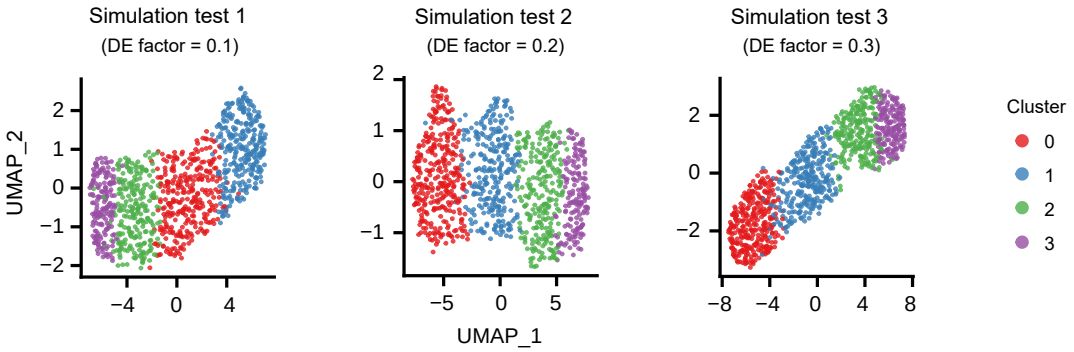
Supplemental Figures

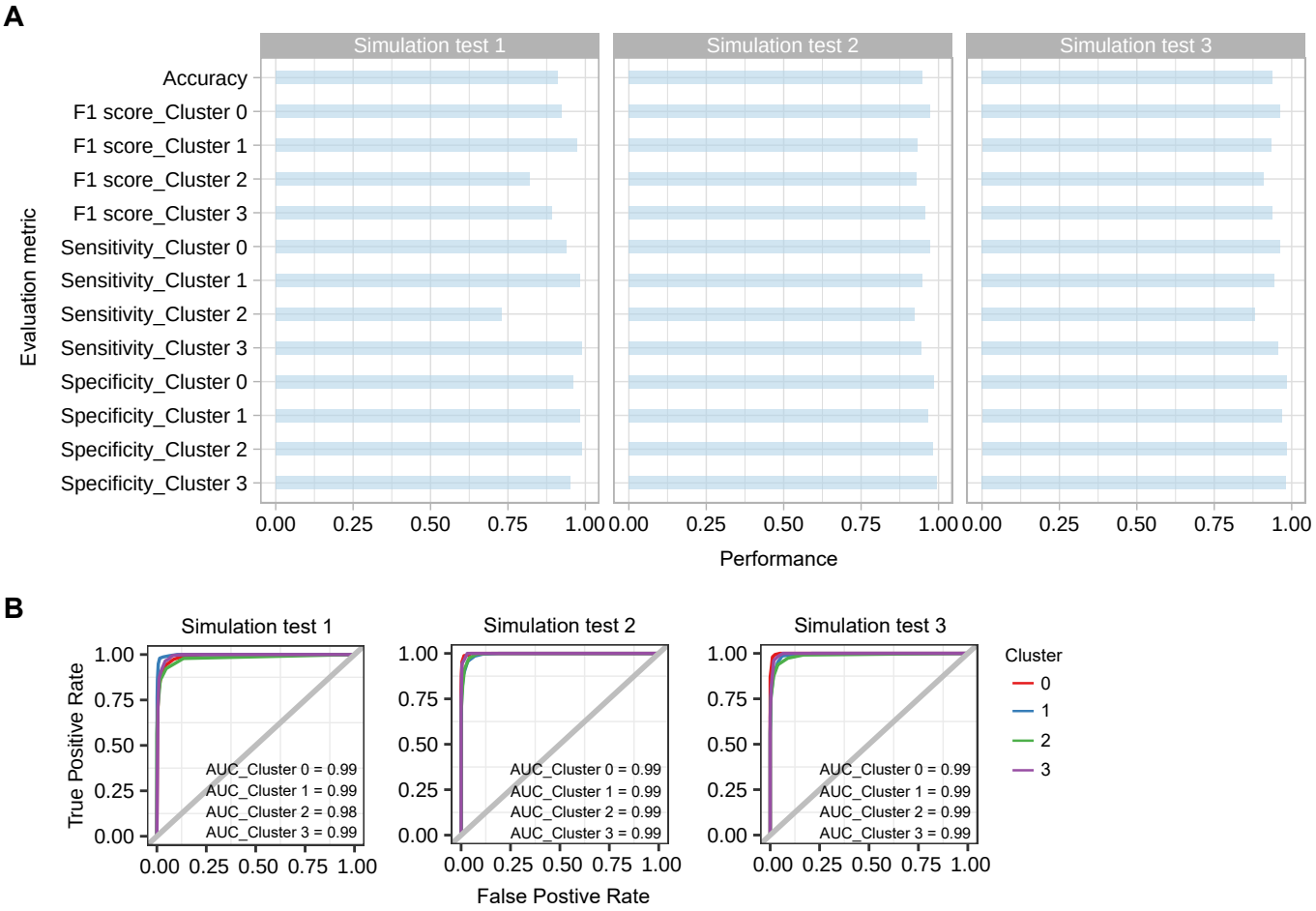
Deciphering context-specific gene programs from single-cell and spatial transcriptomics data with DeCEP

Lin Li[1], Xianbin Su[1], Ze-Guang Han[1, *]

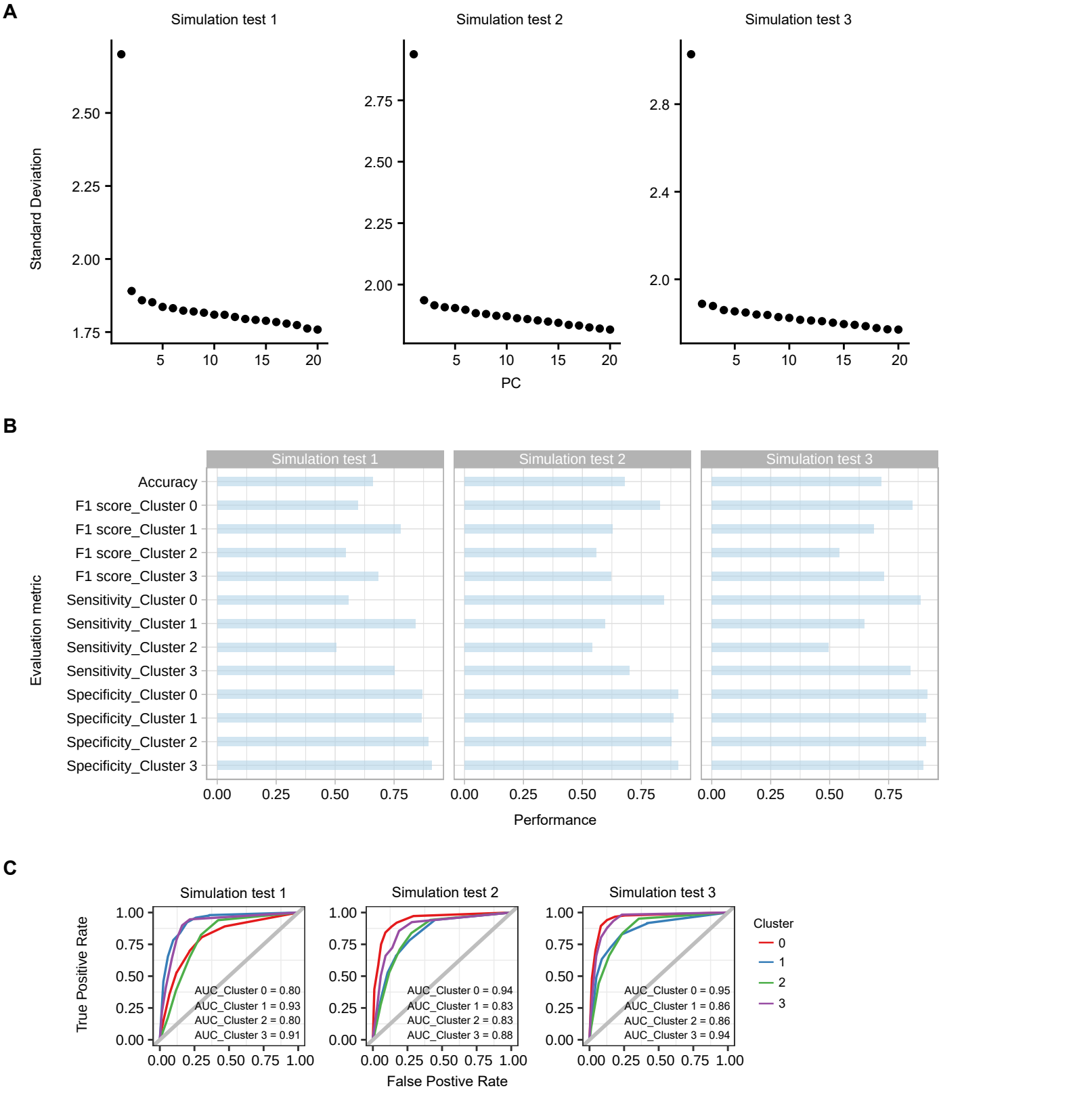
Supplemental Fig. S1-S28



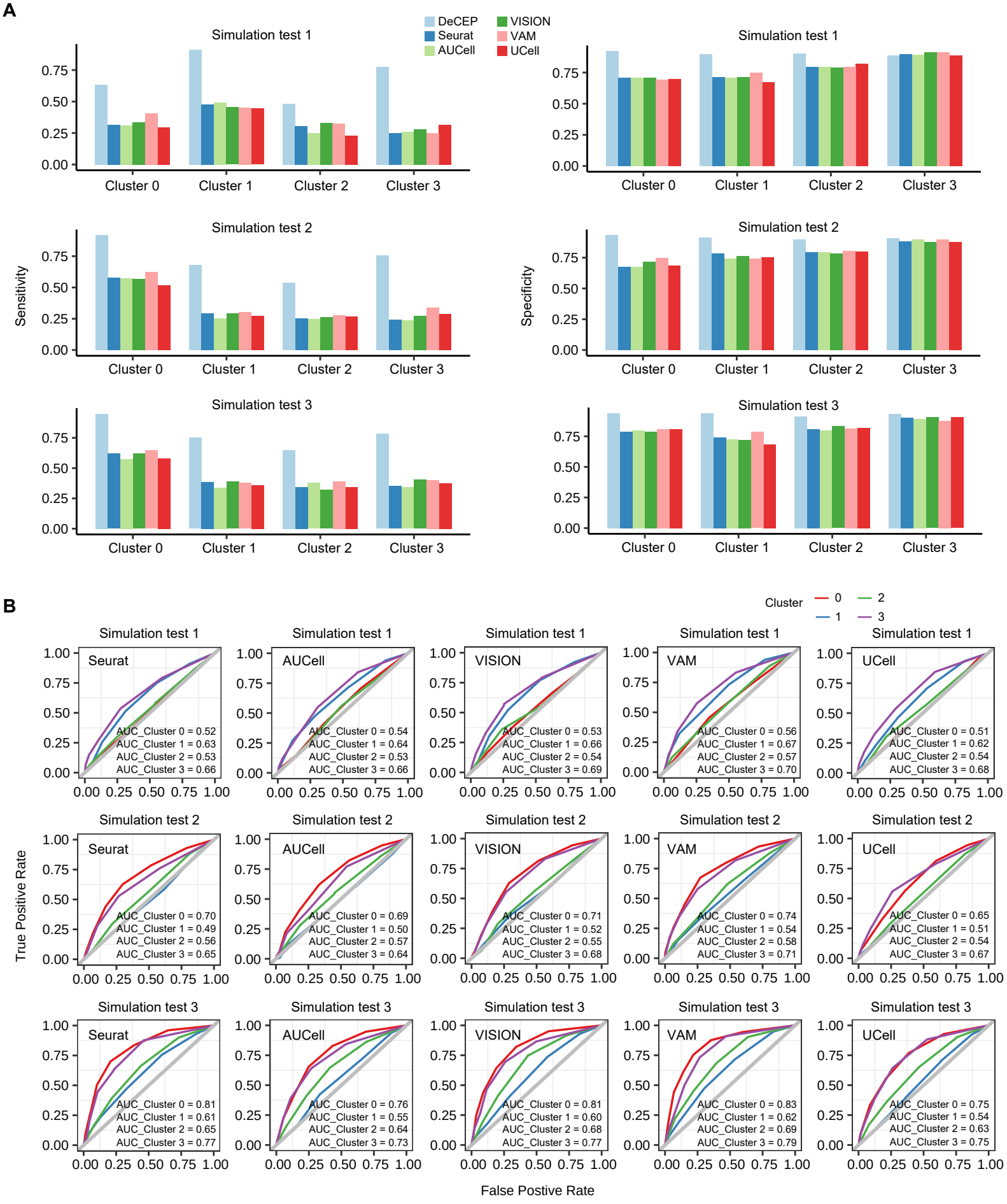
Supplemental Fig. S1. UMAP visualization of simulated cells from simulation tests 1 to 3. From left to right, the differential expression (DE) factors are 0.1, 0.2, and 0.3, respectively, representing increasing gene expression differences between cells.



