Supplementary Figures

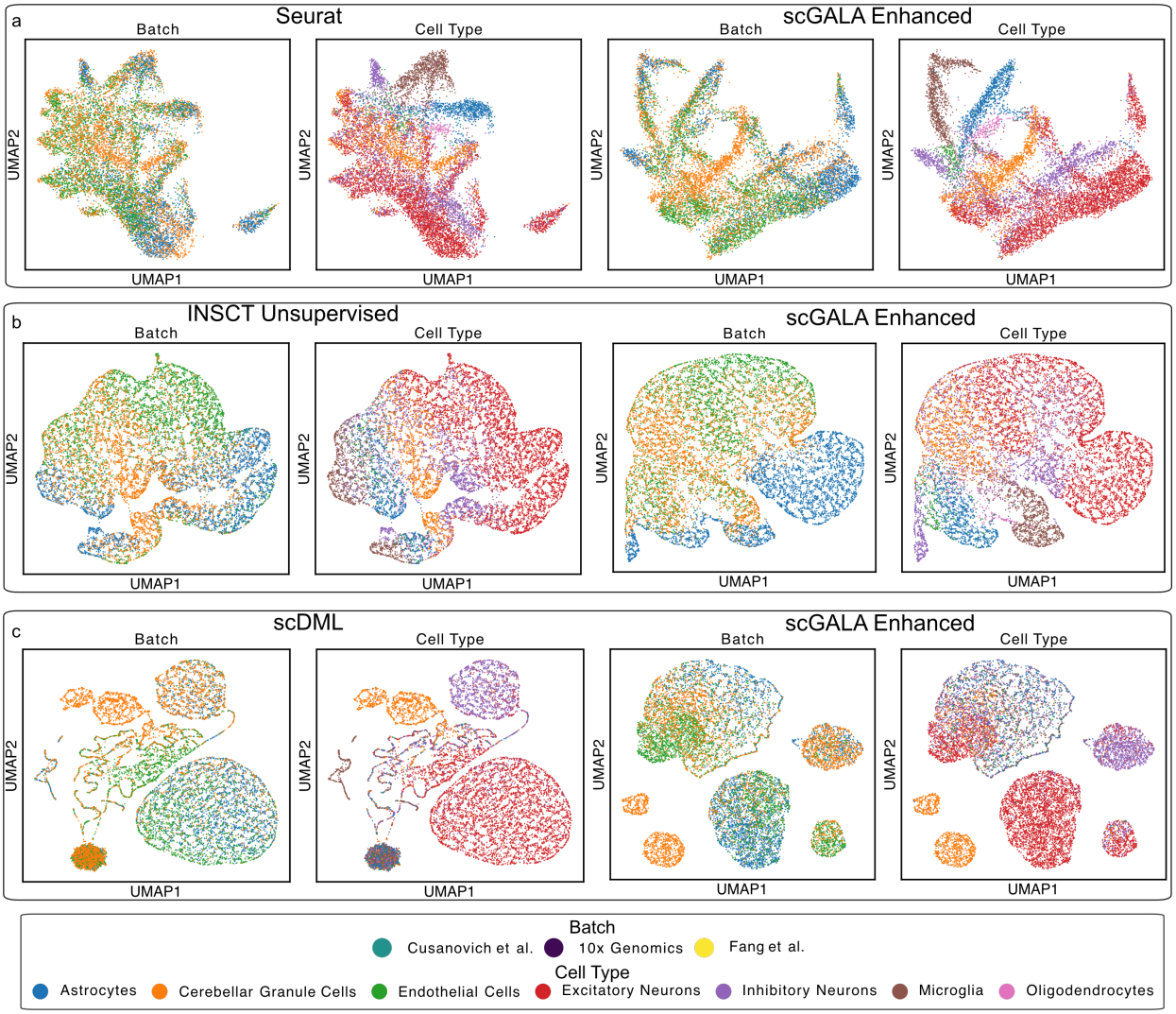

Fig. S1 UMAP Visualizations Colored by Batch (left) and Cell Type (right) Comparing Batch Correction Results between Existing Methods and Their scGALA Enhanced Versions. a-c, UMAPs comparing batch correction results of Seurat, INSCT Unsupervised, scDML and their scGALA-enhanced versions. The improved cell type isolation highlights the advancement produced by integrating scGALA into existing methods.

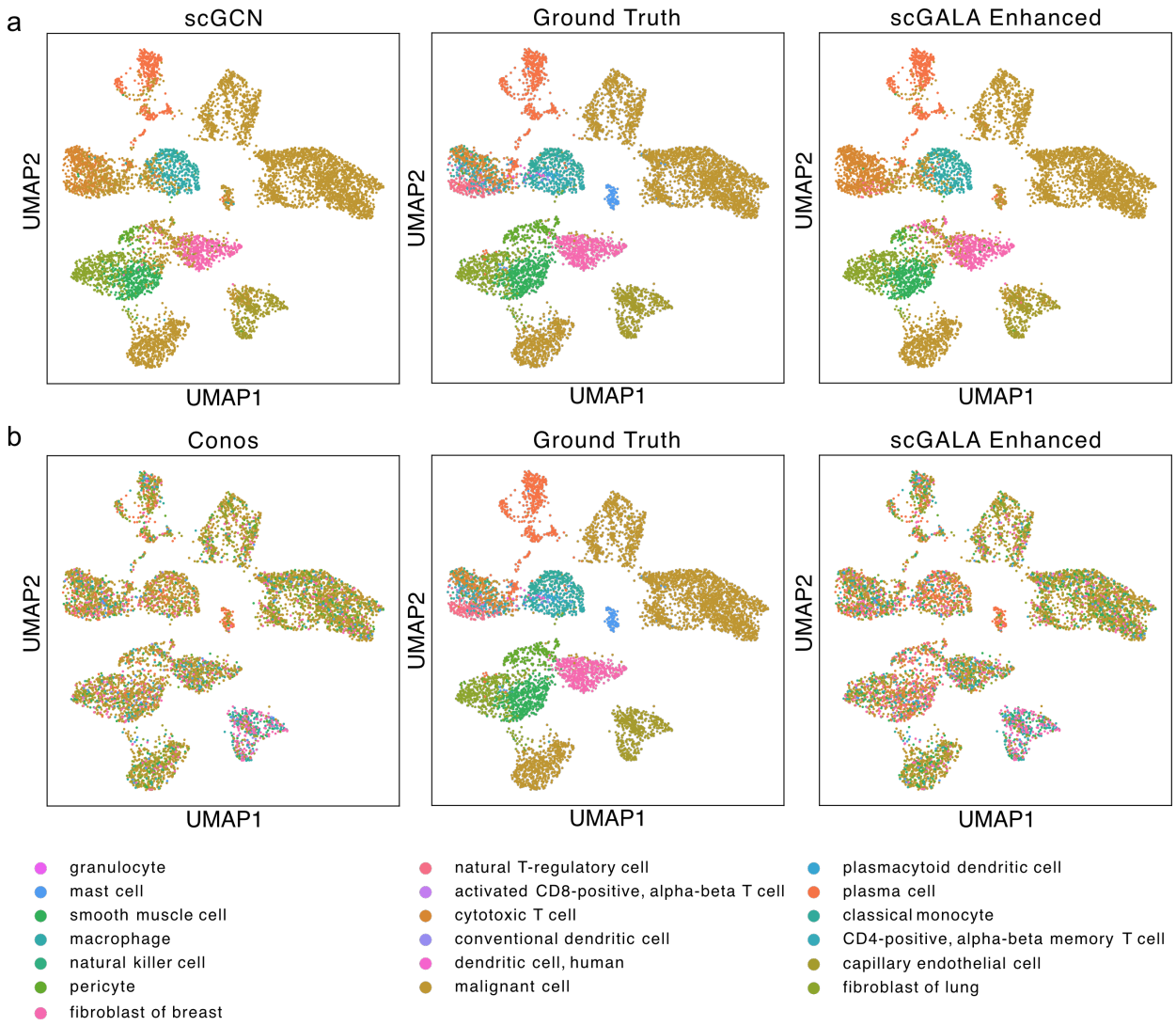

Fig. S2 UMAP Visualizations Colored by Transferred Cell Type Labels Comparing the Label Transfer Results between Existing Methods and Their scGALA Enhanced Versions. a-b, UMAPs comparing the label transfer results of scGCN, Conos and their scGALA-enhanced versions. Better correspondence with the Ground Truth label indicates better label transfer performance.

