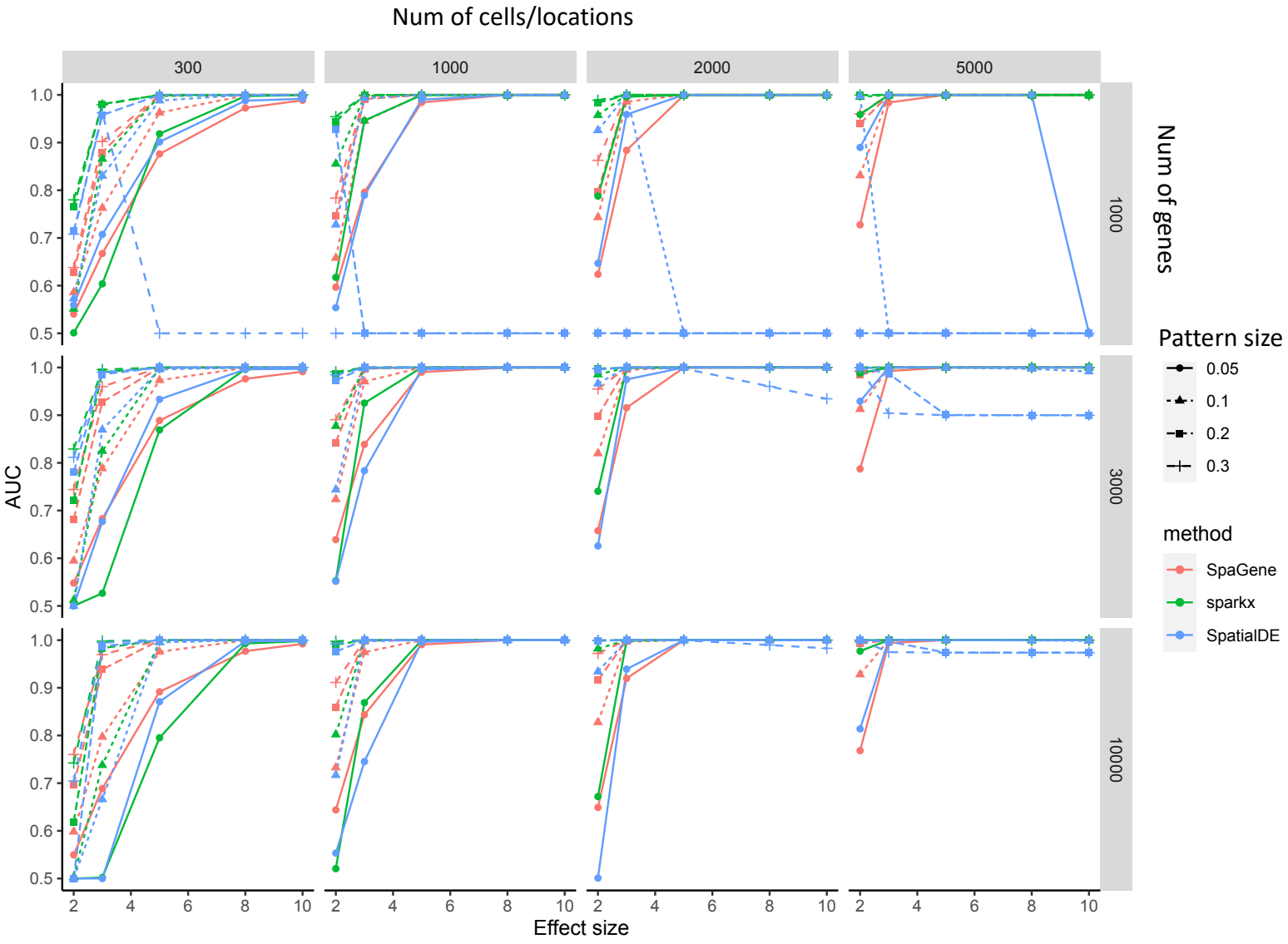
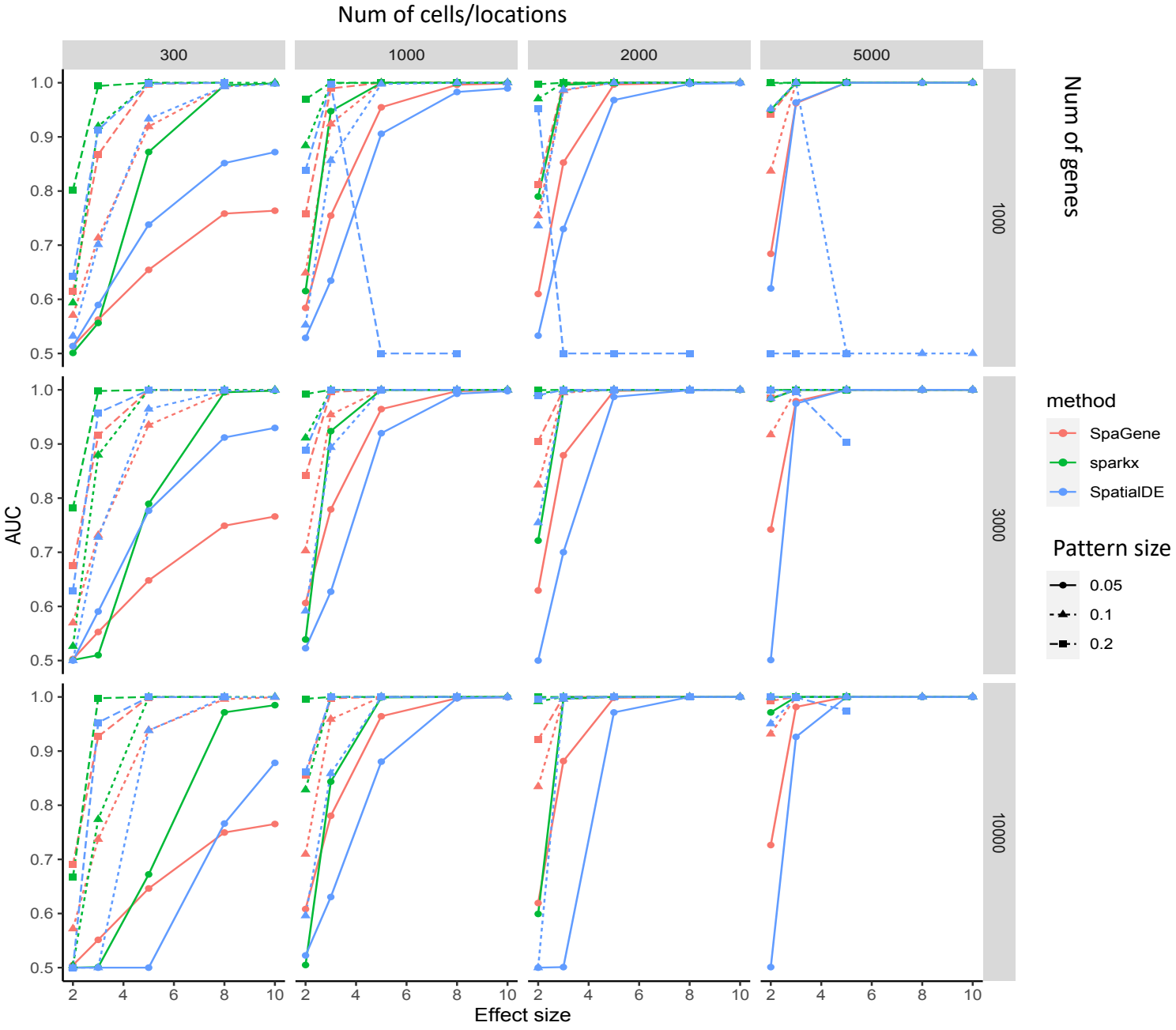
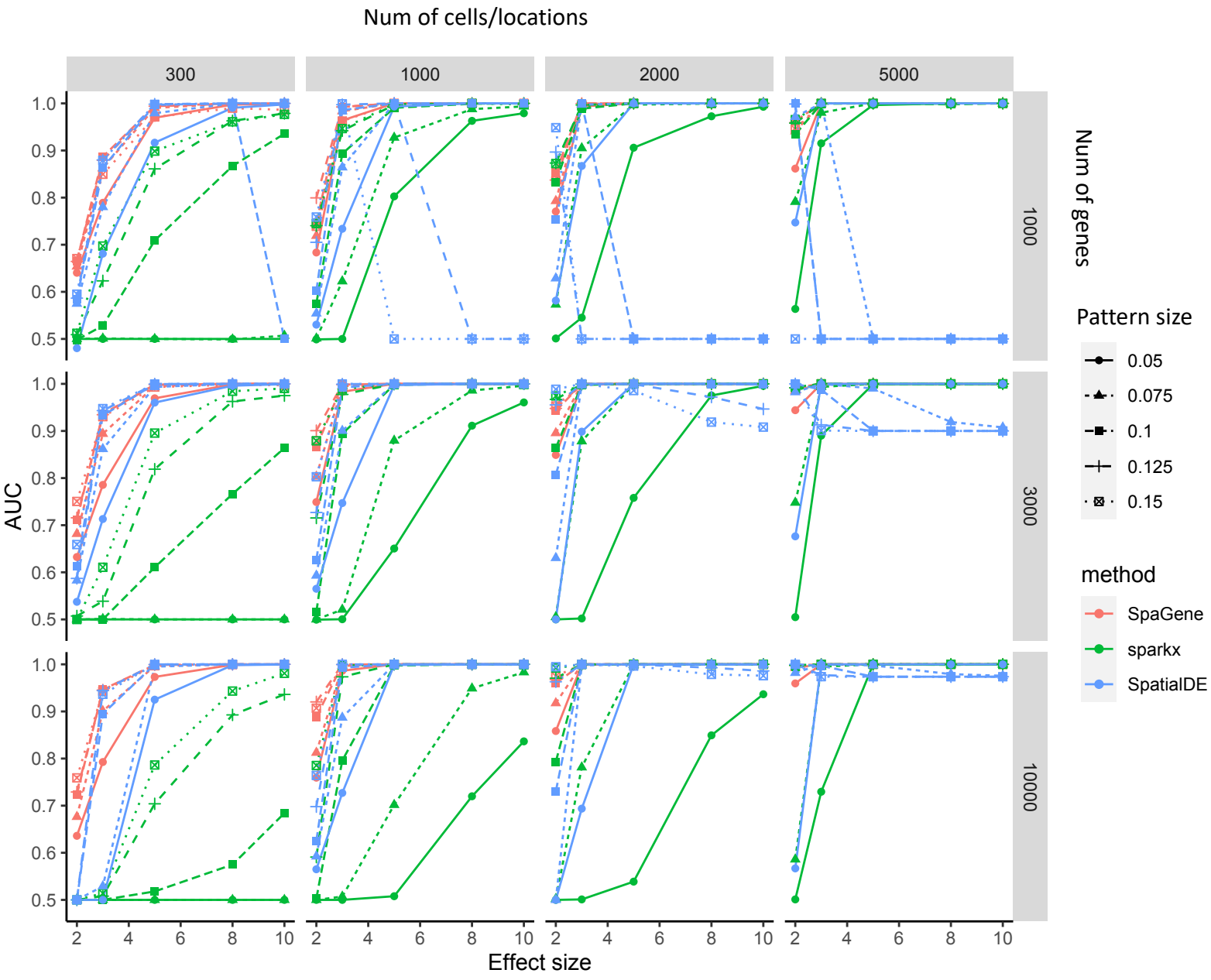
Supplemental Figures



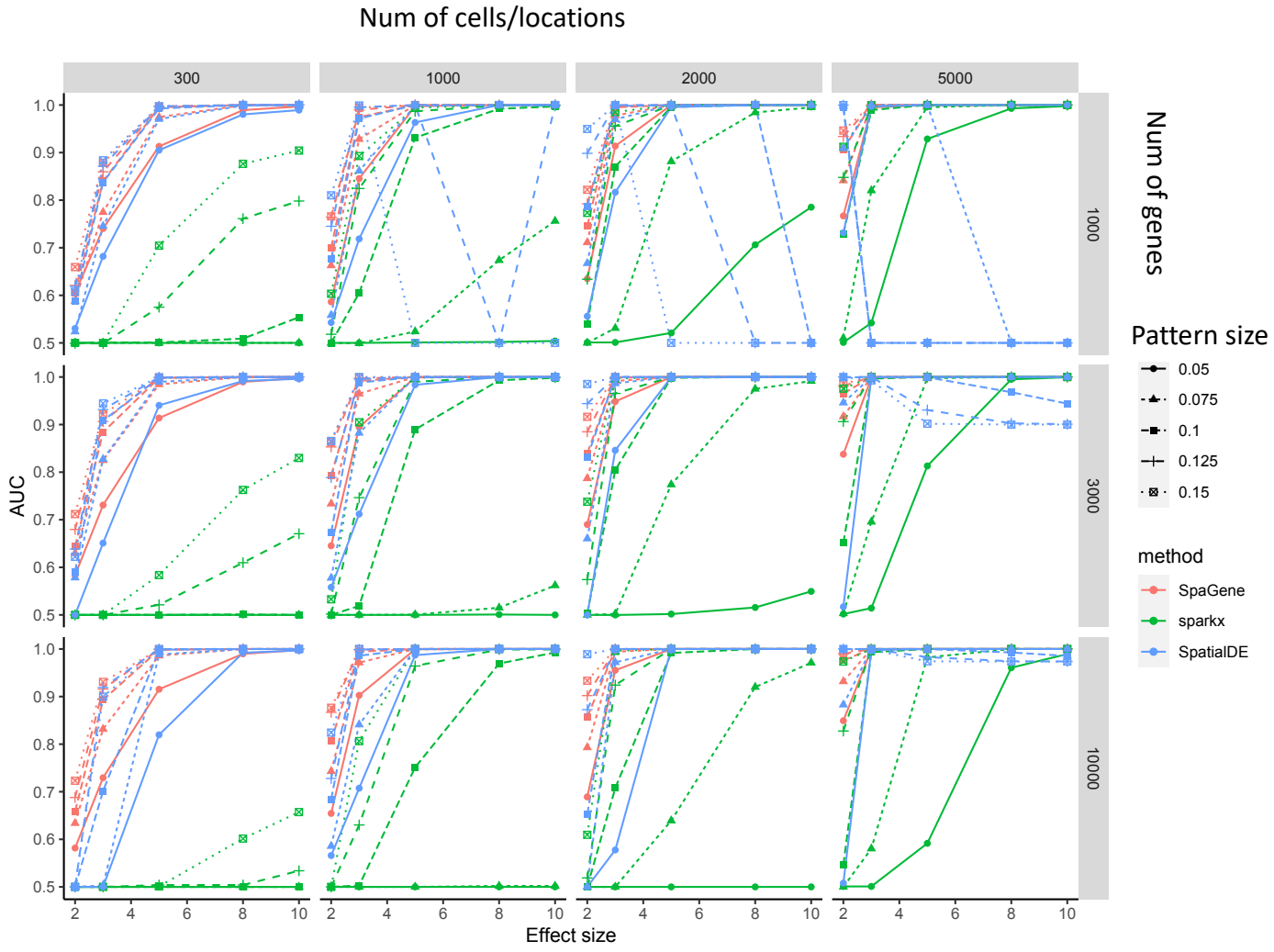
Supplemental Fig. S1. AUC plots of SpaGene (red), SPARK-X (green) and SpatialDE (blue) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a hotspot pattern, which were generated from negative binomial distributions.



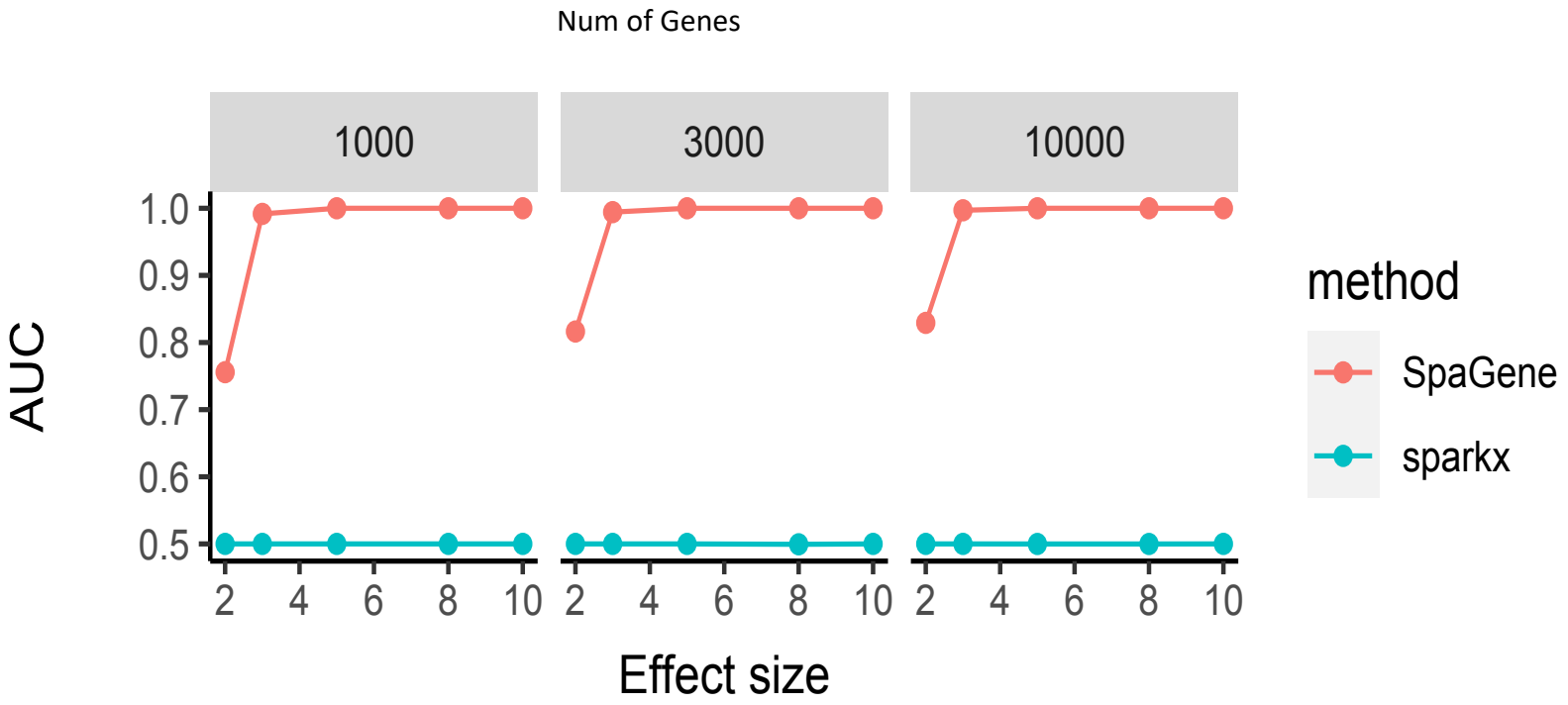
Supplemental Fig. S2. AUC plots of SpaGene (red), SPARK-X (green) and SpatialDE (blue) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a streak pattern, which were generated from negative binomial distributions.