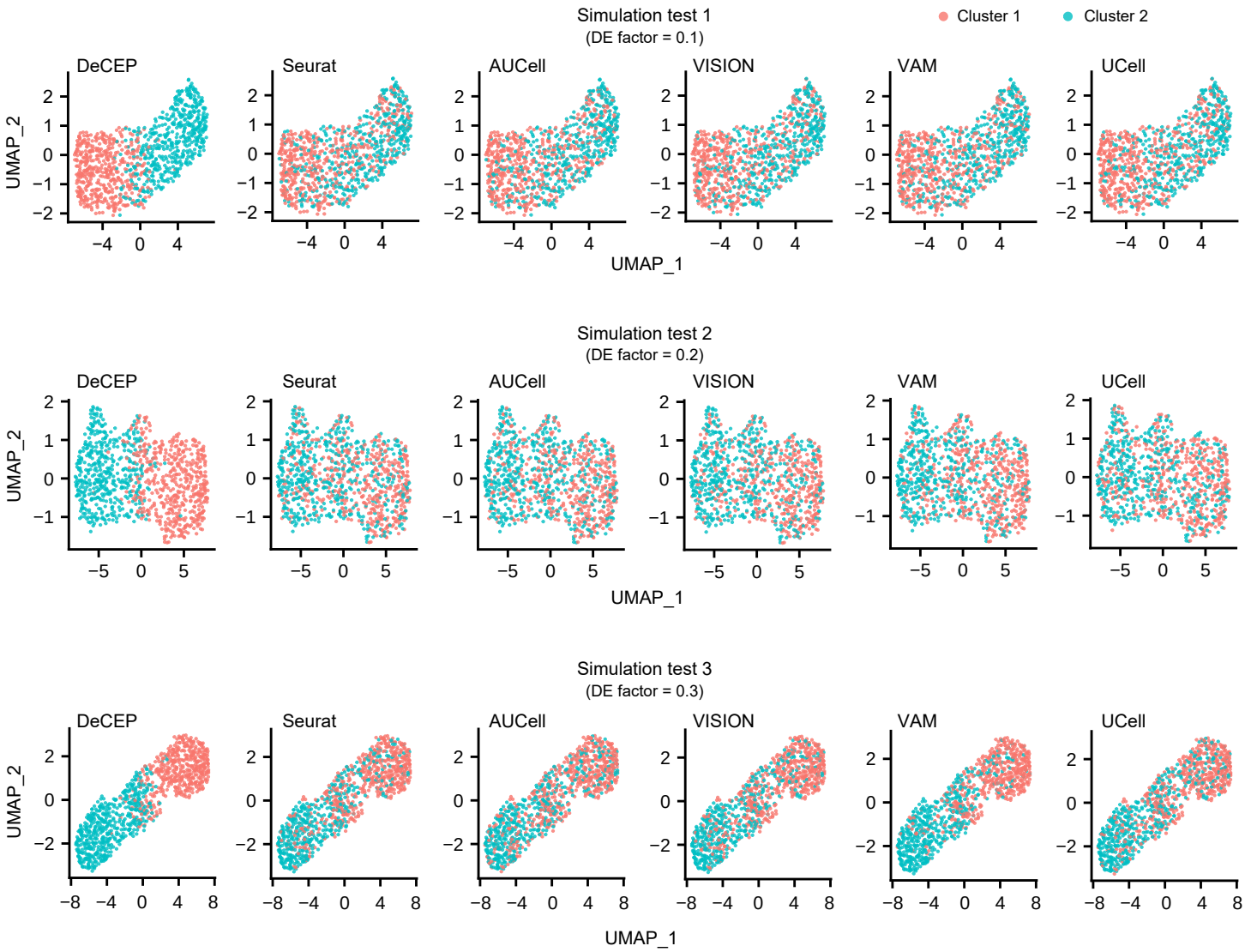
Supplemental Fig. S2. Multiple evaluation metrics demonstrate the strong performance of classification using principal components (PCs) derived from gene expression profiles of highly variable genes identified by the standard Seurat v4 pipeline. (A) Accuracy, F1 score, sensitivity, and specificity of PC-based classification in simulation tests 1-3. (B) ROC curves and corresponding AUC values for PC-based classification of each cell cluster in simulation tests 1-3.



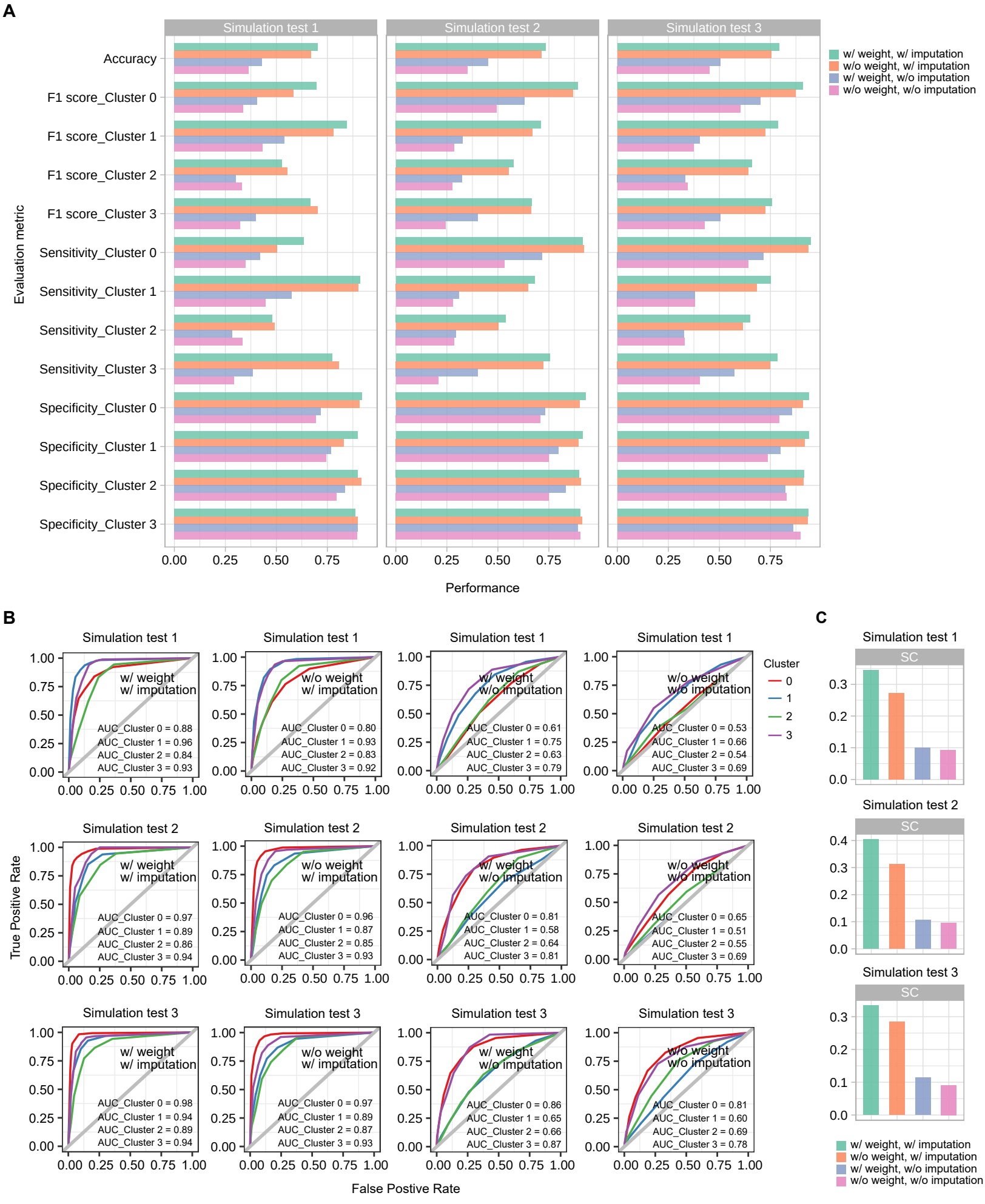
Supplemental Fig. S3. Evaluation of principal components (PCs) derived from gene expression profiles generated using simulated gene programs in simulation tests 1-3. (A) Elbow plots of the standard deviations of the PCs across simulation tests 1-3. (B) Accuracy, F1 score, sensitivity, and specificity of the PCs in simulation tests 1-3. (C) ROC curves and corresponding AUC values of the PCs for each cell cluster in simulation tests 1-3.



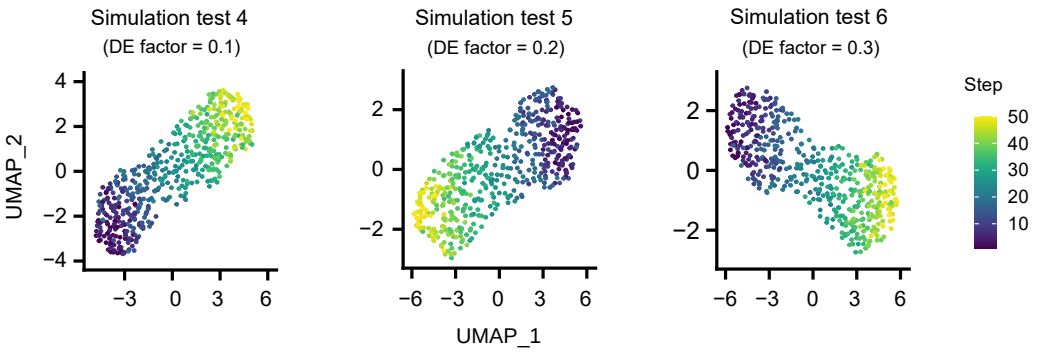
Supplemental Fig. S4. Performance comparison of DeCEP and existing methods on simulated data. (A) Sensitivity and specificity of DeCEP (discrete mode) and five other methods for each cell cluster in simulation tests 1-3. (B) ROC curves and corresponding AUC values of existing methods for each cell cluster in simulation tests 1-3.



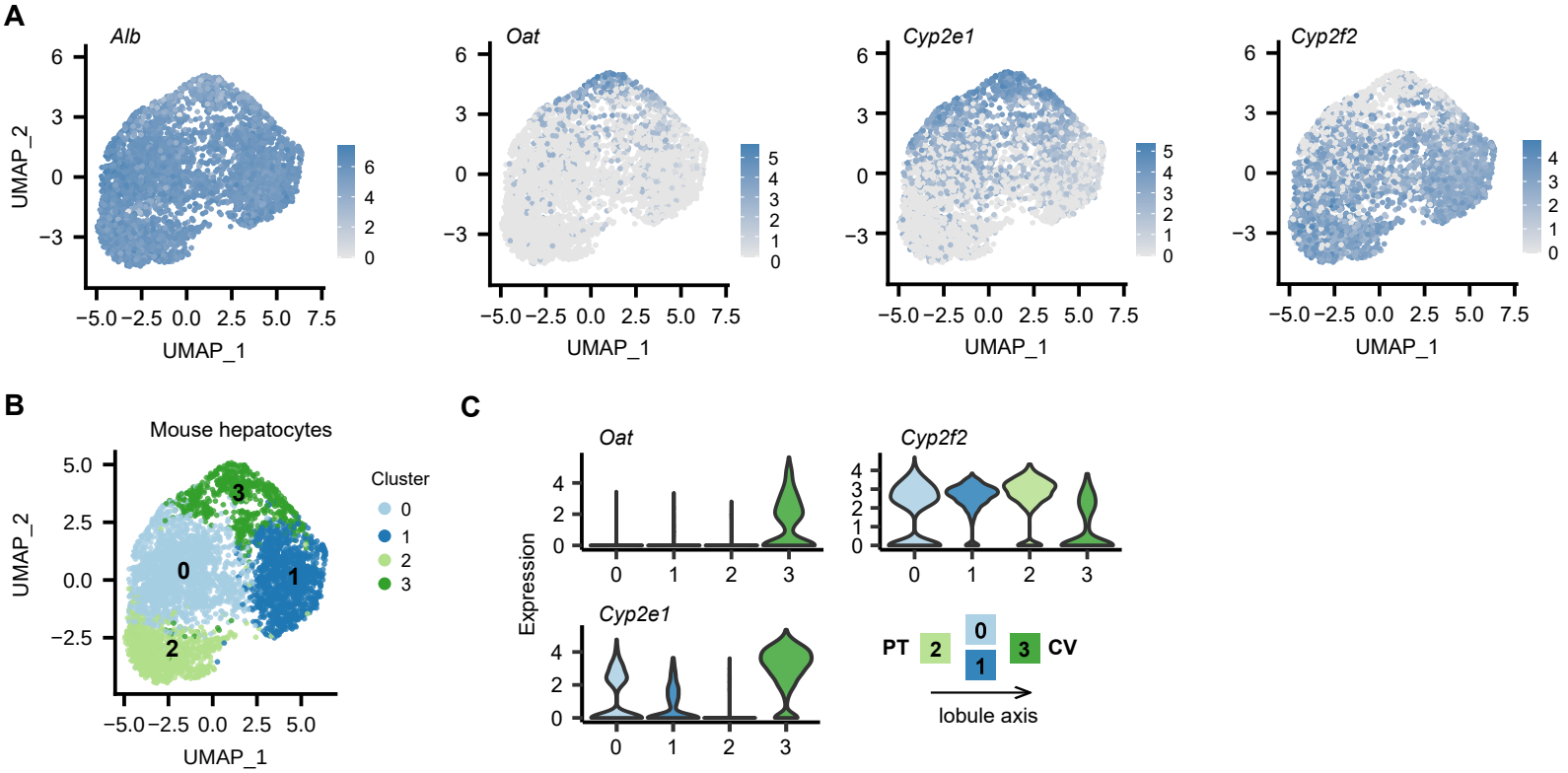
Supplemental Fig. S5. UMAP visualization of simulated cells from simulation tests 1 to 3, with cells colored by clusters identified through *k*-means clustering ($k = 2$).