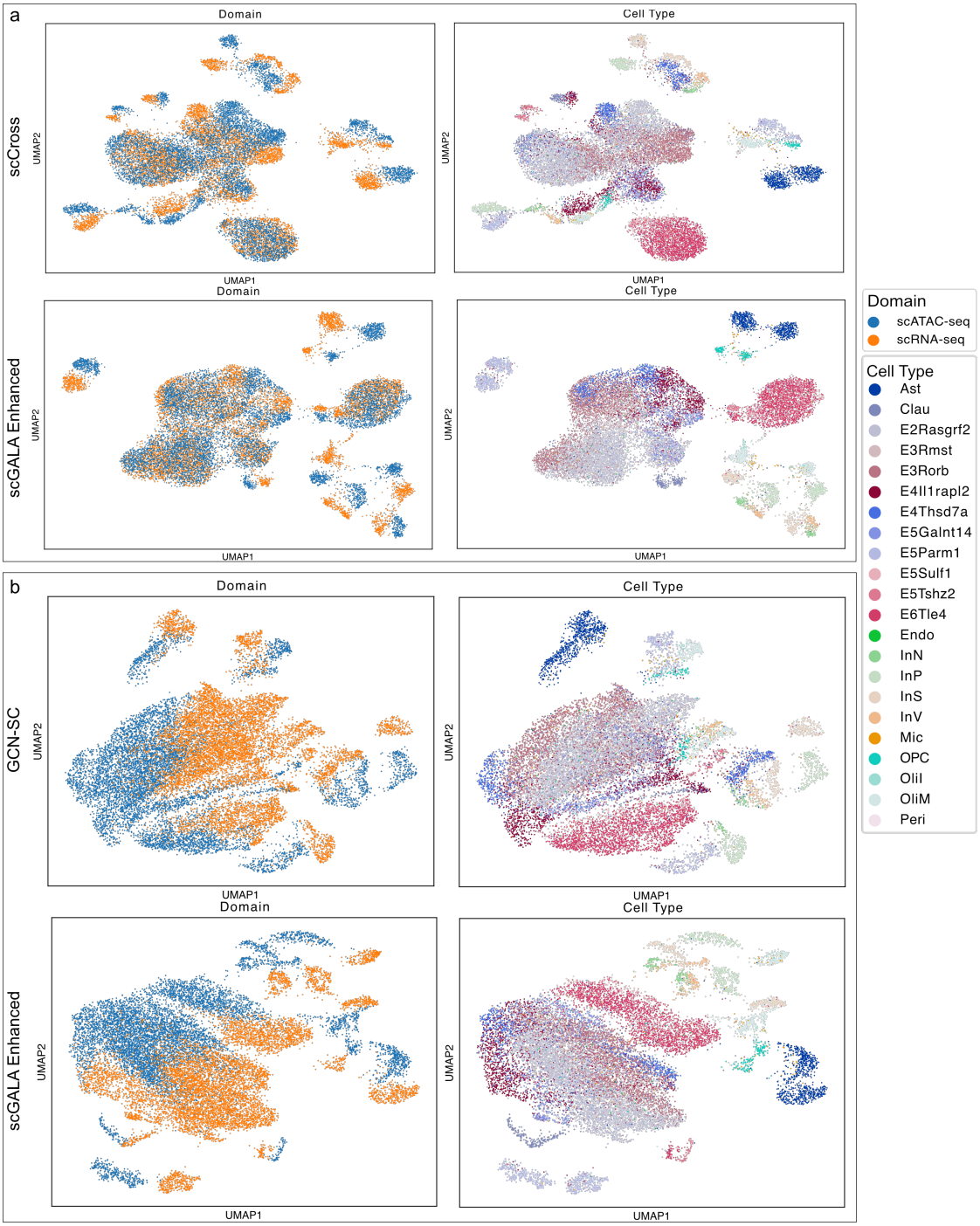

Fig. S3 UMAP Visualizations Colored by Modality Domain (left) and Joint Cell Type Labels (right) Comparing the Multiomics Integration (RNA + ATAC) Results between Existing Methods and Their scGALA Enhanced Versions. a-b, UMAPs comparing the multiomics integration results of scCross, GCN-SC and their scGALA-enhanced versions. The improved cell type isolation highlights the advancement produced by integrating scGALA into existing methods.

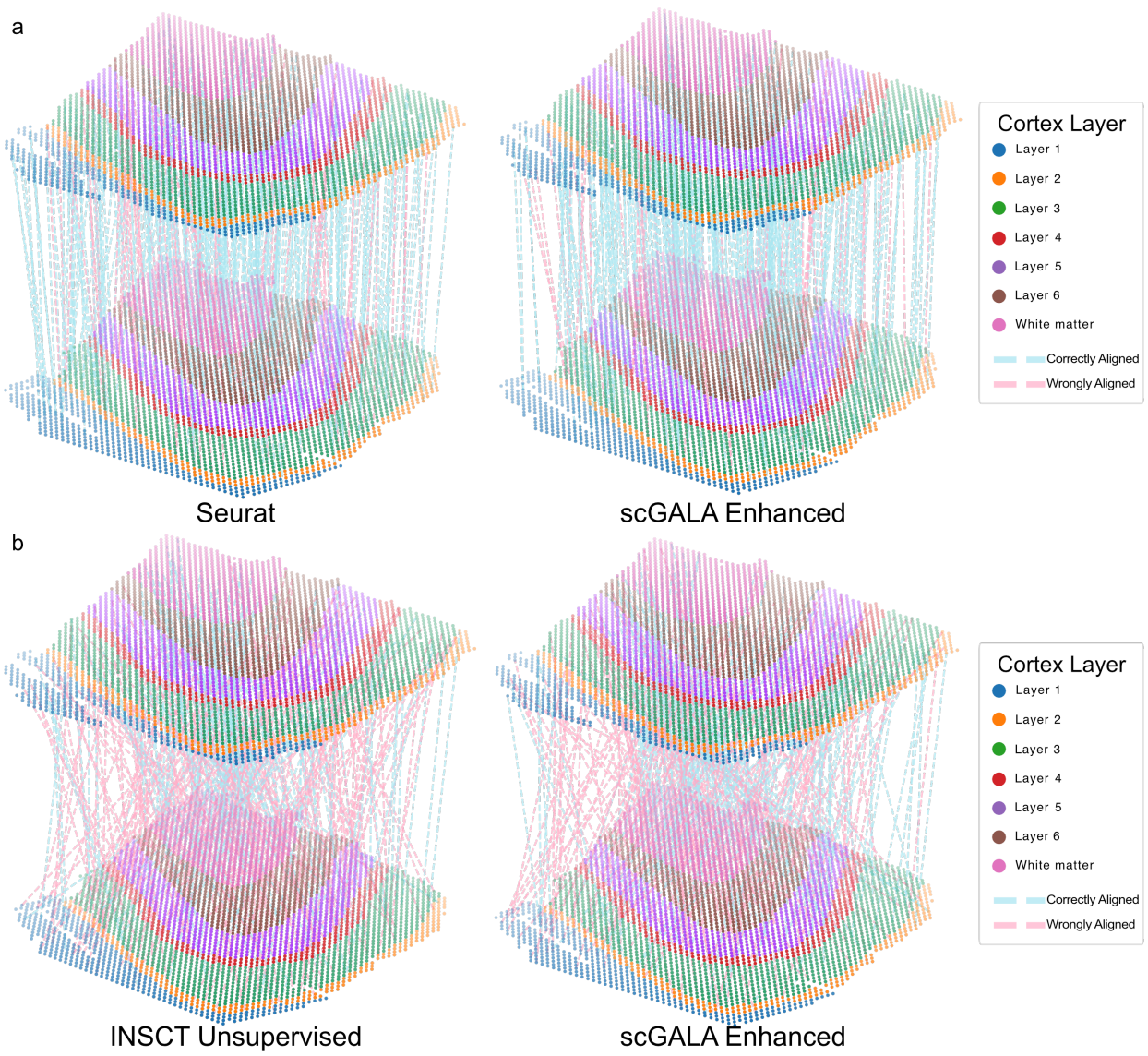

Fig. S4 Spatial Alignment of Tissue Slices: Spatial Coordinate Plots (X/Y/Z tissue positions) Colored by Spatial Region Labels Comparing the Alignments Results between Existing Methods and Their scGALA Enhanced Versions. a-b, Spatial coordinate plots (X/Y/Z tissue positions) comparing the spatial alignment results of Seurat, INSCT Unsupervised and their scGALA-enhanced versions. The improved alignment accuracy highlights the advancement produced by integrating scGALA into existing methods.

