model's ability to capture spatial tissue organization, we performed spatial domain identification on two spatial transcriptomics datasets. For the mouse cortex slice with predefined anatomical annotations, all 33 genes were used during inference; cell embeddings were generated in spatial mode by integrating neighboring transcriptomes and averaging across the local neighborhood, followed by construction of a k-nearest neighbor graph and Leiden clustering. The clustering resolution was optimized from 0.1 to 1.0 (step size 0.1) by maximizing the combined Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI) relative to ground truth. For the LC5-M human lung cancer dataset, which lacks predefined annotations but contains tumor, peri-tumoral transition, and non-malignant compartments, we used the top 500 highly variable genes consistent with model training, while baselines used the standard top 2,000 genes; embeddings were again generated in spatial mode, and a k-nearest neighbor graph was used to accommodate broader tissue structure and tumor heterogeneity before applying Leiden clustering at a fixed resolution of 0.1. Performance was compared against two foundation models (scGPT-spatial and NicheFormer) and three expert spatial methods (SCAN-IT[44], SEDR[59], and SpaGCN[60]). While ARI and NMI confirmed higher clustering accuracy on the mouse cortex dataset, analysis of LC5-M revealed that only our model resolved a distinct peri-tumoral transitional domain. Gene Ontology enrichment analysis identified significant activation of acute-phase response, inflammatory signaling, complement activation, coagulation regulation, and copper ion detoxification pathways within this region. Cell type composition analysis further revealed marked enrichment of immune populations. Spatial gradient analysis using SLOPER[61] demonstrated elevated expression of metallothionein family genes with directional streamlines delineating the spatial organization of these molecular programs.

Cell-Type Annotation

For quality estimation on cell embedding in supervised classification, we took the top 2,000 most variable genes from each dataset and extracted cell embeddings from the corresponding models. Three pretrained baseline models (scGPT, Geneformer, and scFoundation) were applied as baselines with respect to OmniCell. ReLU activation and layer normalization were used to train the AdamW optimizer on a three-layer multilayer perceptron (learning rate 1×10^{-4}, 20 epochs) in order to learn the cell types in the embeddings. Assorted as 5-fold stratified cross-validation of five benchmark datasets (Brain, Hippocampus, Cortex, Immune, Great_Apes), reciprocal cross-batch validation (training in first batch and testing on second batch) and 80/20 stratified train–test split of large sparse datasets (e.g., zheng68k) were used. The classification performance was determined by accuracy and macro-averaged F1-score with mean and

variance calculated across folds or batches. To maintain reproducibility, all experiments were performed on fixed random seeds.

Spatial transcriptomic deconvolution

Cell embeddings derived from the foundation model were used to perform spatial cell-type deconvolution, following a strategy conceptually similar to Tangram[62]. Single-cell reference profiles and spatial transcriptomic data were aligned based on their shared genes. The single-cell embeddings served as the reference space, and spatial embeddings were projected onto this space through an optimization procedure to estimate the proportional contribution of each cell type to each spatial location. The resulting proportion matrix defined the inferred cellular composition per spot, from which the dominant cell type was assigned based on the highest contribution weight. Two spatial foundation models—scGPT-spatial and Nicheformer—were included as baselines for comparison with OmniCell. Deconvolution performance was evaluated using accuracy and macro-averaged F1 score.

Benchmarking against competing methods

Comparisons against foundation models

To thoroughly assess the OmniCell model, we performed benchmarking on various foundational single-cell and spatial omics models (scGPT, Geneformer, scFoundation, scGPT-spatial, Nicheformer). The extraction of embeddings for each foundational model as presented by the above-in-text implementation described below: For scGPT we used the scGPT_human pre-trained weights from the official repository and obtained cell embeddings for the model pre-trained on complete human data using the pipeline described in the original paper. The Geneformer-V2-316M model (from Hugging Face's official repository) has been utilized for Geneformer: to obtain embeddings from the penultimate layer and to group gene-level embeddings into single-cell representations by mean pooling; for scFoundation we obtain pre-trained weights (models.ckpt) from the official repository and to estimate embeddings using the standard inference procedure presented by the authors; For the scGPT-spatial model we extracted the scGPT_spatial_v1 model weights specifically pretrained on spatial transcriptomics data. For Nicheformer, we applied pre-trained model parameters that were trained on the SpatialCorpus-110M dataset. Embeddings provided by the base models were treated as task specific inputs.

Comparisons against spatial domain detection methods

We also compared OmniCell with specialized spatial domain detection methods (SpaGCN v1.2.7, SEDR, and SCAN-IT) to evaluate its performance in spatial clustering tasks. For SpaGCN, we used the official implementation version, utilizing the default hyperparameters proposed in the original paper. Training was

performed using a spatial graph built from the spatial coordinates of cells in an attempt to find spatially coherent regions. The official implementation was applied for SEDR, adhering to their routine workflow for producing low-dimensional representations necessary for spatial clustering by combining profiles of gene expression with spatial locations. The official SCAN-IT implementation followed the recommended spatial domain identification process to identify tissue domains and cellular microenvironments through the combination of gene expression and spatial information.

Code availability

OmniCell is packaged and distributed as an open-source, publicly available repository at BGIResearch/omnicell

Acknowledgement

We would like to express our gratitude to the Stomics Cloud platform (https://cloud.stomics.tech/) for providing GPU computational resources. We also appreciate the insightful discussions and contributions from our research group colleagues.

Author contributions

L.S.S. conceptualized the study. P.J.S. was responsible for the framework design and tool implementation. P.J.S., Q.P., D.Y.T., H.Y.Z. and L.B.L. performed data analysis and model evaluation, while L.B.L. was responsible for building the baseline model. T.F., L.A.D., C.L., J.W., W.H.R., L.S.K, D.Z.Q., and F.S.S. provided key suggestions. P.J.S., Q.P., and H.Y.Z. wrote the manuscript. Z.Y., L.Y.X., L.S.S., and X.X. supervised the study.