A
Simulation test 1
Simulation test 2
Simulation test 3
Accuracy
F1 score_Cluster 0
F1 score_Cluster 1
F1 score_Cluster 2
F1 score_Cluster 3
Sensitivity_Cluster 0
Sensitivity_Cluster 1
Sensitivity_Cluster 2
Sensitivity_Cluster 3
Specificity_Cluster 0
Specificity_Cluster 1
Specificity_Cluster 2
Specificity_Cluster 3
Evaluation metric
0.00 0.25 0.50 0.75
Performance
w/ weight, w/ imputation
w/o weight, w/ imputation
w/ weight, w/o imputation
w/o weight, w/o imputation
B
Simulation test 1
Simulation test 2
Simulation test 3
w/ weight w/ imputation
w/o weight w/ imputation
w/ weight w/o imputation
w/o weight w/o imputation
AUC_Cluster 0 = 0.88
AUC_Cluster 1 = 0.96
AUC_Cluster 2 = 0.84
AUC_Cluster 3 = 0.93
AUC_Cluster 0 = 0.80
AUC_Cluster 1 = 0.93
AUC_Cluster 2 = 0.83
AUC_Cluster 3 = 0.92
AUC_Cluster 0 = 0.61
AUC_Cluster 1 = 0.75
AUC_Cluster 2 = 0.63
AUC_Cluster 3 = 0.79
AUC_Cluster 0 = 0.53
AUC_Cluster 1 = 0.66
AUC_Cluster 2 = 0.54
AUC_Cluster 3 = 0.69
AUC_Cluster 0 = 0.97
AUC_Cluster 1 = 0.89
AUC_Cluster 2 = 0.86
AUC_Cluster 3 = 0.94
AUC_Cluster 0 = 0.96
AUC_Cluster 1 = 0.87
AUC_Cluster 2 = 0.85
AUC_Cluster 3 = 0.93
AUC_Cluster 0 = 0.81
AUC_Cluster 1 = 0.58
AUC_Cluster 2 = 0.64
AUC_Cluster 3 = 0.81
AUC_Cluster 0 = 0.65
AUC_Cluster 1 = 0.51
AUC_Cluster 2 = 0.55
AUC_Cluster 3 = 0.69
AUC_Cluster 0 = 0.98
AUC_Cluster 1 = 0.94
AUC_Cluster 2 = 0.89
AUC_Cluster 3 = 0.94
AUC_Cluster 0 = 0.97
AUC_Cluster 1 = 0.89
AUC_Cluster 2 = 0.87
AUC_Cluster 3 = 0.93
AUC_Cluster 0 = 0.86
AUC_Cluster 1 = 0.65
AUC_Cluster 2 = 0.66
AUC_Cluster 3 = 0.87
AUC_Cluster 0 = 0.81
AUC_Cluster 1 = 0.60
AUC_Cluster 2 = 0.69
AUC_Cluster 3 = 0.78
True Positive Rate
False Postive Rate
Cluster
0
1
2
3
C
Simulation test 1
Simulation test 2
Simulation test 3
SC
w/ weight, w/ imputation
w/o weight, w/ imputation
w/ weight, w/o imputation
w/o weight, w/o imputation

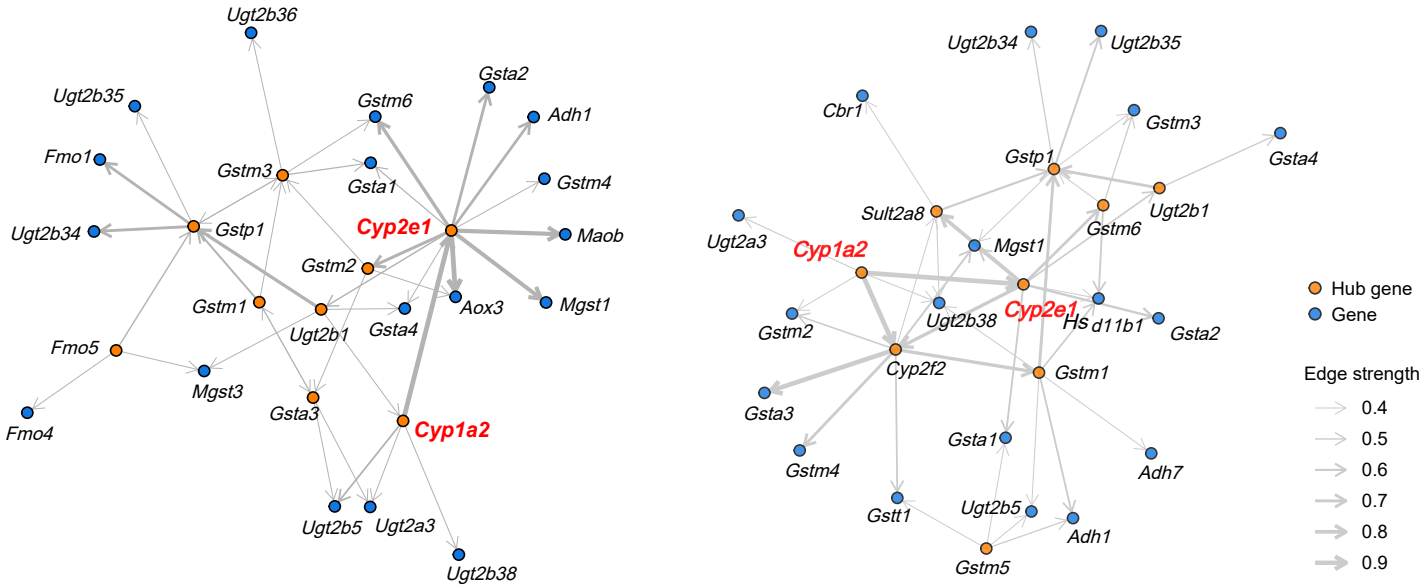
Supplemental Fig. S6. Multiple evaluation metrics indicate that the weighted quantification of context-specific gene programs improves performance on simulated data. (A) Accuracy, F1 score, sensitivity, and specificity for different combinations of the DeCEP framework components in simulation tests 1-3. (B) ROC curves and corresponding AUC values for various DeCEP framework component combinations across each cell cluster. (C) Mean silhouette coefficient (SC) values for different combinations of the DeCEP framework components.



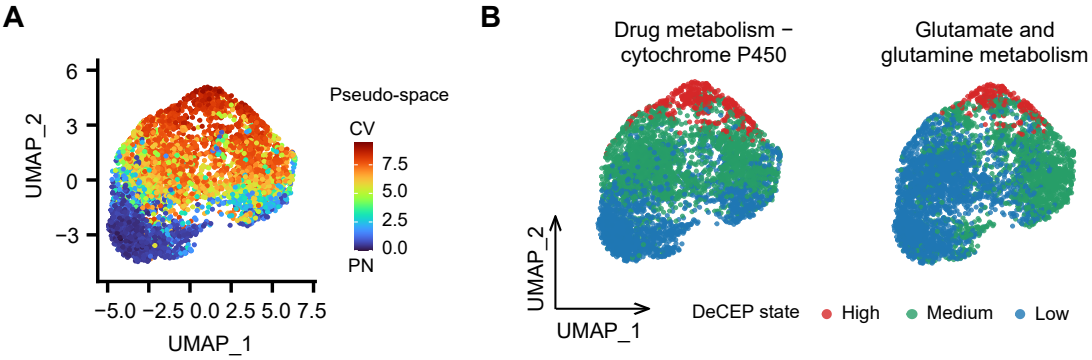
Supplemental Fig. S7. UMAP visualization of simulated cells with 50 dynamic steps in simulation tests 4-6. From left to right, the differential expression (DE) factors are 0.1, 0.2, and 0.3, respectively, indicating the gradually increasing gene expression variability between cells.



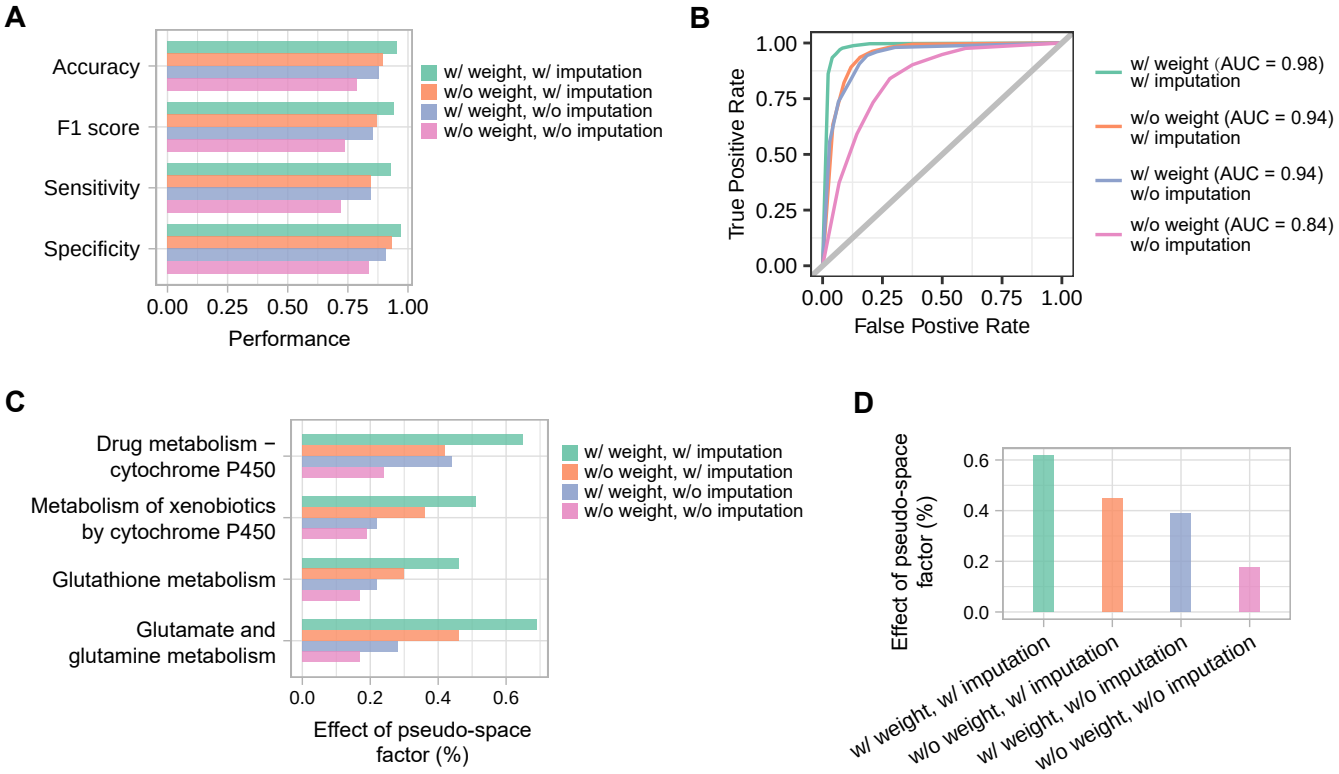
Supplemental Fig. S8. Initial characterization of the normal liver tissue scRNA-seq dataset. (A) Expression levels of *Alb*, *Oat*, *Cyp2e1*, and *Cyp2f2* in hepatocytes. (B) UMAP visualization of hepatocytes derived from the adult mouse liver. Cells are colored by clusters. (C) Expression levels of three zonation marker genes and the resulting inferred order of cell clusters along the lobular axis from the portal triads (PT) to the central vein (CV).