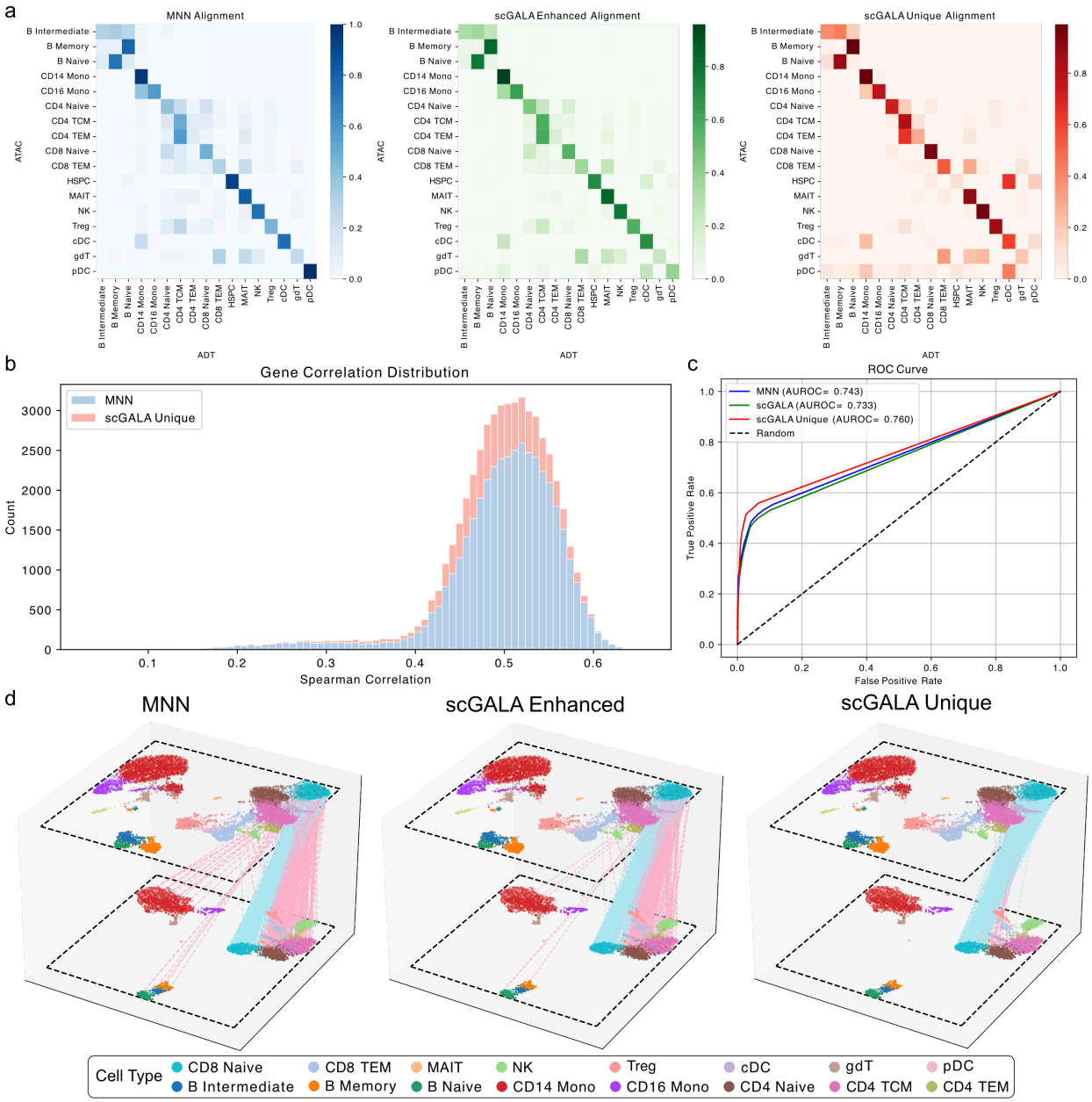

Fig. S5 scGALA's Ability to Unify Multiomics Dataset Demonstrated on the P4 Subsample. a, Cell type-specific alignment precision: Confusion matrices comparing MNN (baseline), scGALA-enhanced (MNN + scGALA), and scGALA-exclusive alignments. Diagonal enrichment highlights scGALA's improved accuracy in triple-omics integration. **b,** Spearman's correlation distribution of aligned cell pairs from MNN (baseline) versus scGALA-exclusive matches. Comparable distributions confirm scGALA maintains alignment quality while expanding coverage. **c,** ROC curves quantifying alignment accuracy for cell type matching, with AUROC values demonstrating scGALA's superior performance in cross-modality integration. **d,** Spatial UMAP visualization of cross-dataset alignment, with connecting lines indicating alignments from MNN (baseline), scGALA-enhanced (MNN + scGALA), and scGALA-exclusive alignments (blue: correct cell type matches; pink: mismatches).

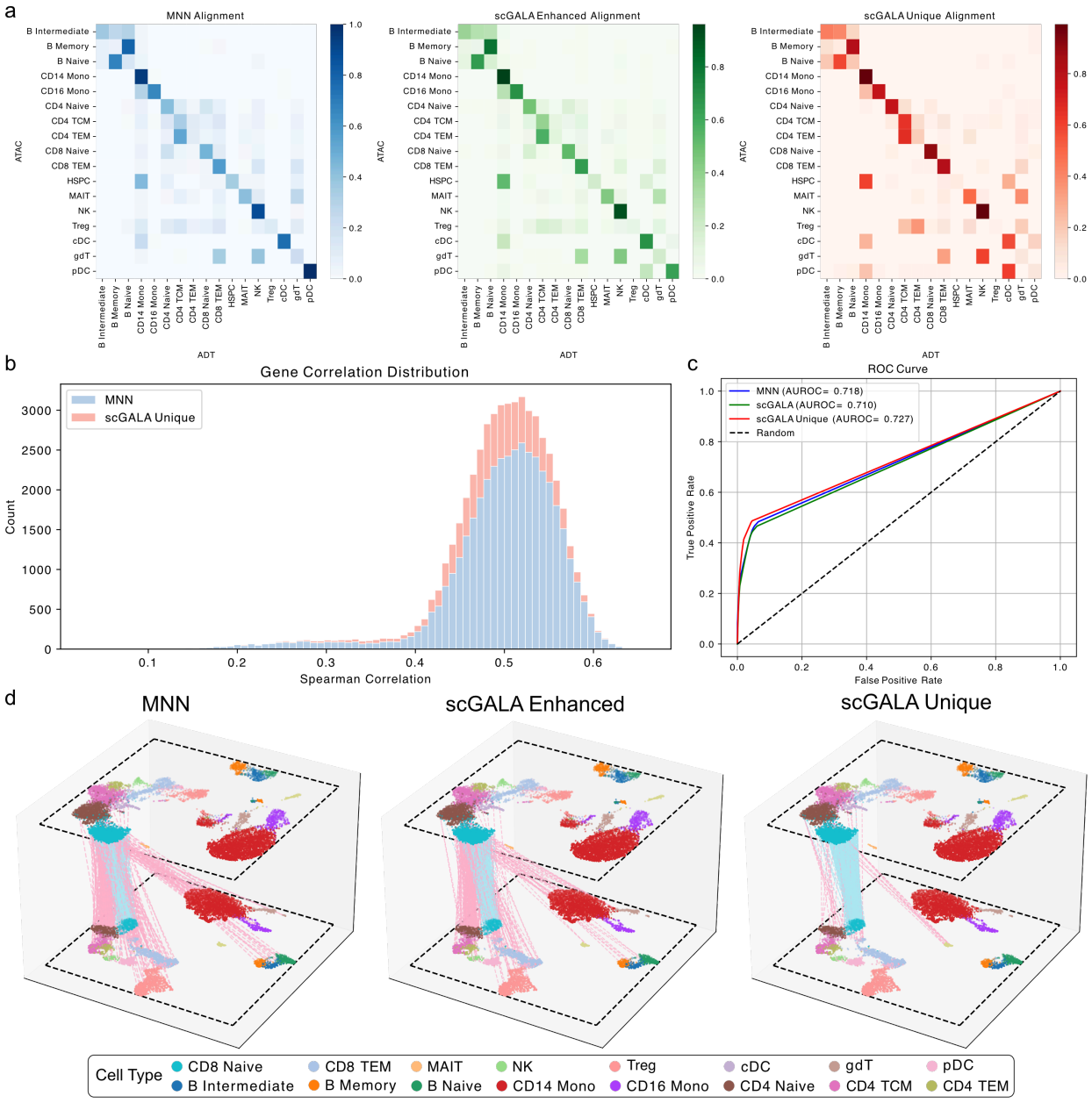

Fig. S6 scGALA's Ability to Unify Multiomics Dataset Demonstrated on the P5 Subsample. a, Cell type-specific alignment precision: Confusion matrices comparing MNN (baseline), scGALA-enhanced (MNN + scGALA), and scGALA-exclusive alignments. Diagonal enrichment highlights scGALA's improved accuracy in triple-omics integration. **b,** Spearman's correlation distribution of aligned cell pairs from MNN (baseline) versus scGALA-exclusive matches. Comparable distributions confirm scGALA maintains alignment quality while expanding coverage. **c,** ROC curves quantifying alignment accuracy for cell type matching, with AUROC values demonstrating scGALA's superior performance in cross-modality integration. **d,** Spatial UMAP visualization of cross-dataset alignment, with connecting lines indicating alignments from MNN (baseline), scGALA-enhanced (MNN + scGALA), and scGALA-exclusive alignments (blue: correct cell type matches; pink: mismatches).