796 **Figure Legend**

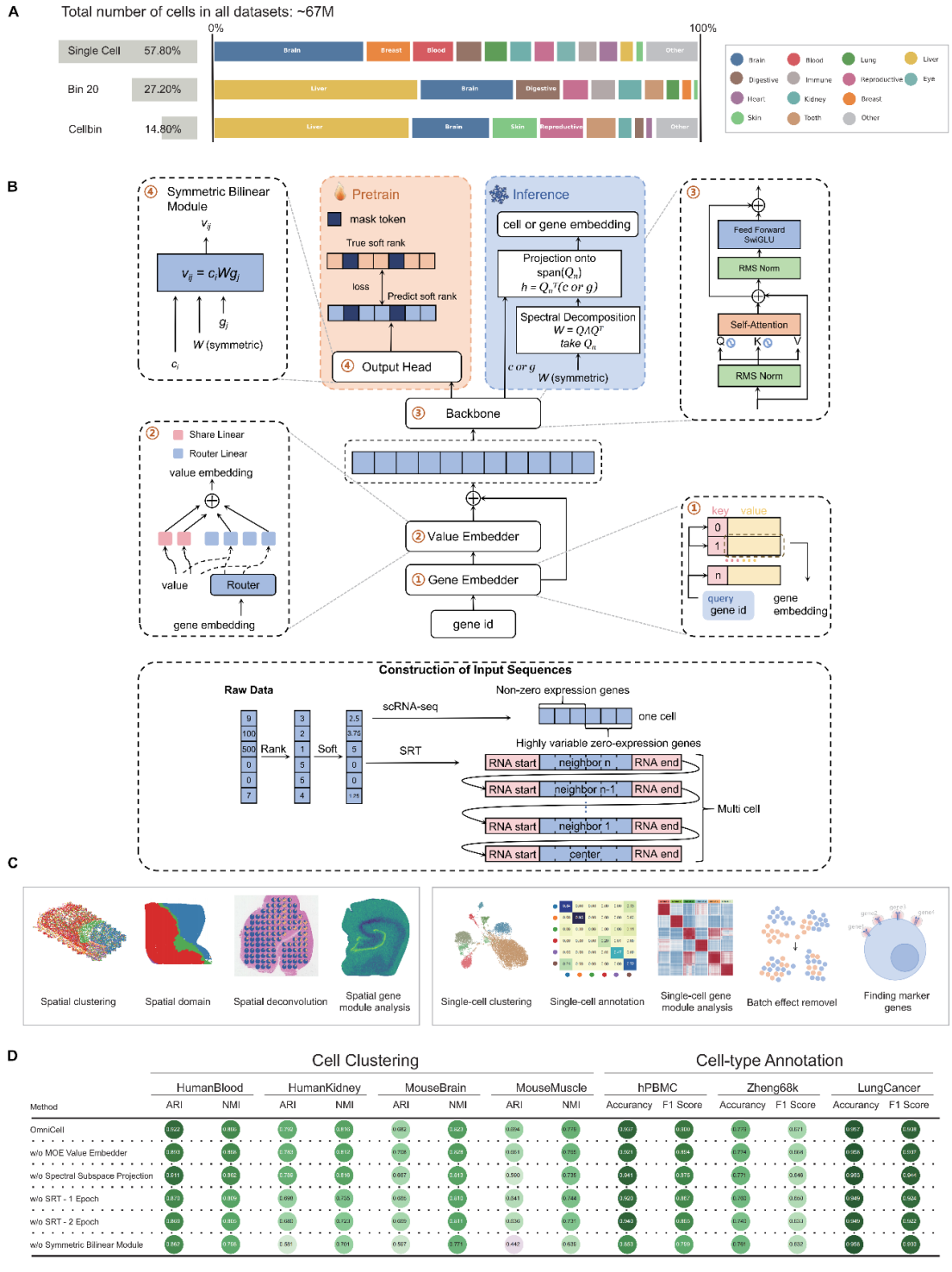


Figure 1. Overview of the OmniCell Model, Pretraining Data, and Evaluation.

An overview of the OmniCell model, its pretraining data, architecture, downstream tasks, and ablation

study. **(A)** Composition of pretraining data. OmniCell was pretrained on 67 million cells comprising single-cell RNA sequencing (scRNA-seq; 57.8%) and Stereo-seq spatial transcriptomics (ST; 42.2%) data across diverse tissues and cell types. **(B)** Model architecture. OmniCell integrates scRNA-seq and ST data within a unified framework. For scRNA-seq, cells are encoded as ordered gene sequences based on expression. For ST, spatial context is incorporated through neighborhood graphs capturing local cellular relationships. Gene expression values are normalized via soft-rank transformation. A mixture-of-experts (MoE) gene-aware value embedding module adaptively encodes expression levels. The architecture comprises 10 Transformer layers, with a symmetric bilinear output module that jointly models cell–gene relationships to generate unified embeddings. **(C)** Downstream tasks. Model performance was assessed on ST-specific tasks (spatial clustering, domain identification, deconvolution, and gene module analysis) and scRNA-seq tasks (cell clustering, cell-type annotation, gene module analysis, batch correction, and marker gene identification). **(D)** Ablation analysis. Systematic removal of model components—including the MoE value embedder, spectral subspace projection, symmetric bilinear module, and ST pretraining data—quantifies their individual contributions to model performance.

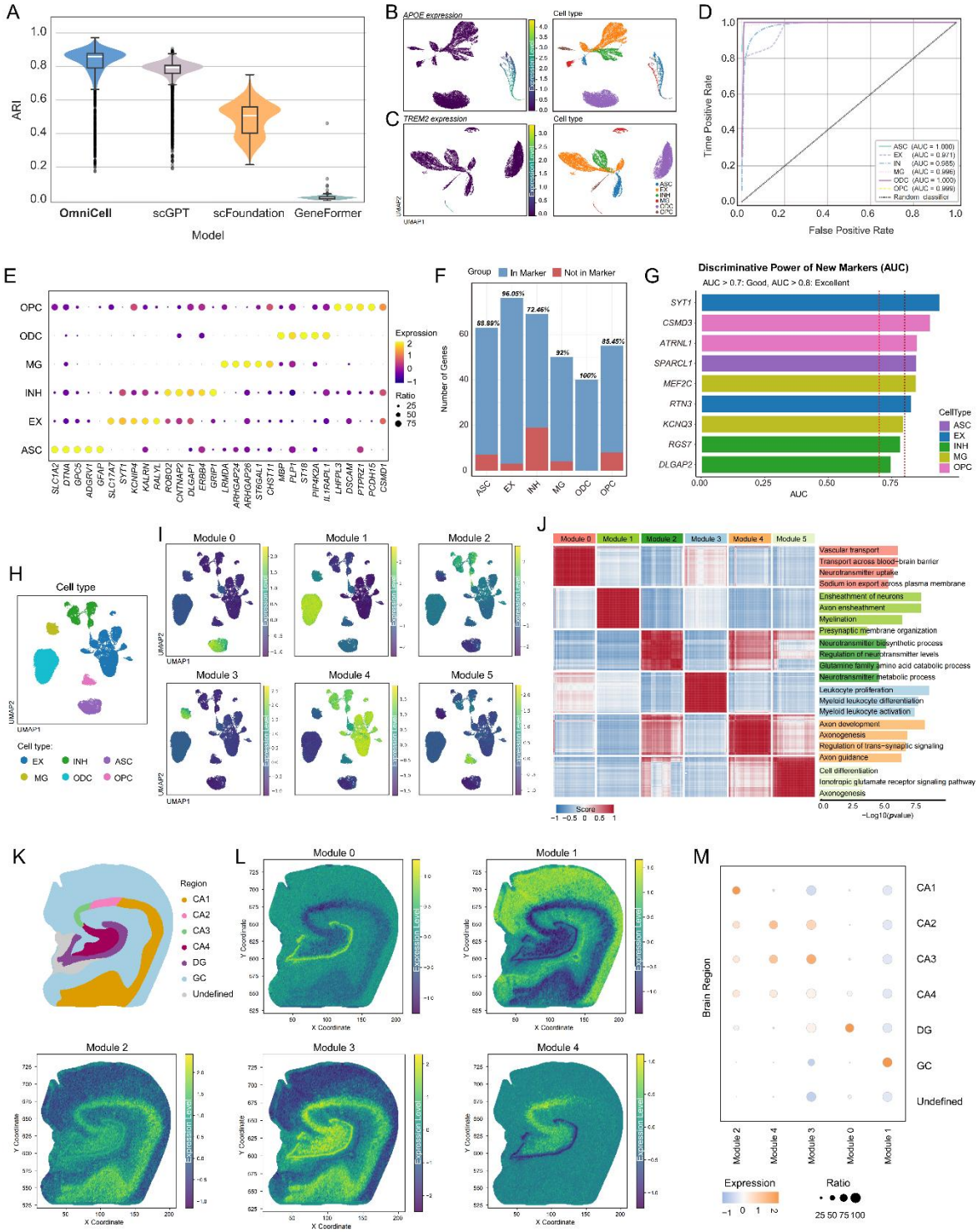


Figure 2. Cell-level and dataset-level gene embeddings capture cell-type-specific signatures and functional modules.