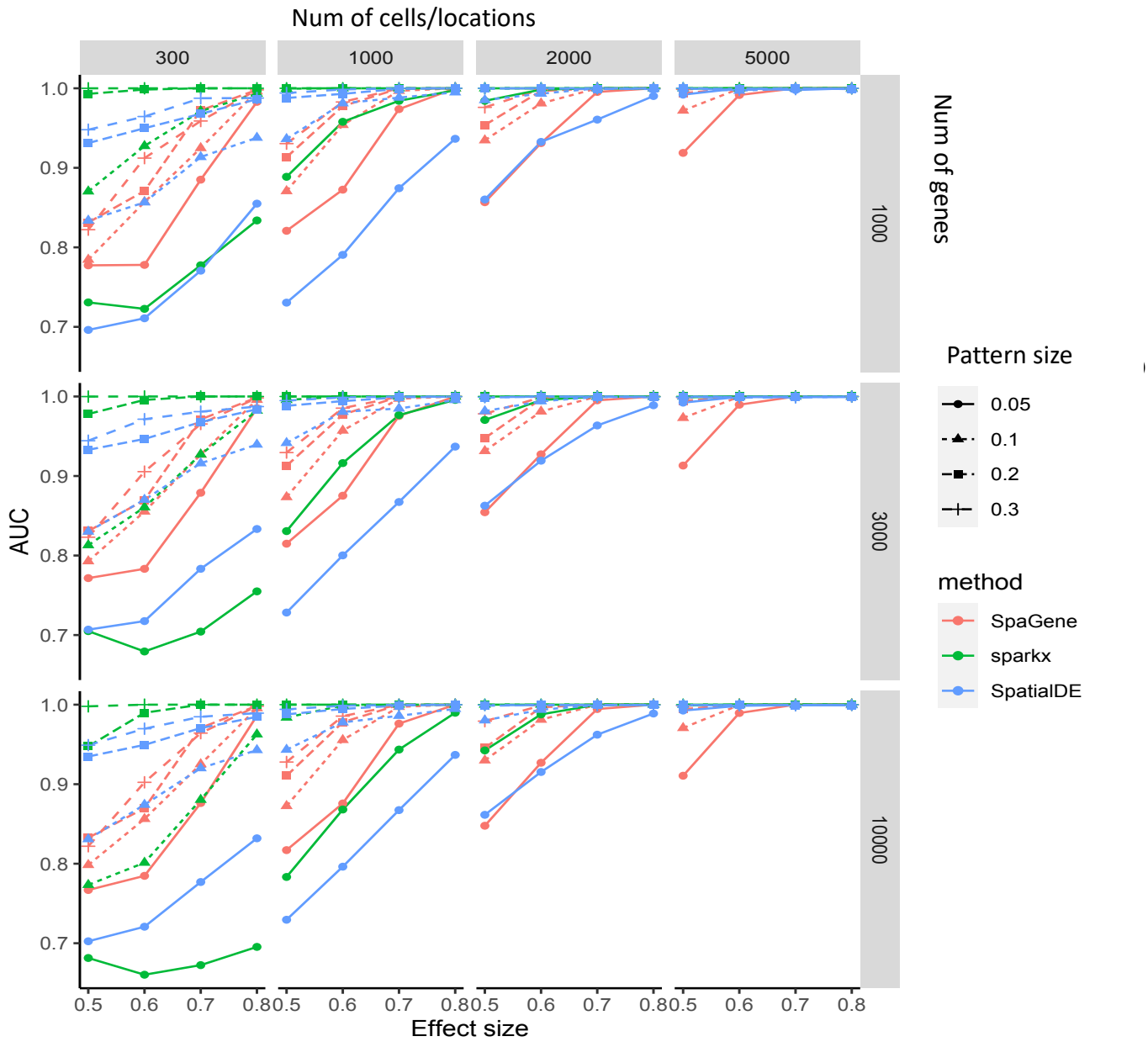
Supplemental Fig. S3. AUC plots of SpaGene (red), SPARK-X (green) and SpatialDE (blue) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a circularity pattern, which were generated from negative binomial distributions.



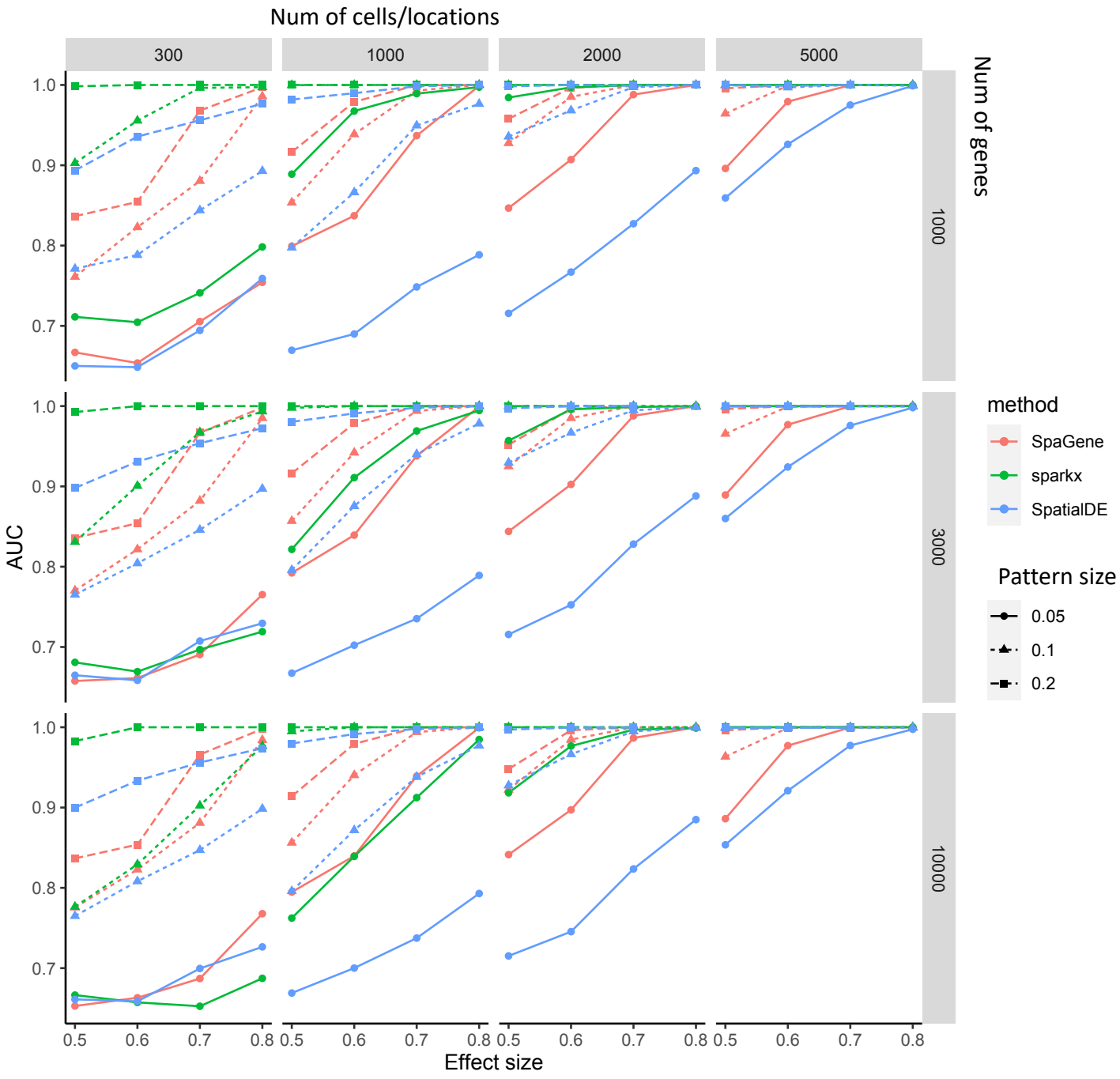
Supplemental Fig. S4. AUC plots of SpaGene (red), SPARK-X (green) and SpatialDE (blue) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a bi-quarter circularity pattern, which were generated from negative binomial distributions.



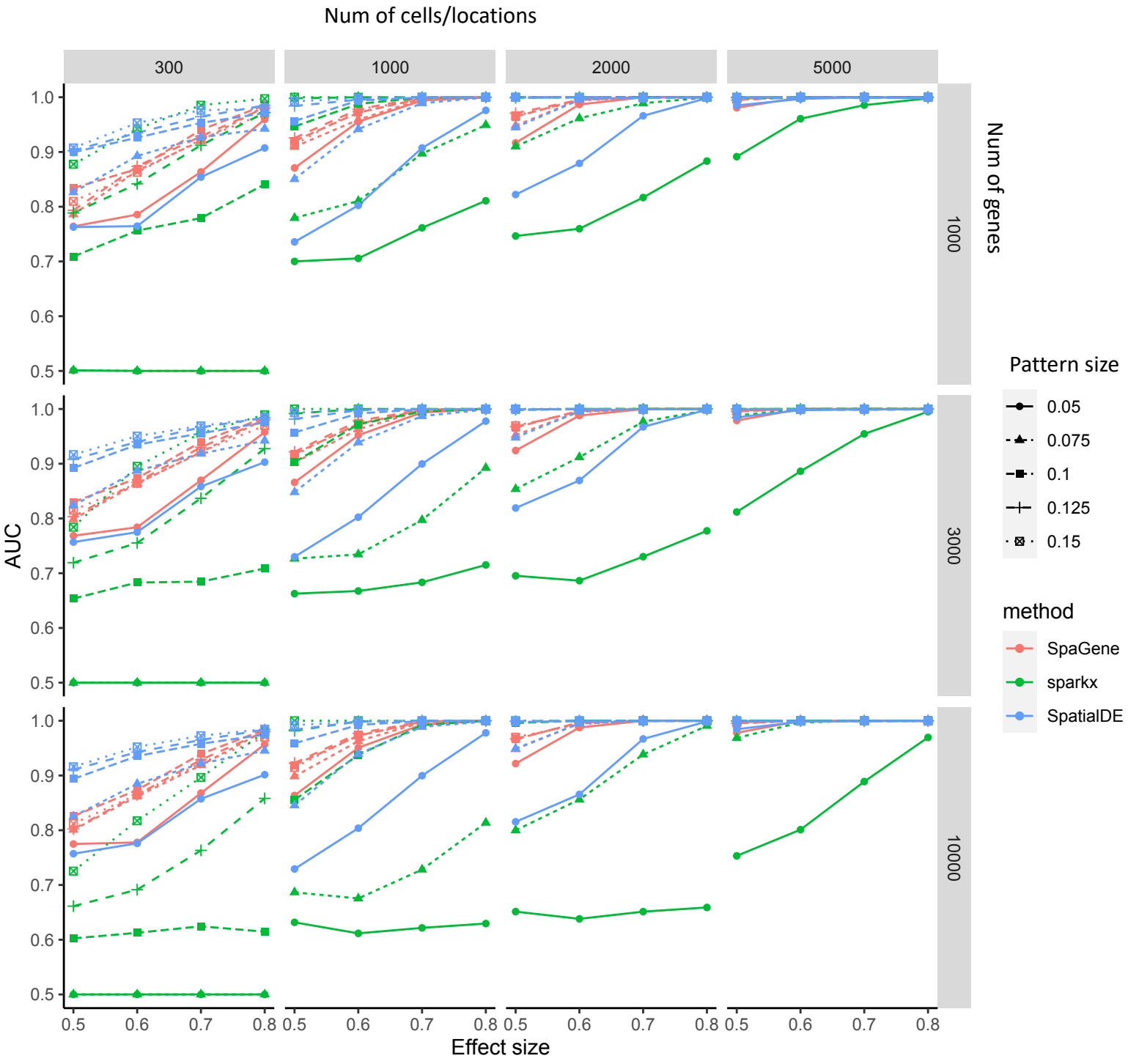
Supplemental Fig. S5. AUC plots of SpaGene (red) and SPARK-X (green) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a purkinje patten, which were generated from negative binomial distributions.



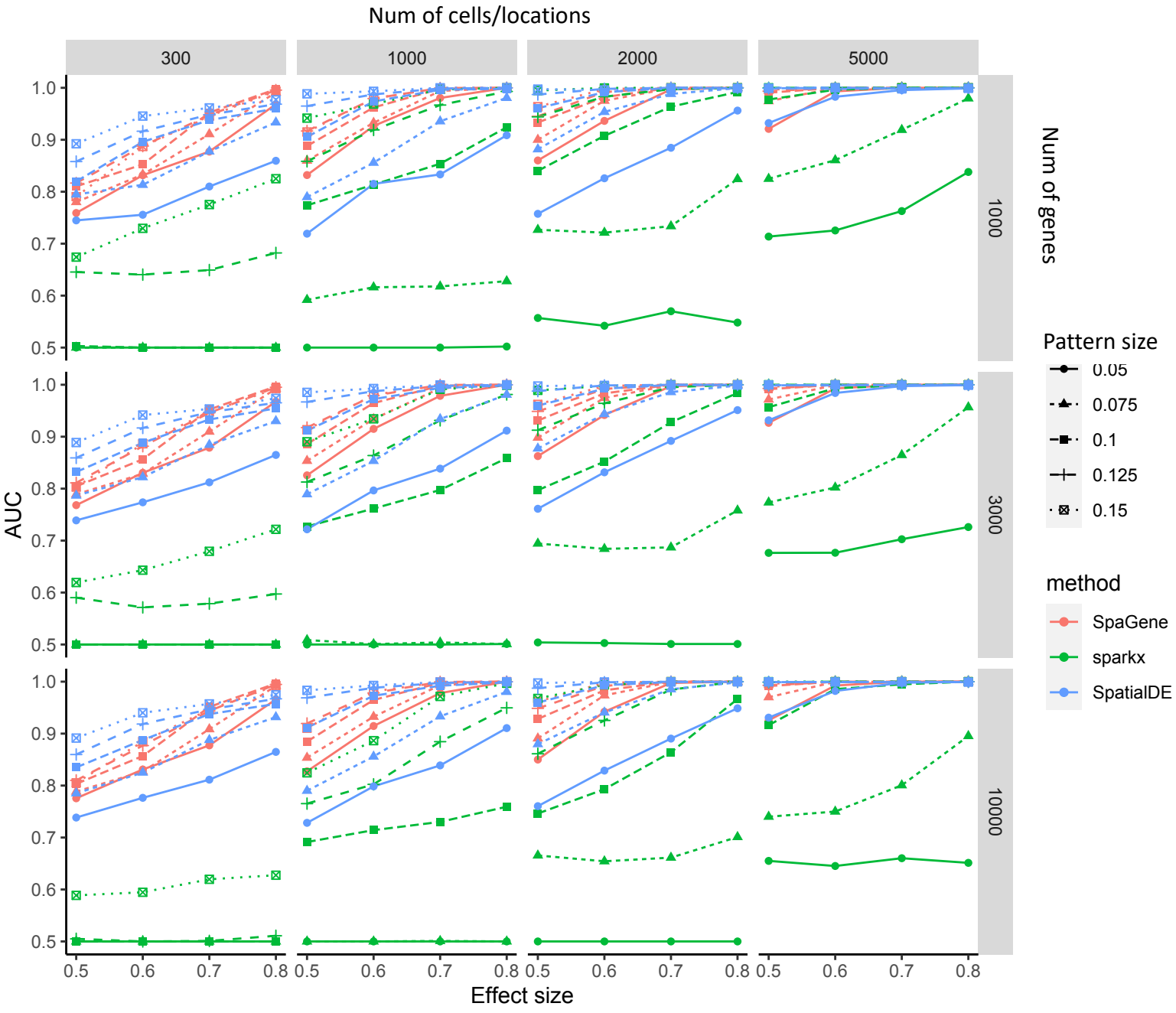
Supplemental Fig. S6 . AUC plots of SpaGene (red), SPARK-X (green) and SpatialDE (blue) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a hotspot pattern, which were generated by resampling real ST MOB data.