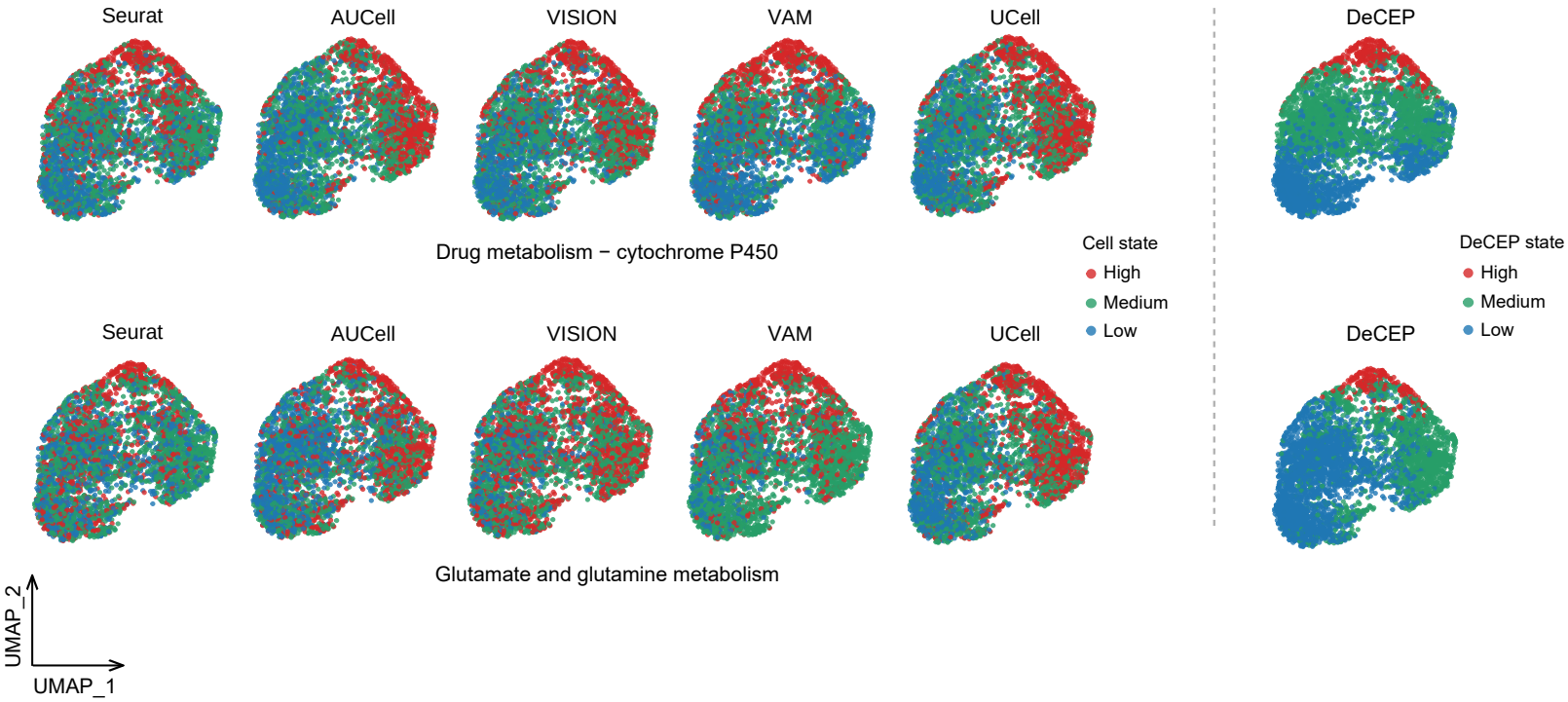
Supplemental Fig. S9. The hepatocyte-specific functional networks of detoxification-related gene programs, where genes in red represent key hub genes related to liver detoxification.



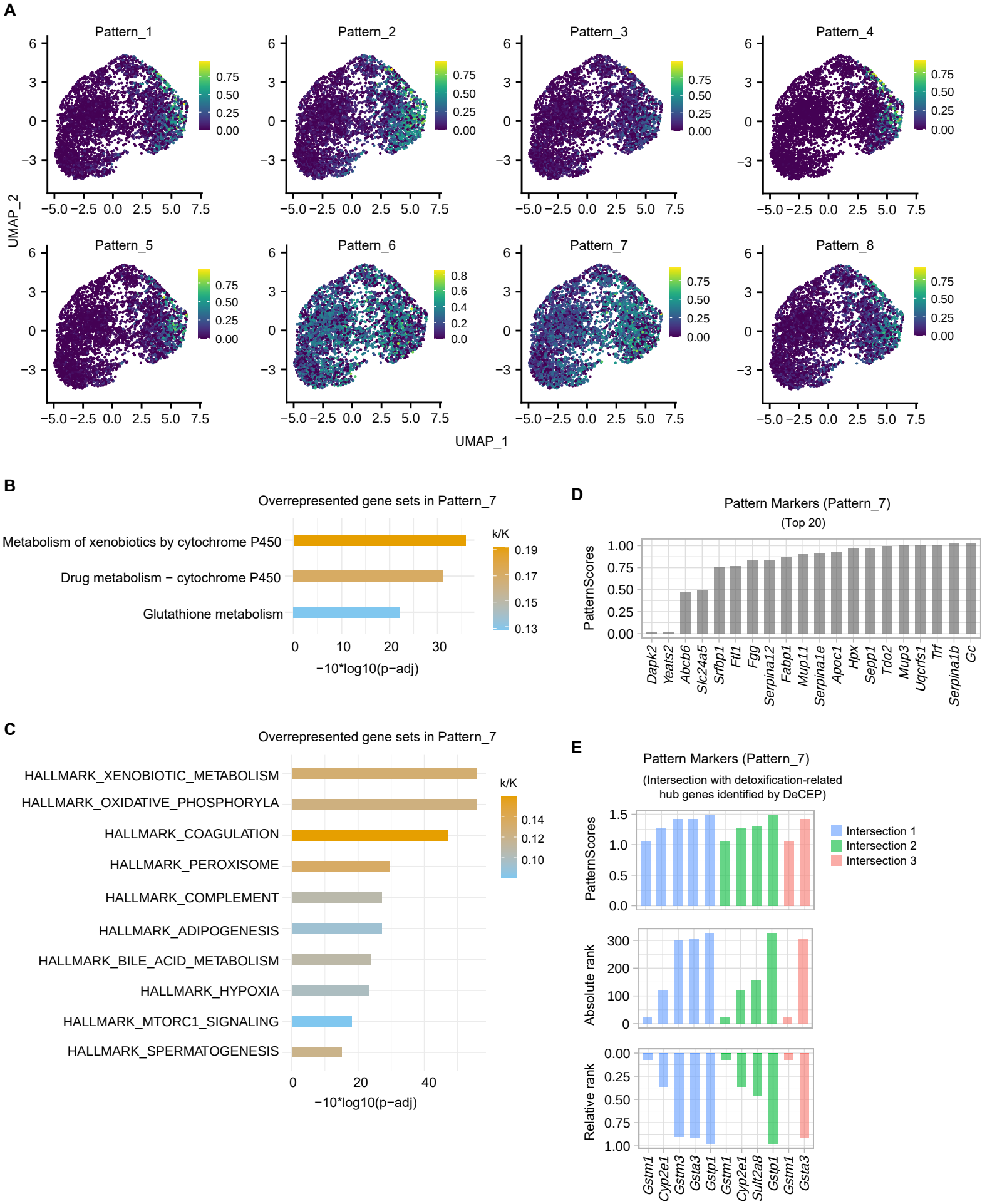
Supplemental Fig. S10. Pseudo-spatial and DeCEP state mapping of hepatocytes by UMAP. (A) UMAP visualization of hepatocytes at the pseudo-spatial level. (B) UMAP visualization of hepatocytes with the DeCEP states.



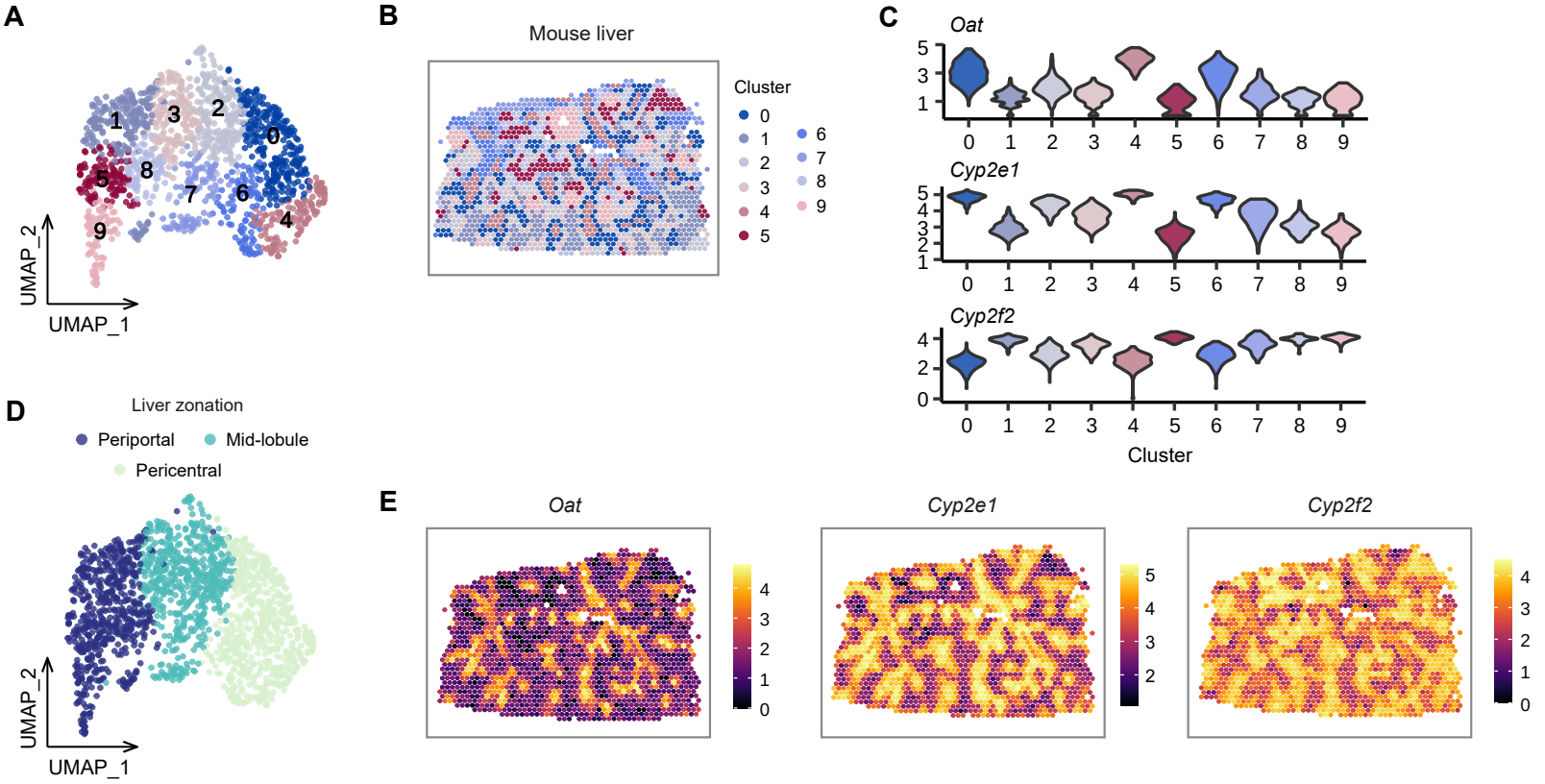
Supplemental Fig. S11. Evaluation metrics demonstrate that the weighted quantification of context-specific gene programs enhances performance on the normal liver tissue dataset. (A) Accuracy, F1 score, sensitivity, and specificity for different combinations of the DeCEP framework components used to characterize liver detoxification-related gene programs. (B) ROC curves and corresponding AUC values for various combinations of DeCEP framework components in characterizing liver detoxification-related gene programs. (C) Associations between individual gene program activity scores and the pseudo-space factor across different DeCEP framework component combinations. (D) Strength of the associations between the overall activity of detoxification-related gene programs and the pseudo-space factor in various DeCEP framework component combinations.