(**A**) Benchmarking of cell-level gene embeddings across foundation models using Adjusted Rand Index (ARI) for clustering concordance with known cell types. (**B-C**) UMAP visualizations of cell-level embeddings for APOE and TREM2 expression (left) and corresponding cell-type annotations (right). (**D**) Assessment of cell-type separability via gene importance vectors. ROC curves depict the discriminative capacity of attention-derived gene importance score vectors across different cell types.(**E**) Dot plot of marker genes identified from cell-level embeddings across six major cell types (ASC, EX, INH, MG, ODC, OPC), with expression ratio (dot size) and level (color). (**F**) Proportion of embedding-derived markers that overlap with cell-type markers generated via the Seurat *FindAllMarkers* function. (**G**) Discriminative power (AUC) of novel embedding-derived marker genes not overlapping with those generated by the Seurat *FindAllMarkers* function. (**H**) UMAP visualization of the Zhou Alzheimer's disease scRNA-seq dataset was colored by cell type. (**I**) UMAP visualization of module expression scores for six embedding-derived gene modules (Modules 0-5). (**J**) Heatmap showing six modules of DEGs classified by OmniCell, with each correlated to distinct biological processes. (**K**) Anatomical parcellation of AD patients' hippocampus showing cornu ammonis subfields (CA1–CA4), dentate gyrus (DG), and granule cell layer (GC). (**L**) Spatial distribution of module expression scores in AD patients' hippocampus spatial transcriptomics data. (**M**) Dot plot of module expression across hippocampal regions, with expression ratio (dot size) and level (color).

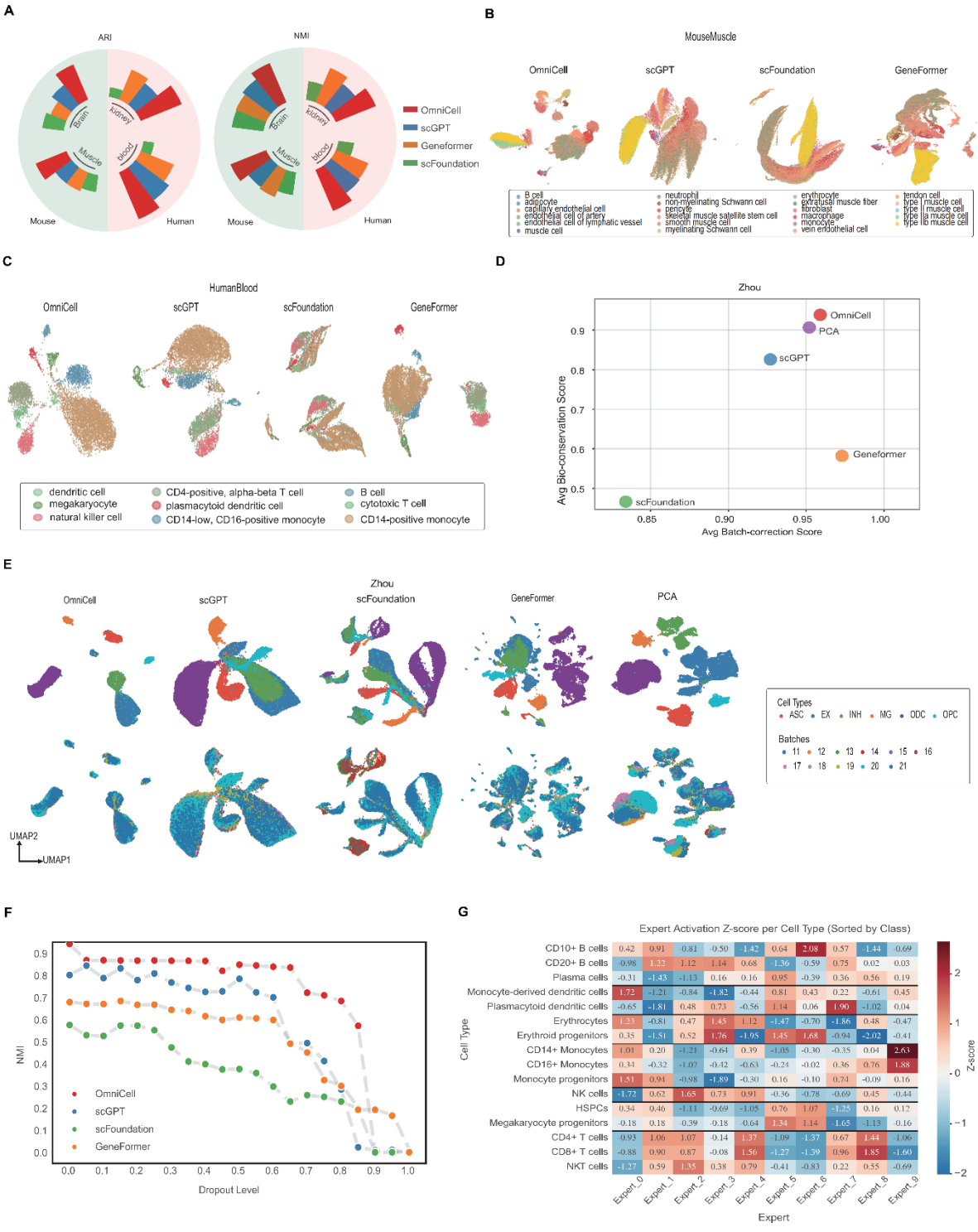


Figure 3. Benchmarking OmniCell for cross-species generalization, batch integration, and expert modularity.

(A) OmniCell consistently achieved the highest clustering performance (ARI and NMI) across all datasets, surpassing existing models. Despite being pretrained solely on human data, OmniCell maintained clear cell-type separability in mouse datasets through homology-based gene projection, demonstrating strong cross-species transferability. **(B–C)** UMAP visualizations of representative human and mouse datasets show that OmniCell embeddings form compact and biologically coherent clusters with superior intra-cluster cohesion and inter-cluster separation compared with scGPT and GeneFormer. **(D)** Scatterplot comparing batch-correction performance (x-axis) against biological conservation (y-axis) on the Zhou dataset. Each point represents a different method. **(E)** UMAP embeddings of the Zhou dataset (11 batches, 6 cell types) generated by five methods. Top row: cells colored by cell type (ASC, EX, INH, MG, ODC, OPC). Bottom row: same embeddings colored by batch (batches 11–21). **(F)** Under progressive random dropout, OmniCell retained the highest clustering stability, indicating noise-resilient and biologically grounded representations. **(G)** Heatmap of mixture-of-experts (MoE) activation patterns across immune cell populations. Rows represent 16 cell types sorted by lineage class; columns denote individual experts (Expert_0 through Expert_9). Color scale indicates activation z-scores, with red reflecting positive and blue reflecting negative enrichment.

