

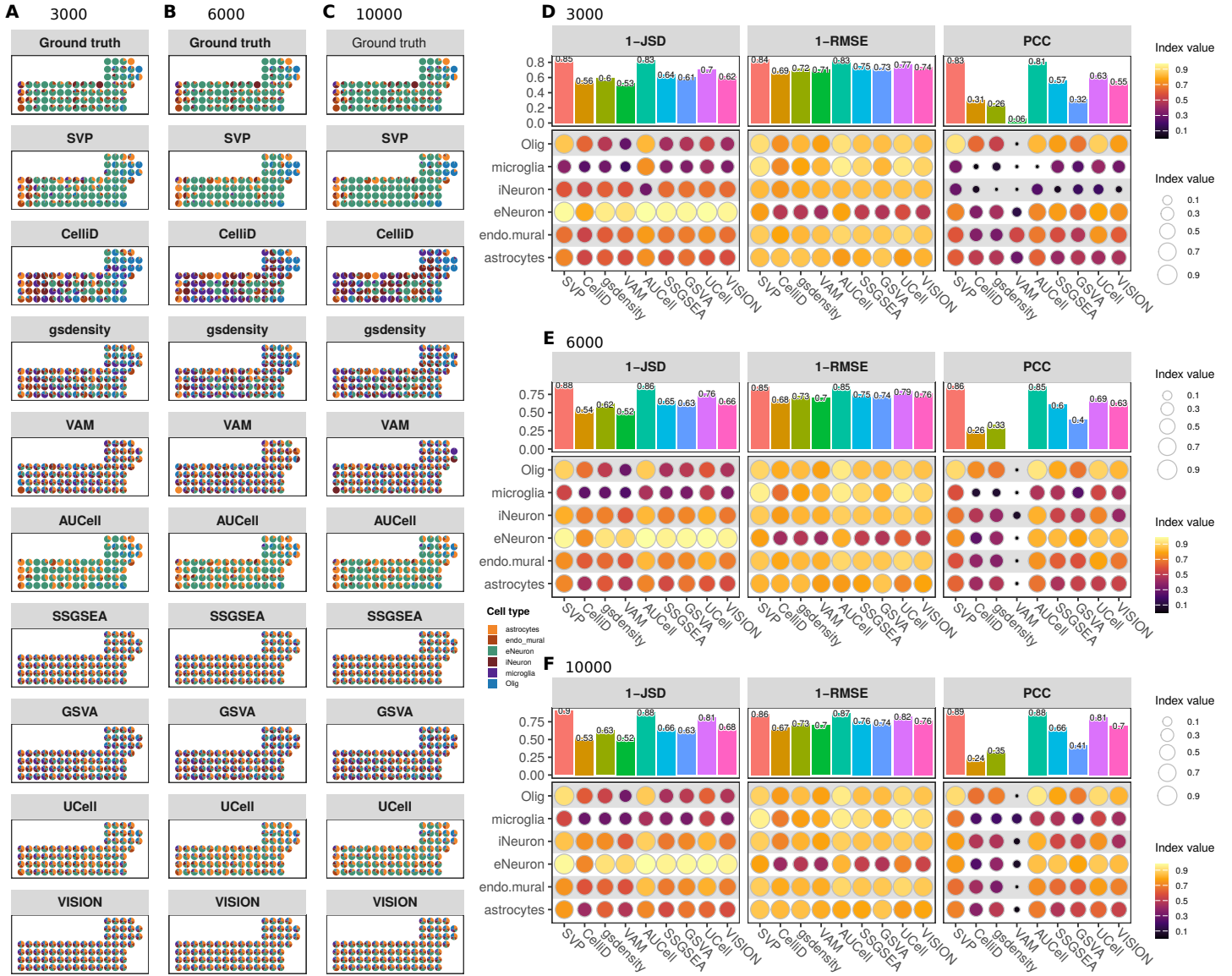
Supplemental File of

Quantitative characterization of cellular states and spatial variable patterns with SVP

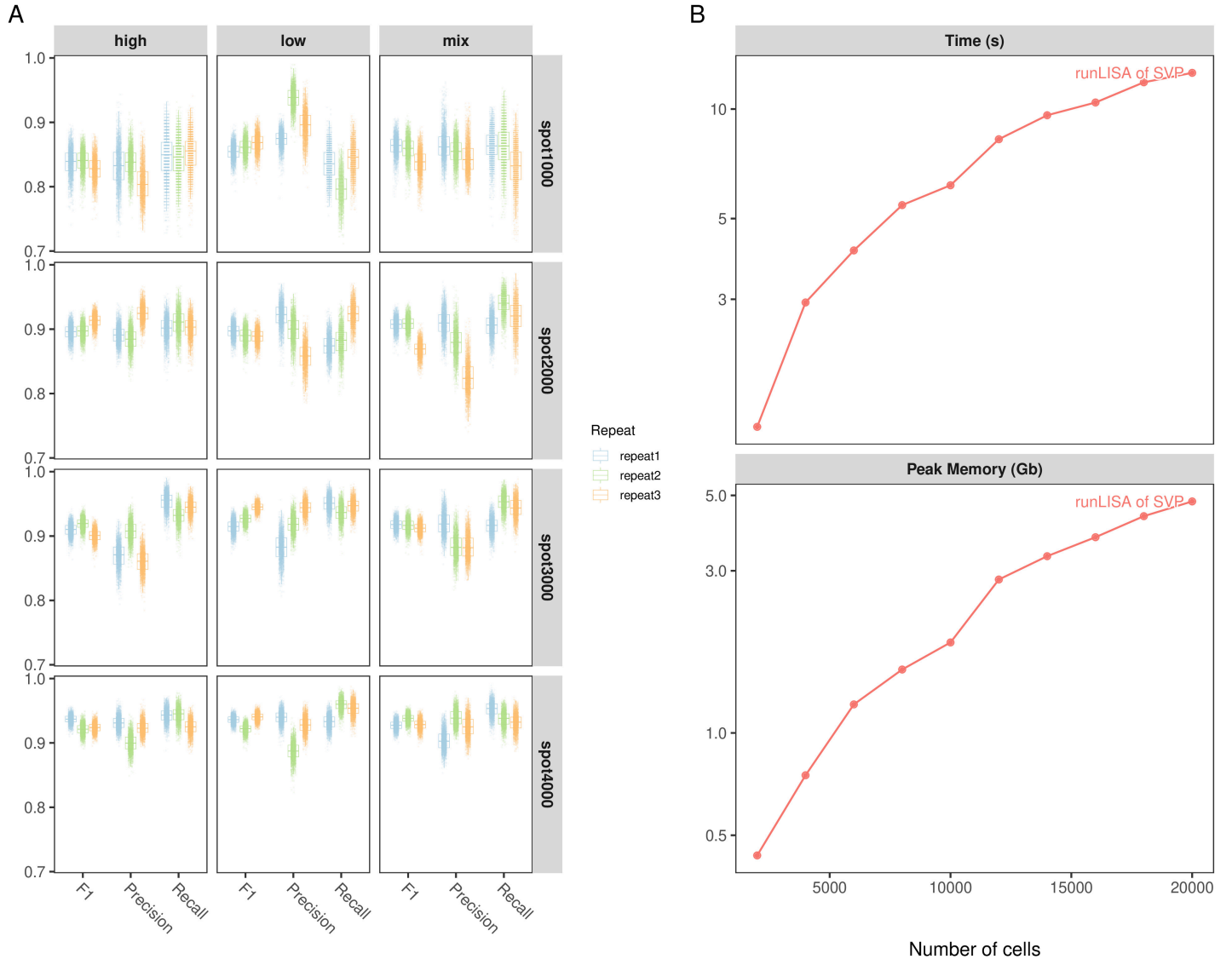
Contents

Supplementary Figures	1
SFig. 1	1
SFig. 2	2
SFig. 3	3
SFig. 4	7
SFig. 5	7
SFig. 6	7
SFig. 7	7
SFig. 8	8
SFig. 9	9
SFig. 10	10
SFig. 11	11
SFig. 12	12
SFig. 13	13
SFig. 14	14
SFig. 15	15
SFig. 16	16
SFig. 17	17
SFig. 18	18
SFig. 19	19
SFig. 20	21
SFig. 21	21
SFig. 22	22
SFig. 23	23
SFig. 24	24
SFig. 25	25