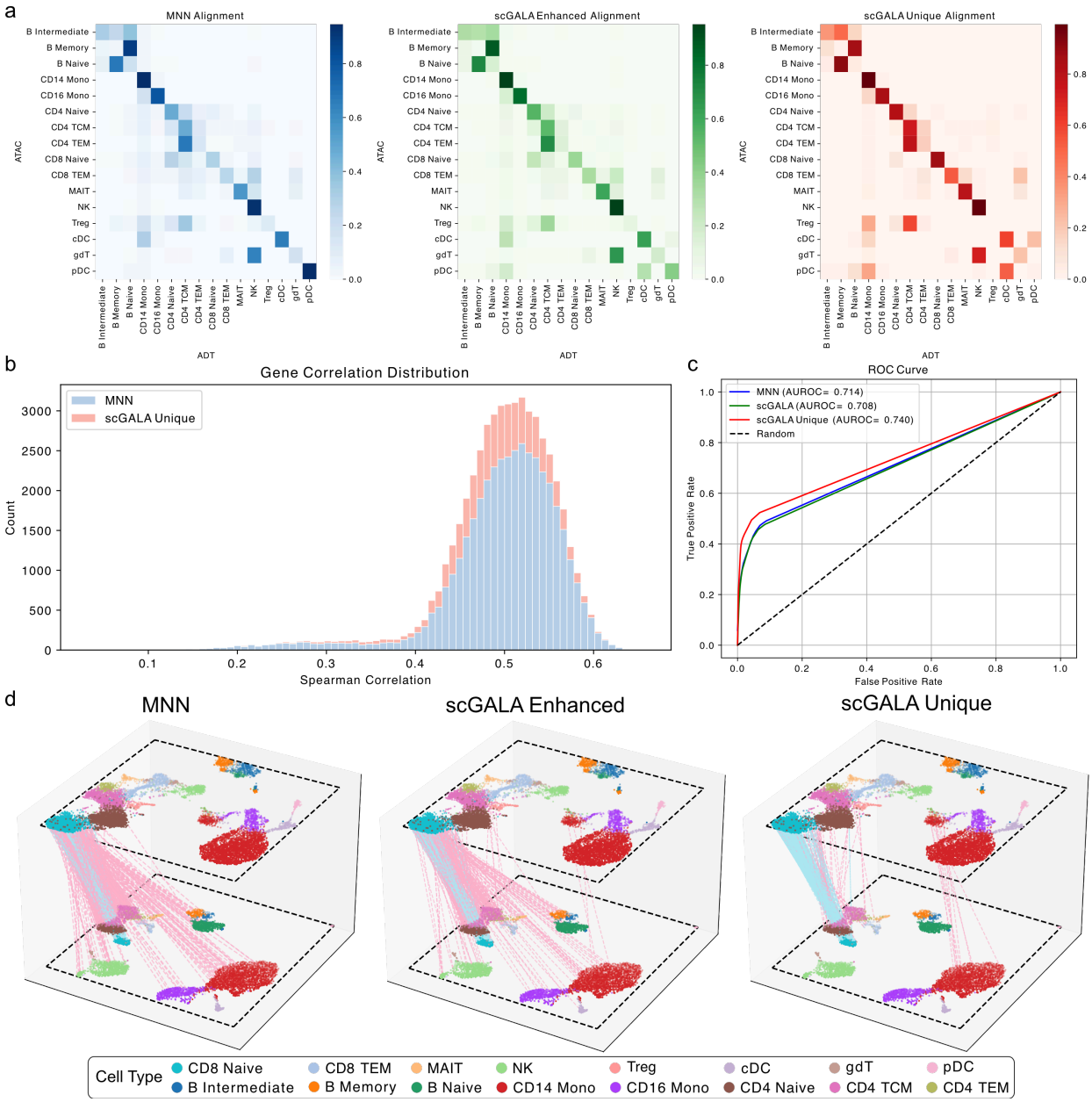

Fig. S7 scGALA's Ability to Unify Multiomics Dataset Demonstrated on the P6 Subsample. a, Cell type-specific alignment precision: Confusion matrices comparing MNN (baseline), scGALA-enhanced (MNN + scGALA), and scGALA-exclusive alignments. Diagonal enrichment highlights scGALA's improved accuracy in triple-omics integration. **b,** Spearman's correlation distribution of aligned cell pairs from MNN (baseline) versus scGALA-exclusive matches. Comparable distributions confirm scGALA maintains alignment quality while expanding coverage. **c,** ROC curves quantifying alignment accuracy for cell type matching, with AUROC values demonstrating scGALA's superior performance in cross-modality integration. **d,** Spatial UMAP visualization of cross-dataset alignment, with connecting lines indicating alignments from MNN (baseline), scGALA-enhanced (MNN + scGALA), and scGALA-exclusive alignments (blue: correct cell type matches; pink: mismatches).

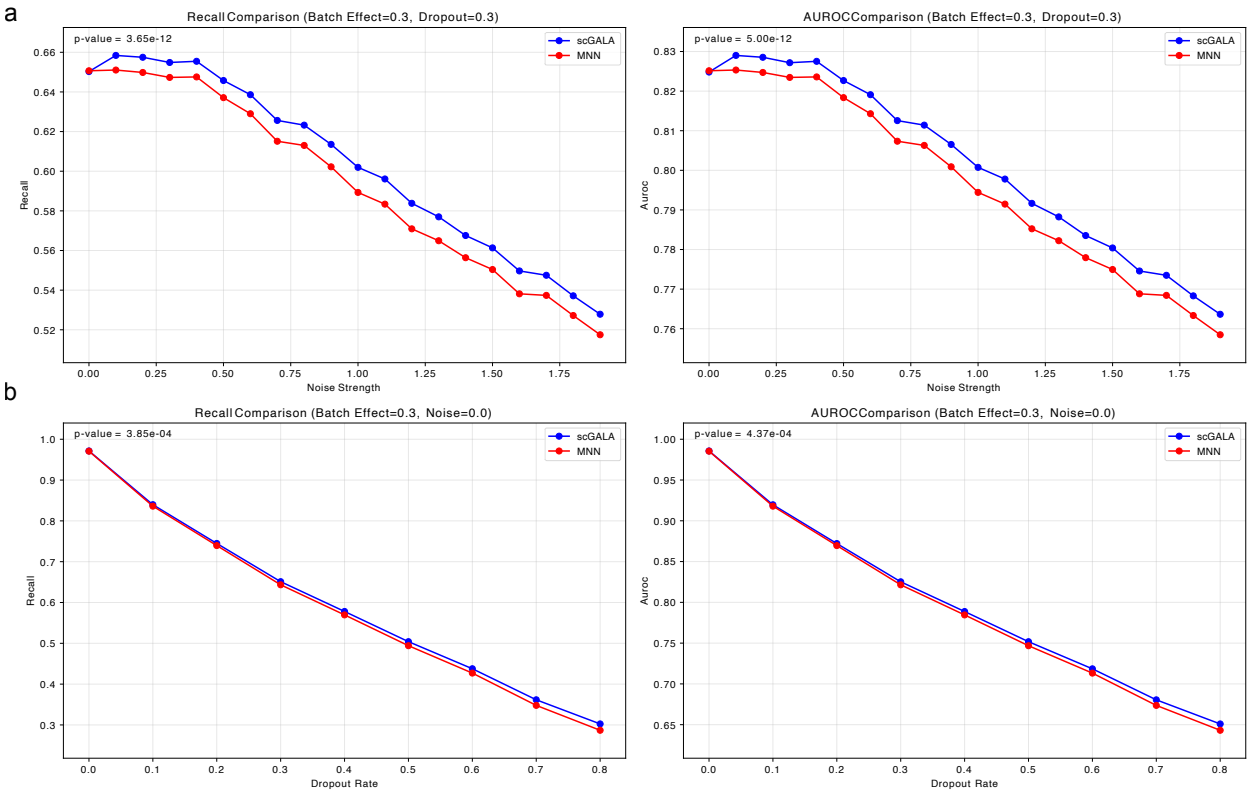

Fig. S8 scGALA's Ability to Unify Multiomics Dataset Demonstrated on the NEAT-seq Dataset. a, Recall and AUROC results of aligned cell pairs from MNN (baseline) versus scGALA-exclusive matches based on ground truth cell barcodes over a variety of noise strength with fixed batch effect and dropout rate. **b,** Recall and AUROC results of aligned cell pairs from MNN (baseline) versus scGALA-exclusive matches based on ground truth cell barcodes over a variety of dropout rate with fixed batch effect and noise strength. *P*-values are calculated using one-sided Student's *t*-test. The consistent higher value indicates the robust performance of scGALA alignment.