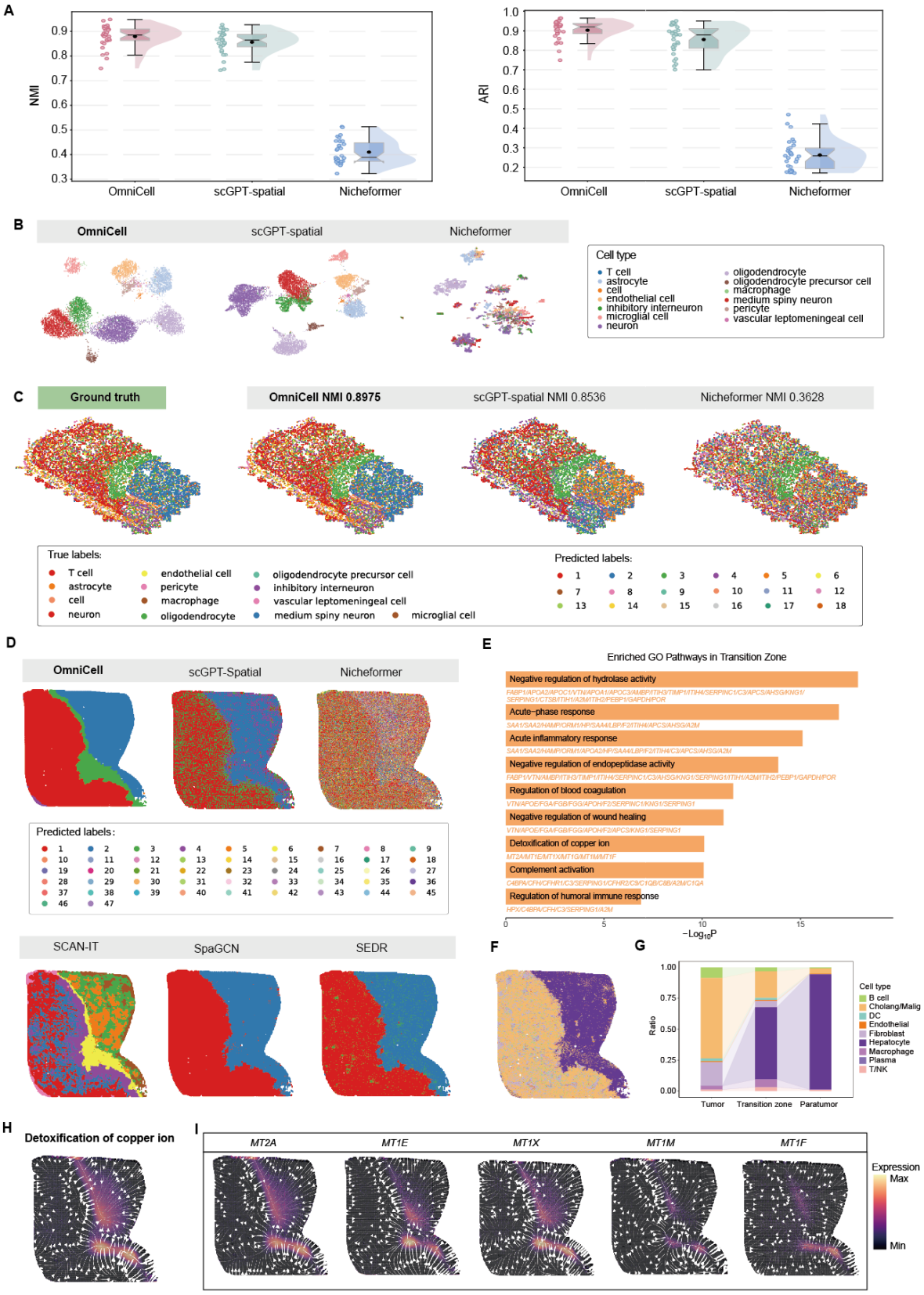


Figure 4. Multi-level spatial evaluation of OmniCell across MERFISH and Stereo-seq datasets.

(**A**) Boxplots comparing clustering performance (NMI and ARI) of OmniCell, scGPT-spatial, and Nicheformer across 31 MERFISH-based mouse brain atlas datasets. (**B**) UMAP visualization of cell embeddings from MERFISH slice 10_0 colored by cell type annotations. (**C**) Spatial coordinate maps comparing ground truth annotations with clustering results from OmniCell (NMI = 0.8975), scGPT-spatial (NMI = 0.8536), and Nicheformer (NMI = 0.3628). (**D**) Spatial segmentation of the Stereo-seq LC5-M liver cancer dataset by six methods (OmniCell, scGPT-Spatial, Nicheformer, SCAN-IT, SpaGCN, and SEDR). (**E**) Gene Ontology enrichment analysis of Cluster 3 (transition zone). (**F**) Spatial distribution of cell types in the Stereo-seq LC5-M liver cancer dataset. (**G**) Cell type composition across tumor, transition zone, and paratumor regions. (**H**) Spatial plot showing the activity of the detoxification of copper ion pathway, where white streamlines indicate the gradient direction. (**I**) Spatial expression patterns of the pathway genes of **H** (*MT2A, MT1E, MT1X, MT1M, MT1F*), with white streamlines depicting expression gradients.