Supplemental Fig. S7 . AUC plots of SpaGene (red), SPARK-X (green) and SpatialDE (blue) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a streak pattern, which were generated by resampling real ST MOB data.

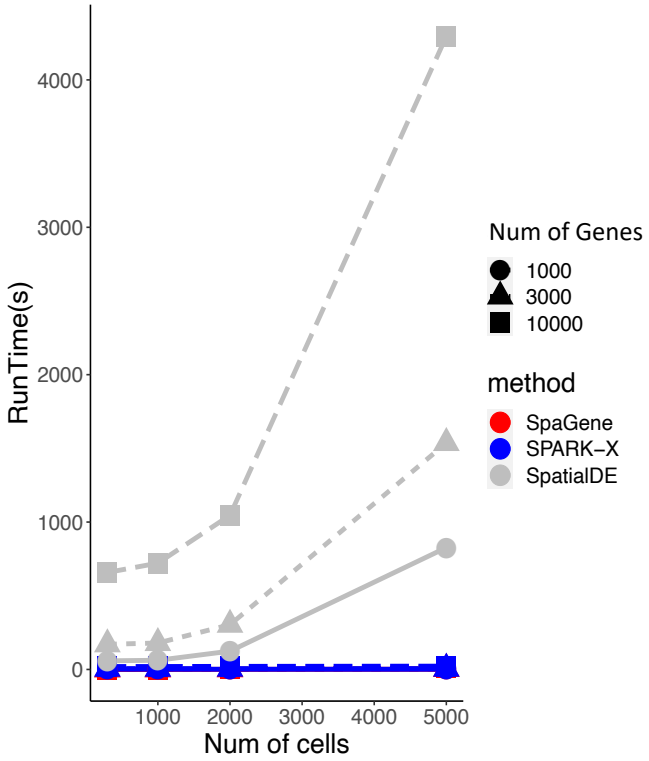


Supplemental Fig. S8 . AUC plots of SpaGene (red), SPARK-X (green) and SpatialDE (blue) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a circularity pattern, which were generated by resampling real ST MOB data.

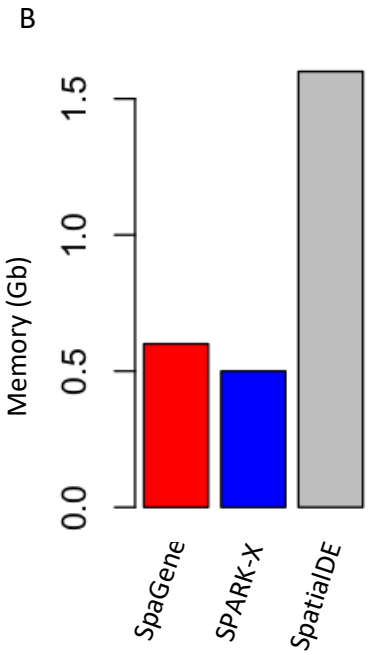


Supplemental Fig. S9 . AUC plots of SpaGene (red), SPARK-X (green) and SpatialDE (blue) in simulated datasets with different effect sizes (x axis) and pattern sizes (point shapes), and varying number of genes and cells/locations. Simulated data have 500 genes with a bi-quarter circularity pattern, which were generated by resampling real ST MOB data.

A



B



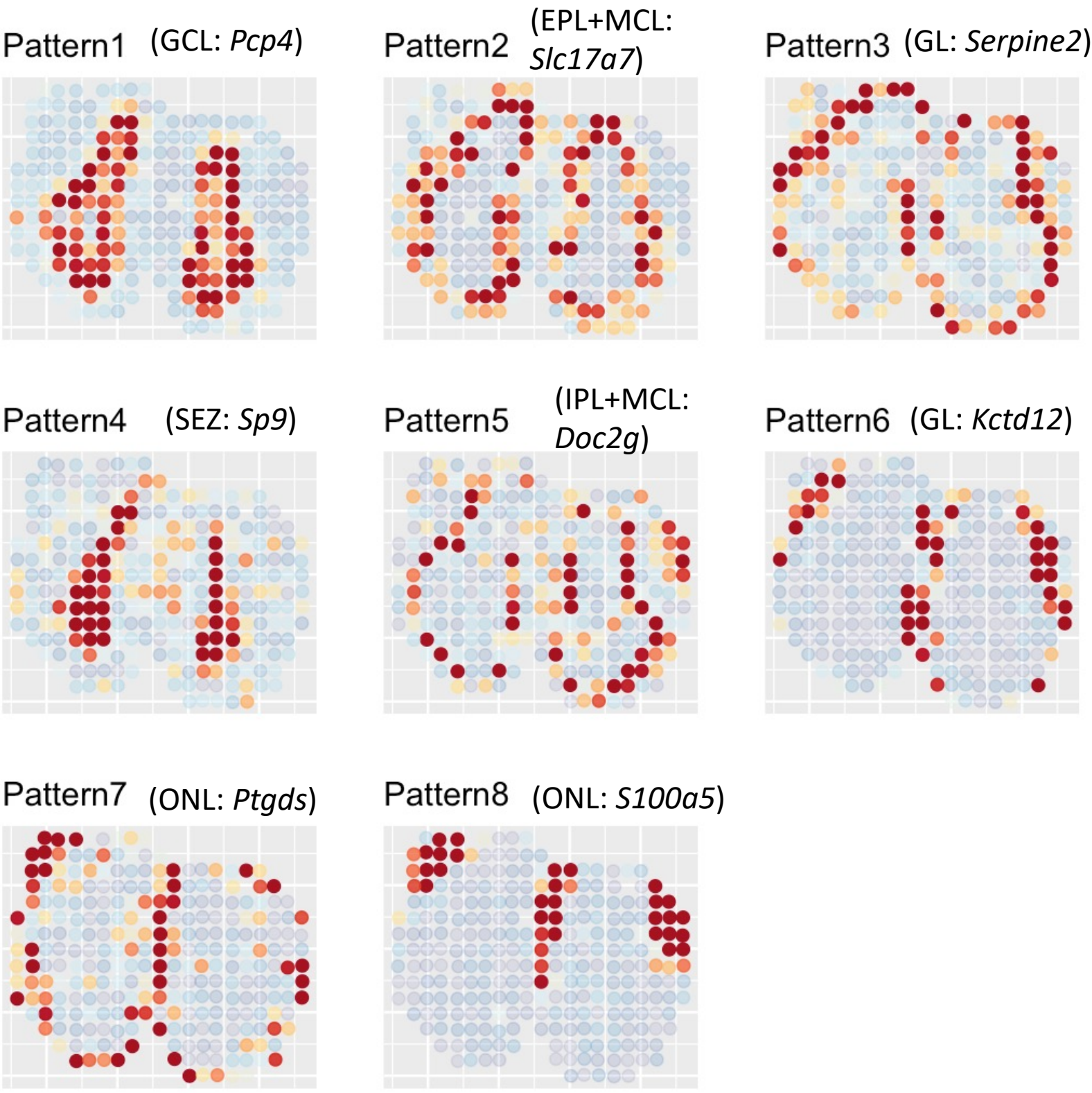
C

Dataset	NumGenes	NumCells	RunTime			Memory (Gb)		
			SpaGene	SPARK-X	SpatialDE	SpaGene	SPARK-X	SpatialDE
ST MOB	16,218	262	2 (s)	34 (s)	928 (s)	0.13	0.13	0.67
MERFISH	161	5,665	0.83 (s)	0.28 (s)	727 (s)	0.04	0.09	0.39
ST BC	14,789	251	1 (s)	26 (s)	877 (s)	0.04	0.09	0.31
10X brain	31,053	2,696	15 (s)	32 (s)	2903 (s)	1.06	0.54	2.85
Cerebellum Slideseq v2	20,141	11,626	0.62 (m)	0.63 (m)	10 (h)#	0.27	0.28	10#
MOB Slideseq v2	21,220	21,724	1.3 (m)	0.64 (m)	40 (h)#	0.48	0.31	20#
HDST	19,950	181,367	3.2 (m)*	0.68 (m)	80 (d) #	2.51*	0.45	200#

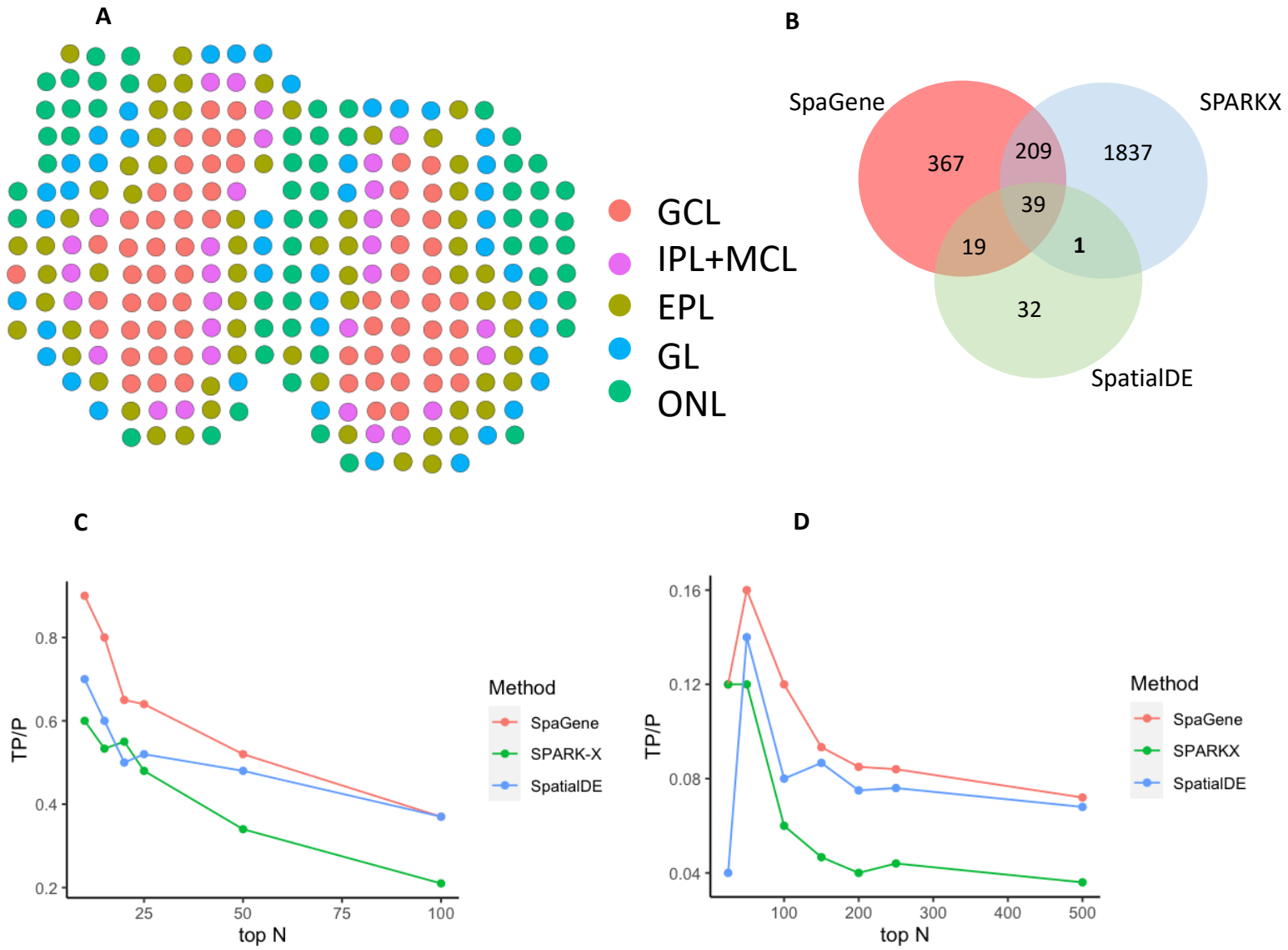
Estimated time and memory;

*: Time and memory estimation on the SpaGene function which adaptively expand NN-search regions

Supplemental Fig. S10. (A) The Runtime of SpaGene (red), SpatialDE (gray) and SPARK-X (blue) on simulated datasets as a function of number of genes and number of locations. (B) The memory usage of SpaGene, SPARK-X and SpatialDE in the simulation dataset with 10,000 genes and 5,000 cells/locations. C) The Runtime and memory usage of SpaGene, SPARK-X and SpatialDE on seven spatial transcriptomics datasets. Computations were carried out using a single thread of an Intel Xeon W-2255 3.7GHz processor with 64G memory.



Supplemental Fig. S11. Eight patterns discovered by SpaGene on ST MOB data, along with the corresponding layer and one representative gene in the parenthesis. SEZ: subependymal zone; GCL: Granular cell layer; IPL: Internal plexiform layer; MCL: Mitral cell layer; EPL: External plexiform layer; ONL: Outer nerve layer; GL: Glomerular layer .

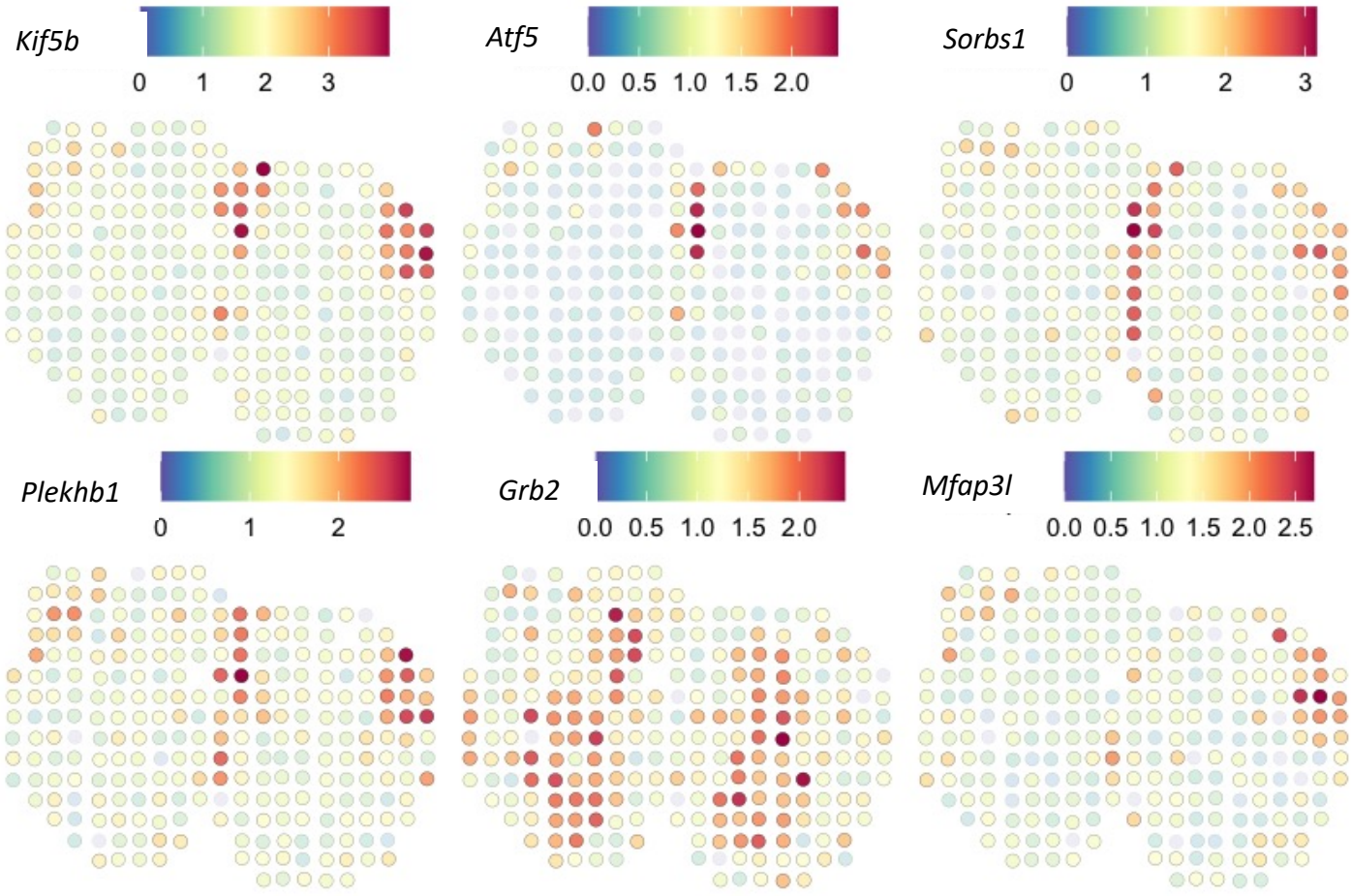


Supplemental Fig. S12. (A) Cell clustering based on transcriptional profiles alone identified five cell types with restricted spatial locations for ST MOB data. B) Overlap between SpaGene, SPARK-X and SpatialDE on ST MOB data. C)True Positive Rates (TPR) in the top N spatially variable genes (N=5,10,15,20,25,50,100) from SpaGene, SPARKX and SpatialDE when cell type specific marker genes were used as true positives. D) TPR in the top N spatially variable genes (N=25,50,100,150,200,250,500) from SpaGene, SPARKX and SpatialDE when differential proteins from five layers in Allen mouse brain atlas were used as true positives.