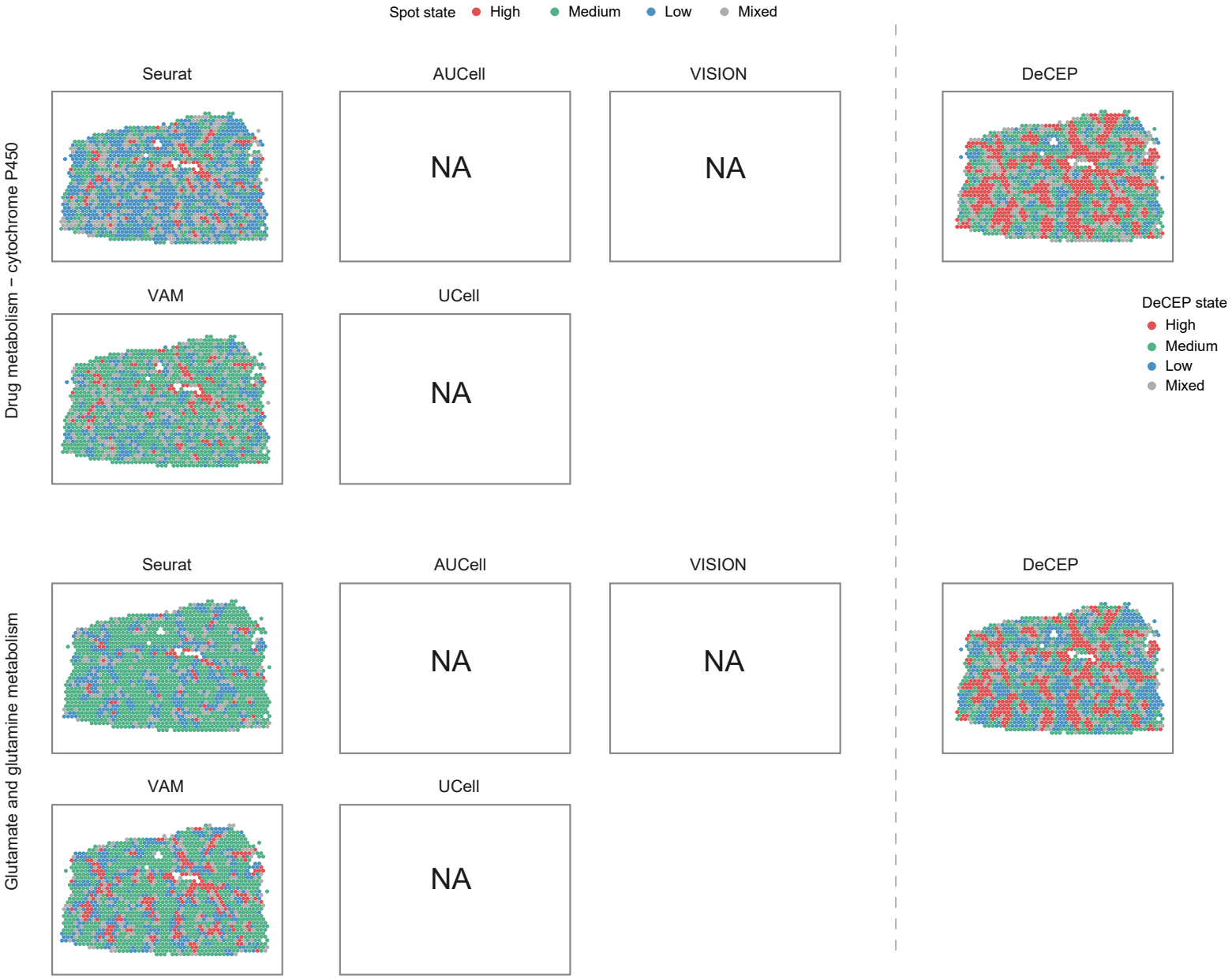
Supplemental Fig. S12. UMAP visualization of hepatocytes with the cell states evaluated by DeCEP and the other five methods.


A
Pattern_1
Pattern_2
Pattern_3
Pattern_4
Pattern_5
Pattern_6
Pattern_7
Pattern_8
UMAP_2
UMAP_1
B
Overrepresented gene sets in Pattern_7
Metabolism of xenobiotics by cytochrome P450
Drug metabolism − cytochrome P450
Glutathione metabolism
k/K
−10*log10(p−adj)
C
Overrepresented gene sets in Pattern_7
HALLMARK_XENOBIOTIC_METABOLISM
HALLMARK_OXIDATIVE_PHOSPHORYLA
HALLMARK_COAGULATION
HALLMARK_PEROXISOME
HALLMARK_COMPLEMENT
HALLMARK_ADIPOGENESIS
HALLMARK_BILE_ACID_METABOLISM
HALLMARK_HYPOXIA
HALLMARK_MTORC1_SIGNALING
HALLMARK_SPERMATOGENESIS
k/K
−10*log10(p−adj)
D
Pattern Markers (Pattern_7)
(Top 20)
PatternScores
Dapk2
Yeats2
Abcb6
Slc24a5
Srfbp1
Ftl1
Fgg
Serpina12
Fabp1
Mup11
Serpina1e
Apoc1
Hpx
Sepp1
Tdo2
Mup3
Uqcrfs1
Trf
Serpina1b
Gc
E
Pattern Markers (Pattern_7)
(Intersection with detoxification-related
hub genes identified by DeCEP)
PatternScores
Absolute rank
Relative rank
Intersection 1
Intersection 2
Intersection 3
Gstm1
Cyp2e1
Gstm3
Gsta3
Gstp1
Gstm1
Cyp2e1
Sult2a8
Gstp1
Gstm1
Gsta3

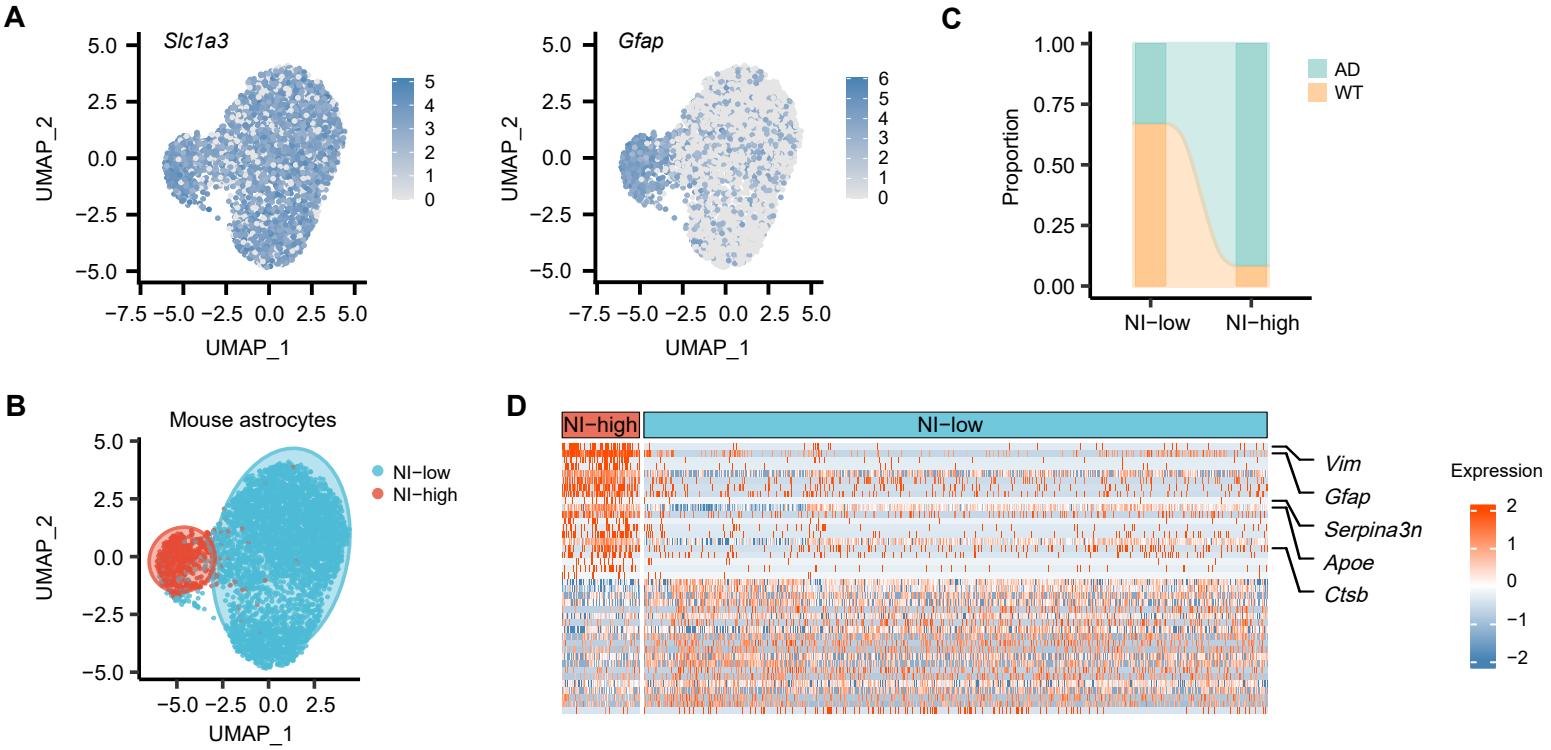
Supplemental Fig. S13. Performance of CoGAPS in characterizing liver detoxification function in the normal liver tissue dataset. (A) UMAP visualization of hepatocytes based on the activity levels of eight patterns factorized by CoGAPS. (B) Gene programs associated with liver detoxification are significantly enriched in Pattern 7. (C) Significant enrichment of the hallmark gene sets in Pattern 7. (D) The top 20 marker genes associated with Pattern 7. (E) Intersection of detoxification-related context-dependent hub genes identified by DeCEP and marker genes associated with Pattern 7 identified by CoGAPS. Intersections 1-3 represent the overlap between context-dependent hub genes related to drug metabolism - cytochrome P450, metabolism of xenobiotics by cytochrome, and glutathione metabolism, as identified by DeCEP, and the marker genes associated with Pattern 7, respectively. The top panel shows the PatternScores of the overlapping genes. The middle panel displays their absolute ranks, and the bottom panel shows their relative ranks among all marker genes for Pattern 7.



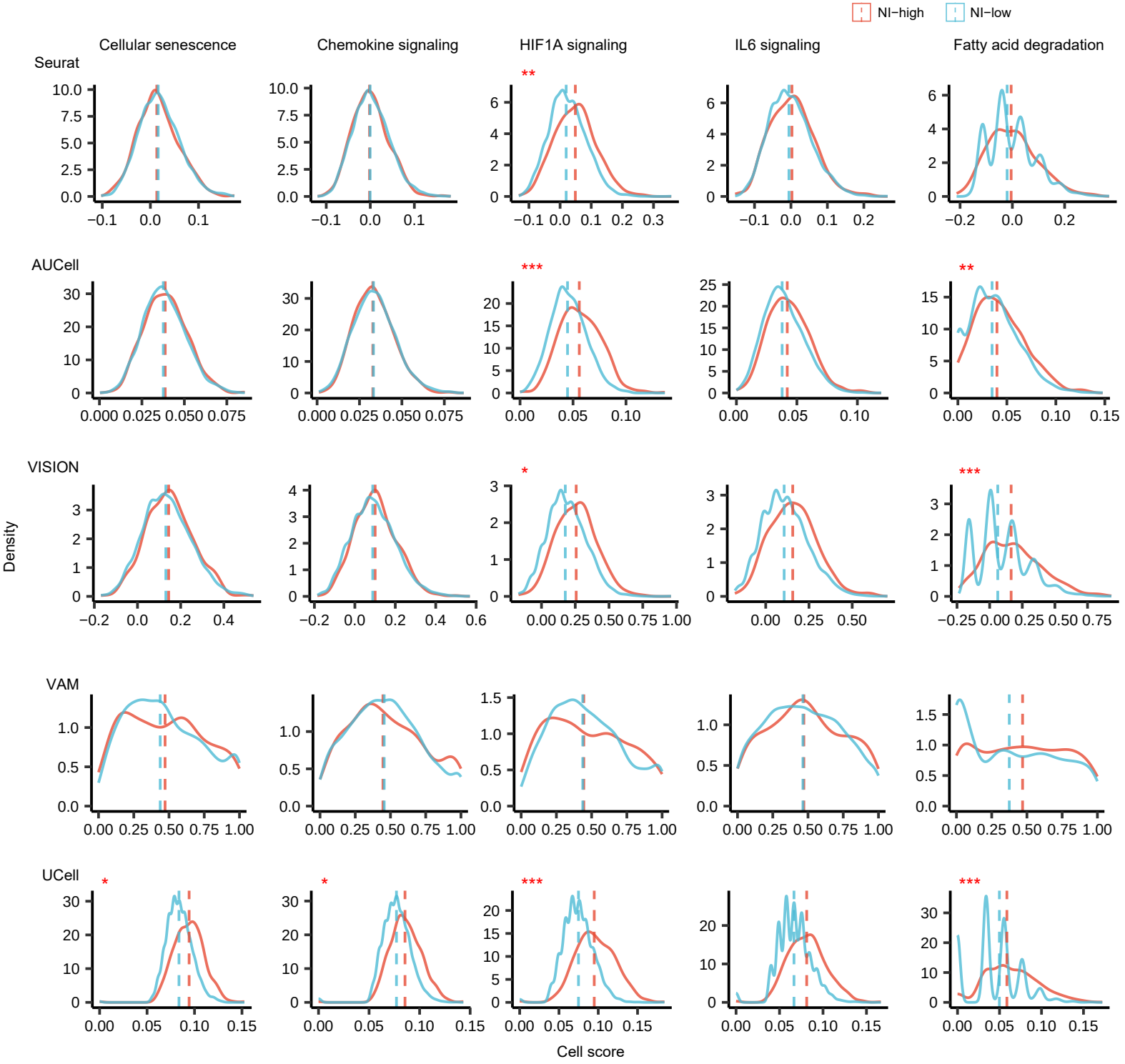
Supplemental Fig. S14. Initial characterization of the adult mouse liver ST data. (A) UMAP visualization of spatial spot embedding. Spots are colored by clusters. (B) Distribution of spatial domains in the liver tissue section. (C) Violin plots of the expression levels of three typical detoxification-related genes across these spatial domains. (D) UMAP visualization of spatial spot embedding. Spots are colored by inferred liver regions. (E) Distribution of the hepatic zonation marker genes *Oat*, *Cyp2e1*, and *Cyp2f2* in the liver tissue section.