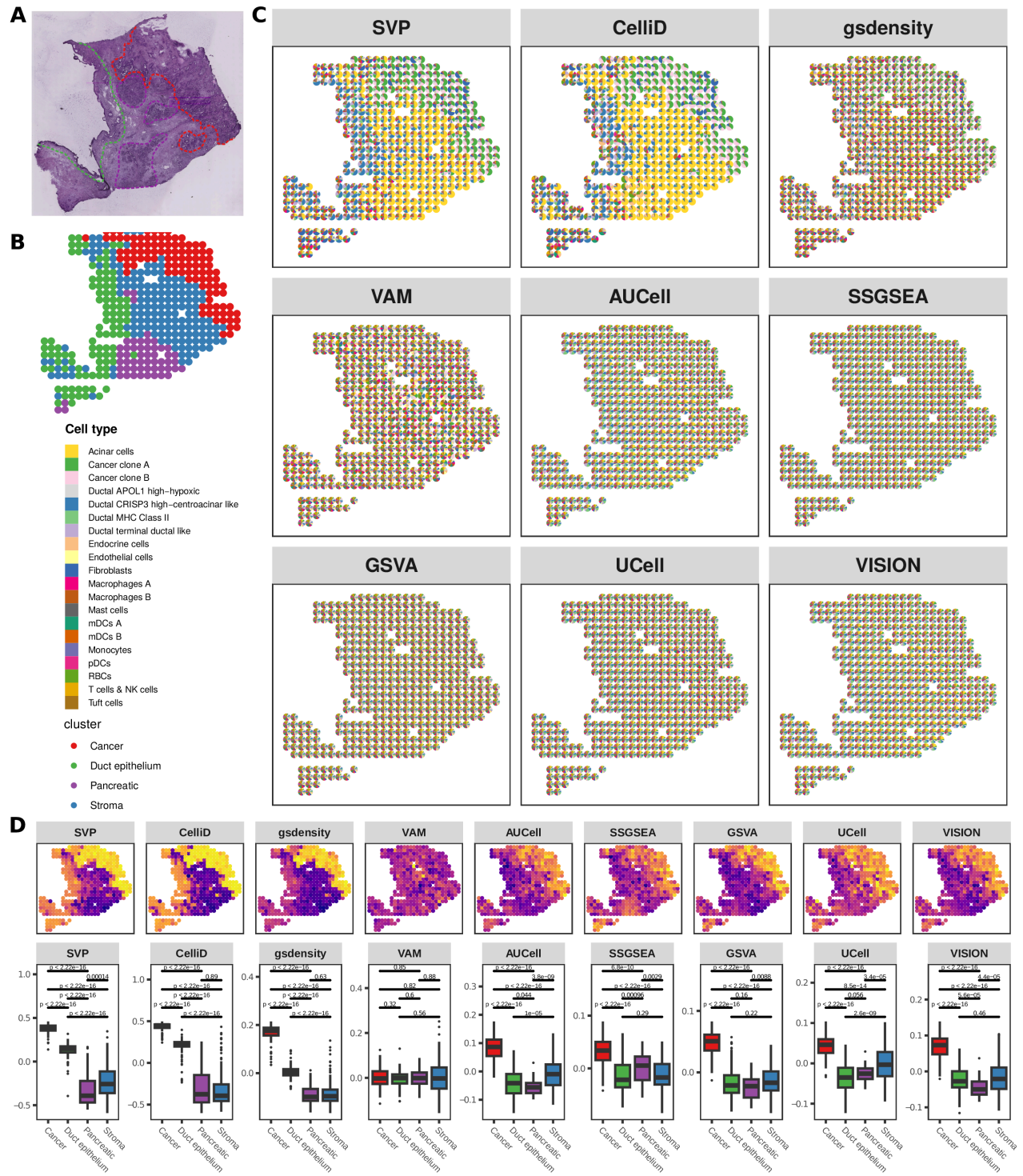
Supplementary Figures



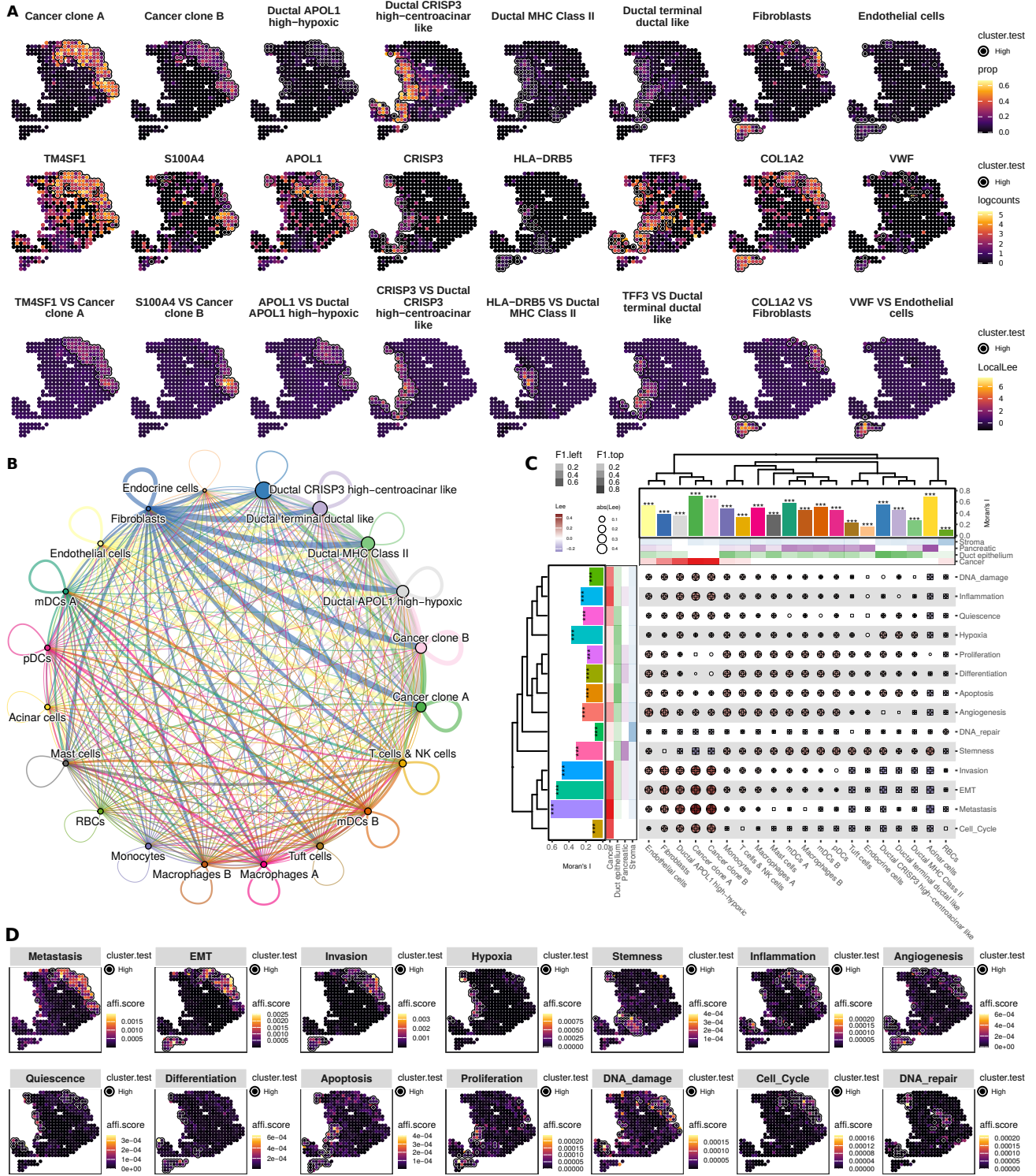
SFig. 1: Comparative results of cell type predictions on mouse somatosensory cortex simulated data using different functional evaluation software. This simulation was constructed based on seqFISH+ sequencing data of the mouse somatosensory cortex, merging cell counts within small square regions measuring around 51um per side to obtain 72 spot microregions. Genes were randomly selected in three groups (3000, 6000, and 10000 genes), resulting in three sets of simulated spatial transcriptome data with 72 spots but varying gene numbers. (A), (B), and (C) show pie charts of actual cell types and those predicted by different software for data sets with different gene counts, respectively. (D), (E), and (F) represent evaluation results of predicted versus actual cell abundance across different gene numbers. The dot plot below shows comparisons at various cell type levels to actual data, while the bar chart above represents comparisons across all cell types. A higher 1-JSD indicates a more consistent distribution of predicted and actual cell abundance, a higher 1-RMSE indicates closer abundance levels, and a higher PCC signifies a stronger correlation between predicted and actual cell abundance, indicating greater prediction accuracy.



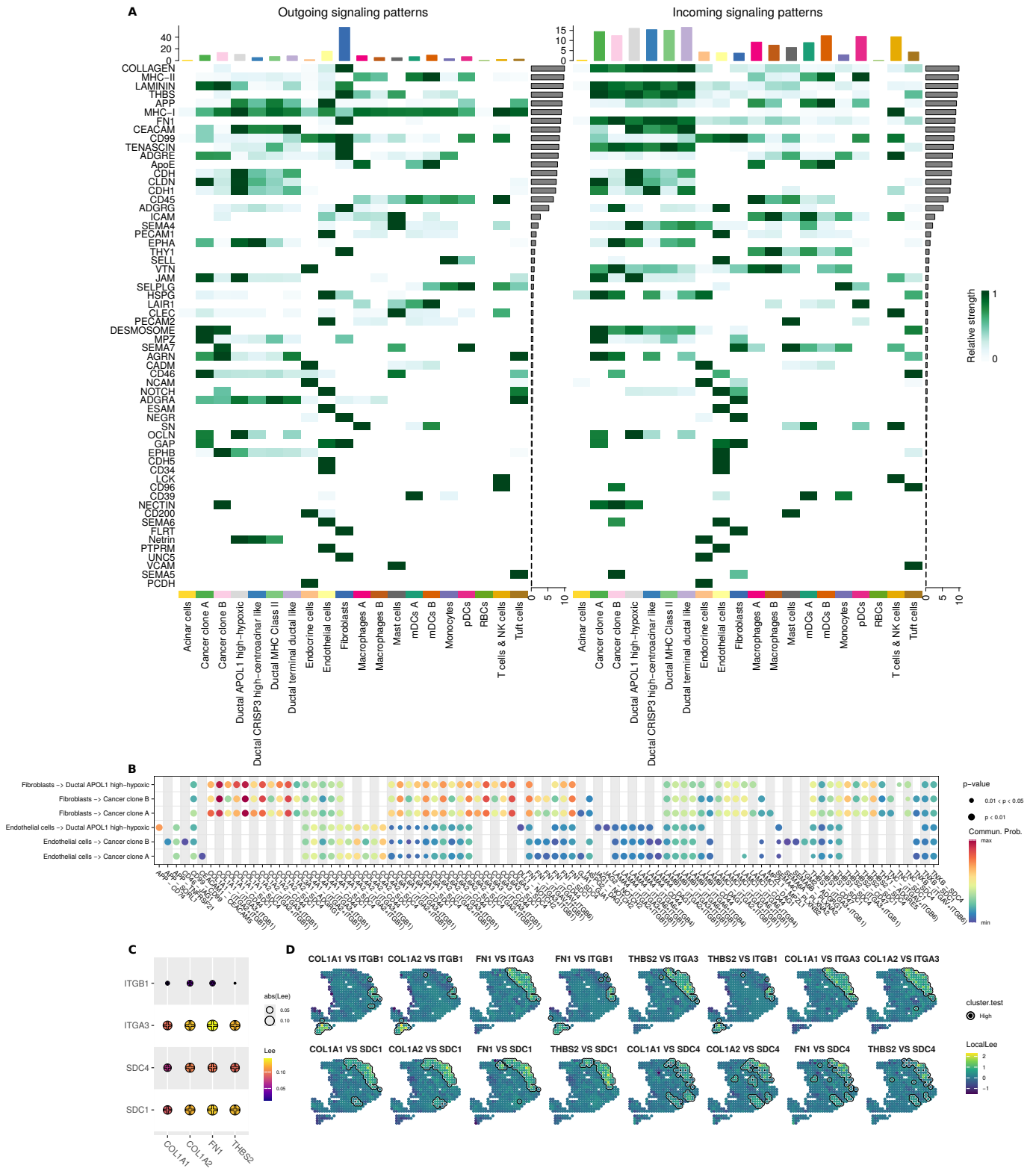
SFig. 2: SVP's prediction results for local spatial clustering regions of SVGs across three type simulated datasets. (A) The x-axis represents different evaluation metrics, with different colors showing repeat samples under the same spot. Panels indicate data with varying spot numbers. The findings indicate that SVP provides good prediction of local spatial clustering regions, with an average F1 score typically above 0.85. (B) The time and memory consumption for local autocorrelation analysis of 1,000 variable features across varying cell counts.



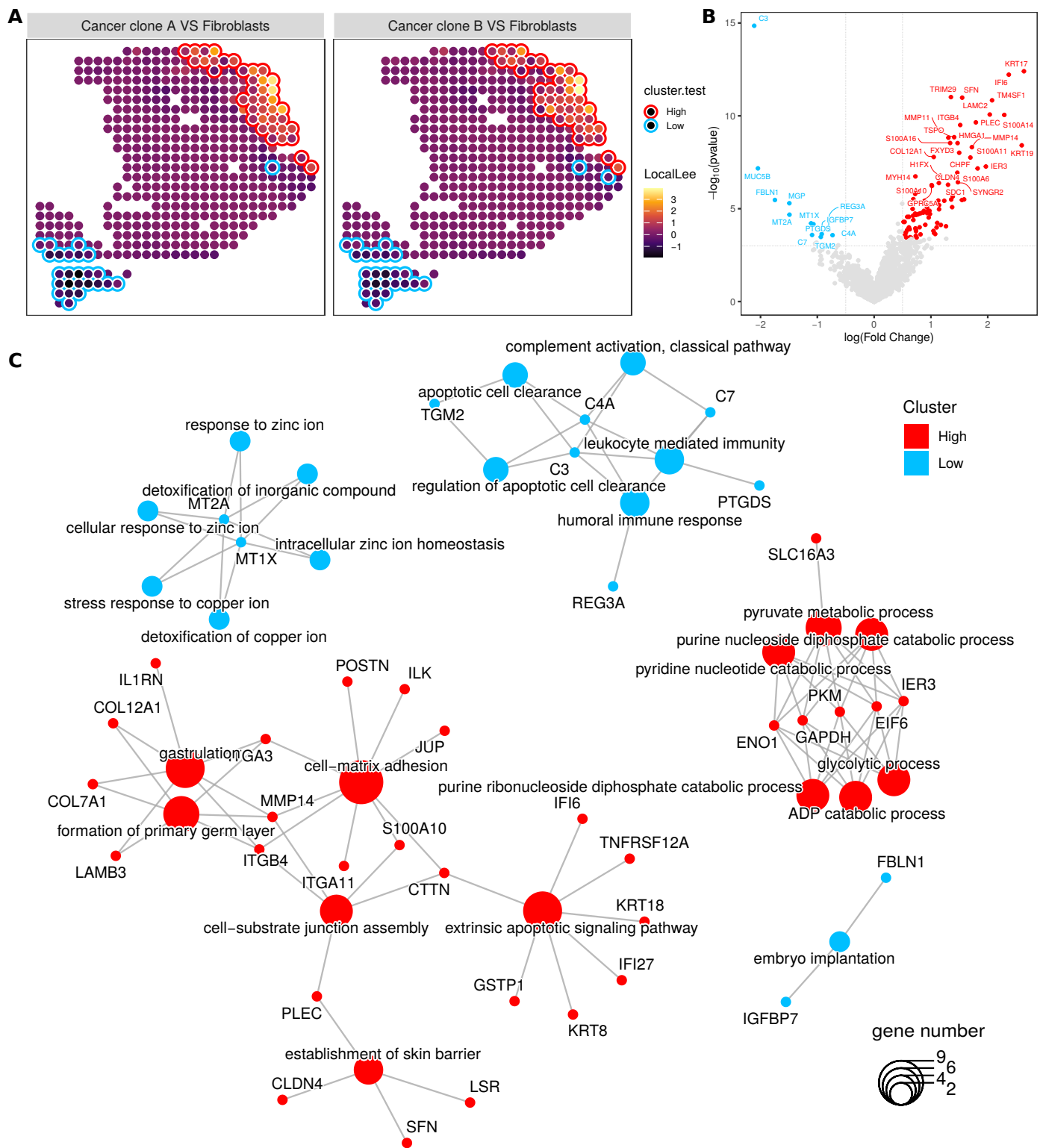
SFig. 3: Comparison of cell type activity predictions in the pancreatic ductal carcinoma microenvironment (sample A) using different functional activity evaluation software. (A) The published HE-stained sections of pancreatic ductal carcinoma with annotations of tumor, ductal, acini, and stroma regions. **(B)** The clustering results of spatial transcriptomic data. **(C)** The pie charts of cell types in the pancreatic ductal carcinoma tumor microenvironment predicted by SVP and different evaluation software. **(D)** The spatial distribution of the first principal component from PCA analysis of cell type activity result predicted by different software, with the bottom chart comparing the first principal component across different annotated regions.