SpaGene: 7e-35
SPARKX: 0.1
SpatialDE: 3e-6

SpaGene: 8e-31
SPARKX: 0.2
SpatialDE: 0.5

SpaGene: 1e-28
SPARKX: 0.6
SpatialDE: 0.02



SpaGene: 2e-25
SPARKX: 1
SpatialDE: 0.2

SpaGene: 2e-23
SPARKX: 7e-8
SpatialDE: 0.3

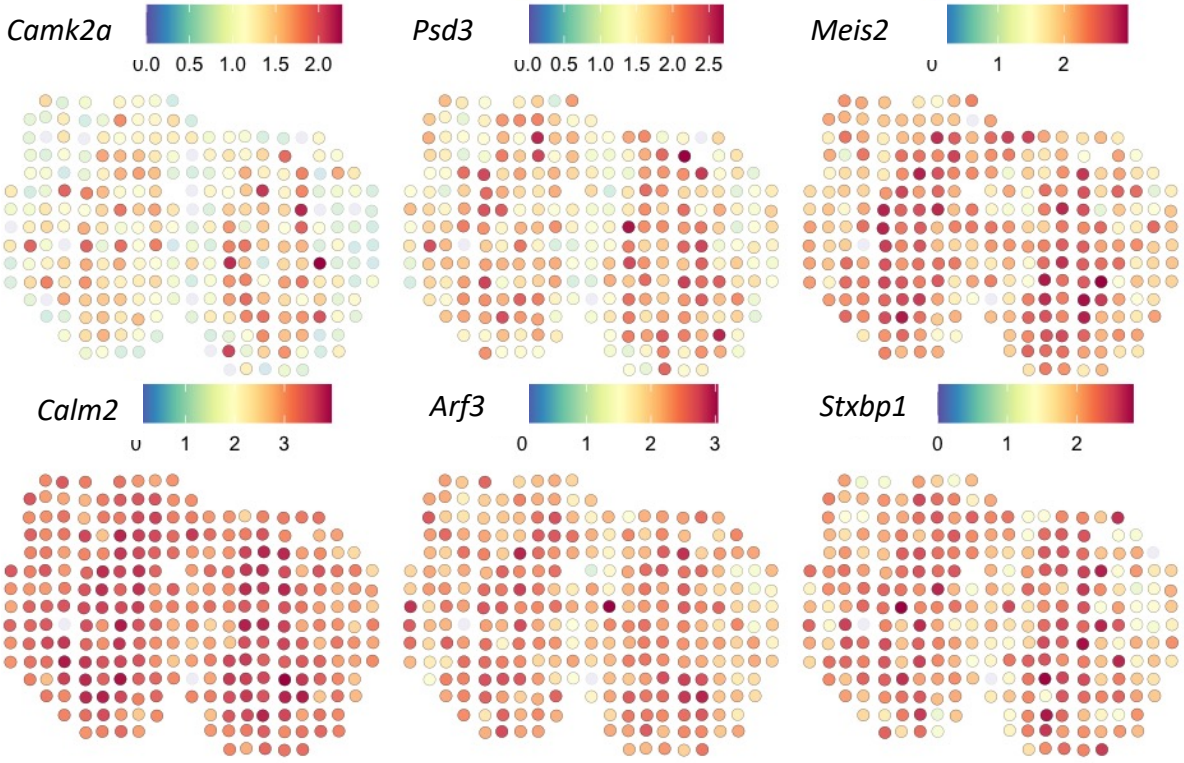
SpaGene: 1e-21
SPARKX: 1
SpatialDE: 0.4

Supplemental Fig. S13. Visualization of spatially variable genes identified to be very significant by SpaGene, but insignificant either by SPARK-X or SpatialDE .

SpaGene: 7e-4
SPARKX: 3e-10
SpatialDE: 0.1

SpaGene: 2e-3
SPARKX: 3e-10
SpatialDE: 0.2

SpaGene: 1e-13
SPARKX: 2e-9
SpatialDE: 0.2

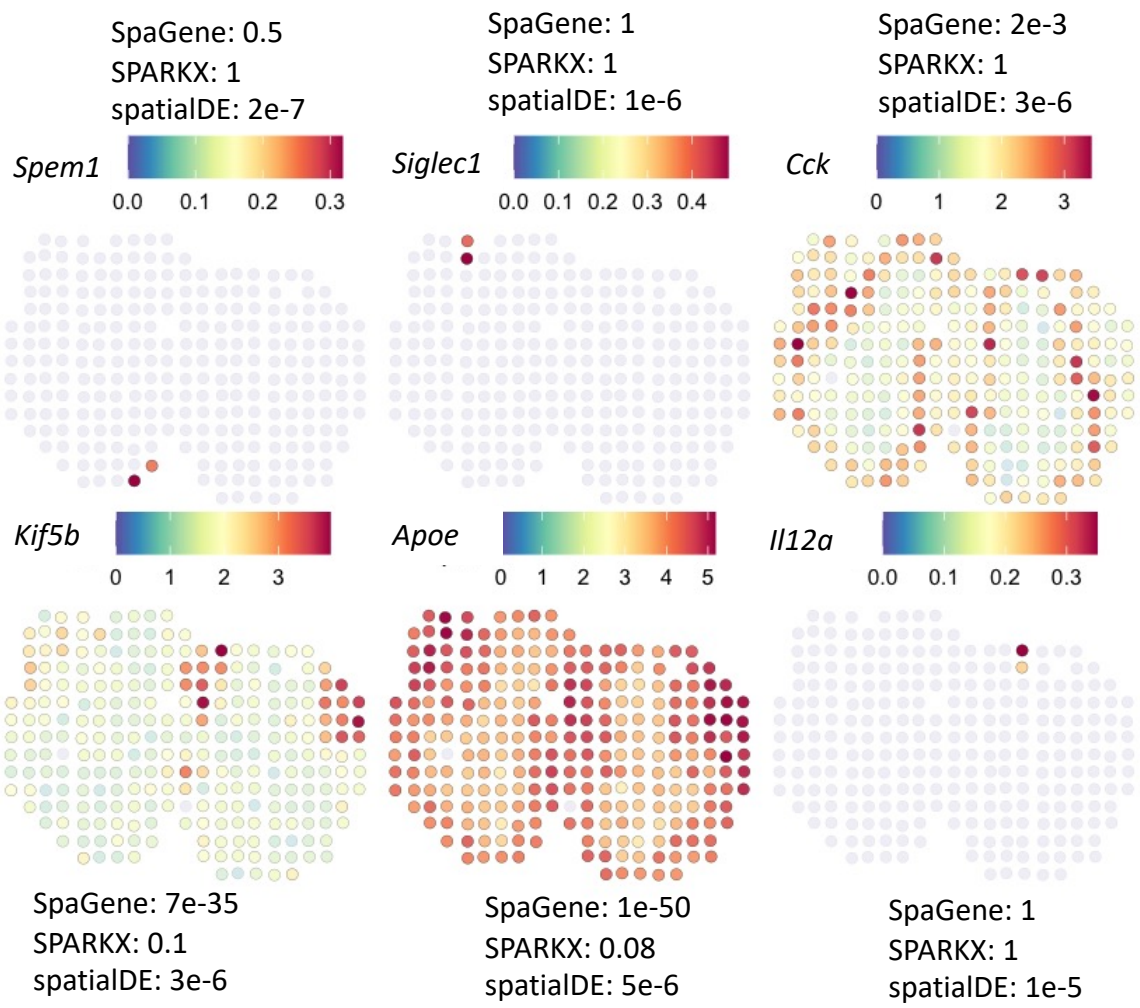


SpaGene: 1e-17
SPARKX: 3e-9
SpatialDE: 0.1

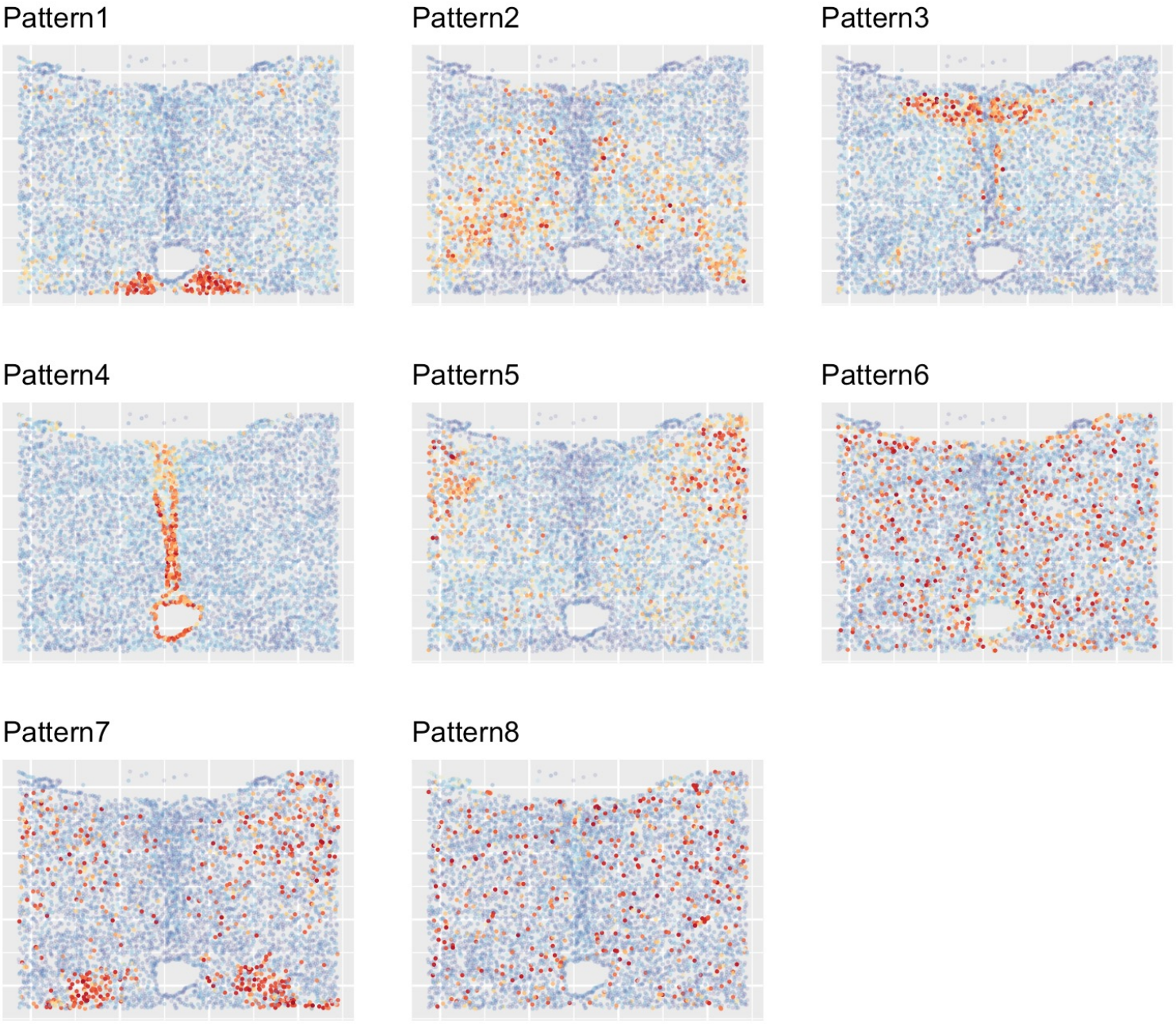
SpaGene: 1e-4
SPARKX: 3e-9
SpatialDE: 0.4

SpaGene: 1e-2
SPARKX: 3e-9
SpatialDE: 1

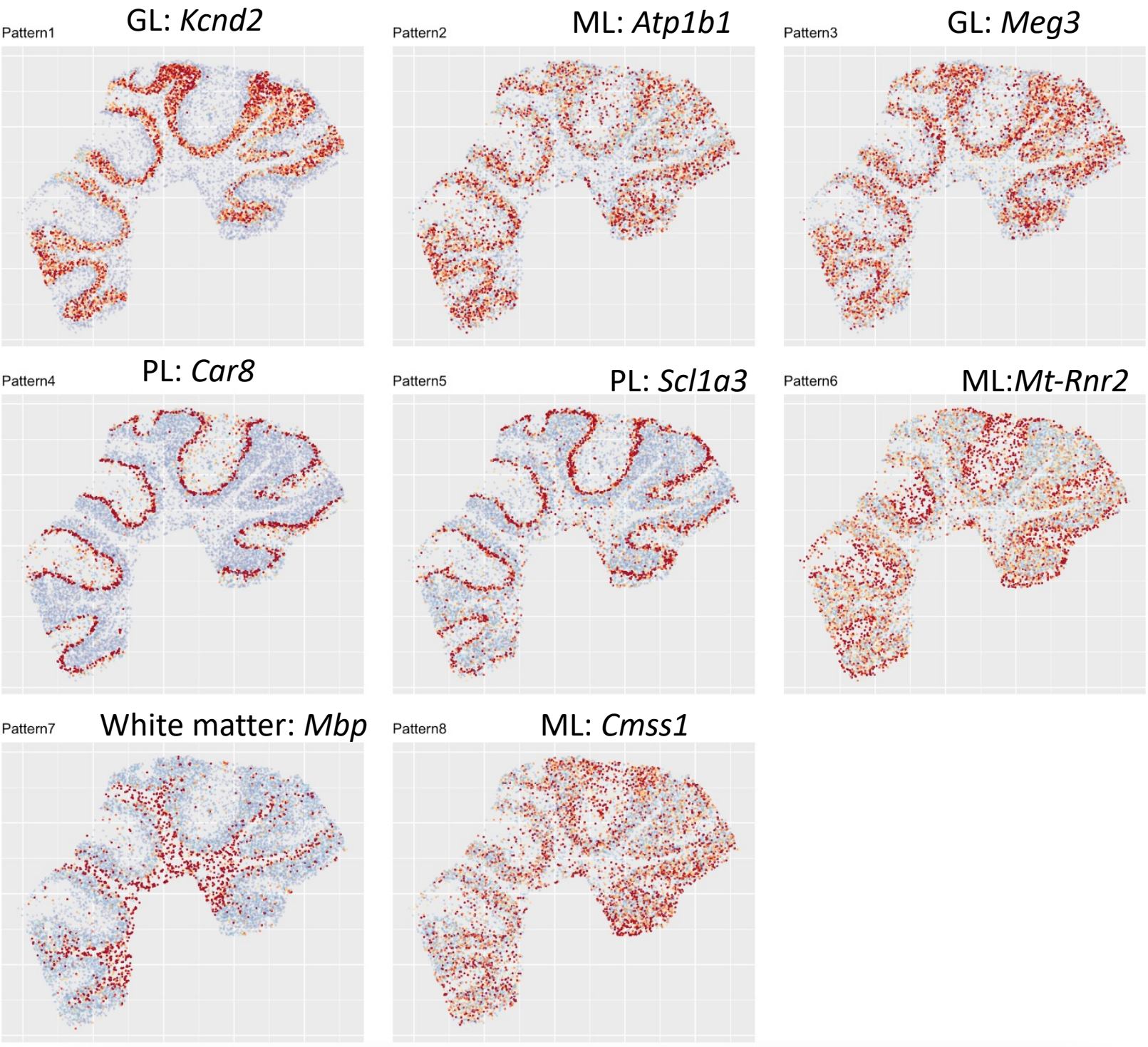
Supplemental Fig. S14. Visualization of spatially variable genes identified to be very significant by SPARK-X, but insignificant either by SpaGene or SpatialDE .