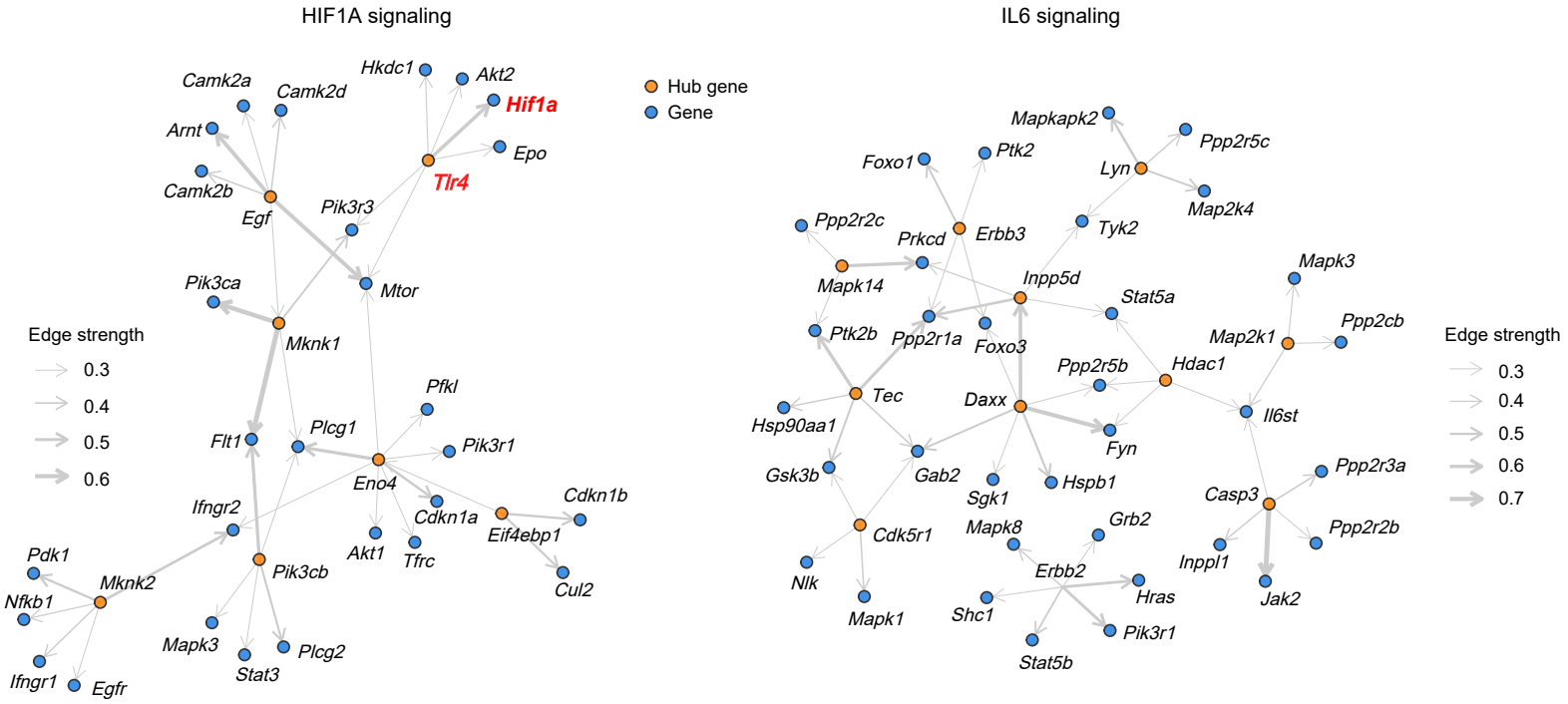
Supplemental Fig. S15. Comparison of gene program activity states identified by DeCEP and existing methods in the liver tissue section. NA indicates that the corresponding method failed to effectively identify the spot states.



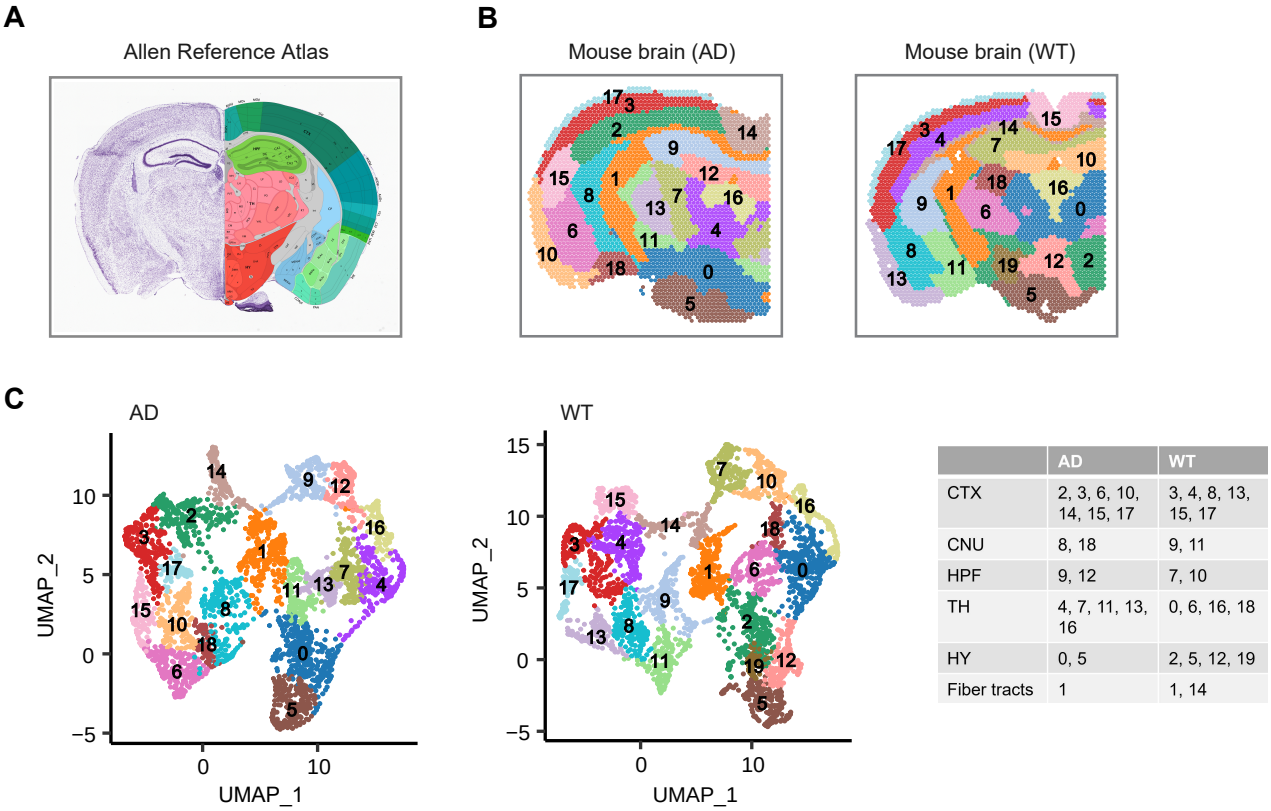
Supplemental Fig. S16. Initial characterization of the Alzheimer's disease scRNA-seq dataset. (A) Expression levels of *Slc1a3* and *Gfap* in astrocytes. (B) UMAP visualization of astrocytes from the hippocampus of Alzheimer's disease (AD) and wild-type (WT) mice. Cells are colored by neuroinflammation (NI) levels. (C) Proportions of astrocytes in AD and WT mice under different inflammatory conditions. (D) Distinct gene expression patterns under NI-high and NI-low conditions.



Supplemental Fig. S17. Density plots of the activity scores for known neuroinflammatory gene programs under NI-high and NI-low conditions calculated by existing methods. The activity scores of NI-high and NI-low astrocytes were compared. The one-sided Wilcoxon rank-sum test: ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

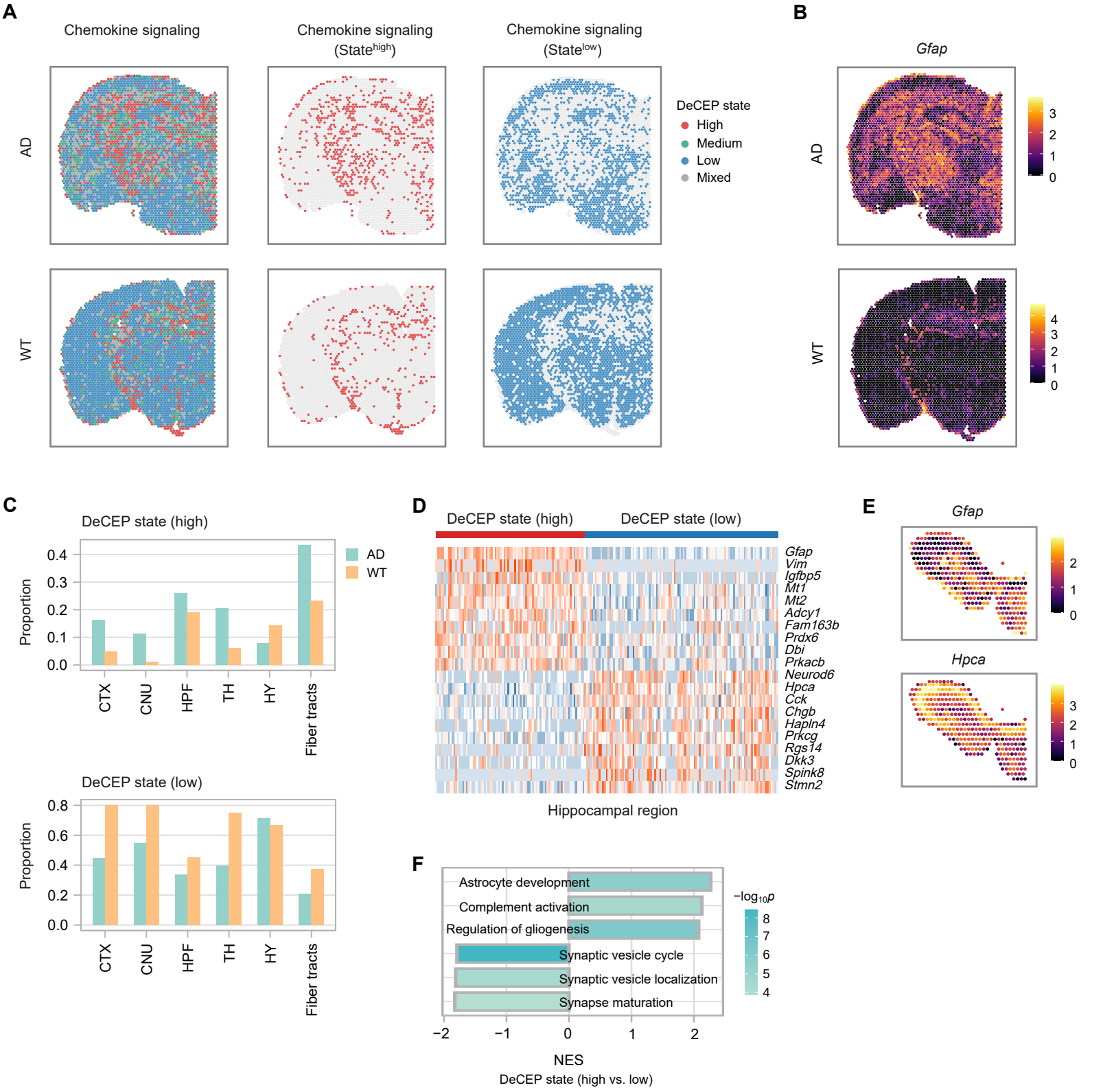


Supplemental Fig. S18. The NI-high-specific functional networks of the neuroinflammatory gene programs.