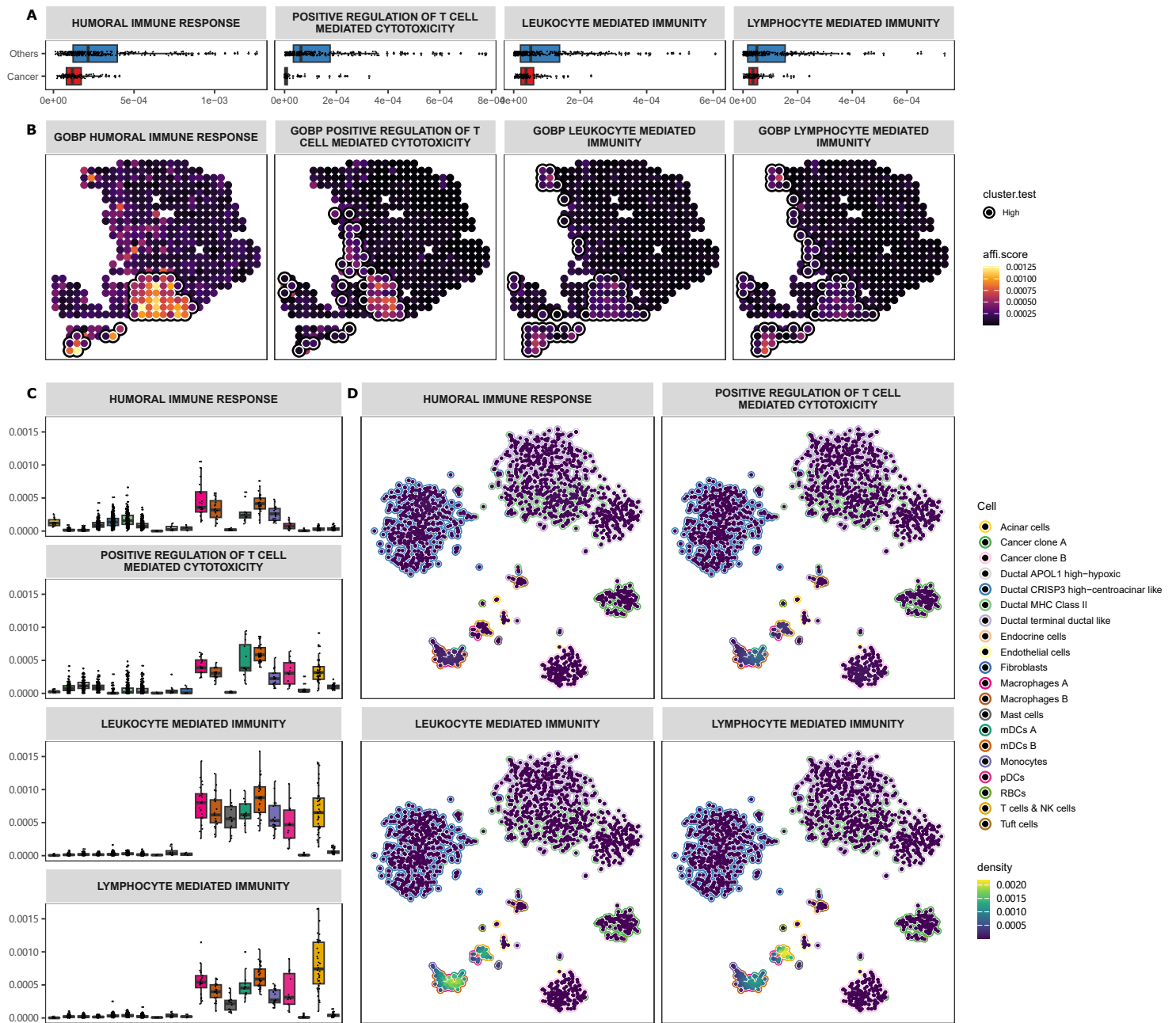
SFig. 5: SVP accurately predicts the co-localization of major cell types and their marker genes within the PADC, as well as their functional states. (A) SVP predicts cell activity for Tumor Cell A, Tumor Cell B, four types of ductal cells (hypoxic ductal cells, CRISP3 high-expressing pancreatic ductal cells, MHC pancreatic ductal cells, and terminal pancreatic ductal cells), fibroblasts, and endothelial cells, along with the expression of their marker genes and significant spatial co-clustering of cell activity and marker genes. The highlighted areas are regions of high clustering identified by local spatial autocorrelation of respective features. **(B)** The CellChat analysis of the number of communications between cell types in single-cell transcriptomics, with points representing different cell types, and line width indicating communication strength between the cells. **(C)** The spatial clustering analysis of predicted cell type activity and CancerSEA tumor cell functions. The dot plot indicates the co-distribution of cells and functions in space. The ribbon heatmaps on the left and top show clustering of functions and cells in tumor, ductal, acini, and stroma regions. Bar charts on the left and top represent spatial clustering of functions and cells, respectively. For detailed explanation, refer to Fig3 (F). **(D)** The spatial distribution of CancerSEA tumor cell functions predicted by SVP, with highlighted areas indicating high clustering of the feature.



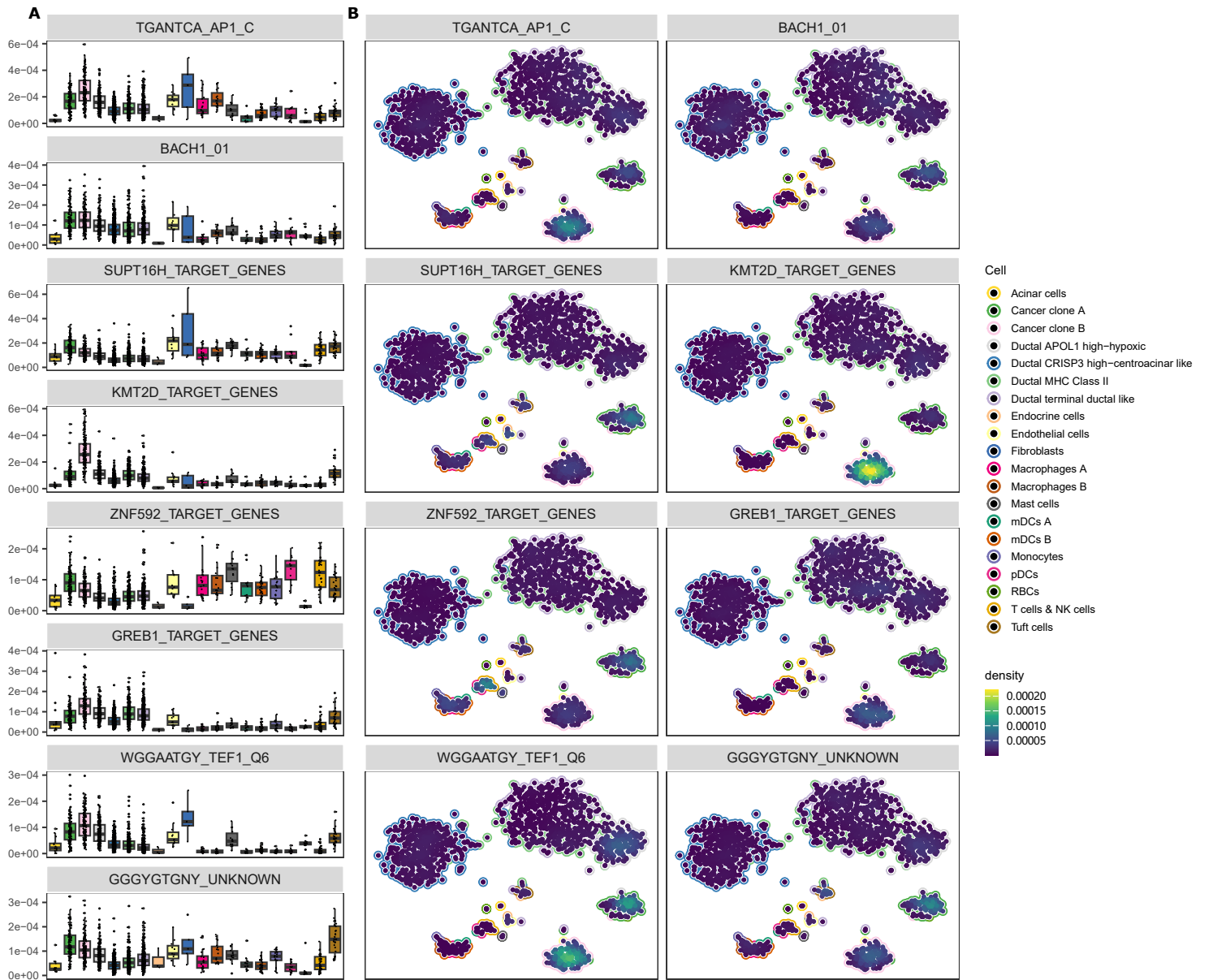
SFig. 6: Single-cell transcriptome cell communication analysis results for pancreatic ductal carcinoma (sample A). (A) The communication intensity between various cell types as predicted by single-cell transcriptomics, with ligand source cells on the left and receptor source cells on the right. The x-axis represents cell types, and the y-axis represents different ligand-receptor pairs. (B) The main significant ligand-receptor pairs involving fibroblasts and endothelial cells with hypoxic ductal cells and tumor cells. The x-axis represents ligand-receptor pairs, the y-axis shows the direction of cell communication, and the color and size of the dots indicate the strength and significance level of the communication, respectively. (C) The spatial co-distribution of major ligand-receptors in the spatial transcriptome, where larger dots indicate greater co-clustering. (D) The co-clustering distribution map of major ligand-receptors in the spatial transcriptome obtained through local bivariate spatial autocorrelation analysis, with highlighted areas indicating primary co-clustering regions. Combining this with previous HE staining annotations reveals that the main co-clustering areas of ligand-receptors are primarily in the tumor region.