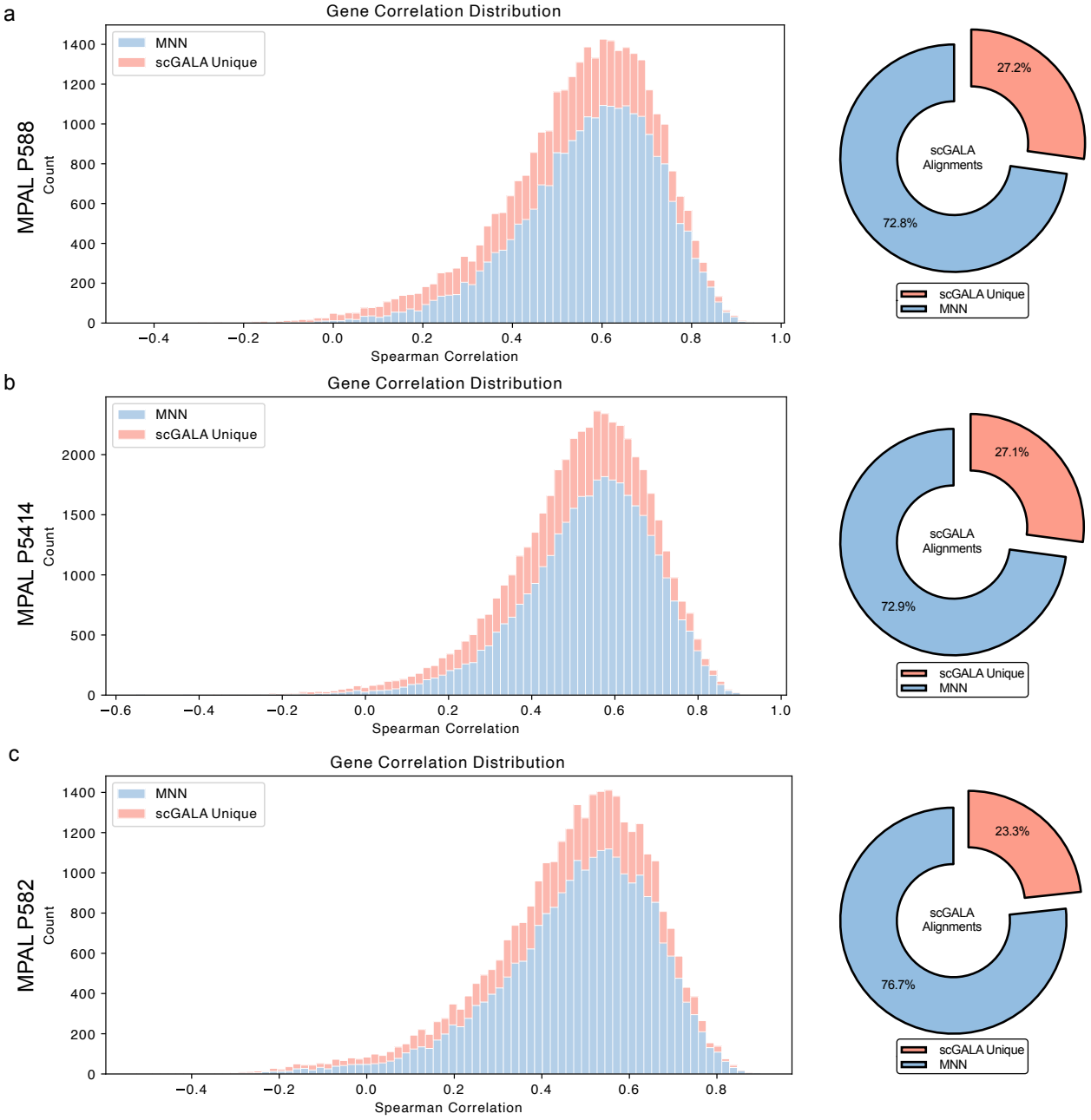

Fig. S9 scGALA Unifies Multiomics Datasets Through Enhanced Cell Alignment Beyond Conventional Matches. a-c, Spearman's correlation distribution of aligned cell pairs from MNN (baseline) versus scGALA-exclusive matches and composition of scGALA alignments (doughnut chart). Comparable distributions demonstrate that scGALA preserves the quality of alignment while expanding the coverage. scGALA identifies new cell pairs (unique to scGALA) while retaining high-confidence MNN matches, enabling comprehensive integration of multiomics data.

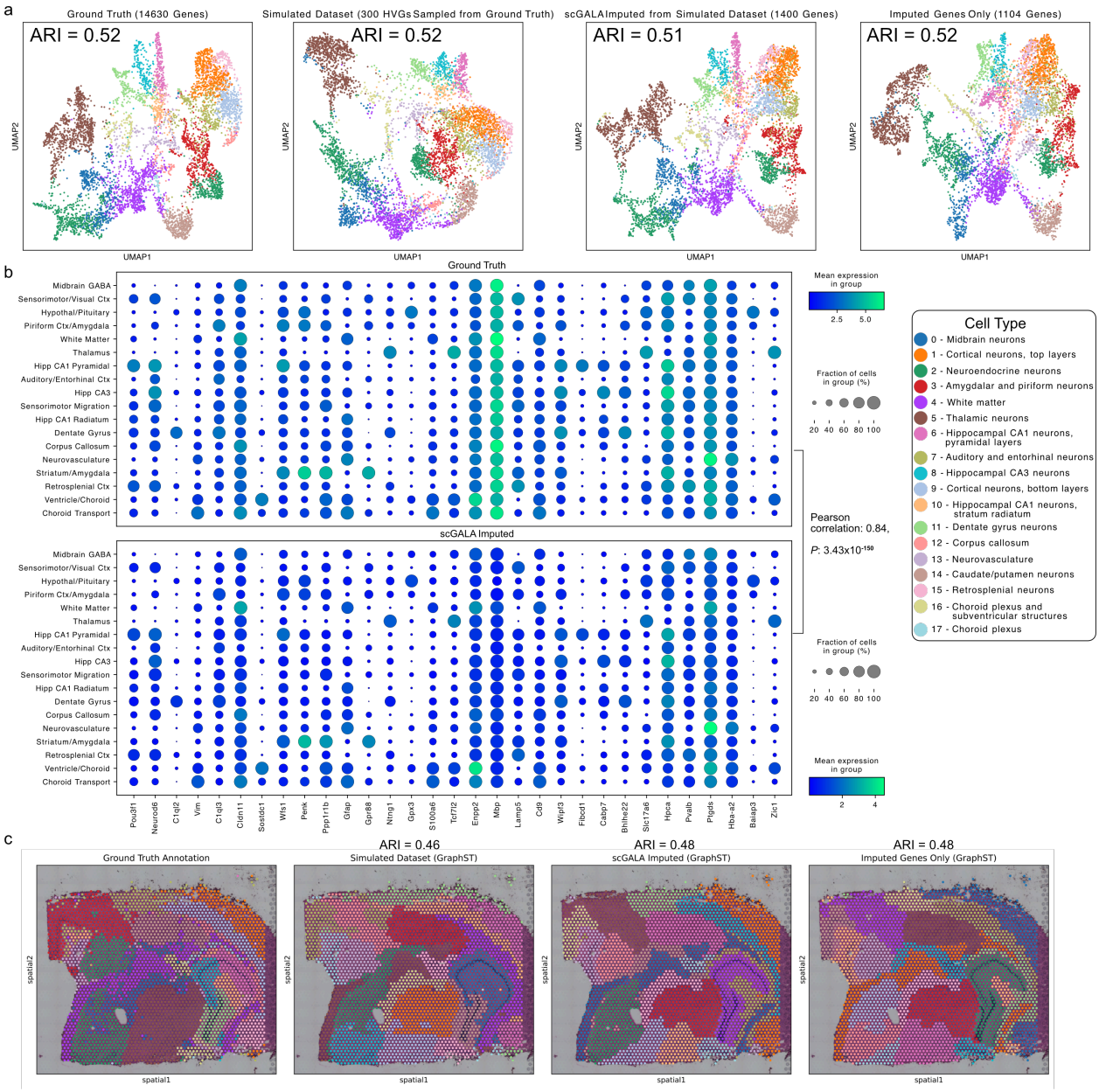

Fig. S10 scGALA Enhances Spatial Transcriptomics Imputation via Alignment-Guided Graph Reconstruction for Multiomics Integration. **a,** Imputation accuracy across datasets: UMAPs compare ground truth (14,630 genes), simulated sparse data (300 high-variance genes), scGALA-imputed data (1,400 genes; using alignment-guided graph reconstruction), and imputed-only genes (1,104 genes). Adjusted Rand Index (ARI) quantifies clustering concordance with ground truth cell types. **b,** Cell type-specific marker gene recovery: Dot plots show Pearson's correlation between imputed and ground truth expression for key marker genes, validating scGALA's precision in preserving biological signals. *P*-value is calculated using two-sided Student's *t*-test. **c,** Spatial domain preservation: Spatial clustering using GraphST on ground truth and scGALA-imputed data (retaining original spatial coordinates). ARI quantifies agreement between scGALA-imputed spatial domains and ground truth annotations (clustering resolution optimized for ground truth labels).