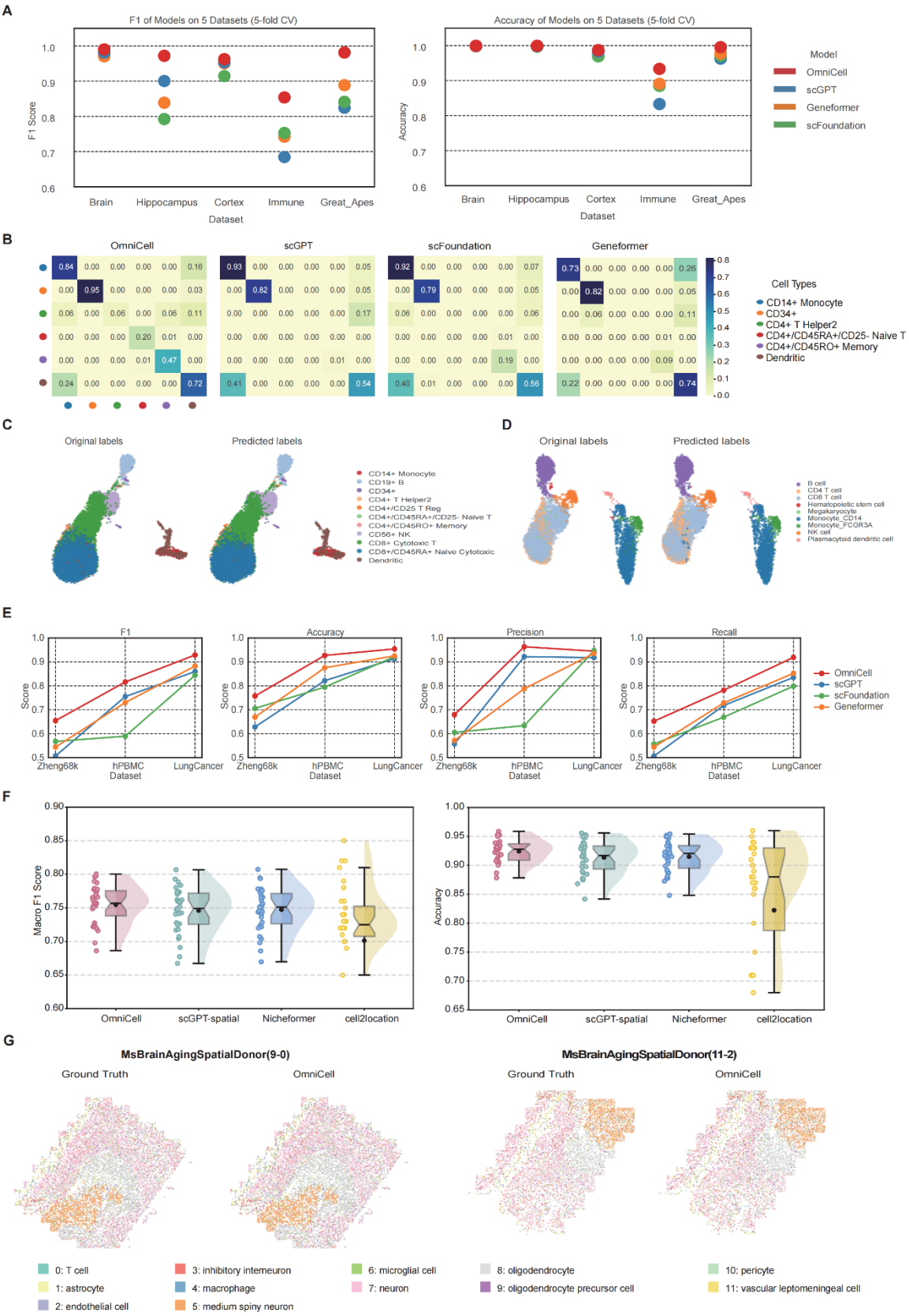
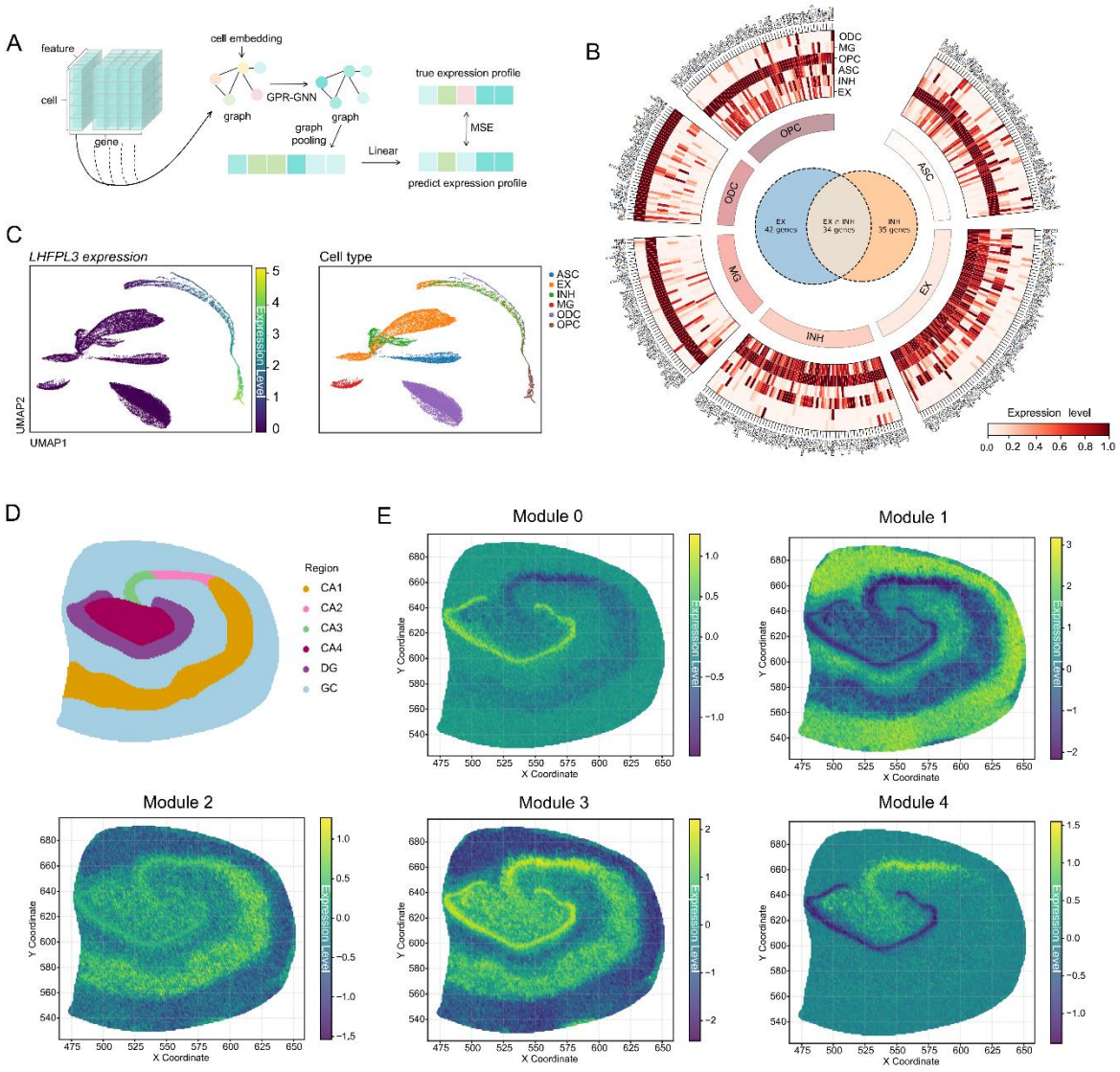


Figure5. Benchmarking OmniCell for single-cell cell-type annotation and spatial transcriptomics deconvolution.

(**A**) Dot plots of cell-type annotation performance across five scRNA-seq datasets (Brain, Hippocampus, Cortex, Immune, Great Apes). F1 score (left) and accuracy (right) were evaluated using 5-fold cross-validation for OmniCell, scGPT, Geneformer, and scFoundation. (**B**) Confusion matrices illustrating OmniCell's accurate annotation of rare cell types in the Zheng68k dataset, compared with other models. (**C**) UMAP projections of the Zheng68k dataset showing original annotations (left) and predicted labels (right). Twelve cell types are displayed, including rare populations such as CD34+ progenitors and dendritic cells. (**D–E**) UMAP visualization of the hPBMC dataset comparing original annotations (left) with predicted labels (right) across eleven cell types. (**E**) Line plots of annotation metrics (F1, Accuracy, Precision, Recall) evaluated on three datasets (Zheng68k, hPBMC, LungCancer). Each line represents one method; error bars indicate variation across folds or batches. (**F**) Boxplots comparing spatial deconvolution performance across 30 spatial transcriptomics datasets. Macro F1 score (left) and accuracy (right) are shown for OmniCell, scGPT-spatial, Nicheformer, and cell2location. (**G**) Spatial reconstructions of mouse brain aging samples (MsBrainAgingSpatialDonor 9-0 and 11-2). Ground truth cell-type maps (left) are compared with OmniCell predictions (right).