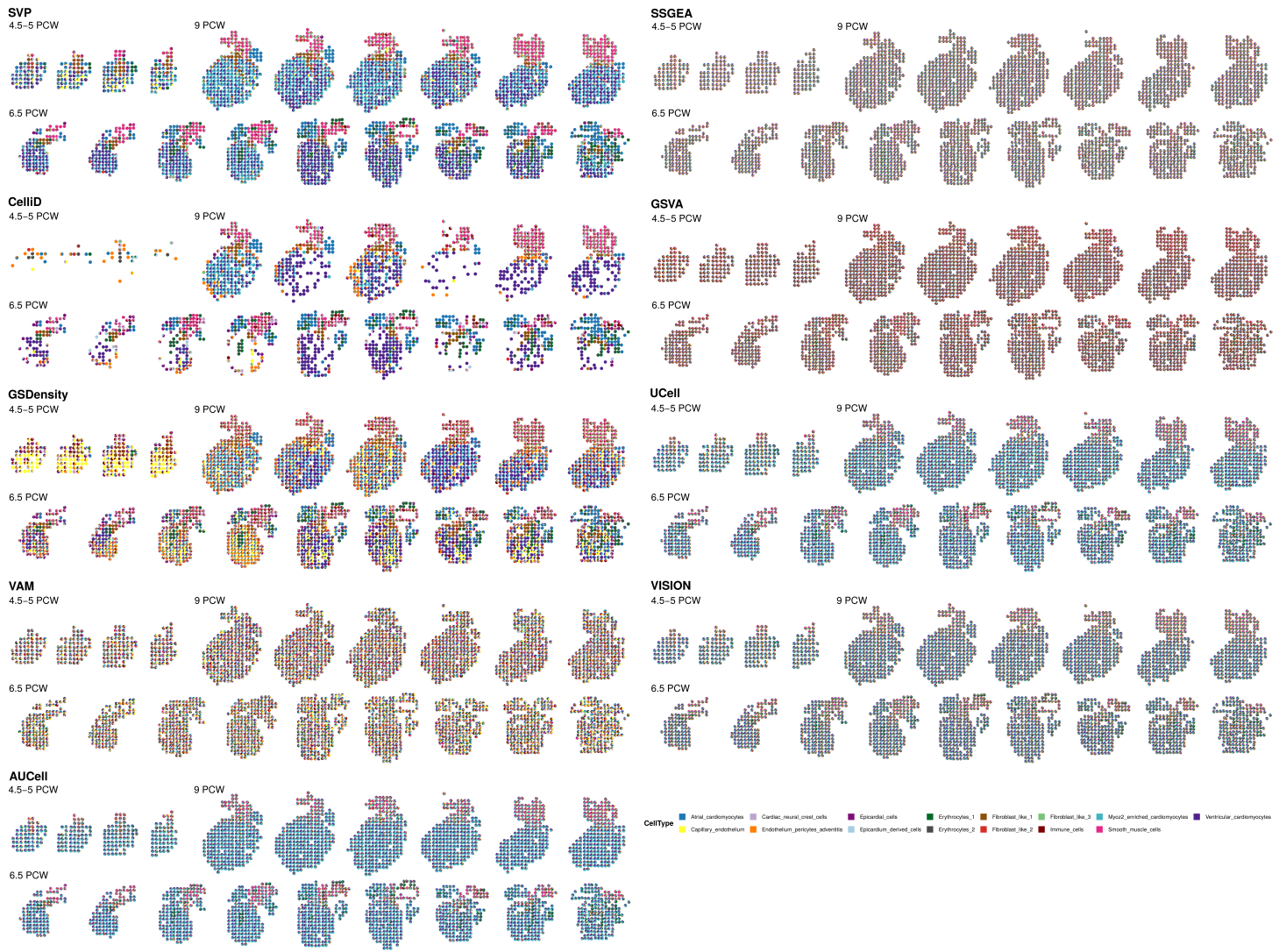
SFig. 7: Spatial distribution of fibroblasts and tumor cells predicted by SVP in PADC sample A, along with differential gene enrichment analysis for high and low aggregation regions. (A) Spatial distribution of fibroblasts and tumor cells predicted by SVP in PADC sample A, the red-highlighted areas represent the aggregation regions (High) of both cell types, while the blue areas denote regions of near mutual exclusivity (Low). **(B)** Volcano plot showing differential genes between the two regions, with red indicating genes upregulated in the high aggregation region and blue representing genes upregulated in the low aggregation region. **(C)** The Gene Ontology (GO) enrichment results for differential genes between the two regions indicate that genes in the red region are primarily enriched in pathways related to fibrosis, cell adhesion, and the extracellular matrix. In contrast, genes in the blue region are mainly enriched in immune-related pathways, such as the complement system.



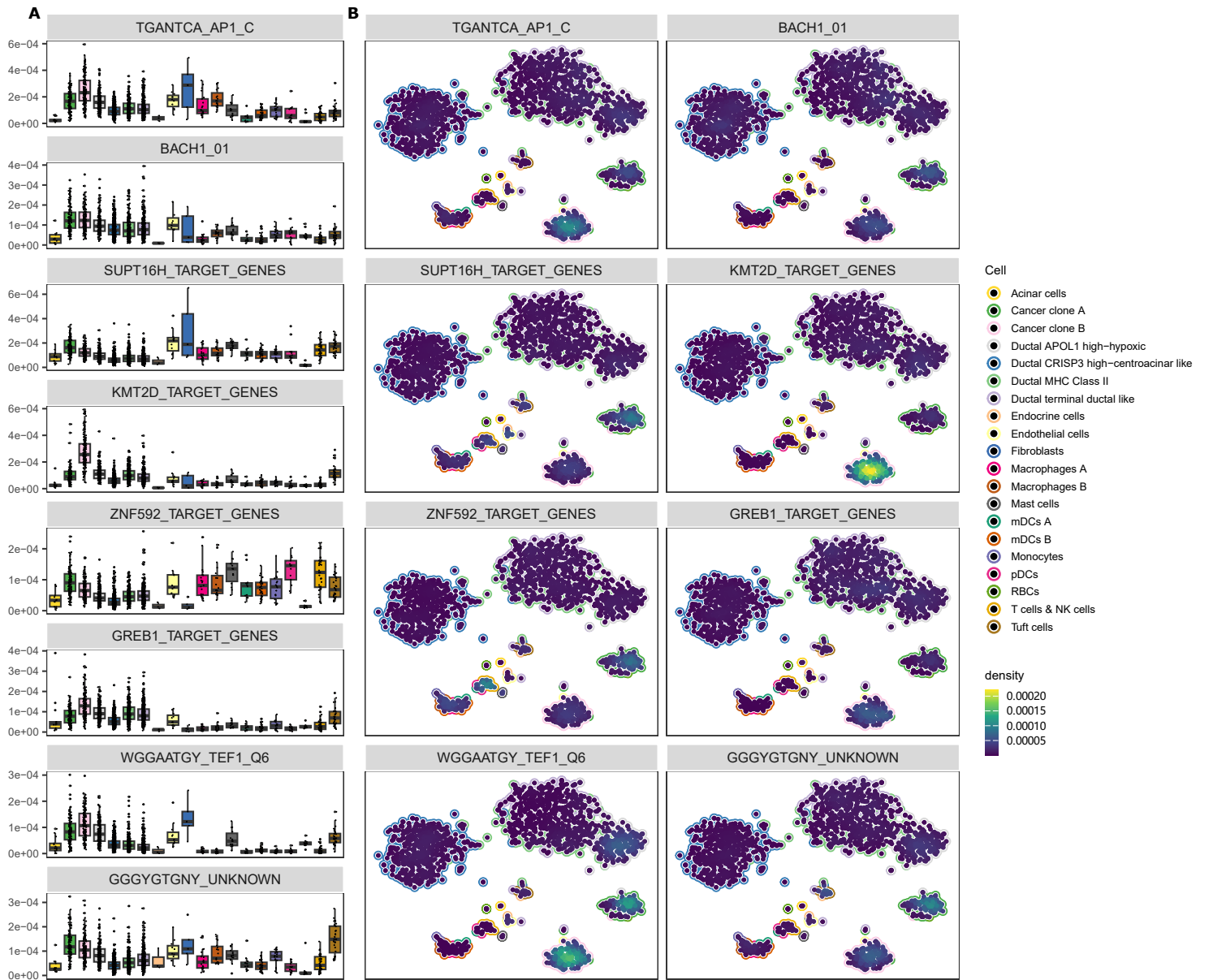
SFig. 8: SVP predicted immune function activity in the spatial and single-cell transcriptomes of pancreatic ductal carcinoma (sample A). (A) Boxplots comparing different immune function activities between tumor and non-tumor regions, showing higher immune function activity in the non-tumor regions. (B) Heatmap showing the spatial distribution of immune function activities. Combining this with HE staining and pathology annotations reveals that immune function activity is mainly concentrated in the lower-left non-tumor area, which includes regions rich in fibroblasts and endothelial cells. (C) and (D) show activity scores of different immune functions across cell types in the single-cell transcriptome data of sample A, indicating that immune function activity is significantly enriched in immune cells, consistent with expectations.



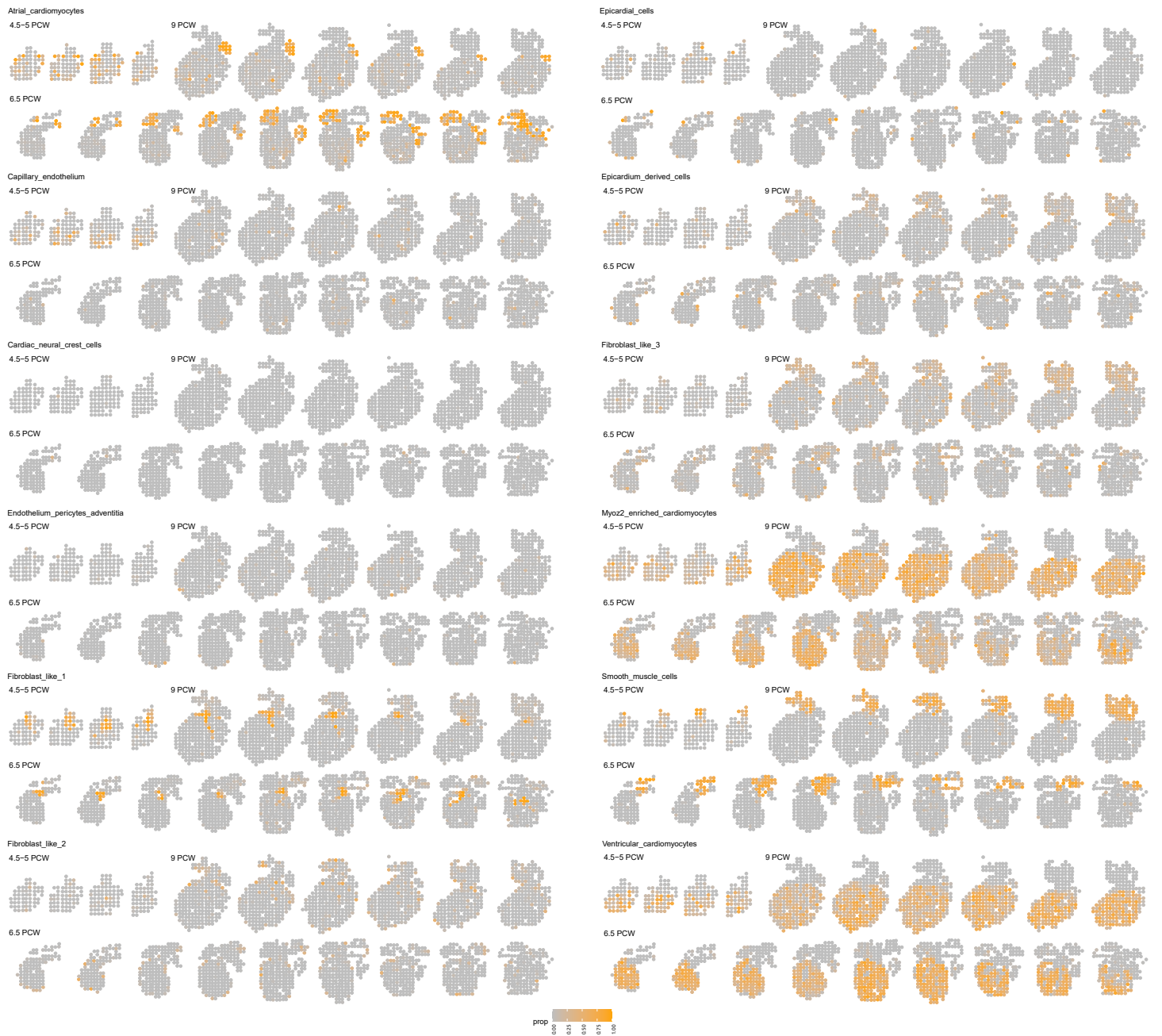
SFig. 9: Results of transcription factor activity related to tumor cells in the single-cell transcriptome of pancreatic ductal carcinoma (sample A), as predicted using SVP. (A) Boxplots showing transcription factor activity related to tumor cells across various cell types. **(B)** Distribution of tumor cell-related transcription factors in the TSNE dimensionality reduction space. The results indicate that tumor-related transcription factor activity is primarily activated in tumor cells, fibroblasts, and endothelial cells.