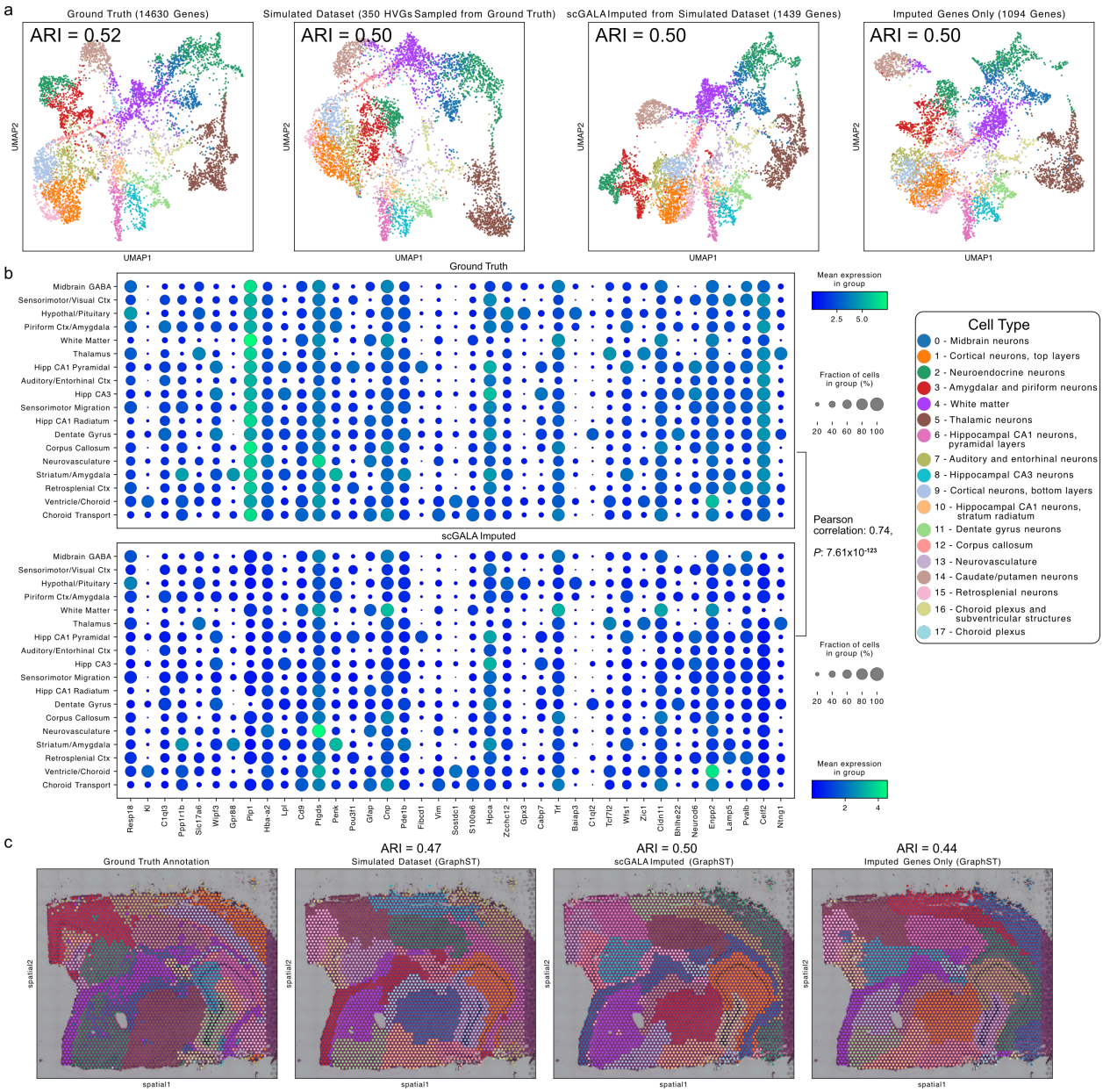

Fig. S11 scGALA Enhances Spatial Transcriptomics Imputation via Alignment-Guided Graph Reconstruction for Multiomics Integration. **a,** Imputation accuracy across datasets: UMAPs compare ground truth (14,630 genes), simulated sparse data (350 high-variance genes), scGALA-imputed data (1,439 genes; using alignment-guided graph reconstruction), and imputed-only genes (1,094 genes). Adjusted Rand Index (ARI) quantifies clustering concordance with ground truth cell types. **b,** Cell type-specific marker gene recovery: Dot plots show Pearson's correlation between imputed and ground truth expression for key marker genes, validating scGALA's precision in preserving biological signals. *P*-value is calculated using two-sided Student's *t*-test. **c,** Spatial domain preservation: Spatial clustering using GraphST on ground truth and scGALA-imputed data (retaining original spatial coordinates). ARI quantifies agreement between scGALA-imputed spatial domains and ground truth annotations (clustering resolution optimized for ground truth labels).