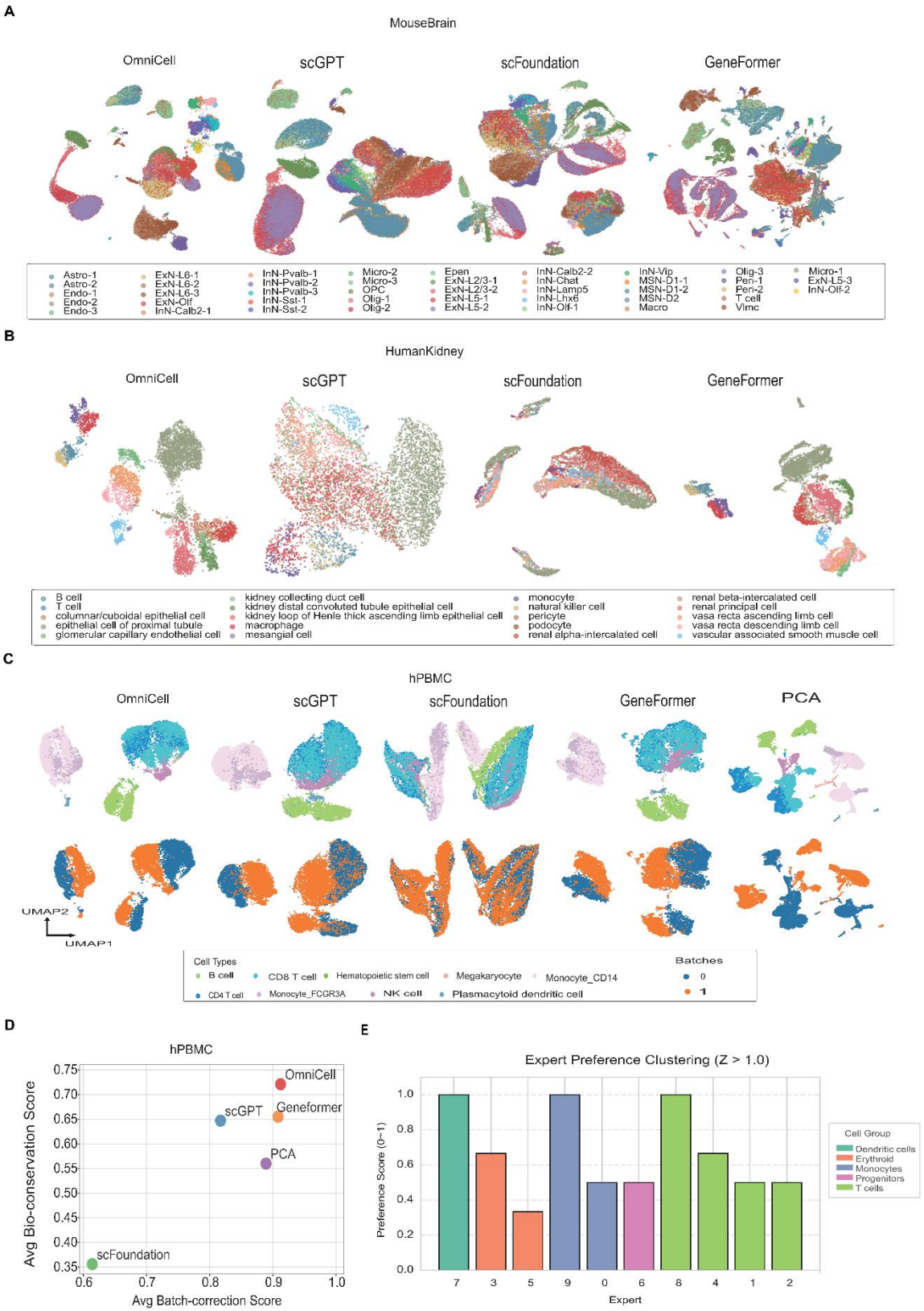
802



Supplementary Figure 2. Extended analysis of gene embeddings and module characterization.

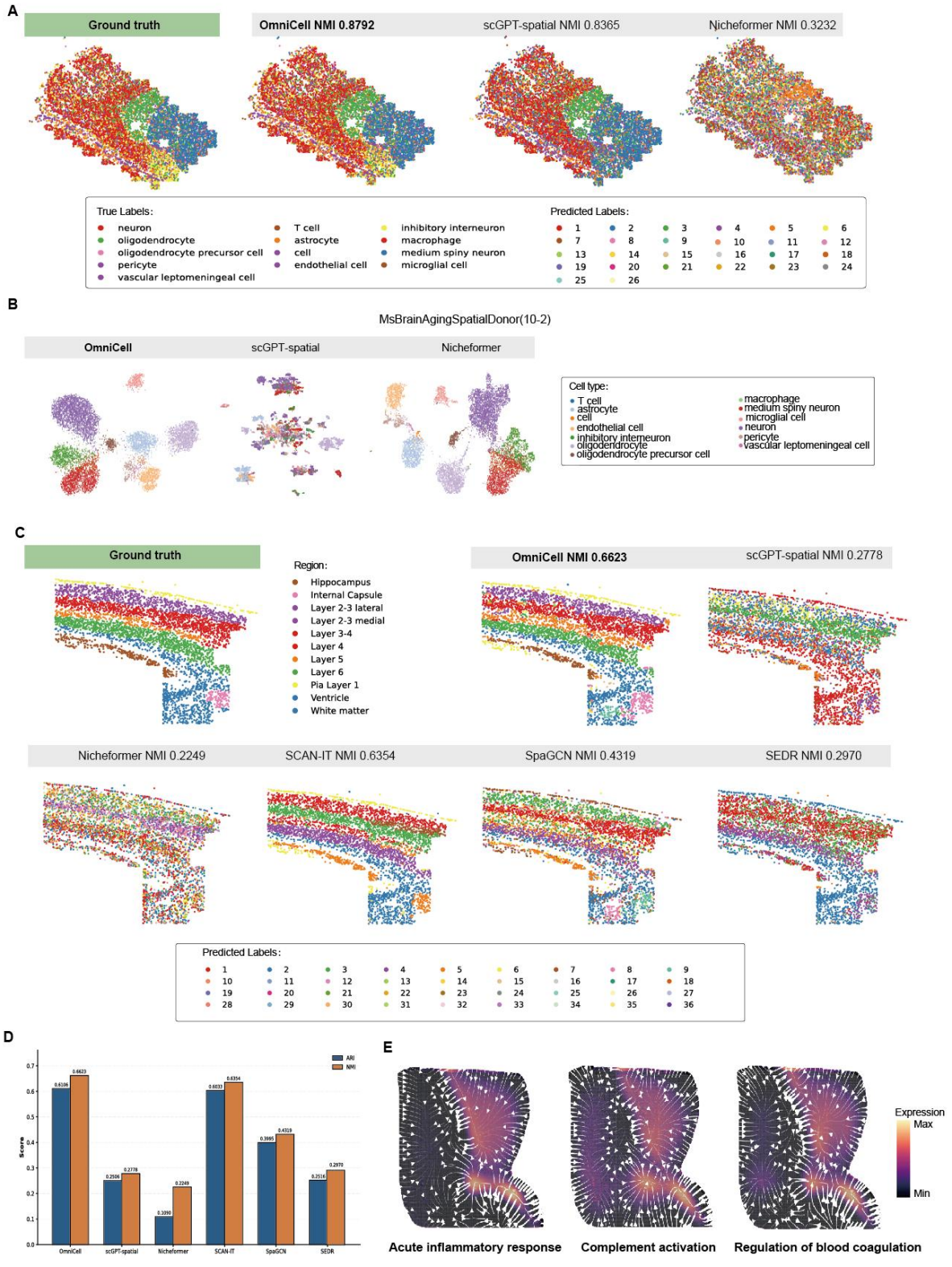
(**A**) Workflow for generating cell-level and dataset-level gene embeddings using OmniCell Transformer encoder, GPR-GNN graph propagation through cell-similarity networks, and linear aggregation across cells. (**B**) The radial heatmap demonstrates clear segregation of these markers, while the central Venn diagram highlights both unique and shared genes between excitatory and inhibitory lineages (42 EX-specific, 35 INH-specific, 34 shared). (**C**) UMAP visualizations of cell-level embeddings for *LHFPL3* expression (left) and corresponding cell-type annotations (right). (**D**) Anatomical parcellation of normal human hippocampus showing cornu ammonis subfields (CA1–CA4), dentate gyrus (DG), and granule cell layer (GC). (**E**) Spatial distribution of module expression scores in normal human hippocampus spatial transcriptomics data.