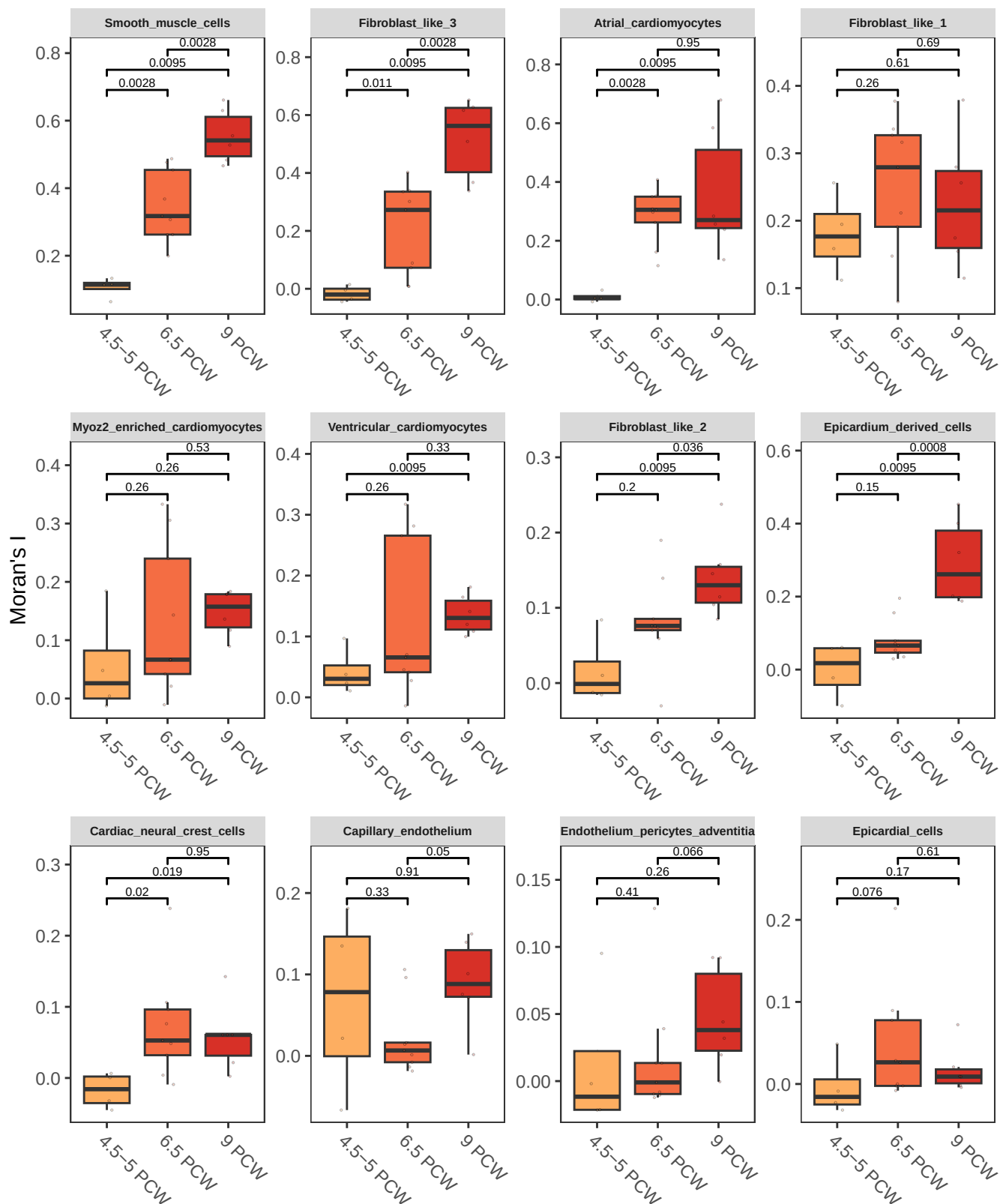
SFig. 10: Prediction results of cell type activity in human heart spatial transcriptome data at different developmental stages using different functional evaluation software. Different colors represent various cell types.



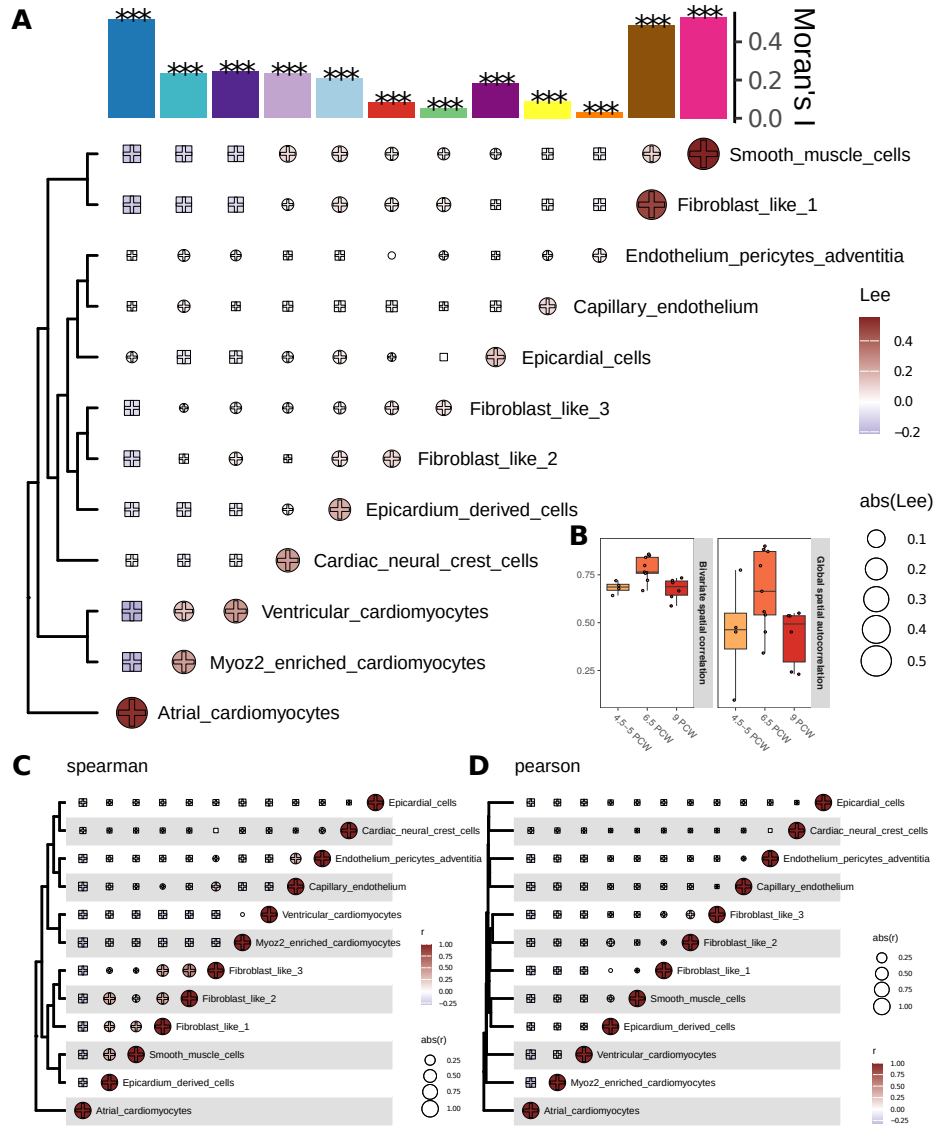
SFig. 11: Spatial distribution map of each heart cell type annotated by ISS at 6.5 PCW from the original data. This can be used for comparison with the cell activity results predicted by the spatial transcriptome. Two types of blood cells and immune cells have been removed.



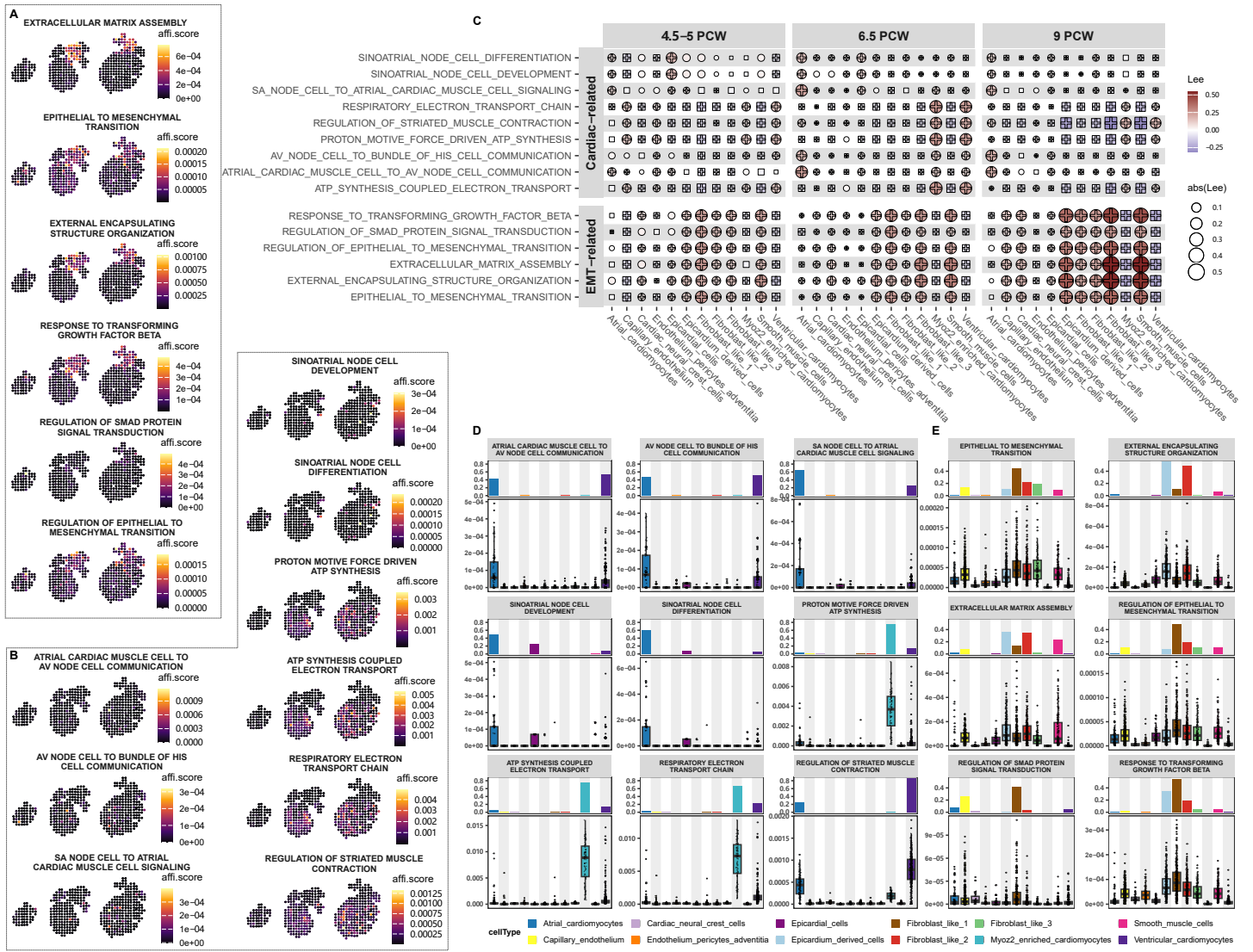
Sfig. 12: Predicted activity abundance of different cell types in the human heart at various developmental stages using SVP. The color of the dots represents cell activity abundance.



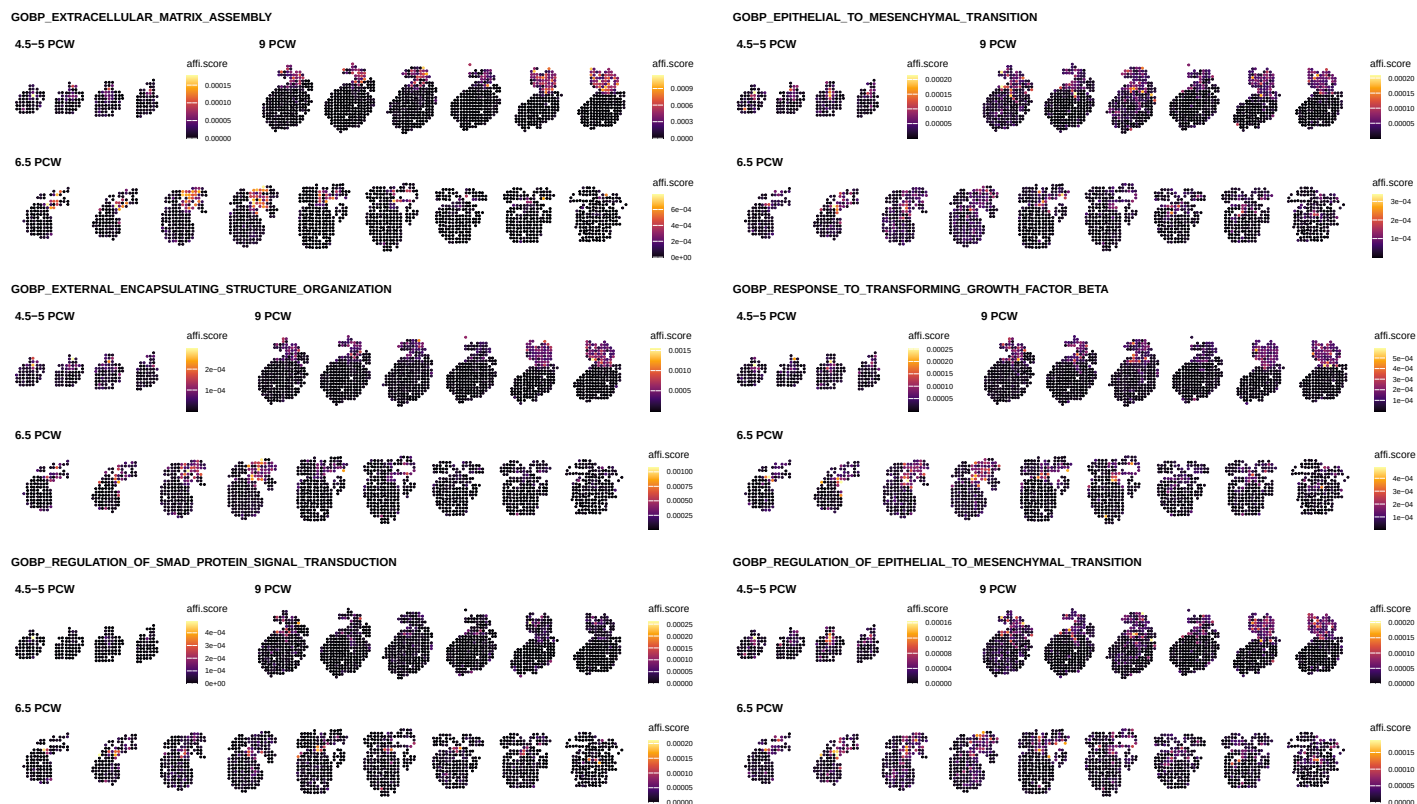
SFig. 13: Moran's I index for different cell types in the human heart at various developmental stages, as calculated by SVP. The x-axis represents different developmental stages, and the y-axis represents the Moran's I index. The closer this index is to 1, the more clustered that cell type is in space. Statistical analysis is based on the wilcoxon rank-sum test results.