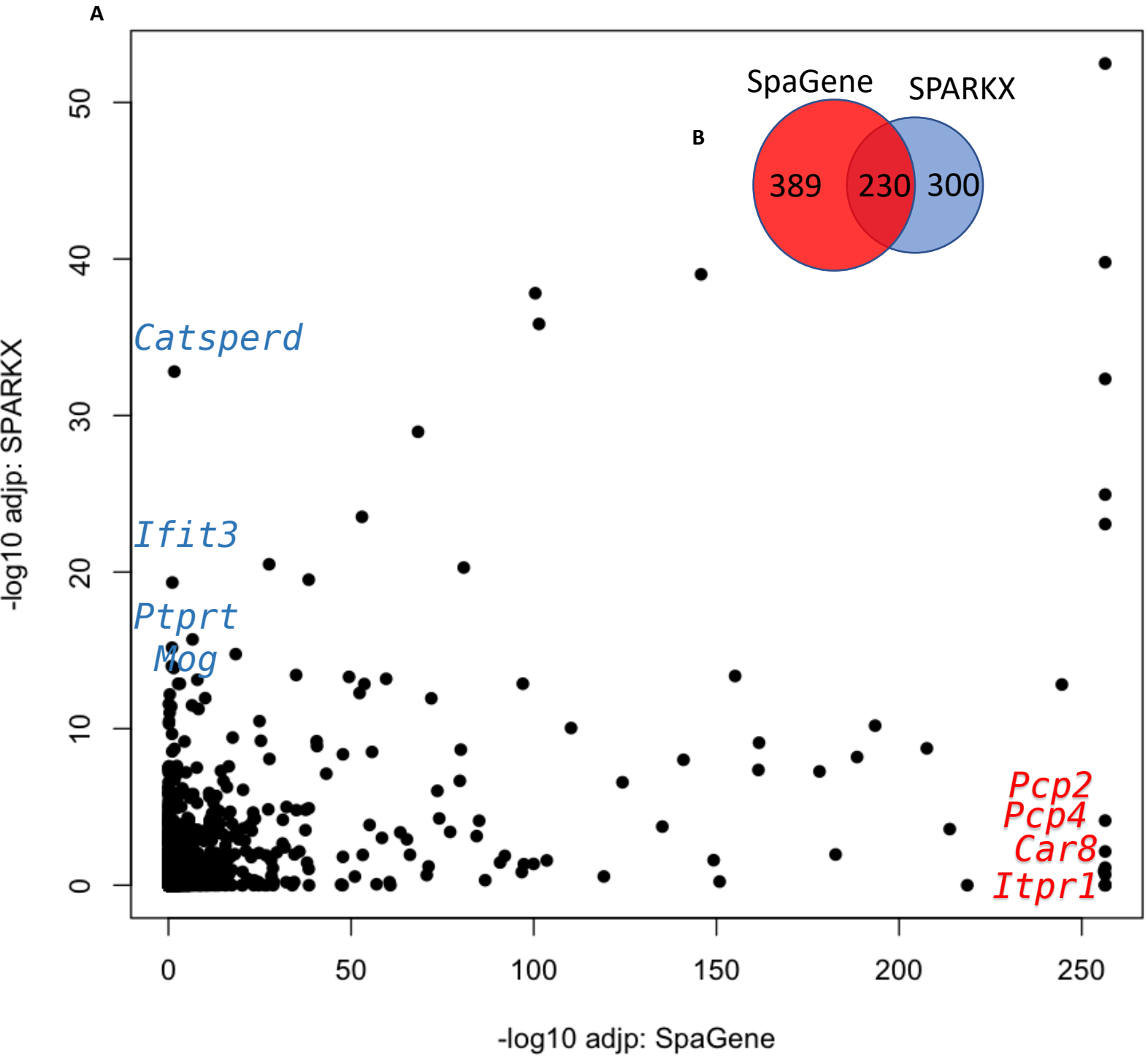
Supplemental Fig. S15. Visualization of spatially variable genes identified to be very significant by SpatialDE, but insignificant either by SPARK-X or SpaGene .



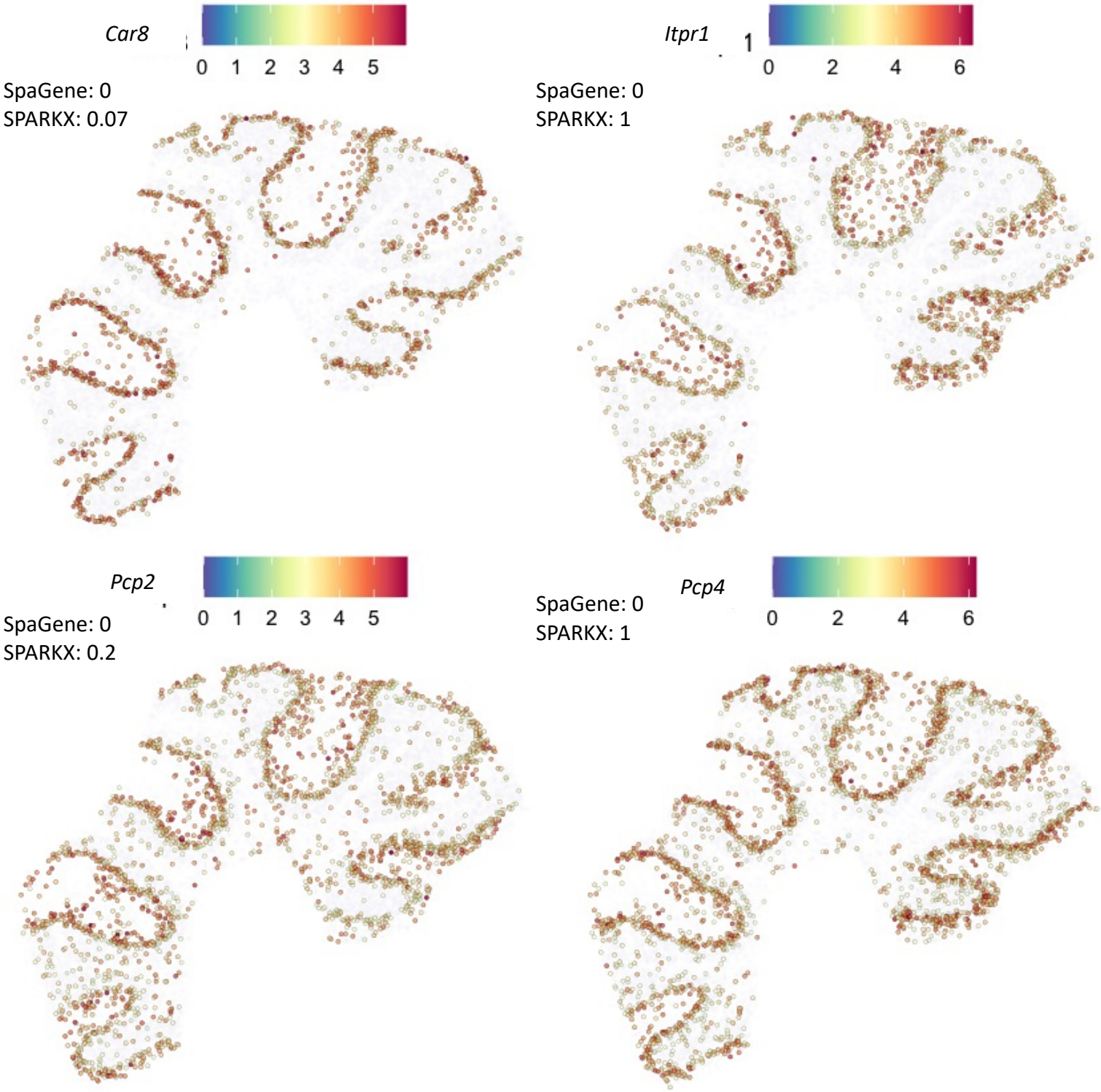
Supplemental Fig. S16. Eight patterns discovered by SpaGene on MERFISH data.



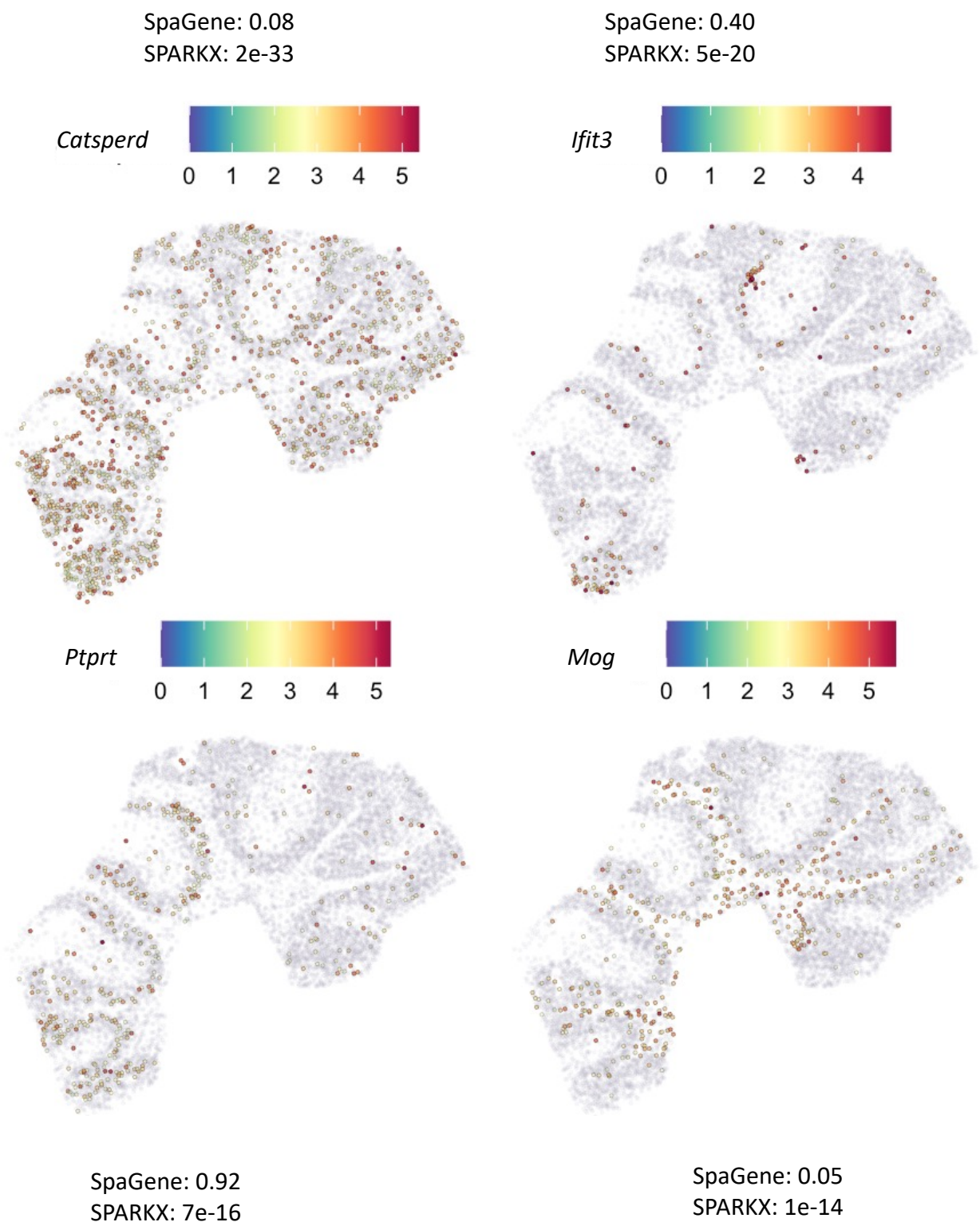
Supplemental Fig. S17. Eight patterns discovered by SpaGene on Slideseq V2 mouse cerebellum data, along with the corresponding layer and one representative gene. GL: Granular layer; ML: Molecular layer; PL: Purkinje layer



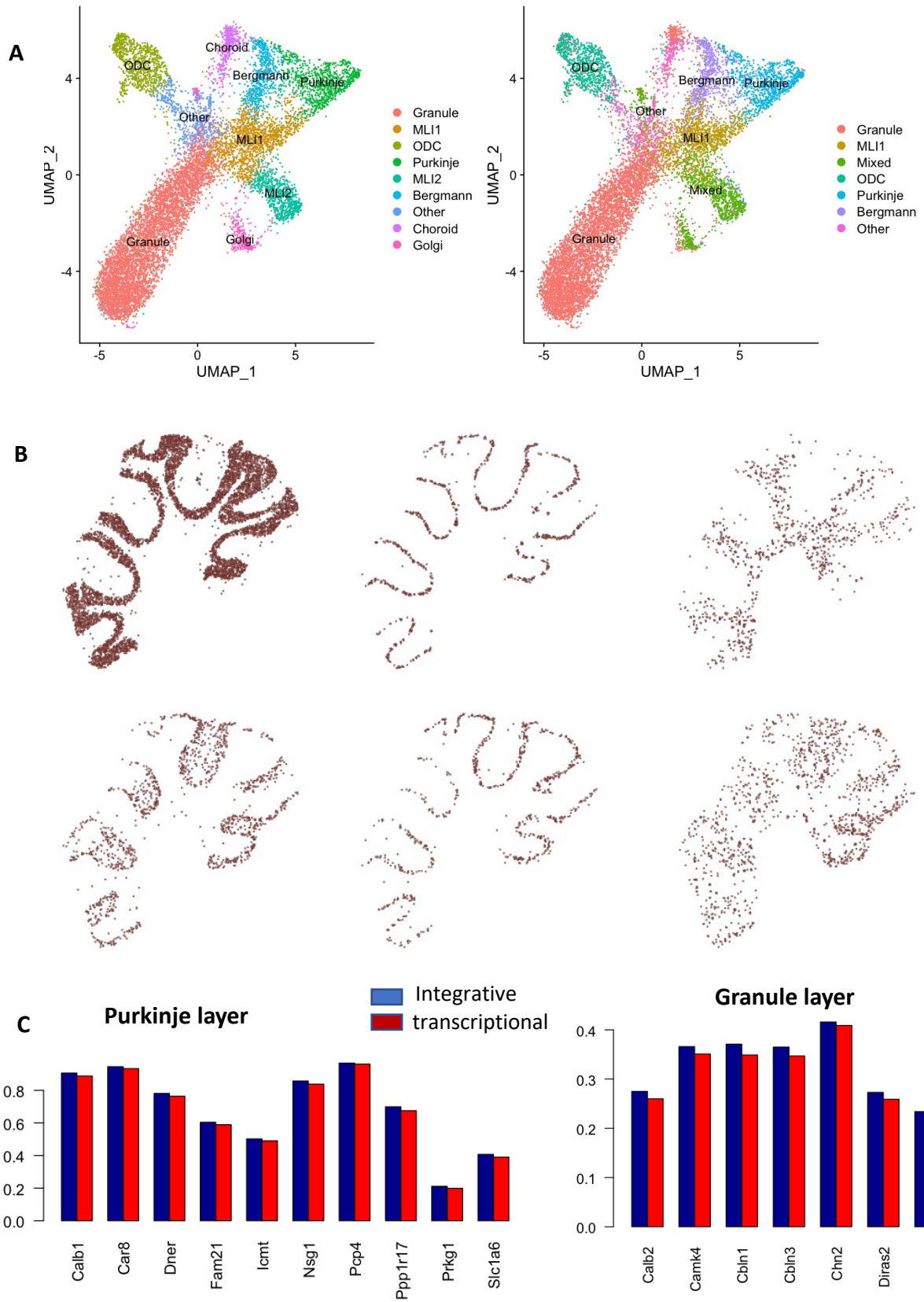
Supplemental Fig. S18. A) Scatter plot of adjusted pvalues between SpaGene and SPARK-X. Those genes identified very significant by one method but not detected by the other method were highlighted in red. For example, *Catsperd*, *Ifit3, Mog* and *Ptprt* were identified very significant by SPARK-X but not by SpaGene, while *Car8*, *Itpr1, Pcp2* and *Pcp4* are identified very significant in SpaGene but not by SPARK-X. B) Overlap between SpaGene and SPARK-X.



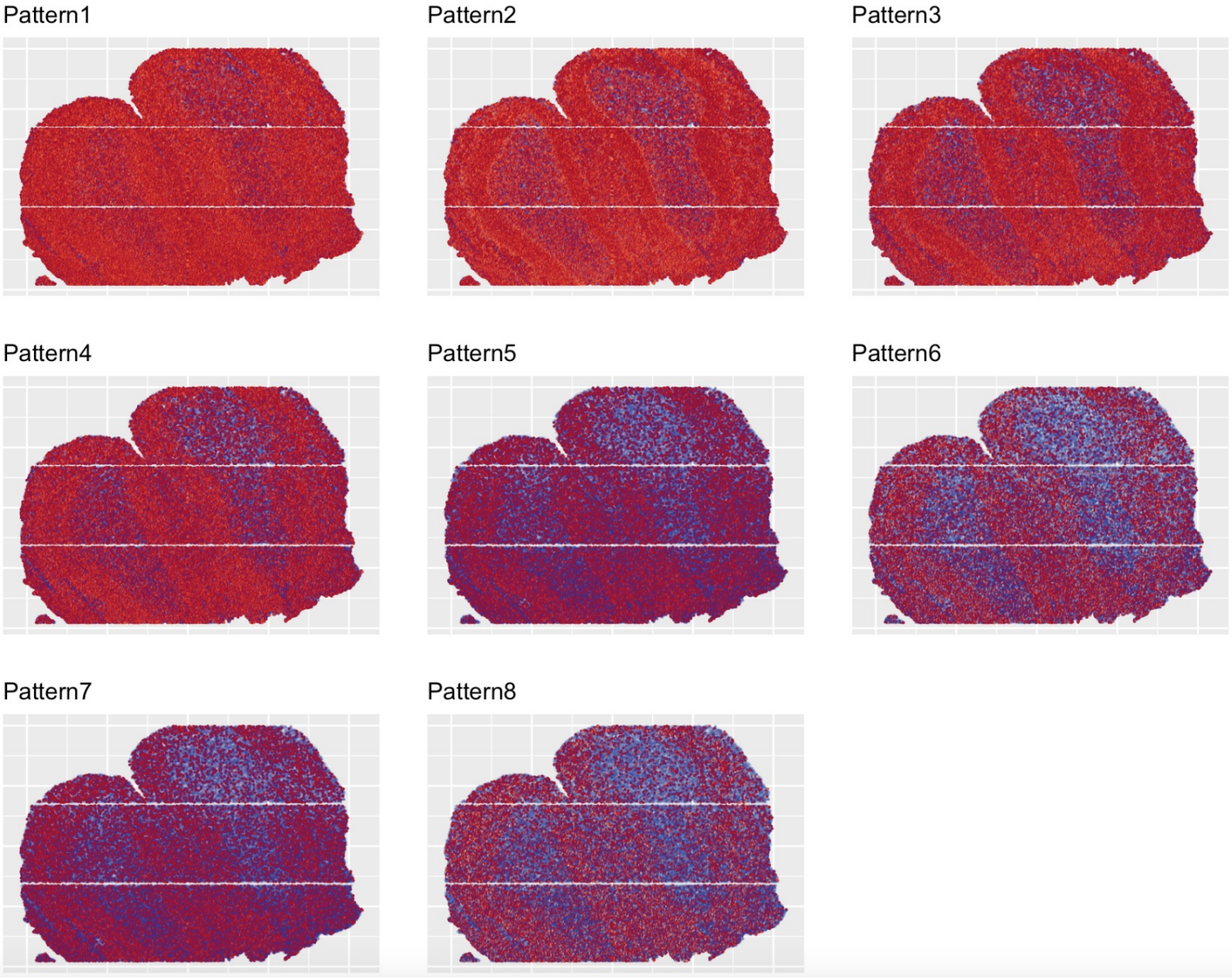
Supplemental Fig. S19. Visualization of spatially variable genes identified to be very significant by SpaGene, but insignificant by SPARK-X.