803



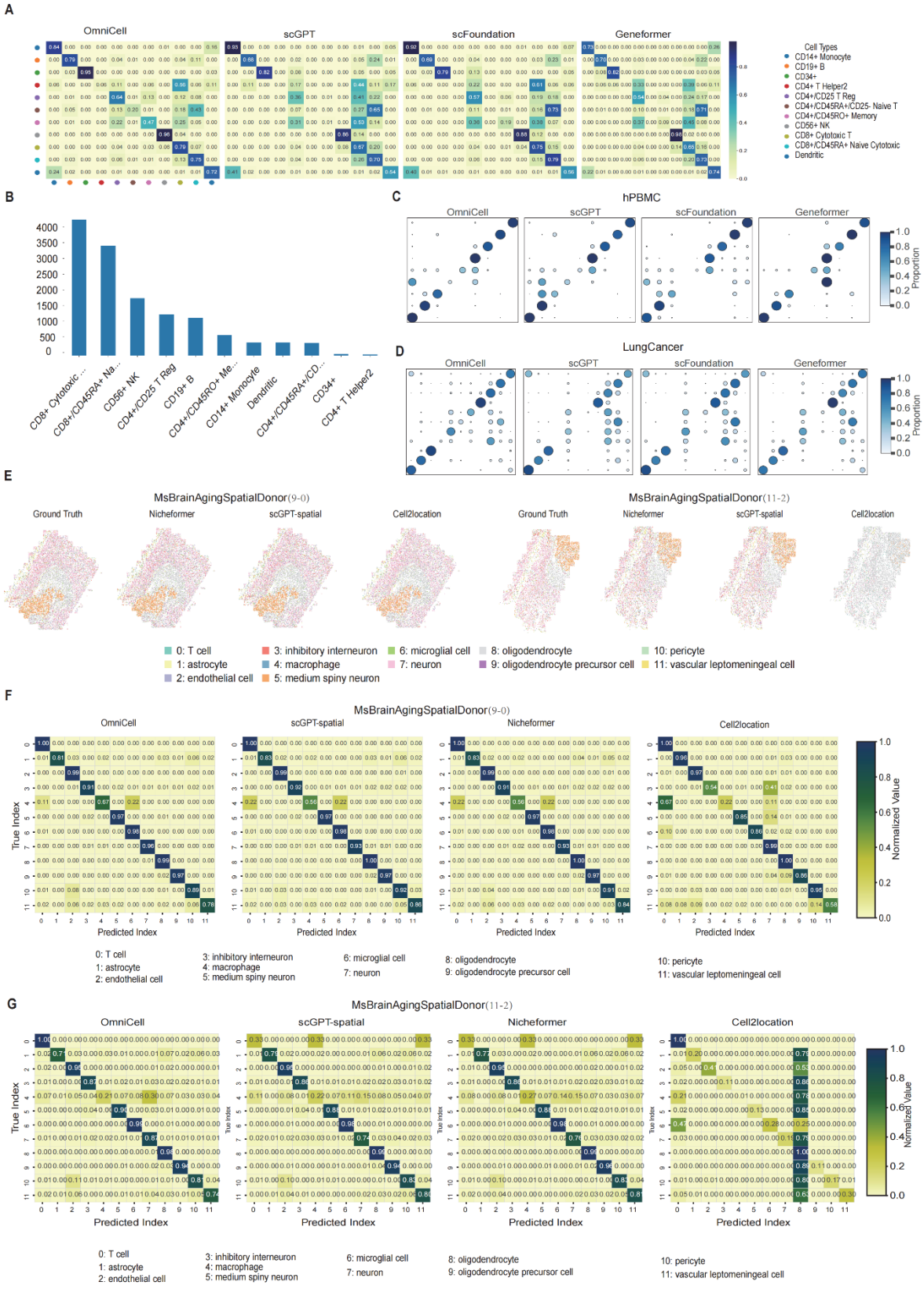
Supplementary Figure 3. Extended evaluation of OmniCell generalization and expert modularity. (**A**) UMAP projections of the MouseBrain dataset comparing embeddings from OmniCell, scGPT, scFoundation, and Geneformer. (**B**) UMAP projections of the HumanKidney dataset across the same four methods. (**C**) UMAP embeddings of the hPBMC dataset generated by five methods (OmniCell, scGPT, scFoundation, Geneformer, PCA). Upper row: cells colored by nine cell types . Lower row: same embeddings colored by batch identity. (**D**) Scatterplot of batch integration performance on the hPBMC dataset. Average batch-correction score (x-axis) is plotted against average bio-conservation score (y-axis) for each method. (**E**) Bar chart of expert preference clustering (z-score threshold > 1.0) across ten experts. Bars are colored by cell lineage group: dendritic cells, erythroid, monocytes, progenitors, and T cells.

804



Supplementary Figure 4. Extended spatial evaluation of OmniCell across MERFISH and Stereo-seq datasets.

(**A**) Spatial coordinate maps of MERFISH slice 10_2 comparing ground truth annotations with clustering results from OmniCell (NMI = 0.8792), scGPT-spatial (NMI = 0.8365), and Nicheformer (NMI = 0.3232). (**B**) UMAP visualization of cell embeddings from MERFISH slice 10_2 (MsBrainAgingSpatialDonor 10-2) colored by cell type annotations, comparing OmniCell, scGPT-spatial, and Nicheformer. (**C**) Spatial domain delineation on a mouse cortex dataset with gold-standard regional annotations. Ground truth shows anatomical regions including Hippocampus, Internal Capsule, cortical Layers 2 – 6, Pia Layer 1, Ventricle, and White matter. Comparison across methods demonstrates that OmniCell (NMI = 0.6623) best recapitulates cortical layer architecture, outperforming SCAN-IT (NMI = 0.6354), SpaGCN (NMI = 0.4319), SEDR (NMI = 0.2970), scGPT-spatial (NMI = 0.2778), and Nicheformer (NMI = 0.2249). (**D**) Quantitative comparison of ARI and NMI across all methods for spatial domain reconstruction. (**E**) Spatial plots showing the activity of enriched pathways in the transition zone, including acute inflammatory response, complement activation, and regulation of blood coagulation, where white streamlines indicate gradient directions.



Supplementary Figure 5. Extended benchmarking OmniCell for single-cell cell-type annotation and spatial transcriptomics deconvolution.

(**A**) Full confusion matrices for OmniCell, scGPT, scFoundation, and Geneformer on the Zheng68k dataset, showing OmniCell's superior accuracy and clearer diagonal dominance across all major and rare cell types. (**B**) Bar chart of cell-type frequency distribution in the Zheng68k dataset. (**C–D**) Bubble plots comparing predicted versus ground truth cell-type proportions on the hPBMC (**C**) and LungCancer (**D**) datasets. Dot size corresponds to cell-type proportion; color intensity reflects prediction accuracy. Each row represents a method; each column represents a cell type. **(E–G**) Spatial deconvolution evaluation on MERFISH mouse brain atlas slices. (**E**) Spatial reconstructions of slice 9_0 comparing ground truth annotations with predictions from OmniCell, scGPT-spatial, Nicheformer, and cell2location. (**F**) Corresponding confusion matrices for slice 9_0. (**G**) Spatial reconstructions and confusion matrices for slice 11_2. Cell types include neuronal, glial, vascular, and immune populations as annotated in the reference atlas.

806

Reference

1. Liao, J., et al., *Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics.* Trends Biotechnol, 2021. **39**(1): p. 43-58.
2. Lee, J., M. Yoo, and J. Choi, *Recent advances in spatially resolved transcriptomics: challenges and opportunities.* BMB Rep, 2022. **55**(3): p. 113-124.
3. Gulati, G.S., et al., *Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics.* Nat Rev Mol Cell Biol, 2025. **26**(1): p. 11-31.
4. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. **8**: p. 14049.
5. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.
6. Picelli, S., et al., *Full-length RNA-seq from single cells using Smart-seq2.* Nat Protoc, 2014. **9**(1): p. 171-81.
7. Hagemann-Jensen, M., et al., *Single-cell RNA counting at allele and isoform resolution using Smart-seq3.* Nat Biotechnol, 2020. **38**(6): p. 708-714.
8. Hilts, K.E., et al., *Impact of Medicaid expansion on smoking prevalence and quit attempts among those newly eligible, 2011-2019.* Tob Prev Cessat, 2021. **7**: p. 16.
9. Moncada, R., et al., *Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas.* Nat Biotechnol, 2020. **38**(3): p. 333-342.
10. Chen, K.H., et al., *RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells.* Science, 2015. **348**(6233): p. aaa6090.
11. Eng, C.L., et al., *Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH.* Nature, 2019. **568**(7751): p. 235-239.
12. Wang, X., et al., *Three-dimensional intact-tissue sequencing of single-cell transcriptional states.* Science, 2018. **361**(6400).
13. Chen, X., et al., *Efficient in situ barcode sequencing using padlock probe-based BaristaSeq.* Nucleic Acids Res, 2018. **46**(4): p. e22.
14. Stahl, P.L., et al., *Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.* Science, 2016. **353**(6294): p. 78-82.
15. Rodriques, S.G., et al., *Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.* Science, 2019. **363**(6434): p. 1463-1467.
16. Chen, A., et al., *Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays.* Cell, 2022. **185**(10): p. 1777-1792 e21.
17. Cui, H., et al., *scGPT: toward building a foundation model for single-cell multi-omics using generative AI.* Nat Methods, 2024. **21**(8): p. 1470-1480.
18. Hao, M., et al., *Large-scale foundation model on single-cell transcriptomics.* Nat Methods, 2024. **21**(8): p. 1481-1491.
19. Theodoris, C.V., et al., *Transfer learning enables predictions in network biology.* Nature, 2023. **618**(7965): p. 616-624.
20. Szalata, A., et al., *Transformers in single-cell omics: a review and new perspectives.* Nat Methods, 2024. **21**(8): p. 1430-1443.
21. Wang, C., et al., *scGPT-spatial: Continual Pretraining of Single-Cell Foundation Model for Spatial Transcriptomics.* bioRxiv, 2025.