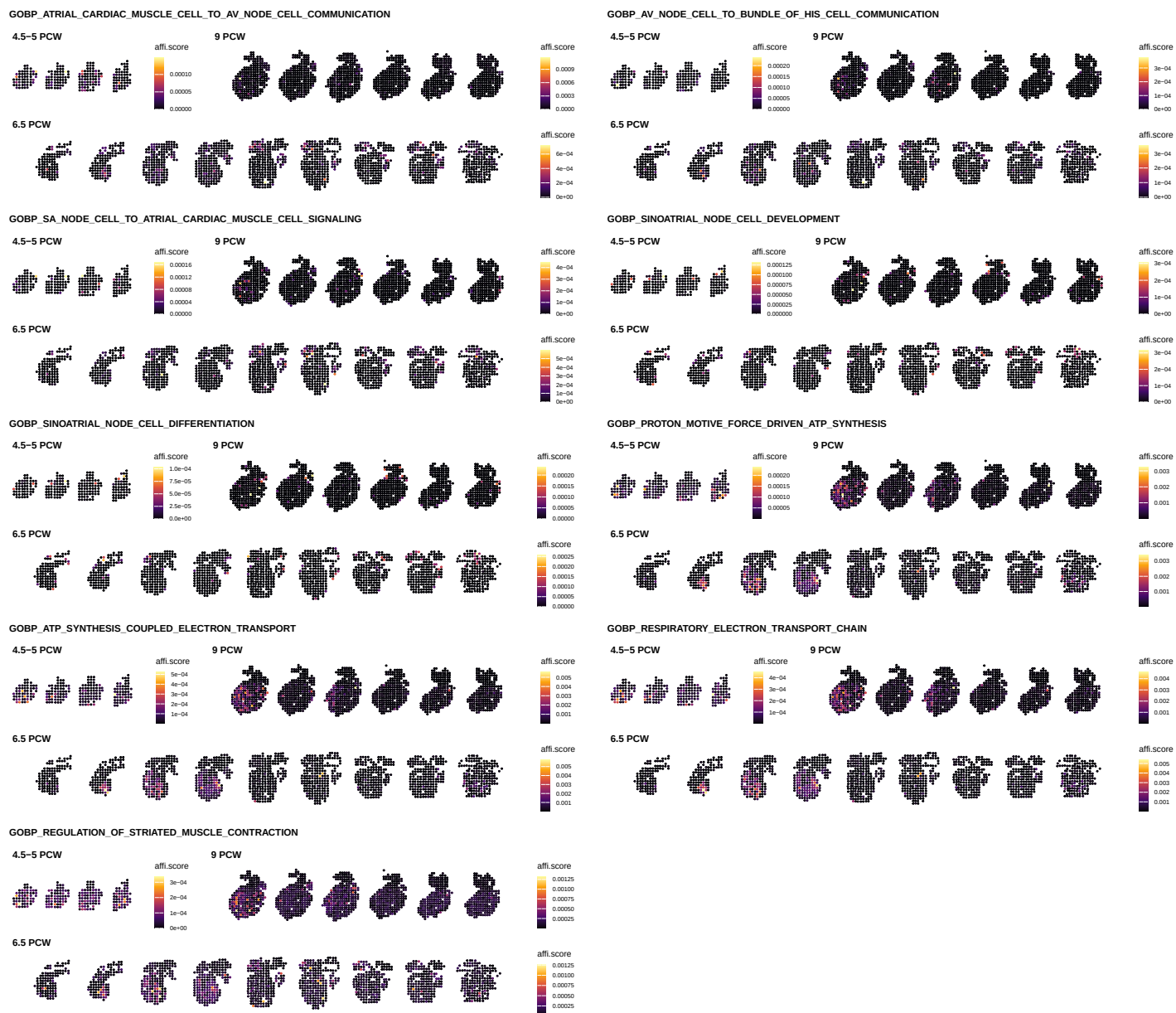
SFig. 14: Spatial co-distribution of different human heart cells at 6.5 PCW obtained by ISS sequencing and comparison with spatial co-distribution results of different heart cells obtained from spatial transcriptomics. (A) The spatial co-distribution of different human heart cells at 6.5 PCW as obtained by ISS sequencing. The dot plot represents the degree of spatial co-distribution between cells, with an index ranging from -1 to 1. Larger positive values (bigger dots) indicate a stronger spatial co-clustering of two features, values approaching zero suggest no spatial autocorrelation, while larger negative values (bigger square dots) suggest mutual spatial exclusion between features. (B) The correlation between the spatial distribution of human heart cells at 6.5 PCW obtained by ISS sequencing and spatial distribution across spatial transcriptome samples at different developmental stages. (C) and (D) show correlation maps of different human heart cells at 6.5 PCW using traditional correlation analysis methods. The results indicate that spatial bivariate correlation analysis through SVP reveals clear spatial co-clustering between MYOZ2 myocardial cells and ventricular cells, as well as between atrial cells and epicardial cells, which traditional correlation methods fail to detect.



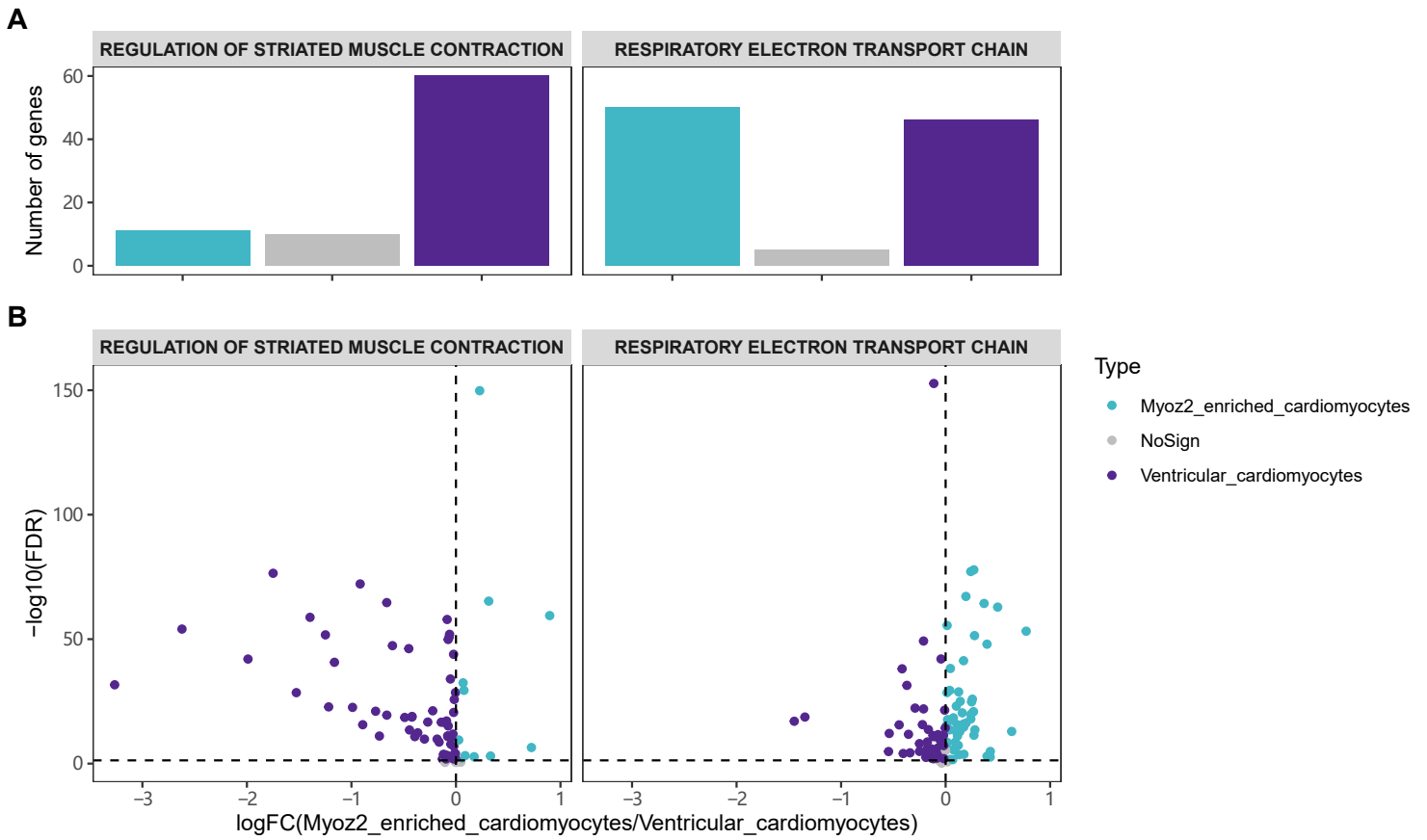
SFig. 15: SVP accurately predicts the dynamic changes in cell functions during human heart development. (A) The spatial distribution of extracellular matrix-related functional activity at different developmental stages of the human heart as predicted by SVP. **(B)** The spatial distribution of SVP-predicted functional activities related to cardiac muscle cells, such as those associated with the sinoatrial and atrioventricular nodes, across different developmental stages. **(C)** The heatmap of the spatial co-distribution of different heart cells with extracellular matrix and cardiac muscle cell-related functions at various developmental stages as predicted by SVP. **(D)** The boxplot comparing the activity of cardiac muscle cell-related functions in different cell types within the 6.5 PCW human heart single-cell transcriptome as predicted by SVP, the bar chart is the F1 value to measure the feature enrichment in a cell cluster or spatial domain after local spatial autocorrelation analysis (see methods). **(E)** The boxplot comparing the activity of extracellular matrix-related functions in different cell types within the 6.5 PCW human heart single-cell transcriptome as predicted by SVP, the bar chart is akin to (D).



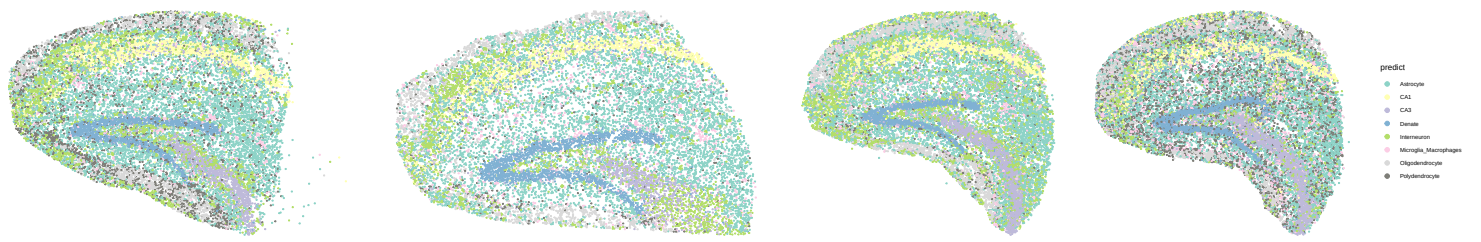
SFig. 16: Spatial distribution map of cell pathways related to extracellular matrix functions in human heart cells at different developmental stages.