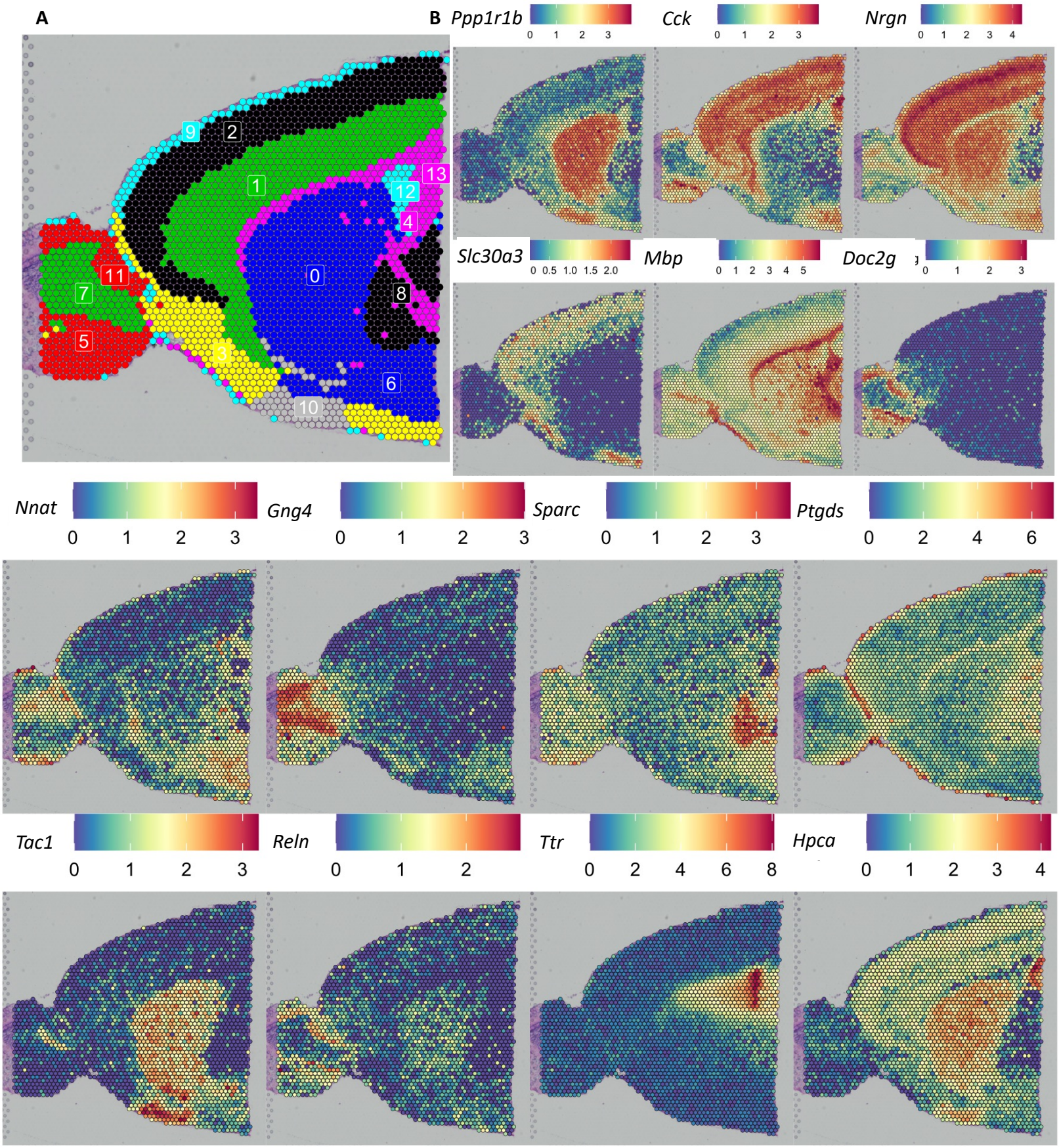
Supplemental Fig. S20. Visualization of spatially variable genes identified to be very significant by SPARK-X, but insignificant by SpaGene.



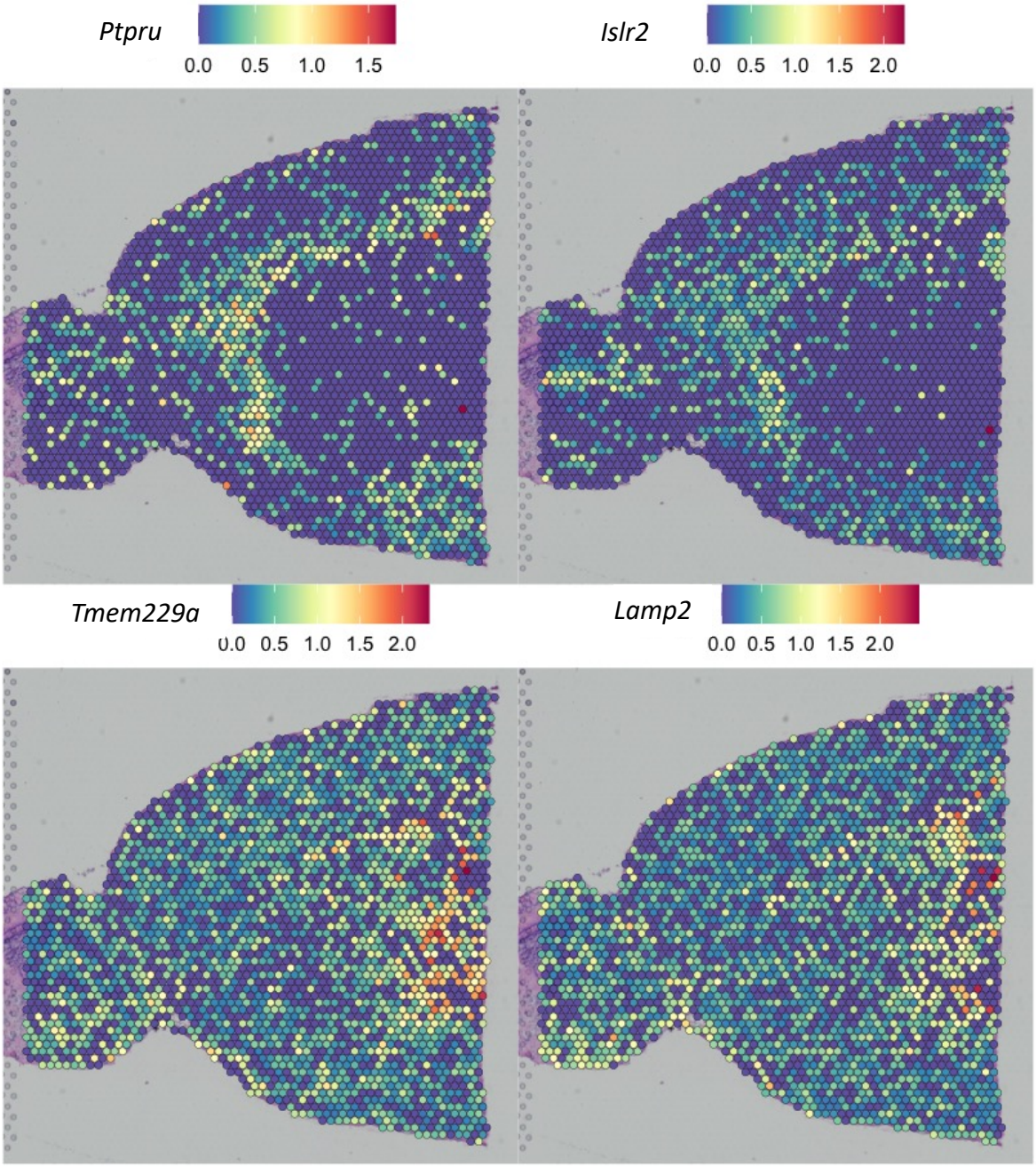
Supplemental Fig. S21. (A) Cell clustering based on the top 2000 transcriptionally variable genes (left) and the 619 spatially variable genes (right). (B) six cell clusters based on the 619 spatially variable genes are specifically located in different layers and white matter. (C) Cell clustering based on the 2000 integrative features has a higher percentage of spots expressing those marker genes than the one based on the top 2000 transcriptionally variable features.



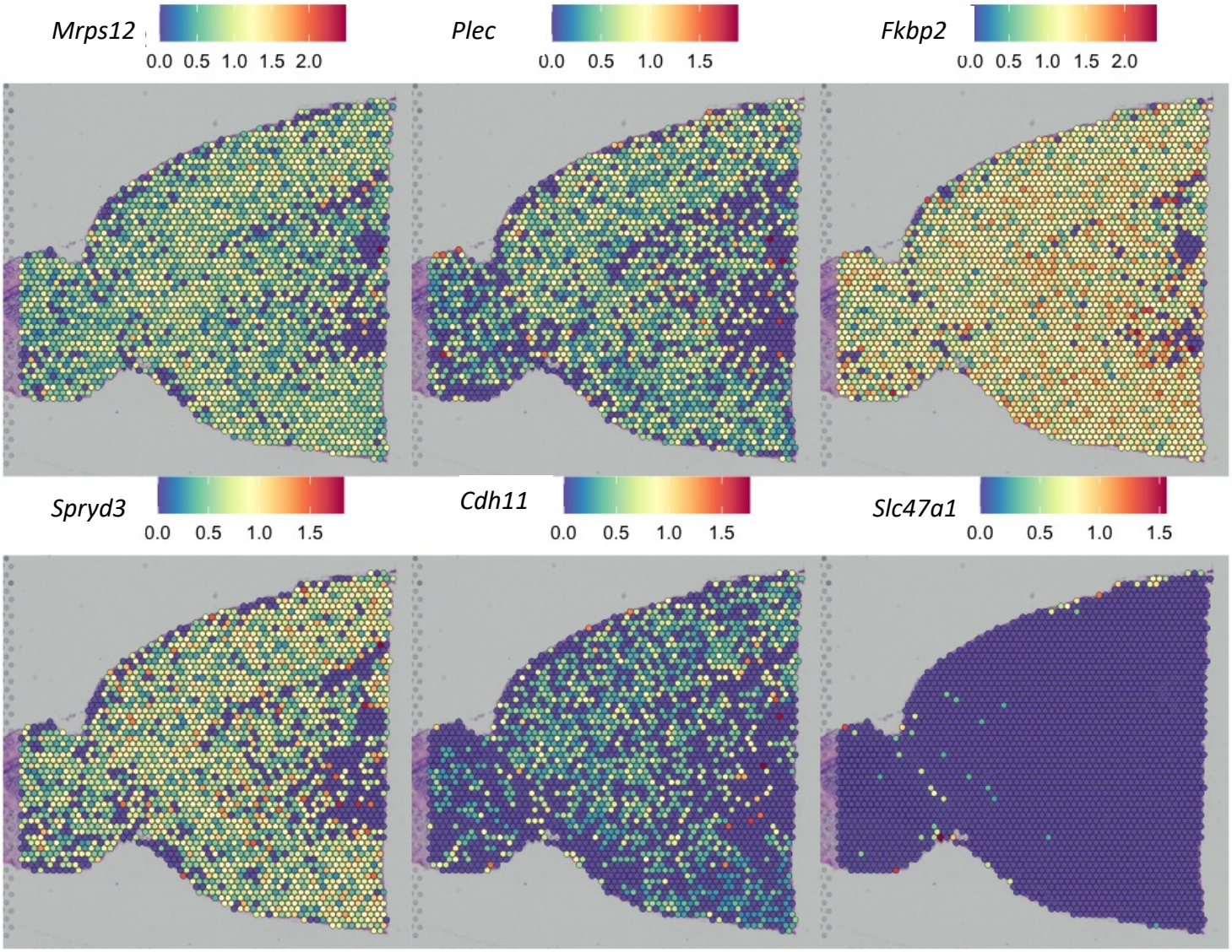
Supplemental Fig. S22. Eight patterns discovered by SpaGene on HDST data.



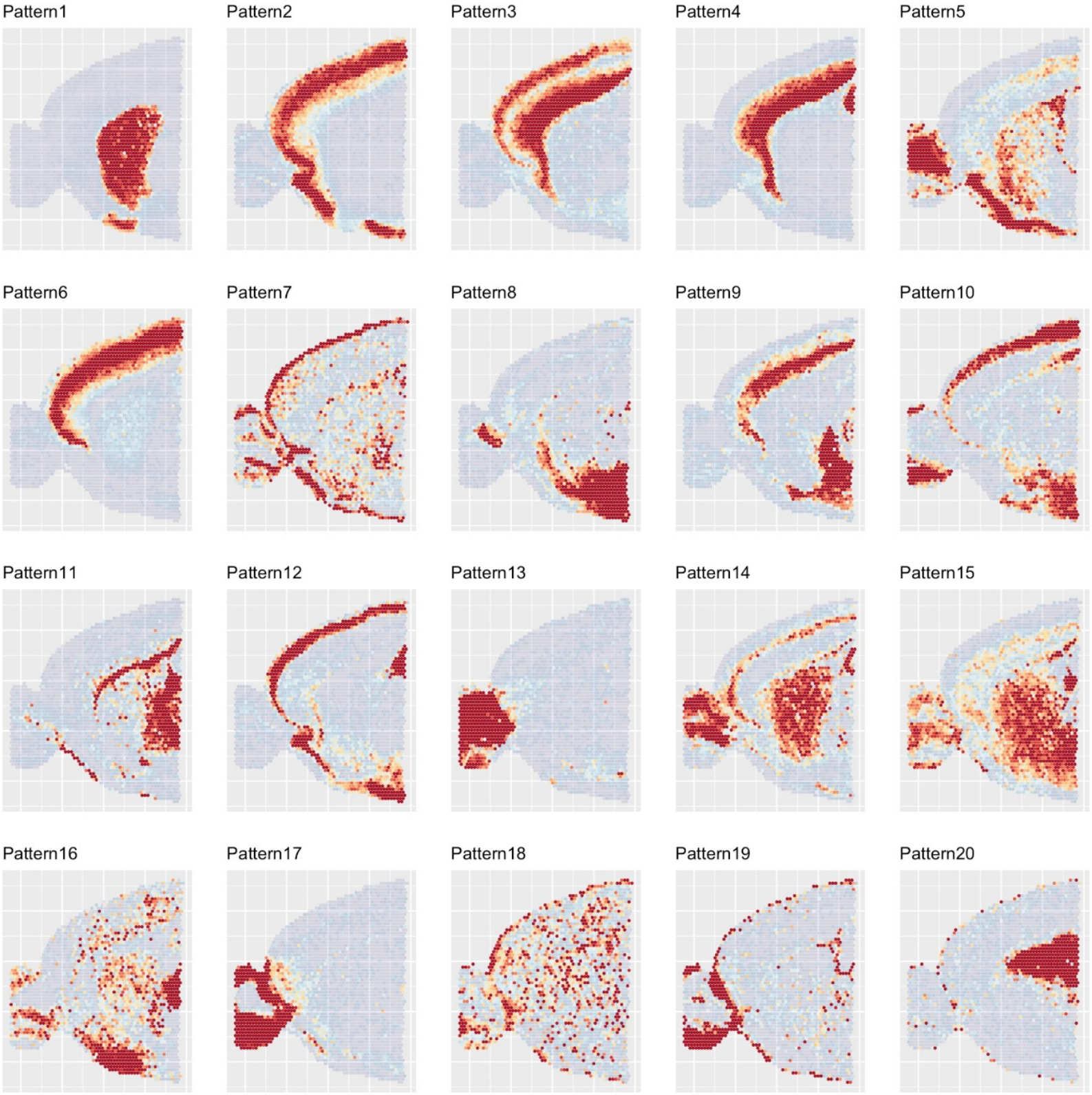
Supplemental Fig. S23. (A) Spatially unaware cell clustering on 10X mouse brain data (B) Visualization of spatially variable genes identified by SpaGene.



Supplemental Fig. S24. Visualization of spatially variable genes identified to be very significant by SpaGene, but insignificant by SPARK-X or SpatialDE.