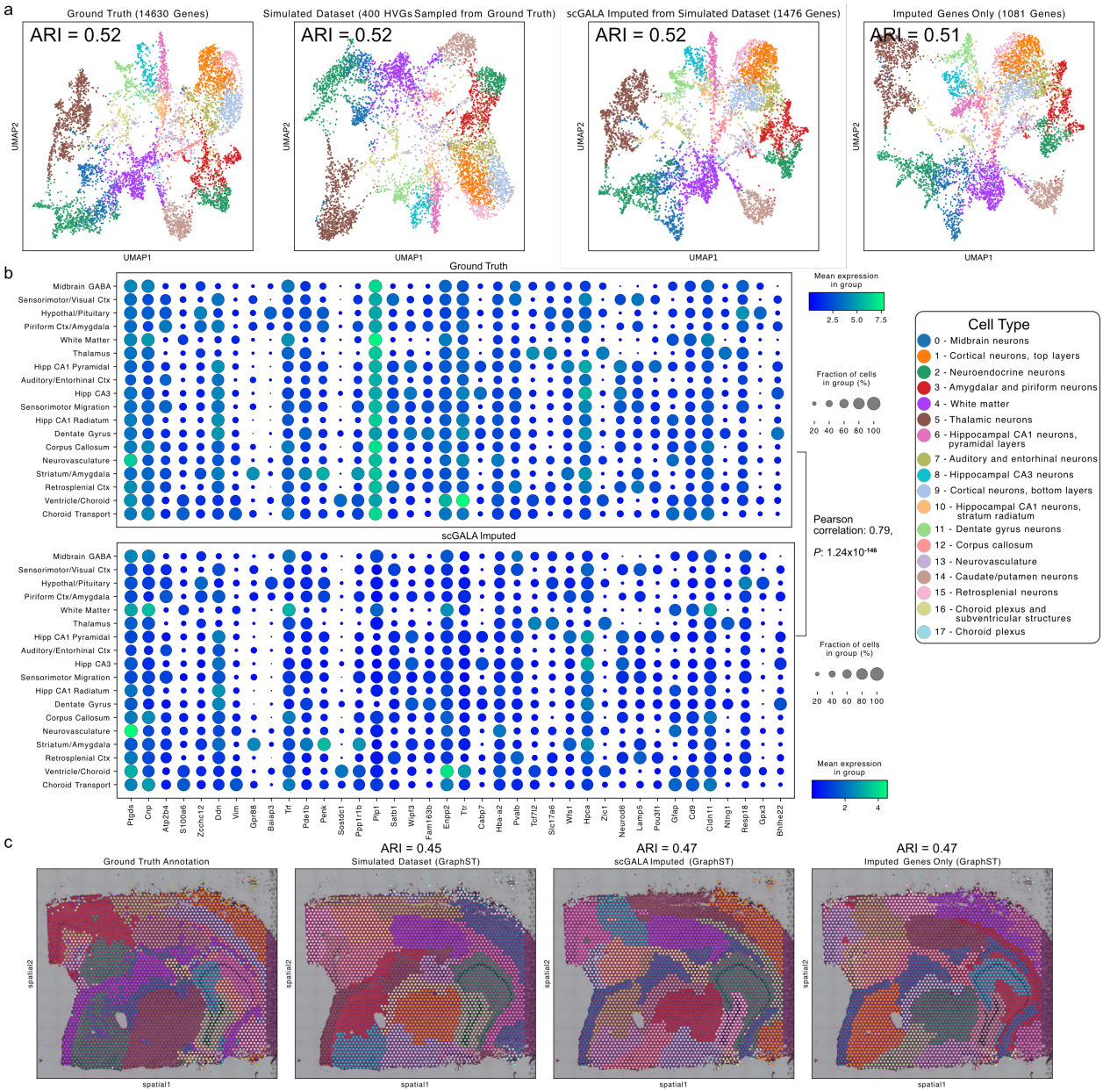

Fig. S12 scGALA Enhances Spatial Transcriptomics Imputation via Alignment-Guided Graph Reconstruction for Multiomics Integration. **a,** Imputation accuracy across datasets: UMAPs compare ground truth (14,630 genes), simulated sparse data (400 high-variance genes), scGALA-imputed data (1,476 genes; using alignment-guided graph reconstruction), and imputed-only genes (1,081 genes). Adjusted Rand Index (ARI) quantifies clustering concordance with ground truth cell types. **b,** Cell type-specific marker gene recovery: Dot plots show Pearson's correlation between imputed and ground truth expression for key marker genes, validating scGALA's precision in preserving biological signals. *P*-value is calculated using two-sided Student's *t*-test. **c,** Spatial domain preservation: Spatial clustering using GraphST on ground truth and scGALA-imputed data (retaining original spatial coordinates). ARI quantifies agreement between scGALA-imputed spatial domains and ground truth annotations (clustering resolution optimized for ground truth labels).

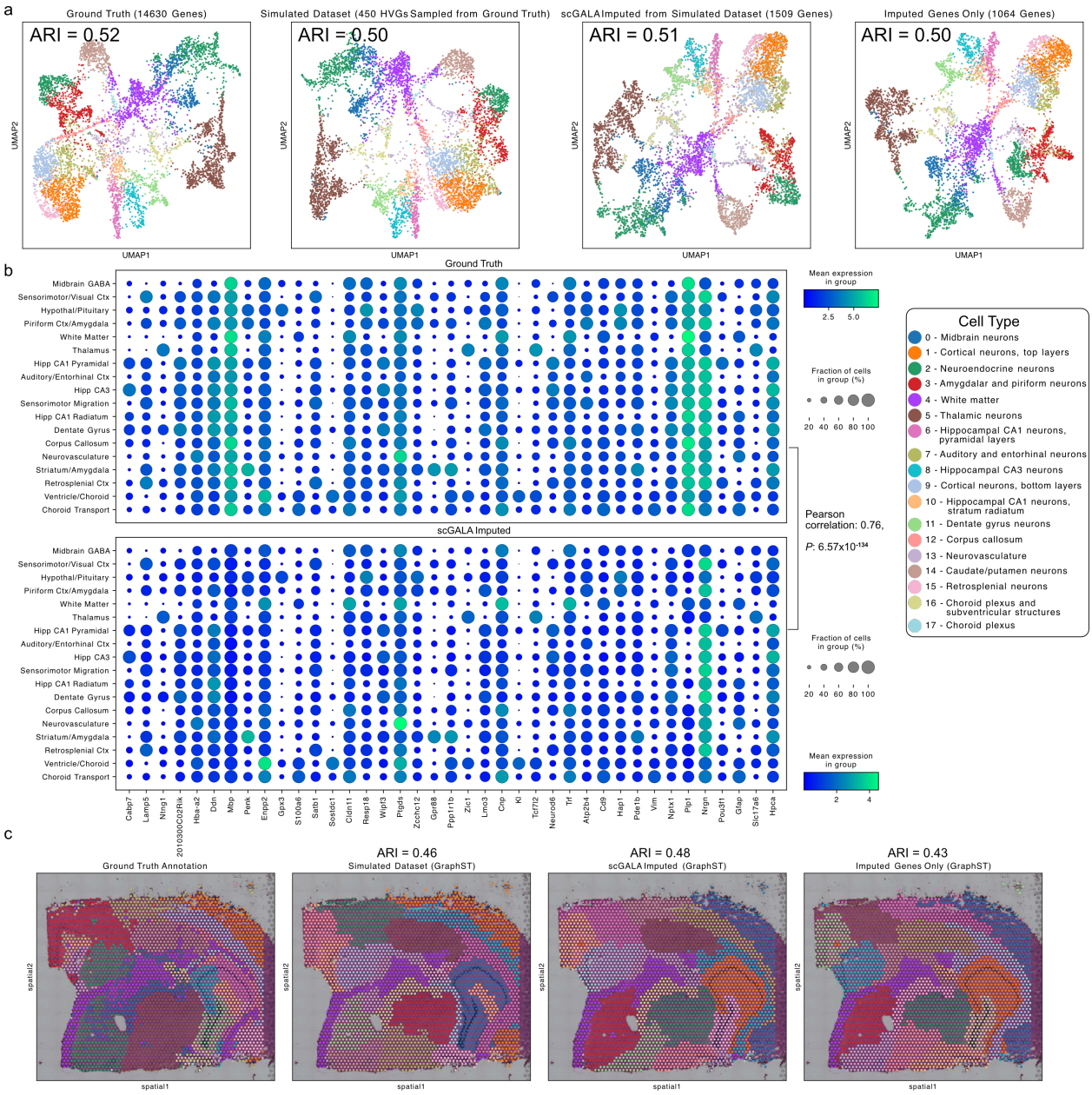

Fig. S13 scGALA Enhances Spatial Transcriptomics Imputation via Alignment-Guided Graph Reconstruction for Multiomics Integration. **a,** Imputation accuracy across datasets: UMAPs compare ground truth (14,630 genes), simulated sparse data (450 high-variance genes), scGALA-imputed data (1,509 genes; using alignment-guided graph reconstruction), and imputed-only genes (1,064 genes). Adjusted Rand Index (ARI) quantifies clustering concordance with ground truth cell types. **b,** Cell type-specific marker gene recovery: Dot plots show Pearson's correlation between imputed and ground truth expression for key marker genes, validating scGALA's precision in preserving biological signals. *P*-value is calculated using two-sided Student's *t*-test. **c,** Spatial domain preservation: Spatial clustering using GraphST on ground truth and scGALA-imputed data (retaining original spatial coordinates). ARI quantifies agreement between scGALA-imputed spatial domains and ground truth annotations (clustering resolution optimized for ground truth labels).

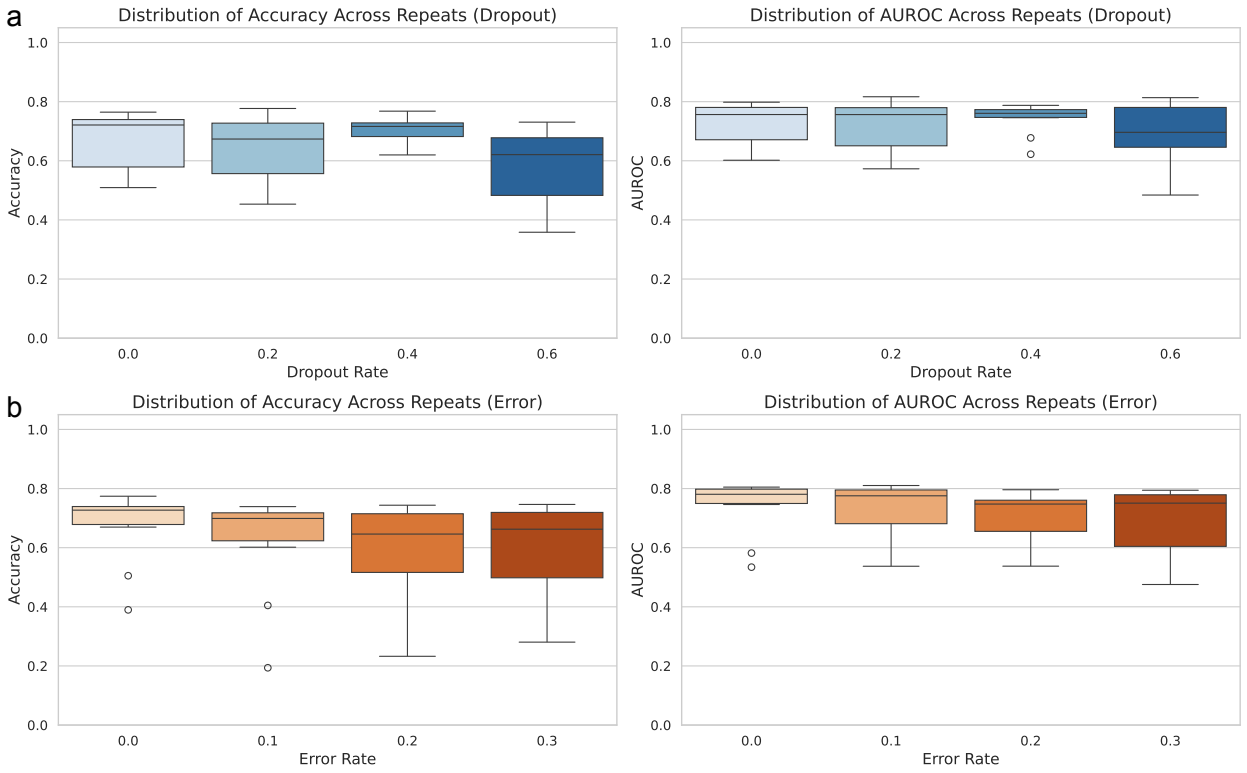

Fig. S14 scGALA Maintains Stable Performance under Diverse Initial Alignments. **a,** Alignment accuracy and AUROC based on the cell type label under a range of dropout effects imposed on the initial MNN alignments. The small decrease in performance under a large dropout ratio of 0.6 highlights the robustness of the scGALA alignment. **b,** Alignment accuracy and AUROC based on the cell type label under a range of rates of random errors added into the initial MNN alignments. The small decrease in performance under a large error rate of 0.3 indicates the robustness of the scGALA alignment.