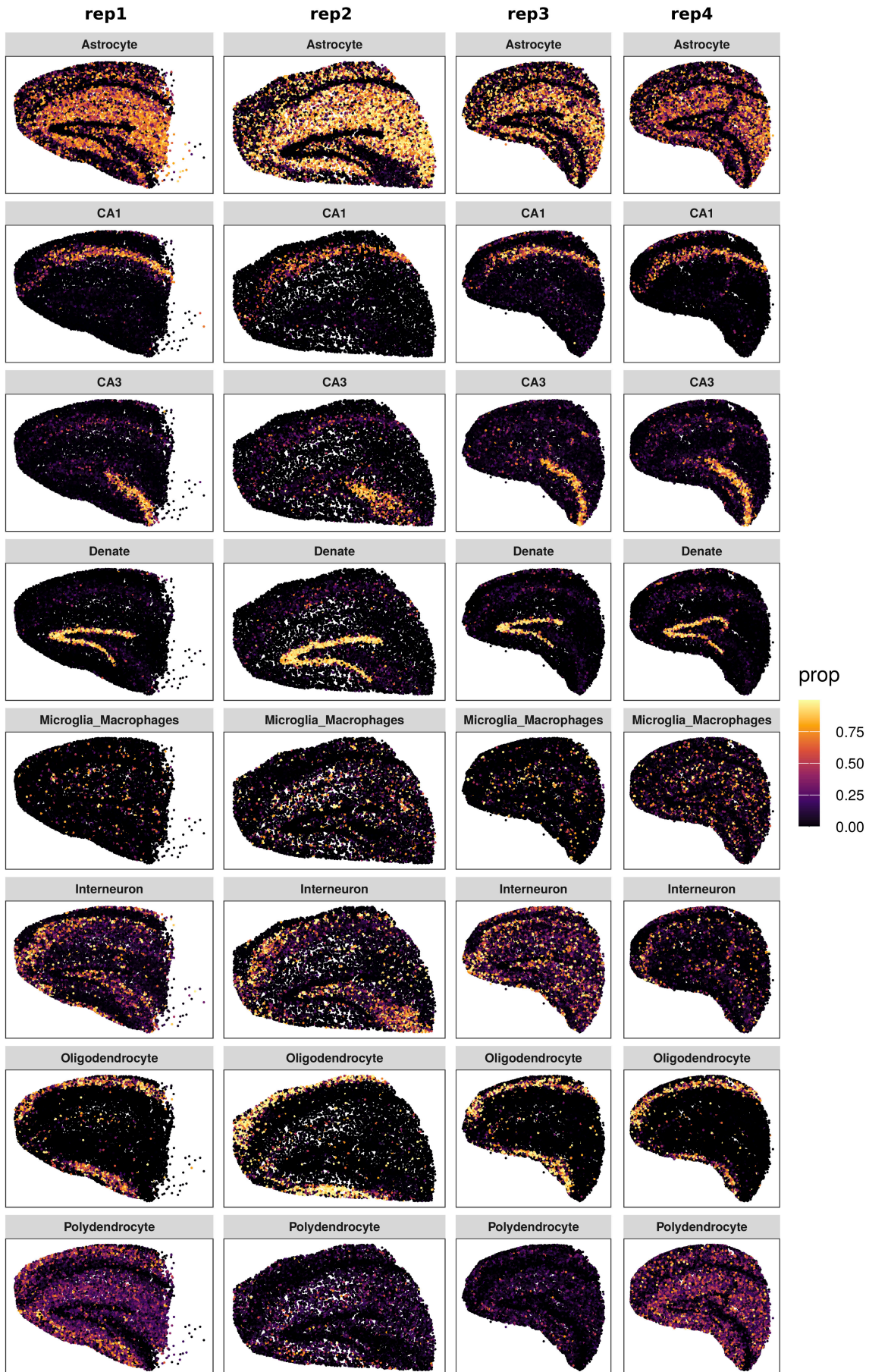
SFig. 17: Spatial distribution map of cell pathways related to cardiac muscle cell functions in human heart cells at different developmental stages.



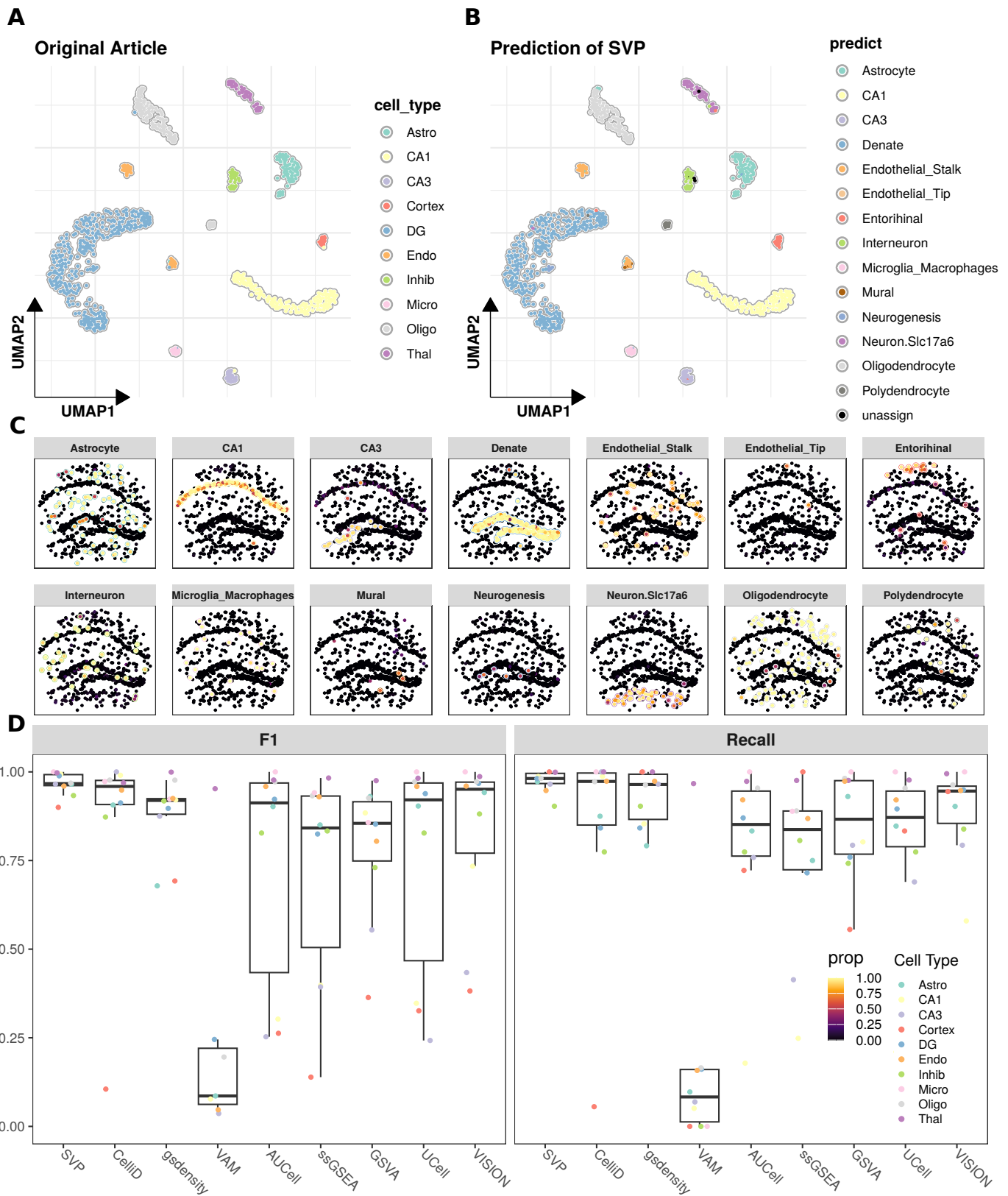
SFig. 18: The comparative gene results between MYOZ2-enriched cardiomyocytes and ventricular cardiomyocytes regarding electron transport and smooth muscle contraction functions are as follows. (A) The number of genes significantly upregulated for these two functions in both cell types. **(B).** The volcano plot of differential genes for these functions in both cell types. The results indicate that ventricular cells have higher expression of genes related to smooth muscle contraction, whereas the MYOZ2-enriched cardiomyocytes cells show higher expression of genes involved in electron transport.



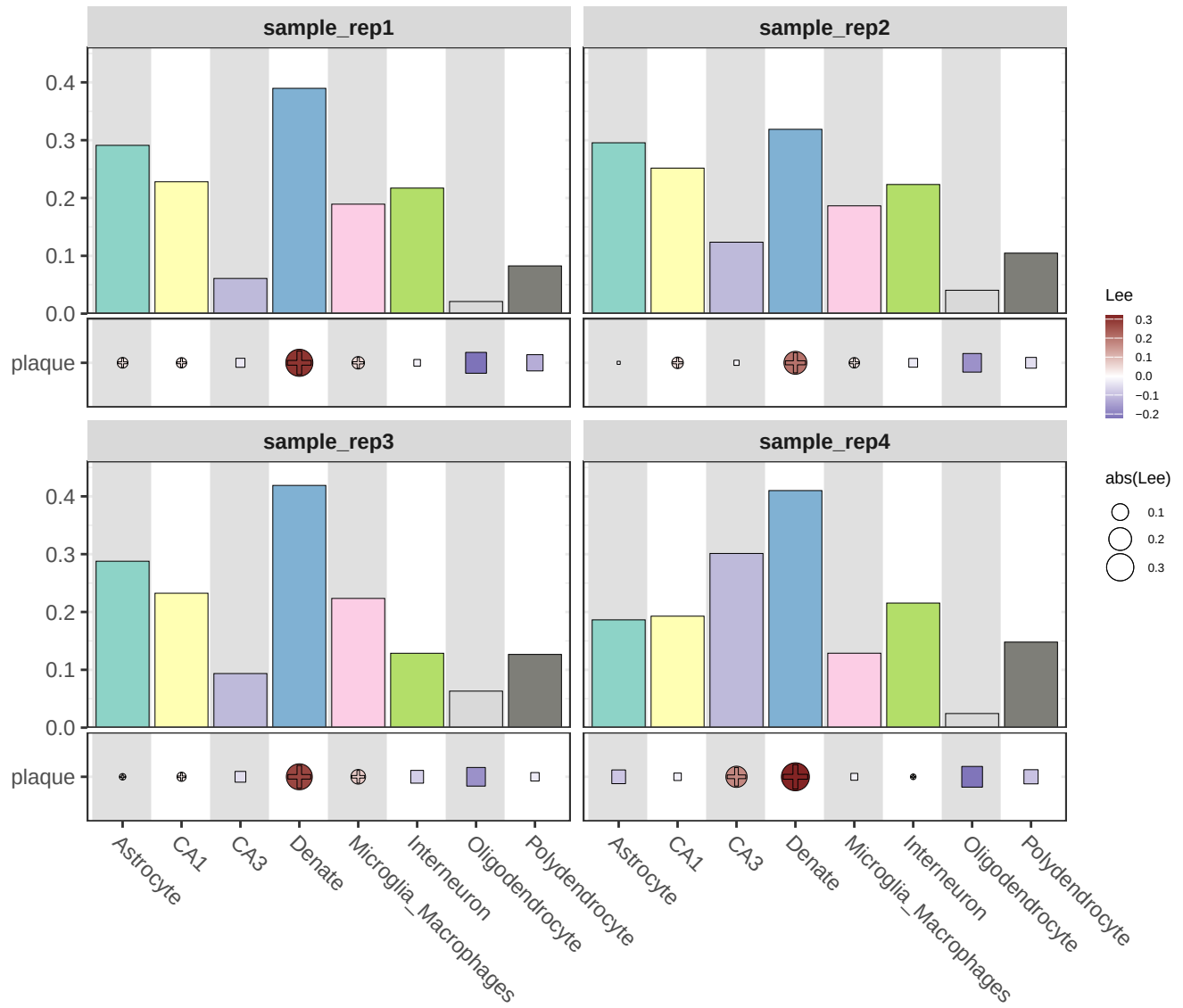
SFig. 19: Cell type annotation of spatial transcriptomic data across all AD model mouse hippocampal slices. The primary cell types predicted by SVP for each cell were utilized as the final annotation results.