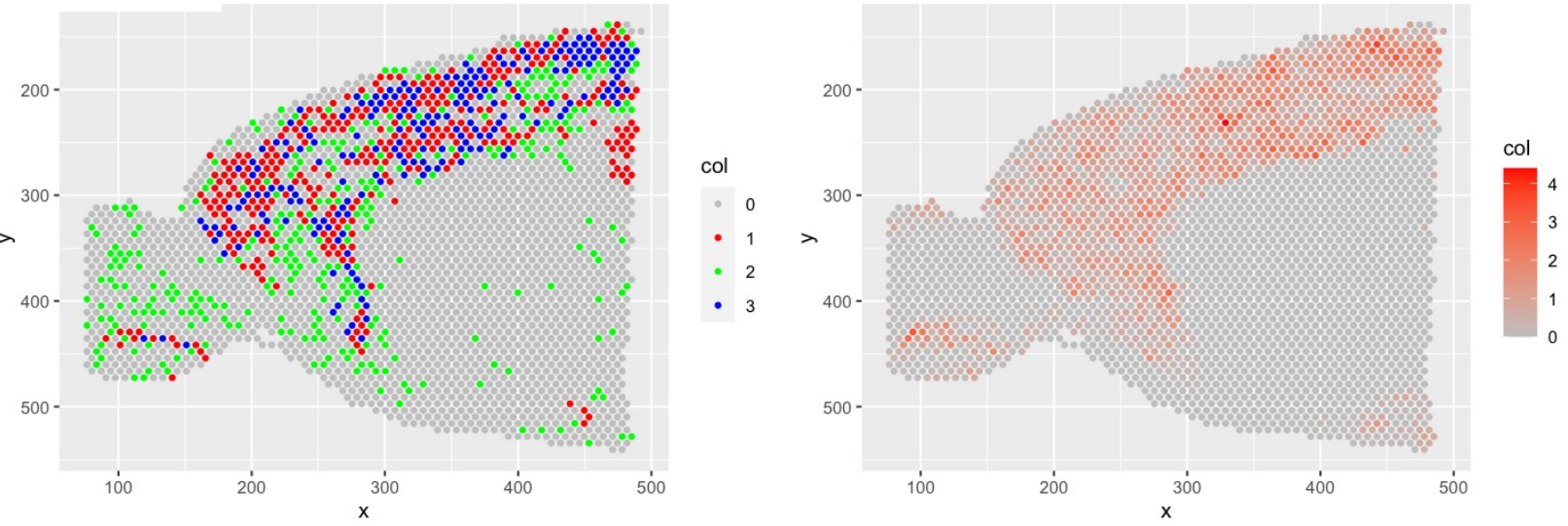


Supplemental Fig. S25. Visualization of spatially variable genes identified to be very significant by SPARK-X, but insignificant by SpaGene.

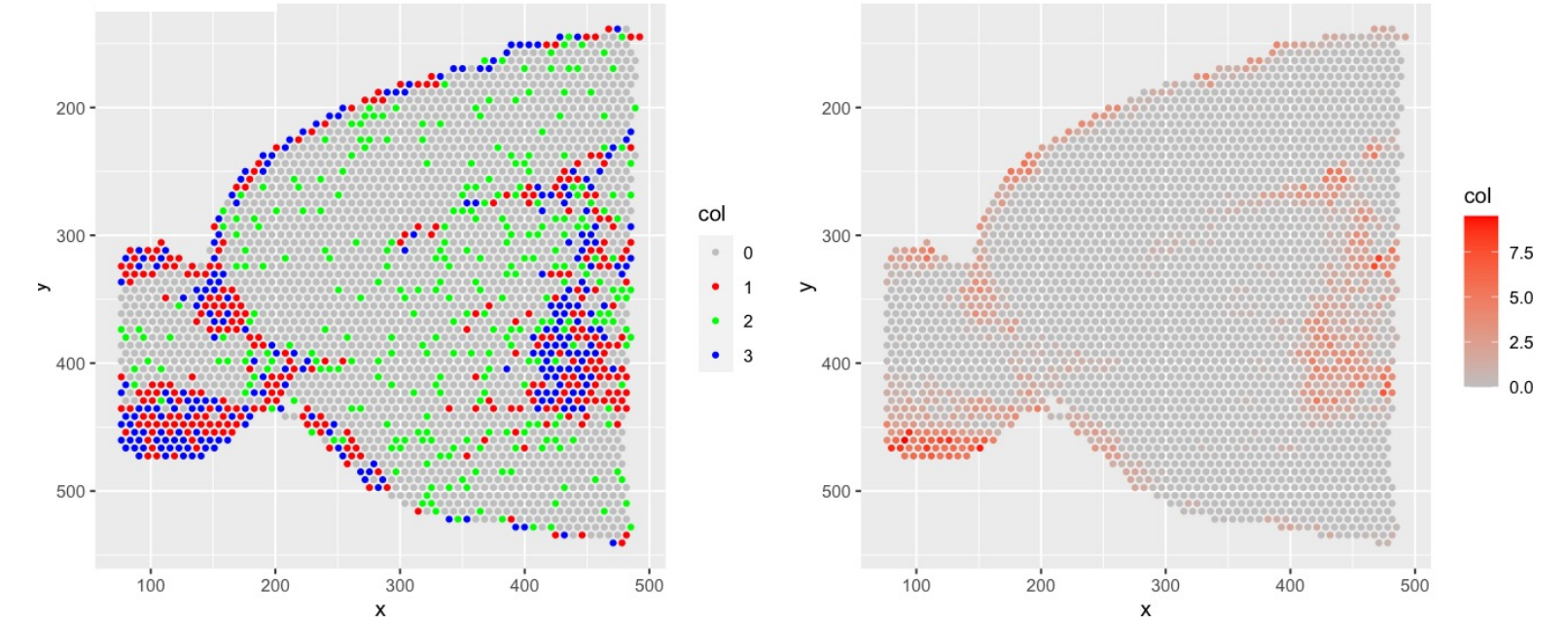


Supplemental Fig. S26. Twenty patterns discovered by SpaGene on 10x brain data.

Cck-Cckbr adjp:6e-222

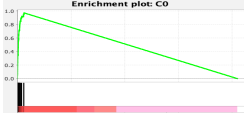


Apoe-Sdc4 adjp:3e-194



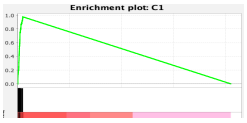
Supplemental Fig. S27. Top significant ligand-receptor interactions identified by SpaGene on 10x mouse brain data.