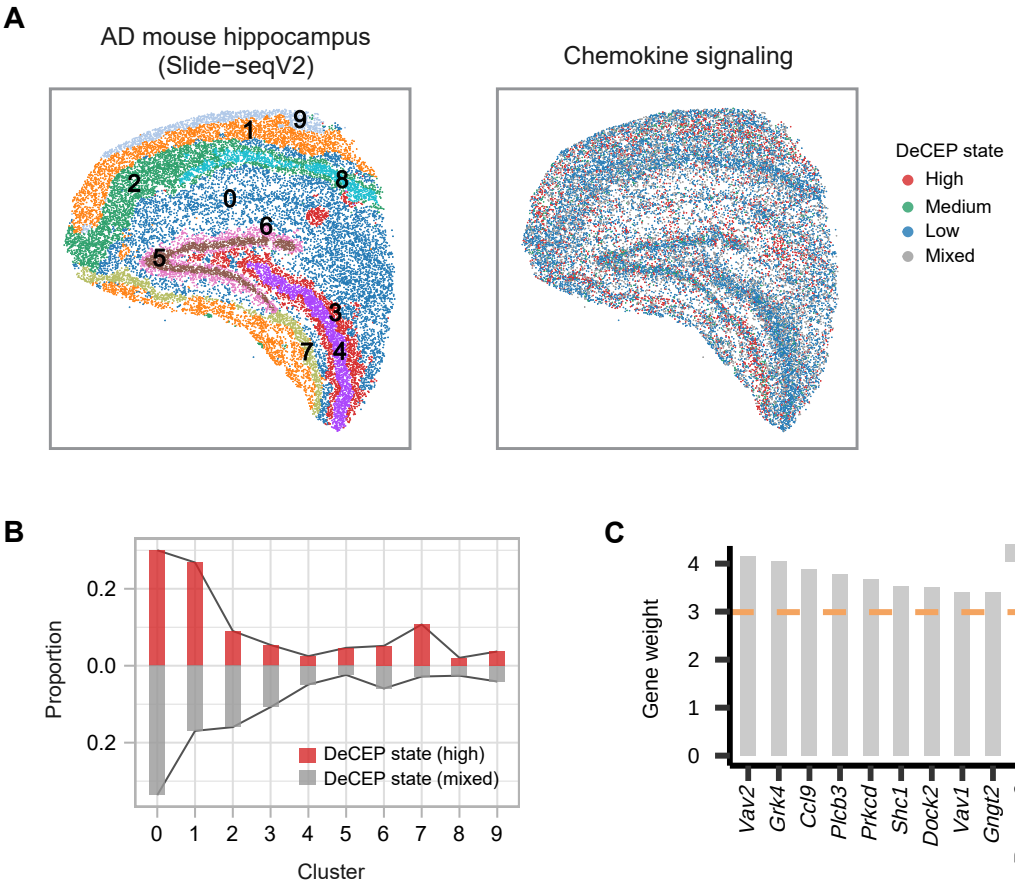


	AD	WT
CTX	2, 3, 6, 10, 14, 15, 17	3, 4, 8, 13, 15, 17
CNU	8, 18	9, 11
HPF	9, 12	7, 10
TH	4, 7, 11, 13, 16	0, 6, 16, 18
HY	0, 5	2, 5, 12, 19
Fiber tracts	1	1, 14

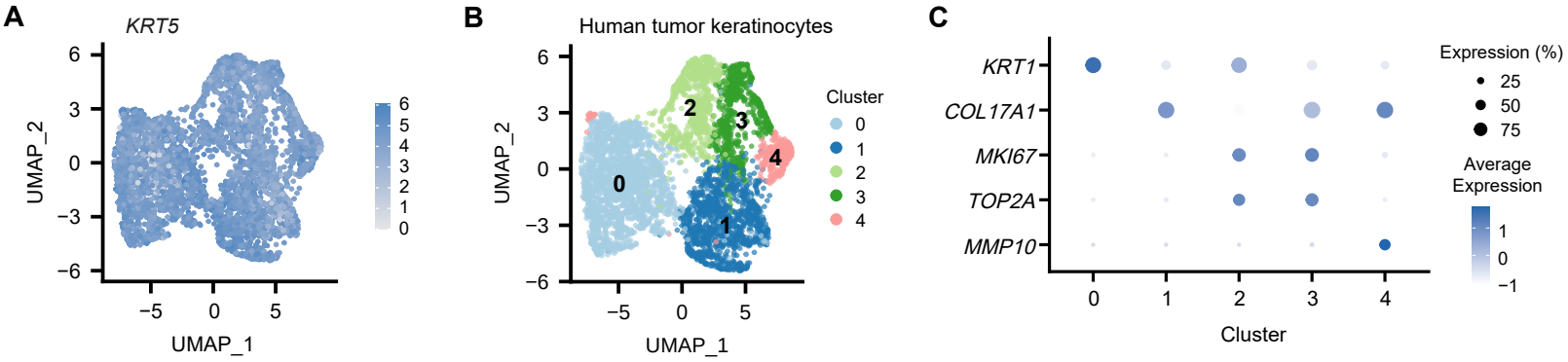
Supplemental Fig. S19. Initial characterization of the mouse brain ST data. (A) Anatomy of the mouse brain from the Allen Brain Atlas. (B-C) Demonstration of spatial domains derived from spatial clustering in brain sections from Alzheimer's disease (AD) and wild-type (WT) mice. The spatial domains were further assigned to six brain regions (cerebral cortex (CTX), cerebral nuclei (CNU), hippocampal formation (HPF), thalamus (TH), hypothalamus (HY), and fiber tracts) based on the Allen Brain Atlas (B) and UMAP visualization (C). The table on the right summarizes the spatial domain assignments corresponding to each brain region.



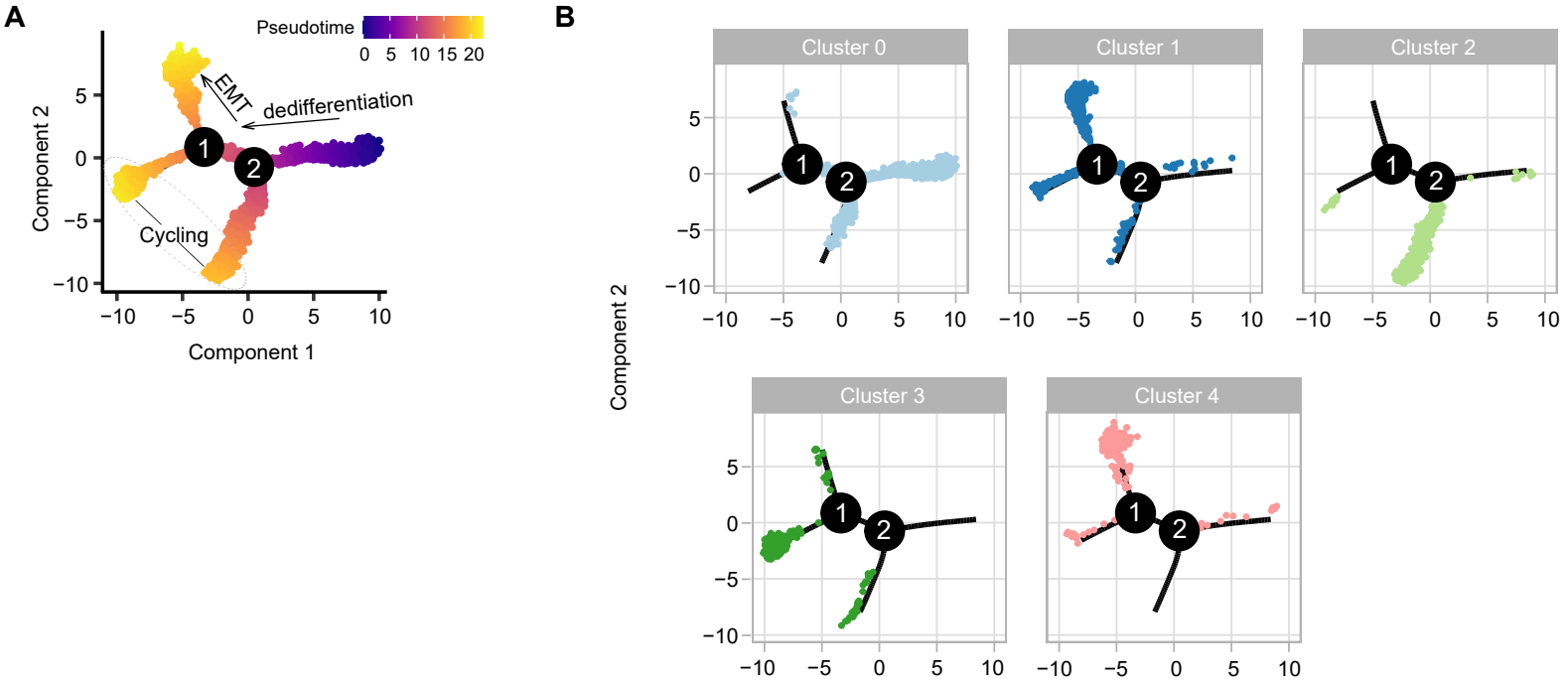
Supplemental Fig. S20 DeCEP reveals spatial distributions of the representative neuroinflammatory gene program in mouse brains. (A) Comparison of the distributions of the DeCEP states across spatial locations in AD and WT mice, respectively. (B) Comparison of *Gfap* expression levels in AD and WT sections. (C) Comparison of the proportions of high and low DeCEP states in six brain regions in the AD and WT sections. (D) Significantly differentially expressed genes in the hippocampal region between high and low DeCEP states in the AD section. (E) Comparison of the expression levels of *Gfap* and *Hpca* in the AD hippocampus. (F) Comparison of functions significantly enriched in high DeCEP states versus low DeCEP states in AD mice.



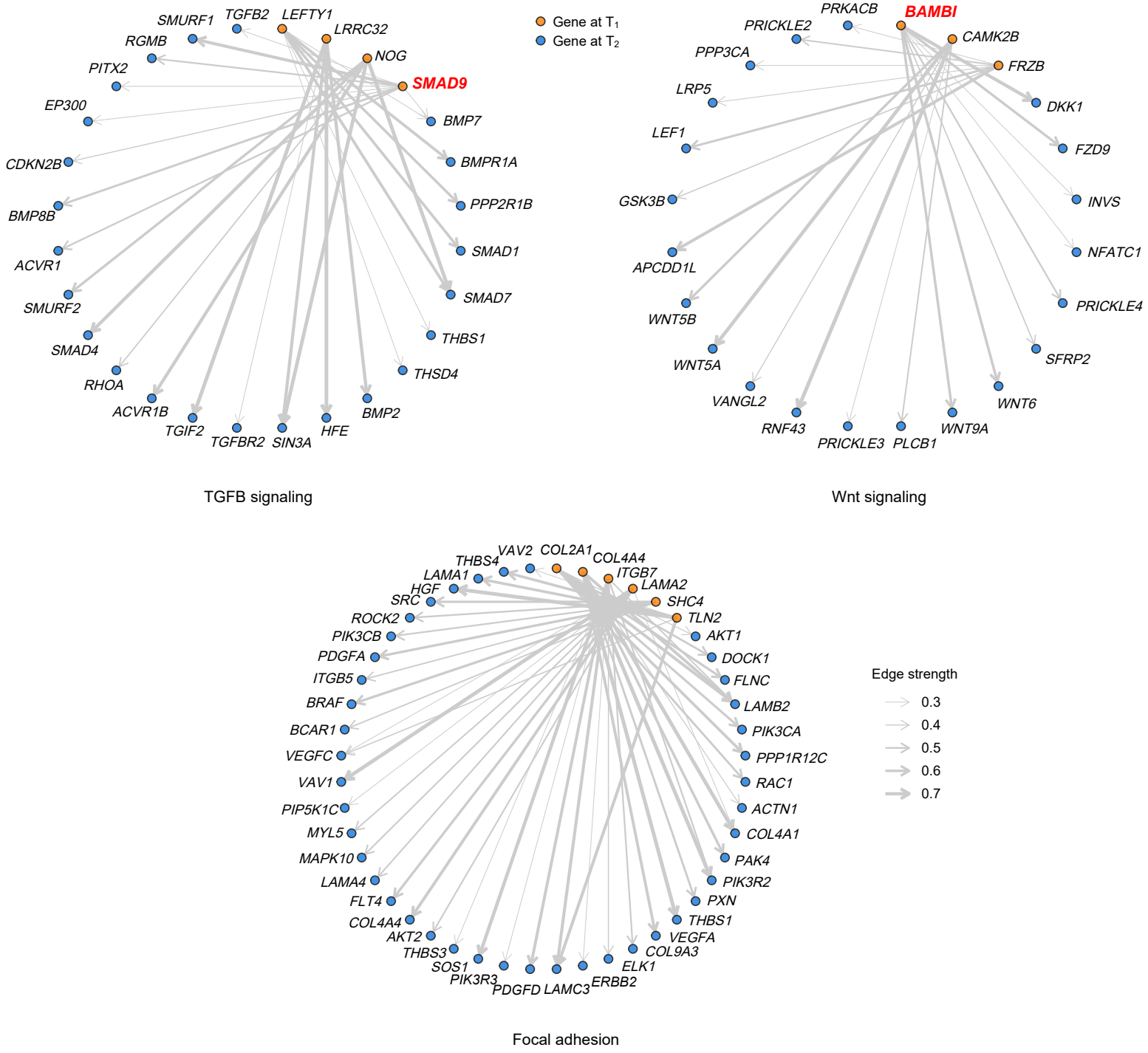
Supplemental Fig. S21. DeCEP further identifies the spatial characteristics of the representative neuroinflammatory gene program in a mouse brain using Slide-seqV2 data. (A) Spatial domains and DeCEP states of AD hippocampus data from the Slide-seqV2 platform, where the numbers represent multiple spatial domains derived from spatial clustering. (B) Comparison of the proportions of high DeCEP states and mixed DeCEP states in different spatial domains of the AD hippocampus. (C) Weights of hub genes in the representative neuroinflammatory gene program with mixed DeCEP states as the ROI.