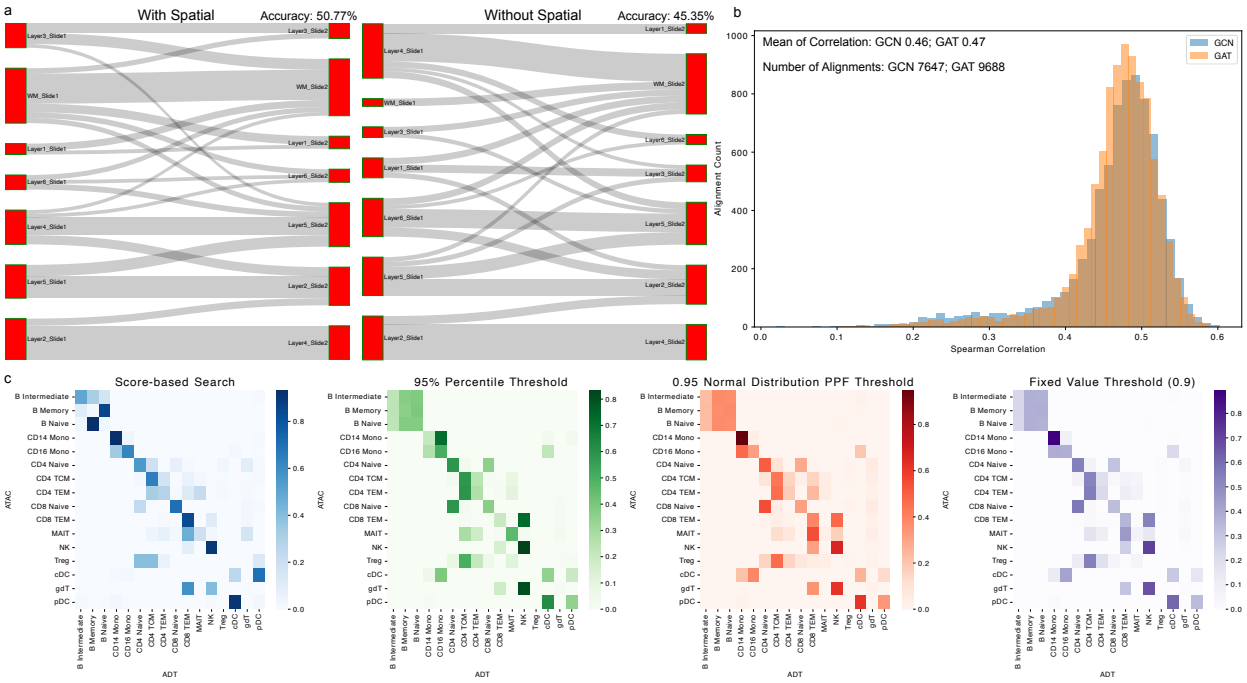

Fig. S15 Visualizations of Ablation Study Comparing Different Options of scGALA Components. a, The sankey diagrams comparing the spatial alignment results of scGALA with and without using spatial information. The experiment was conducted on the Rodent Research-3 dataset. **b,** The spearman correlation distributions comparing the multiplet integration results of scGALA using GAT and GCN. The experiment was conducted between the bone marrow CITE-seq dataset and the P7 subsample of 10X PBMC Multiome dataset. **c,** Cell type-specific alignment precision: Confusion matrices comparing diverse alignment strategies: Score-based Search, 95% Percentile Threshold, 0.95 Normal Distribution PPF Threshold and Fixed Value Threshold (0.9). Diagonal enrichment highlights the improved alignment ability of score-based search in triple-omics integration compared to threshold-based alignments.

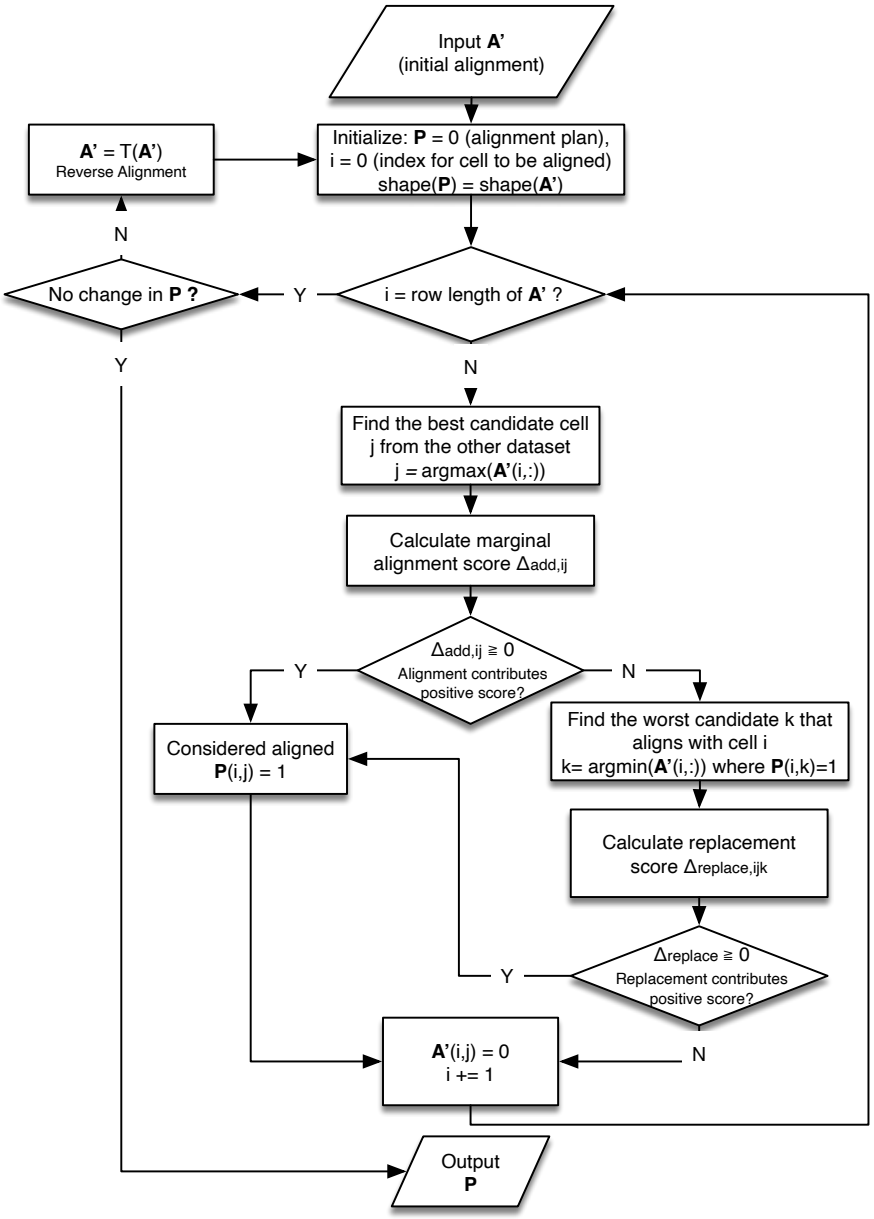

Fig. S16 Diagram of the Score-based Search for Cell Alignment. The score-based search algorithm refines initial cell-cell relationships predicted by the GAT model, enhancing flexibility in matching while preventing over-alignment through a quadratic penalty. It formulates a constrained maximization problem to optimize a binary alignment matrix, balancing many-to-many relationships with alignment control. The iterative process alternates between datasets, calculating marginal alignment scores to accept or replace alignments based on improvement. The final alignment matrix integrates optimized plans with high-confidence anchor pairs, yielding a comprehensive map of cell-cell correspondences across datasets. T() represents the matrix transpose. **A'** represents the edge probability matrix and **P** denotes the alignment plan. Reverse alignment means that the algorithm alternates between datasets to perform alignments from the other direction.