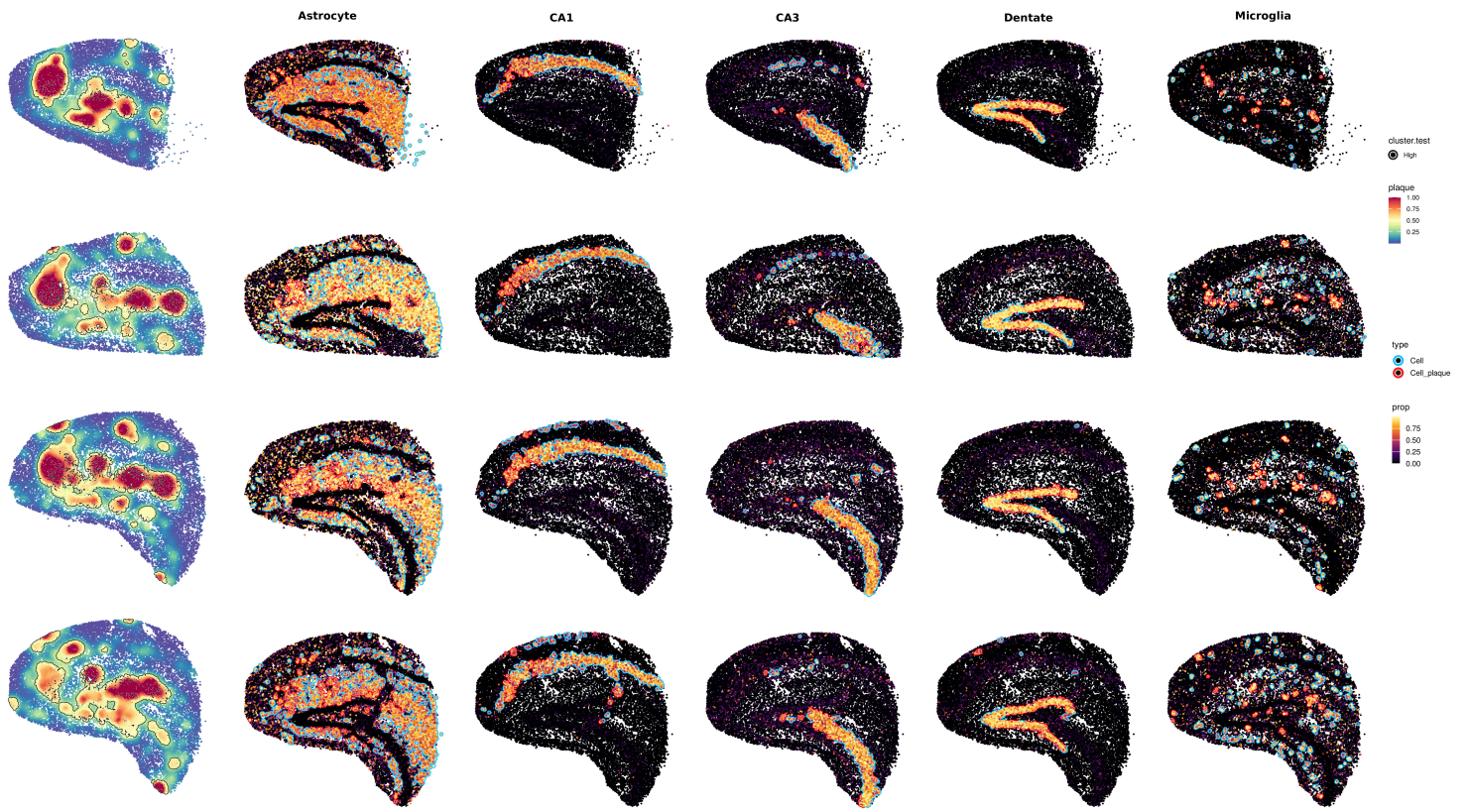
SFig. 20: Spatial distribution maps of each cell type in the spatial transcriptomic data from all AD model mouse hippocampal slices.



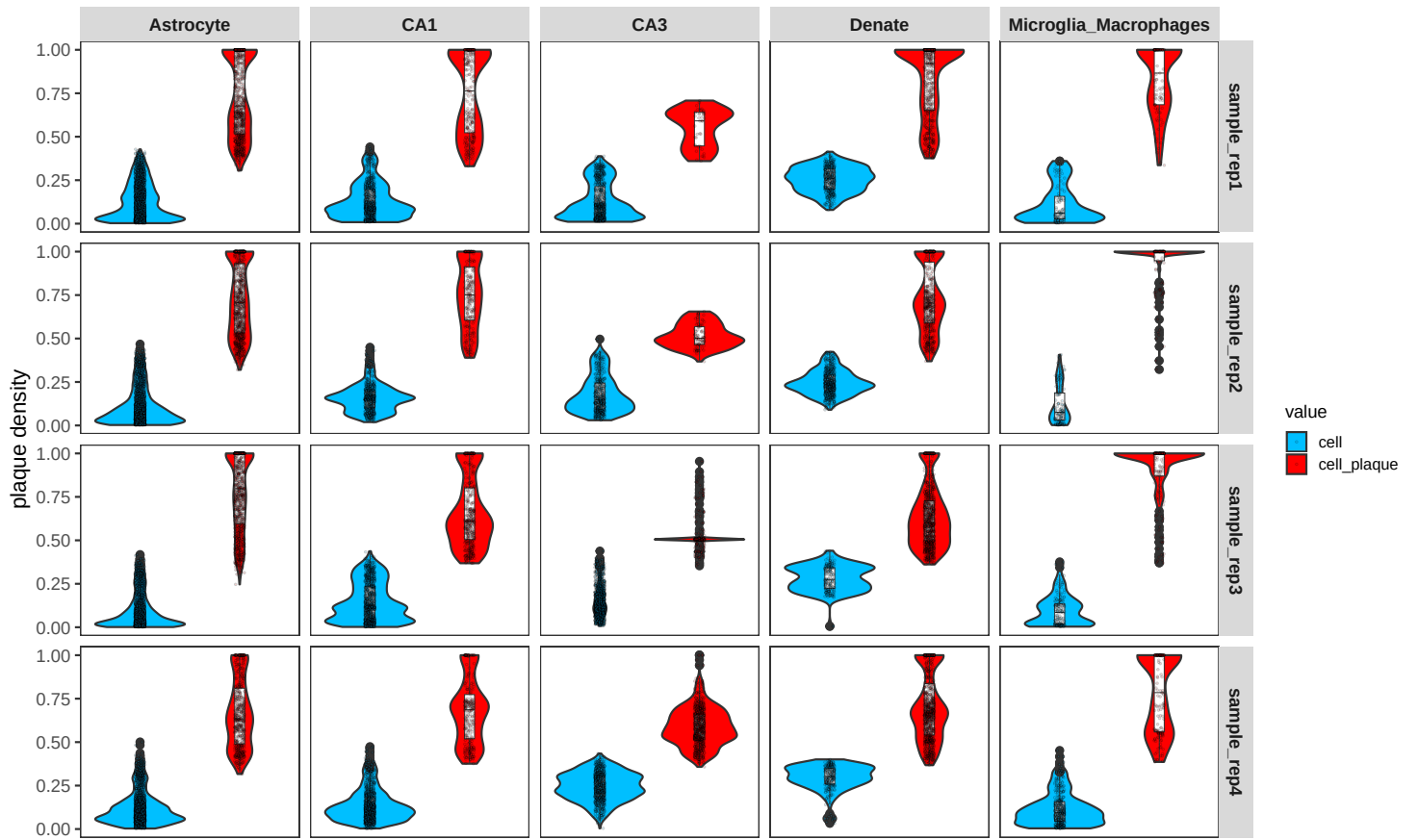
SFig. 21: Cell type predictions using SVP on Slide-Tag spatial transcriptomic data from mouse hippocampus, along with comparative results against other tools. (A) Cell type annotation results from the original publication. **(B)** Cell type annotation results predicted by SVP, which show a high degree of consistency with the original results, while distinguishing Oligodendrocyte into two cell types: Oligodendrocyte and Polydendrocyte. **(C)** Spatial distribution maps of each cell type predicted by SVP, demonstrating accurate spatial placement for each predicted cell type. **(D)** Comparison of prediction results between SVP and other scoring methods.



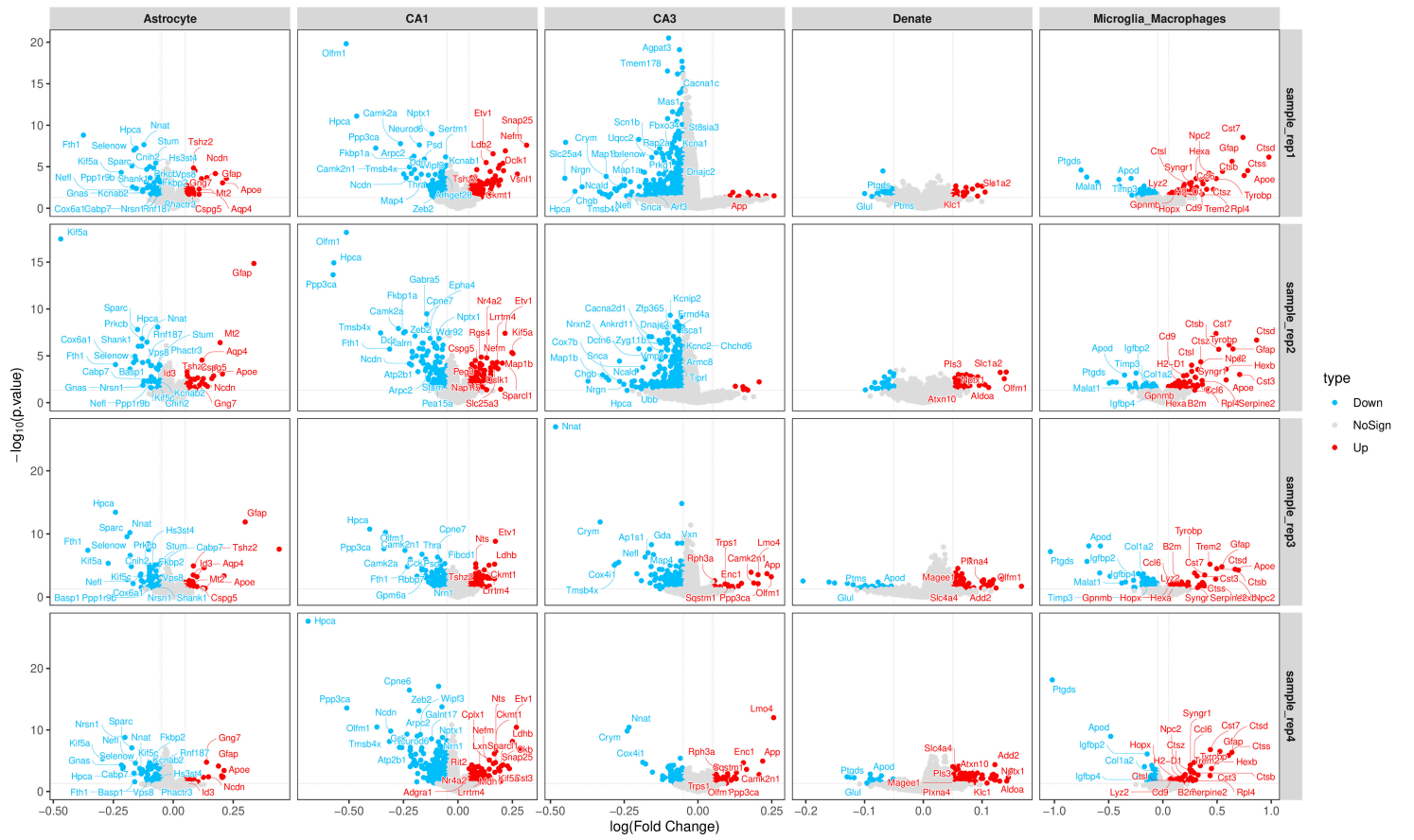
SFig. 22: Spatial aggregation of A beta plaques with cell types predicted by SVP in hippocampal slices from all AD model mice. Heatmap illustrating the spatial co-aggregation of A plaques and various cell types, alongside the enrichment score F1 for each cell type within the A plaque aggregation area, as calculated by the local autocorrelation algorithm (refer to the Methods section)



SFig. 23: The abundance distribution of the main cell types predicted by SVP significantly co-aggregating with A plaques from all AD model mice. red indicating spots where cells and A plaques co-aggregate, and blue indicating spots where only cells aggregate.



SFig. 24: Comparison of A plaque density between spots where cells and A plaques co-aggregate and spots where only cells aggregate from all AD model mice. The x-axis represents the two regions, while the y-axis indicates the density of A beta plaques.



SFig. 25: Differentially expressed genes between spots co-aggregating with A plaques and spots aggregated with each cell type alone from all AD model mice. Red indicates upregulated differential genes in the co-aggregation spots of cells and A plaques, while blue represents upregulated differential genes found exclusively at the cell aggregation spots.