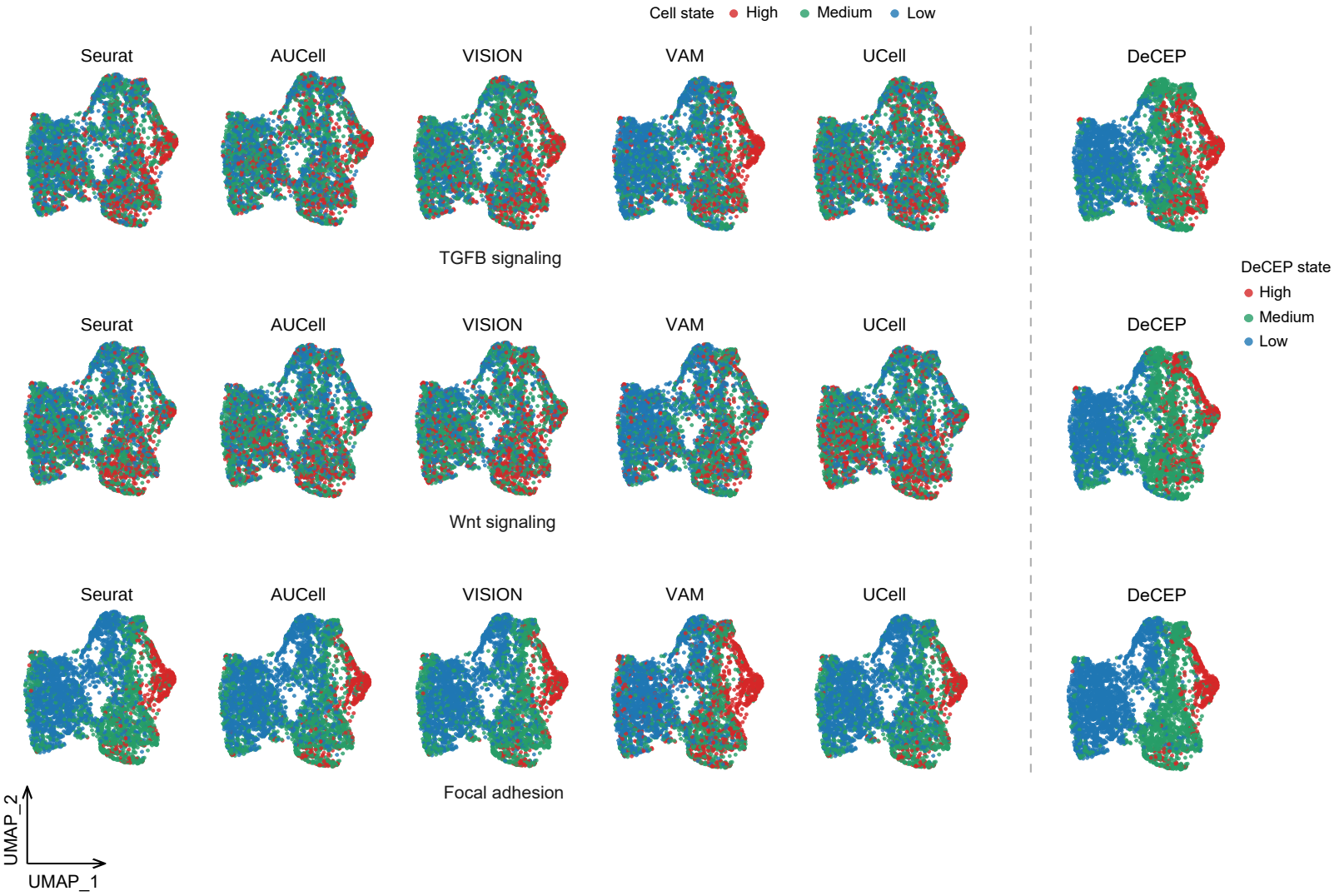
Supplemental Fig. S22. Initial characterization of the cSCC scRNA-seq dataset. (A) UMAP visualization of *KRT5* expression levels in tumor keratinocytes. (B) UMAP visualization of tumor keratinocytes from the human cutaneous squamous cell carcinoma (cSCC) dataset. Cells are colored by clusters. (C) Dot plot of the expression levels of marker genes in different cell clusters.



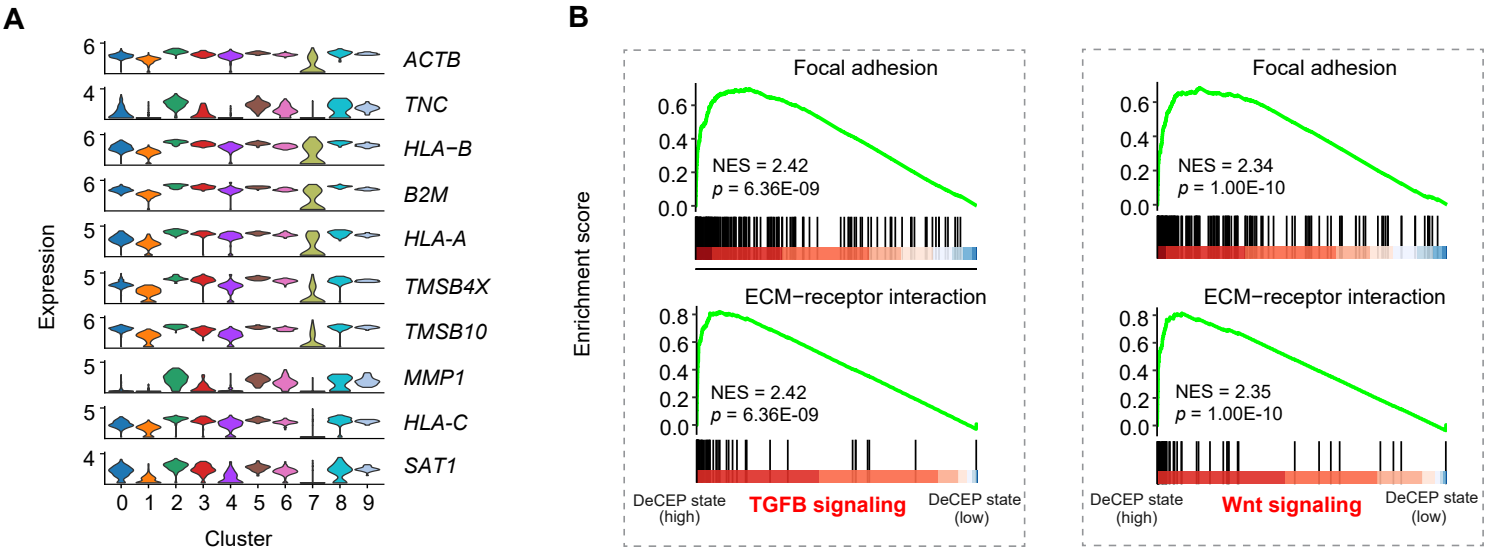
Supplemental Fig. S23. Trajectory inference of tumor keratinocytes. (A) Pseudotime of tumor keratinocytes and their distribution along the trajectory. (B) Distribution of each cell cluster along the trajectory.



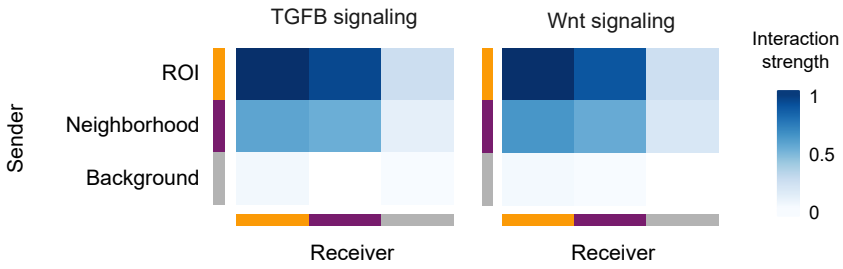
Supplemental Fig. S24. The dynamic functional networks of cancer-related gene programs during tumor progression. T_1 represents the cellular context before the transition points of the gene programs, and T_2 represents the cellular context after the transition points.



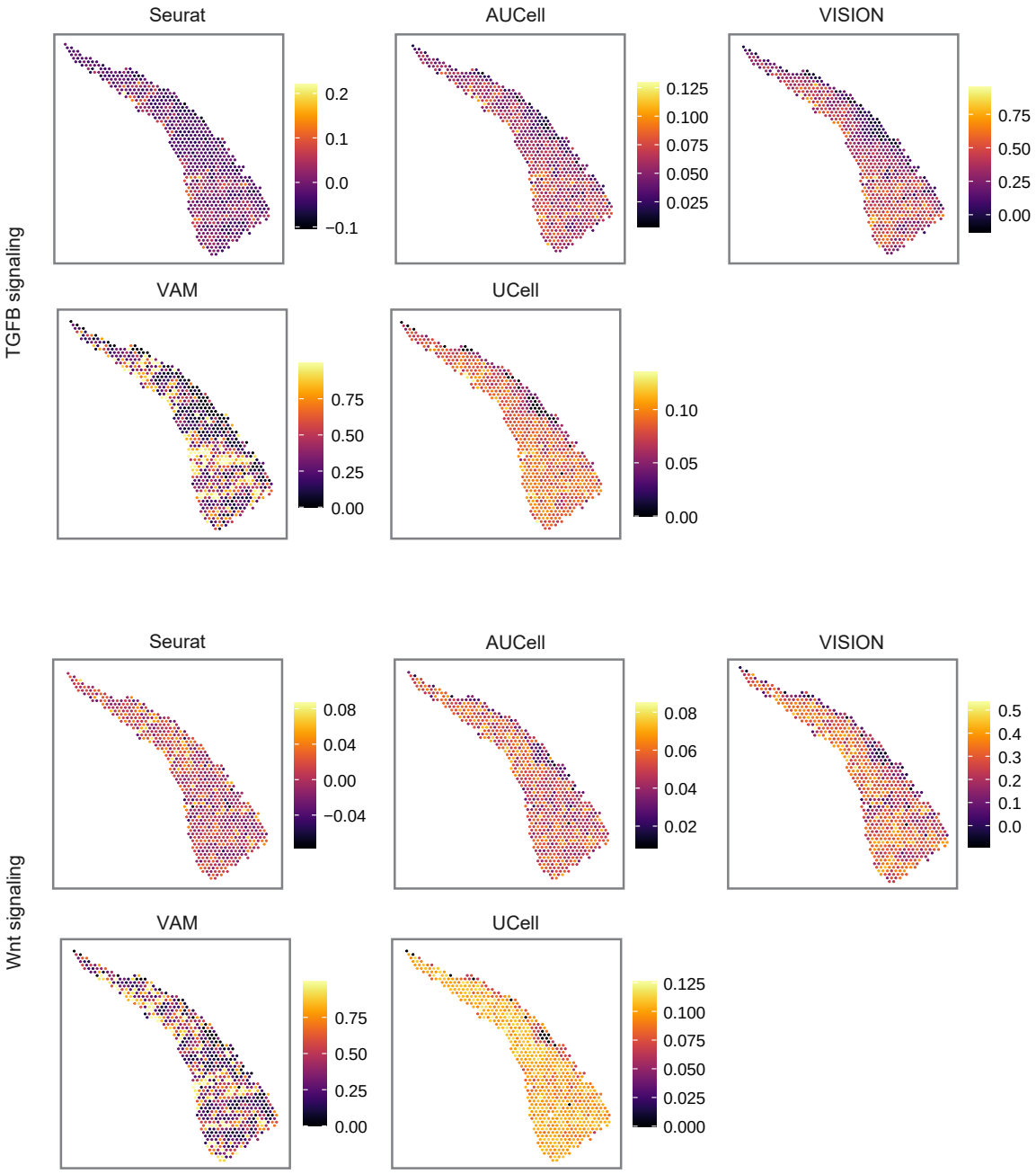
Supplemental Fig. S25 UMAP visualization of tumor keratinocytes with the cell states evaluated by DeCEP and the other five methods.



Supplemental Fig. S26. Characterization of the human cSCC ST data. (A) Violin plots of several genes significantly expressed at high levels simultaneously in clusters 2 and 5. (B) GSEA plots of the top two significantly enriched functions in high DeCEP states of TGFB and Wnt signaling.



Supplemental Fig. S27. Compared with background spots, the ROIs and their corresponding neighborhoods have a higher probability of cell-cell communication.



Supplemental Fig. S28. Gene program activity scores for TGFB and Wnt signaling calculated by existing methods in the human cSCC tissue section.