22. Tejada-Lapuerta, A., et al., *Nicheformer: a foundation model for single-cell and spatial omics.* Nat Methods, 2025.
23. Wen, H., et al., *CellPLM: Pre-training of Cell Language Model Beyond Single Cells.* 2023.
24. Vaswani, A., et al. *Attention Is All You Need*. 2017. arXiv:1706.03762 DOI: 10.48550/arXiv.1706.03762.
25. Zeng, H., et al., *Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures.* Cell, 2012. **149**(2): p. 483-96.
26. Gerfen, C.R. and D.J. Surmeier, *Modulation of striatal projection systems by dopamine.* Annu Rev Neurosci, 2011. **34**: p. 441-66.
27. Eng, L.F., et al., *An acidic protein isolated from fibrous astrocytes.* Brain Res, 1971. **28**(2): p. 351-4.
28. Butovsky, O., et al., *Identification of a unique TGF-beta-dependent molecular and functional signature in microglia.* Nat Neurosci, 2014. **17**(1): p. 131-43.
29. Nave, K.A. and H.B. Werner, *Myelination of the nervous system: mechanisms and functions.* Annu Rev Cell Dev Biol, 2014. **30**: p. 503-33.
30. Spitzer, S.O., et al., *Oligodendrocyte Progenitor Cells Become Regionally Diverse and Heterogeneous with Age.* Neuron, 2019. **101**(3): p. 459-471 e5.
31. Zeisel, A., et al., *Molecular Architecture of the Mouse Nervous System.* Cell, 2018. **174**(4): p. 999-1014 e22.
32. Jones, E.V., et al., *Astrocytes control glutamate receptor levels at developing synapses through SPARC-beta-integrin interactions.* J Neurosci, 2011. **31**(11): p. 4154-65.
33. Chien, E., et al. *Adaptive Universal Generalized PageRank Graph Neural Network*. 2020. arXiv:2006.07988 DOI: 10.48550/arXiv.2006.07988.
34. Wu, L., et al., *An invasive zone in human liver cancer identified by Stereo-seq promotes hepatocyte-tumor cell crosstalk, local immunosuppression and tumor progression.* Cell Res, 2023. **33**(8): p. 585-603.
35. Si, M. and J. Lang, *The roles of metallothioneins in carcinogenesis.* J Hematol Oncol, 2018. **11**(1): p. 107.
36. Wong, D.R., A.S. Hill, and R. Moccia, *Simple controls exceed best deep learning algorithms and reveal foundation model effectiveness for predicting genetic perturbations.* Bioinformatics, 2025. **41**(6).
37. Ahlmann-Eltze, C., W. Huber, and S. Anders, *Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines.* Nat Methods, 2025. **22**(8): p. 1657-1661.
38. Vinas Torne, R., et al., *Systema: a framework for evaluating genetic perturbation response prediction beyond systematic variation.* Nat Biotechnol, 2025.
39. Zhou, Y., et al., *Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease.* Nat Med, 2020. **26**(1): p. 131-142.
40. Wang, P., et al., *Molecular pathways and diagnosis in spatially resolved Alzheimer's hippocampal atlas.* Neuron, 2025. **113**(13): p. 2123-2140 e9.
41. De Simone, M., et al., *A comprehensive analysis framework for evaluating commercial single-cell RNA sequencing technologies.* Nucleic Acids Res, 2025. **53**(2).
42. Zhang, Y., et al., *Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response.* Proc Natl Acad Sci U S A, 2021. **118**(24).
43. Becker, W.R., et al., *Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer.* Nat Genet, 2022. **54**(7): p. 985-995.
44. Cang, Z., et al., *SCAN-IT: Domain segmentation of spatial transcriptomics images by graph neural*

network. BMVC, 2021. **32**.

45. Tran, H.T.N., et al., *A benchmark of batch-effect correction methods for single-cell RNA sequencing data.* Genome Biol, 2020. **21**(1): p. 12.
46. Siletti, K., et al., *Transcriptomic diversity of cell types across the adult human brain.* Science, 2023. **382**(6667): p. eadd7046.
47. Velmeshev, D., et al., *Single-cell analysis of prenatal and postnatal human cortical development.* Science, 2023. **382**(6667): p. eadf0834.
48. Ivanova, E.N., et al., *mRNA COVID-19 vaccine elicits potent adaptive immune response without the acute inflammation of SARS-CoV-2 infection.* iScience, 2023. **26**(12): p. 108572.
49. Jorstad, N.L., et al., *Comparative transcriptomics reveals human-specific cortical features.* Science, 2023. **382**(6667): p. eade9516.
50. Allen, W.E., et al., *Molecular and spatial signatures of mouse brain aging at single-cell resolution.* Cell, 2023. **186**(1): p. 194-208 e18.
51. Codeluppi, S., et al., *Spatial organization of the somatosensory cortex revealed by osmFISH.* Nat Methods, 2018. **15**(11): p. 932-935.
52. Sarkar, A. and M. Stephens, *Separating measurement and expression models clarifies confusion in single-cell RNA sequencing analysis.* Nat Genet, 2021. **53**(6): p. 770-777.
53. Dai, D., et al. *DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models*. 2024. arXiv:2401.06066 DOI: 10.48550/arXiv.2401.06066.
54. Wang, P., et al. *Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution*. 2024. arXiv:2409.12191 DOI: 10.48550/arXiv.2409.12191.
55. Burkhardt, D.B., et al., *Quantifying the effect of experimental perturbations at single-cell resolution.* Nat Biotechnol, 2021. **39**(5): p. 619-629.
56. Moon, K.R., et al., *Visualizing structure and transitions in high-dimensional biological data.* Nat Biotechnol, 2019. **37**(12): p. 1482-1492.
57. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587 e29.
58. Luecken, M.D., et al., *Benchmarking atlas-level data integration in single-cell genomics.* Nat Methods, 2022. **19**(1): p. 41-50.
59. Xu, H., et al., *Unsupervised spatially embedded deep representation of spatial transcriptomics.* Genome Med, 2024. **16**(1): p. 12.
60. Hu, J., et al., *SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network.* Nat Methods, 2021. **18**(11): p. 1342-1351.
61. Wang, A., et al., *Mapping spatial gradients in spatial transcriptomics data with score matching.* bioRxiv, 2025.
62. Biancalani, T., et al., *Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram.* Nat Methods, 2021. **18**(11): p. 1352-1362.