ES=0.99
NES=8.26

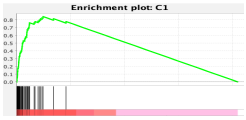


ES=0.97
NES=8.09

ES=0.98
NES=8.28



ES=0.98
NES=8.04

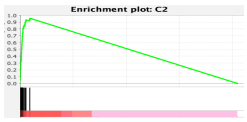


ES=0.84
NES=7.12

ES=0.98
NES=8.13

ES=0.95
NES=7.90

ES=0.91
NES=7.61



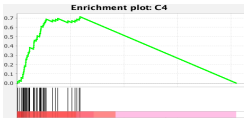
ES=0.96
NES=8.04

ES=0.93
NES=7.88

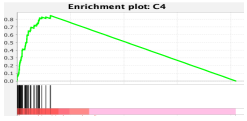
ES=0.78
NES=6.53

ES=0.91
NES=7.54

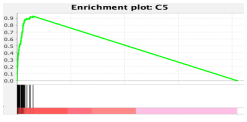
ES=0.96
NES=7.95



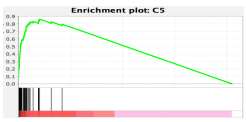
ES=0.71
NES=5.95



ES=0.85
NES=7.04

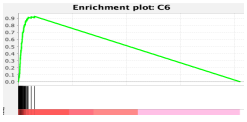


ES=0.93
NES=7.62



ES=0.86
NES=7.36

ES=0.91
NES=7.89



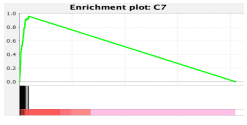
ES=0.93
NES=7.76

ES=0.89
NES=7.34

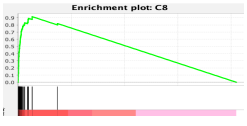
ES=0.93
NES=7.64

ES=0.97
NES=8.26

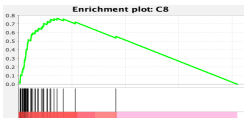
ES=0.97
NES=7.91



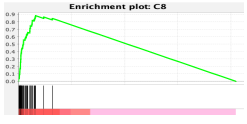
ES=0.96
NES=8.19



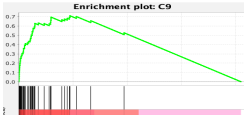
ES=0.92
NES=7.69



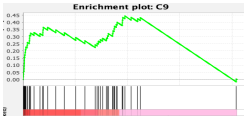
ES=0.77
NES=6.37



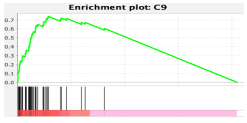
ES=0.88
NES=7.28



ES=0.71
NES=5.89



ES=0.44
NES=3.77



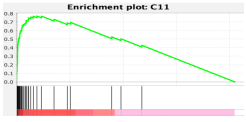
ES=0.74
NES=6.02

ES=0.92
NES=7.85

ES=0.88
NES=7.28

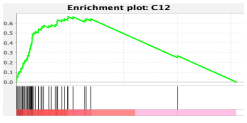
ES=0.92
NES=7.81

ES=0.85
NES=7.22

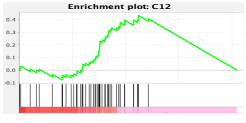


ES=0.77
NES=6.43

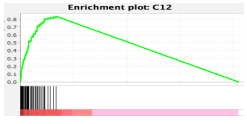
ES=0.94
NES=8.00



ES=0.67
NES=5.64



ES=0.43
NES=3.63



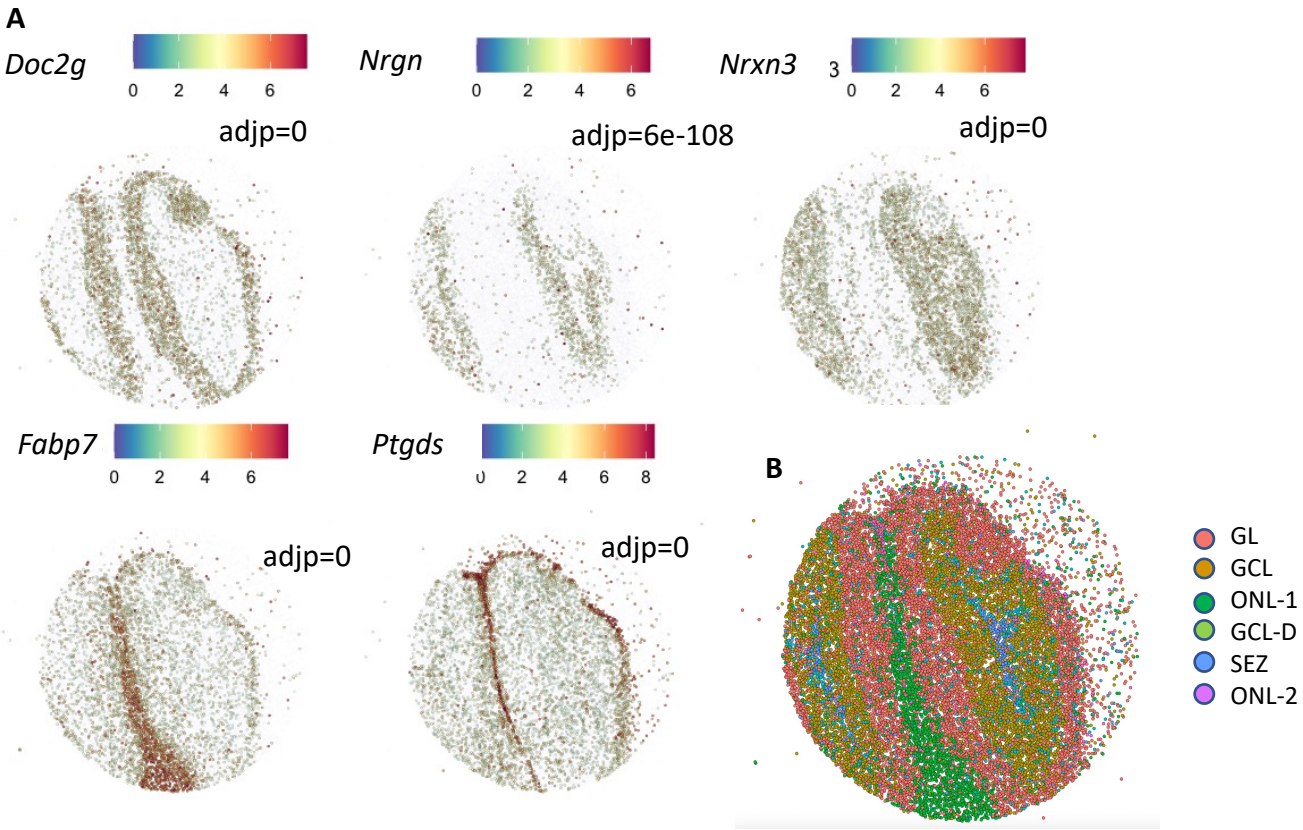
ES=0.84
NES=6.93

ES=0.82
NES=6.89

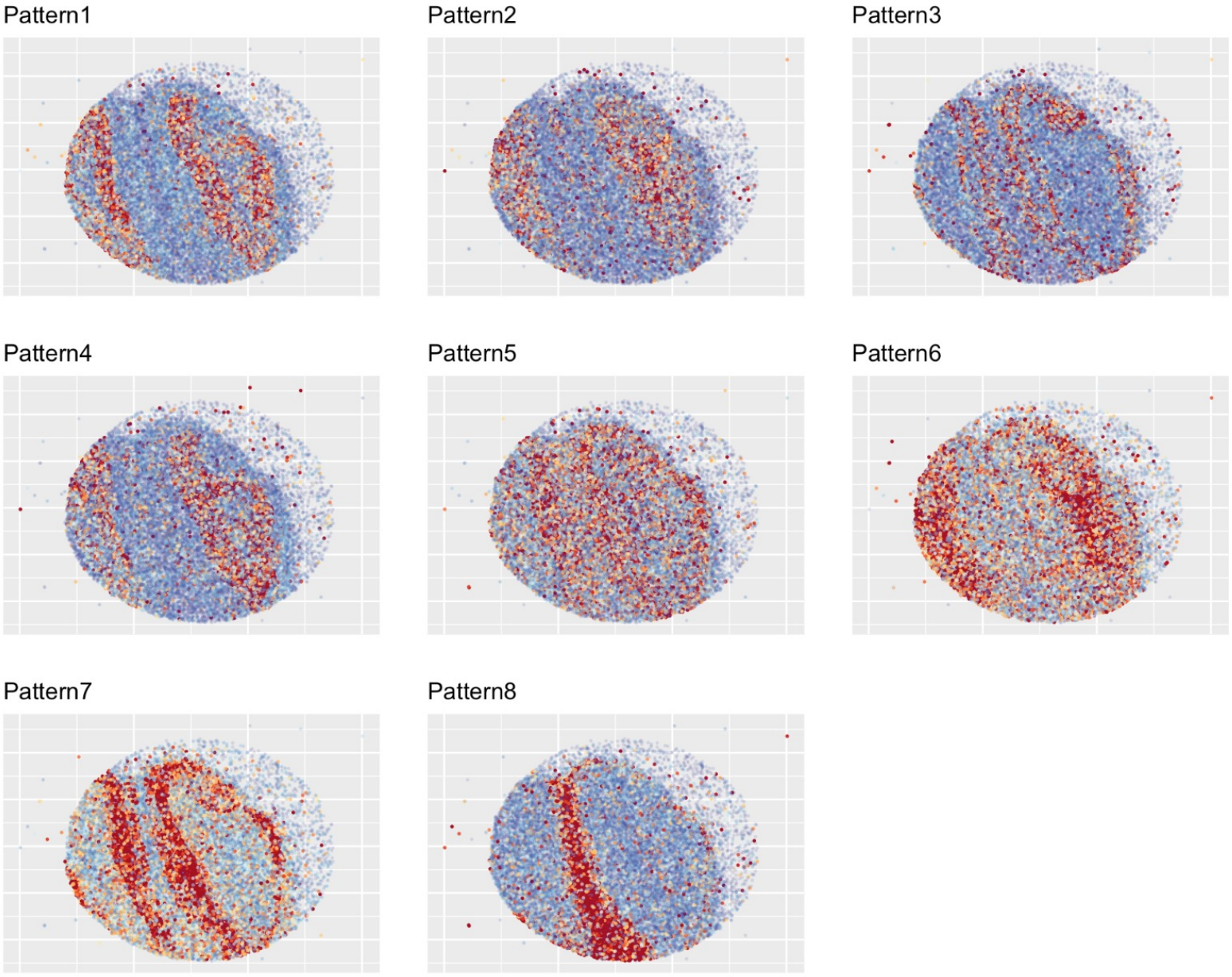
ES=0.68
NES=5.69

ES=0.89
NES=7.30

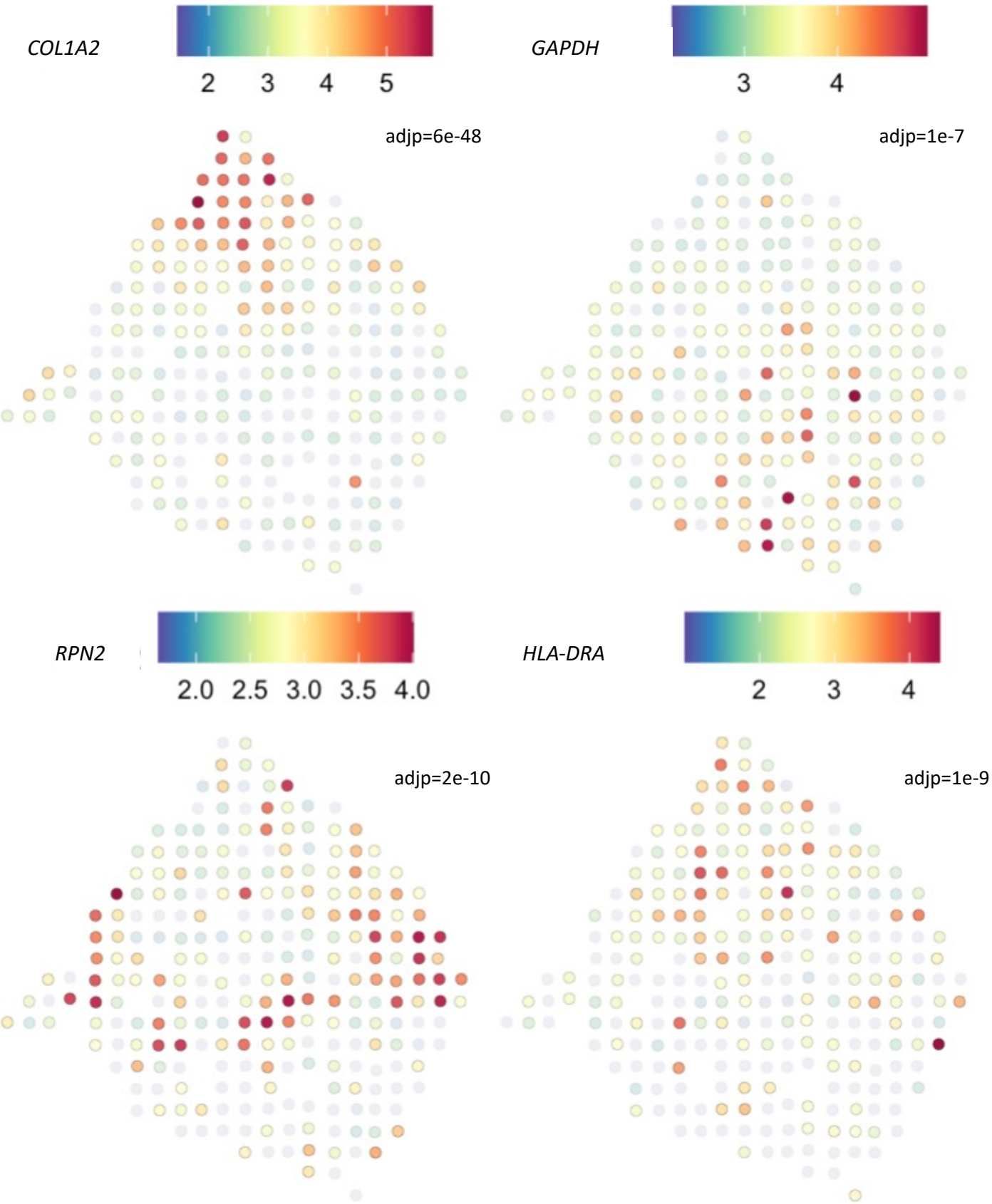
Supplemental Fig. S28. Enrichment scores of markers from 14 location-restricted cell types by SpaGene (left), SPARK-X (middle) and SpatialDE (right).



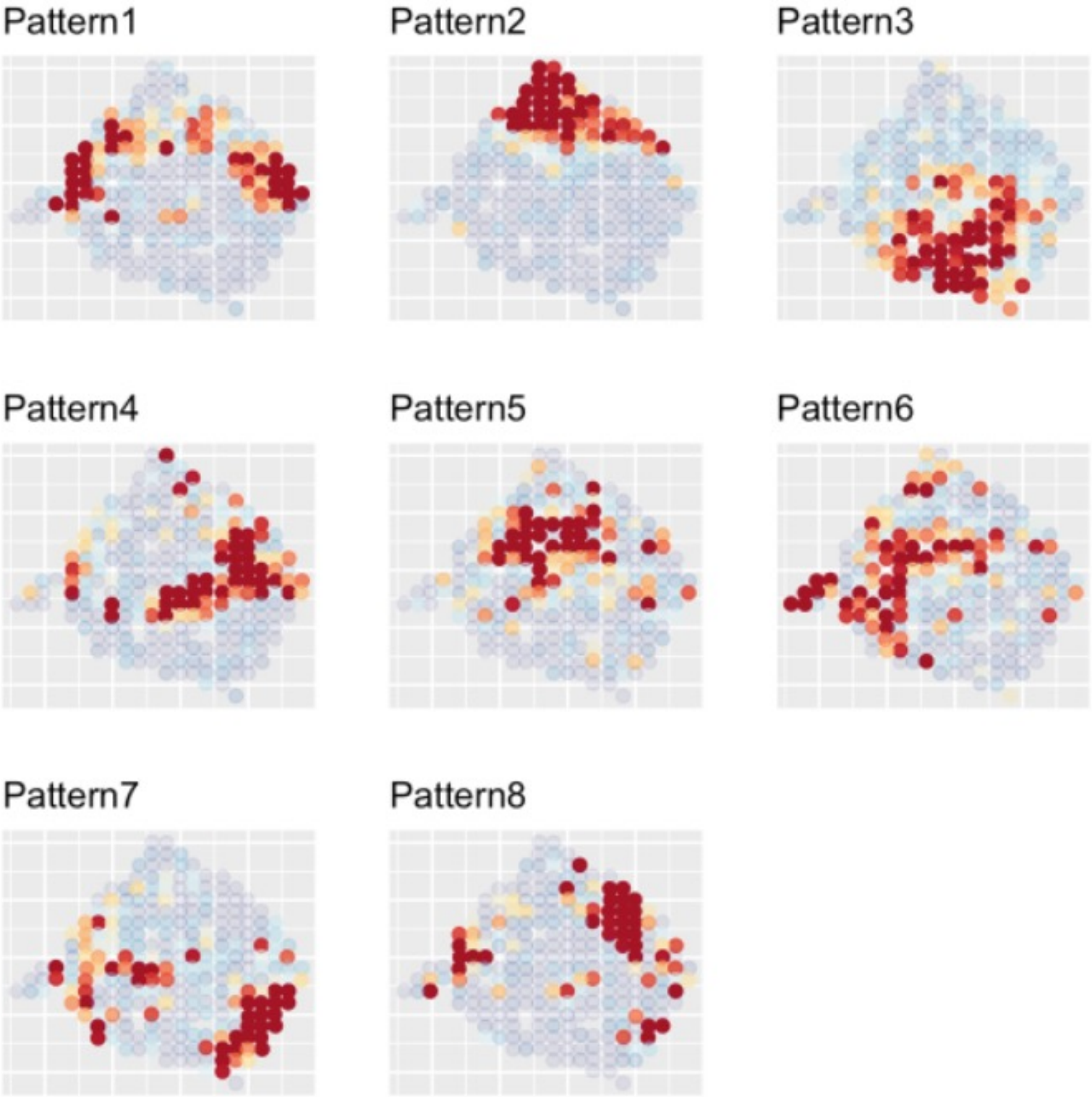
Supplemental Fig. S29. Application of SpaGene on Slideseq v2 MOB data. A) Visualization of spatially variable genes by SpaGene along with adjusted p-values. B) Spatially unaware cell clustering.



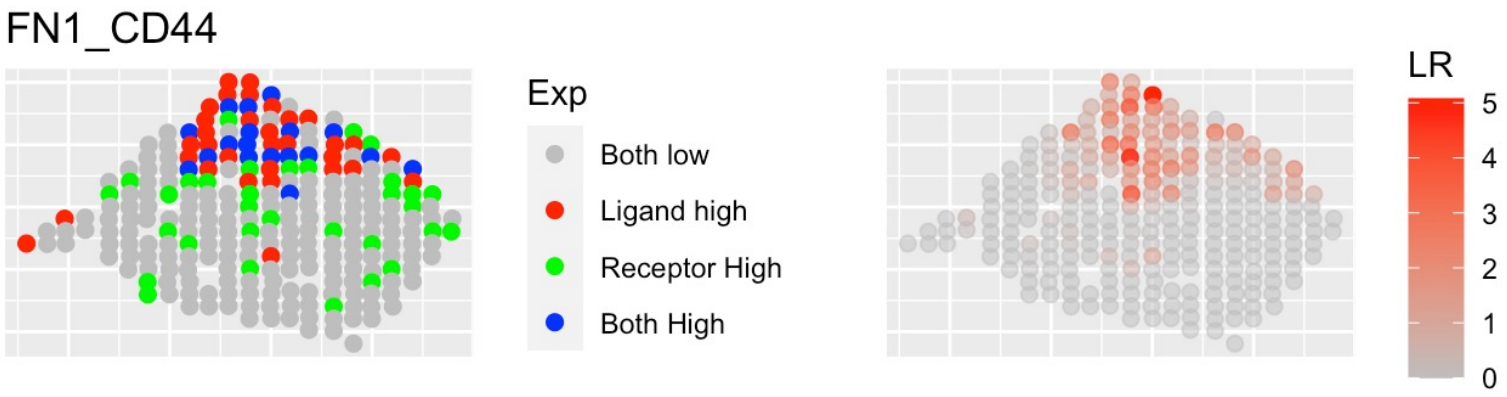
Supplemental Fig. S30. Eight patterns discovered by SpaGene on Slideseq V2 MOB data.



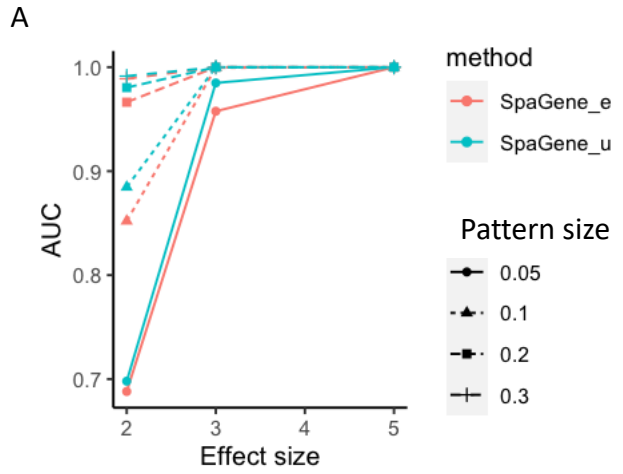
Supplemental Fig. S31. Application of SpaGene on ST breast cancer data. Visualization of four spatially variable genes by SpaGene, along with adjusted p-values.

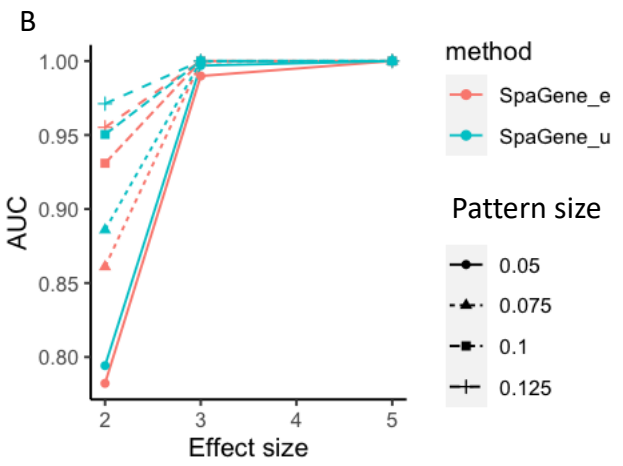


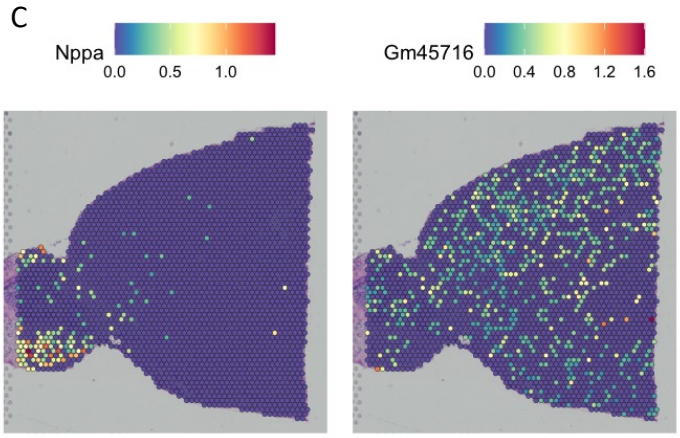
Supplemental Fig. S32. Eight patterns discovered by SpaGene on ST breast cancer data.



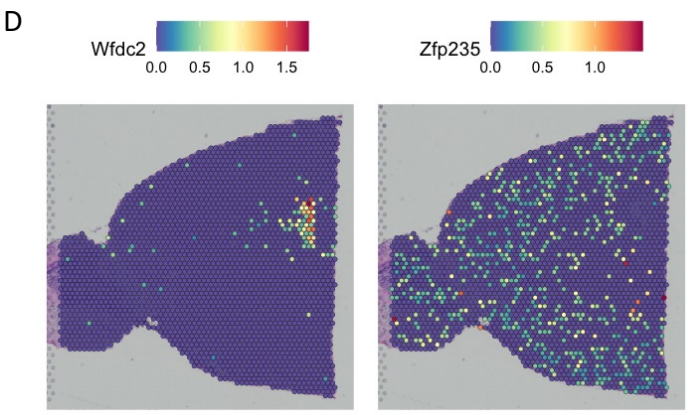
Supplemental Fig. S33. FN1-CD44 interaction identified by SpaGene (adjp=7e-14)





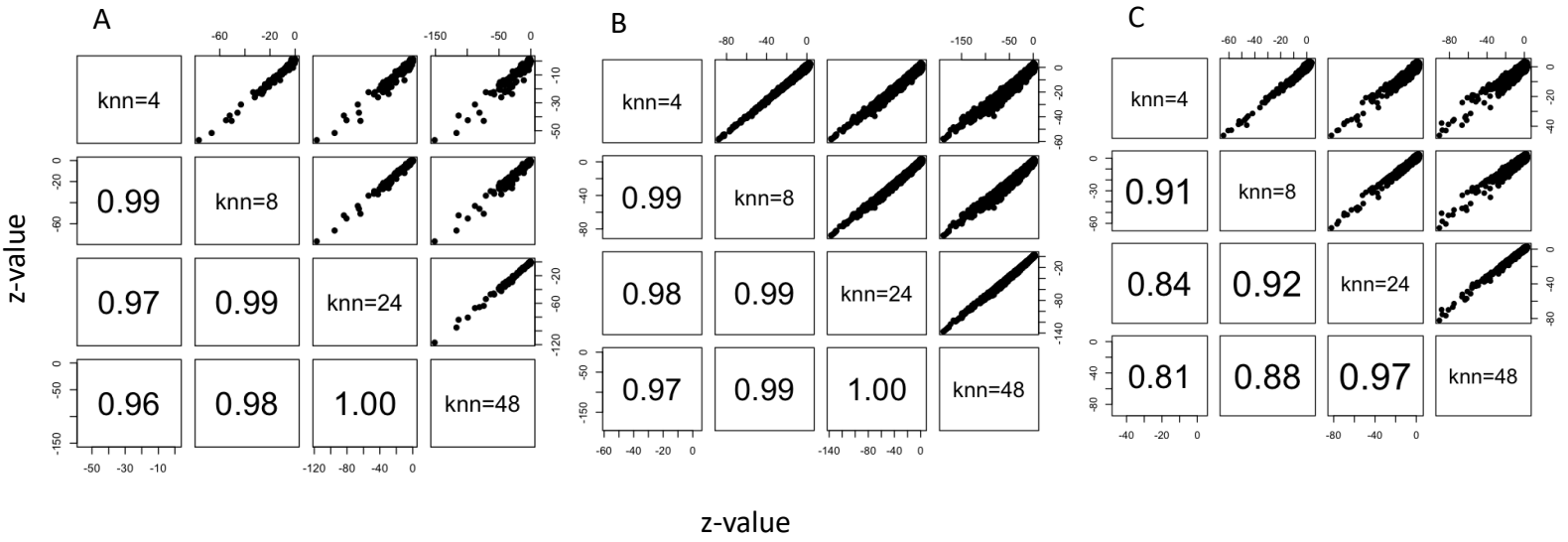


SpaGene_e: *Nppa* (adjp= 6e-9) > *Gm45716* (adjp= 6e-10)
SpaGene_u: *Nppa* (adjp=e-35) < *Gm45716* (adjp= e-8)



SpaGene_e: *Wfdc2* (adjp= 3e-4) > *Zfp235* (adjp= 2e-6)
SpaGene_u: *Wfdc2* (adjp=4e-16) < *Zfp235* (adjp= 3e-4)

Supplemental Fig. S34. Comparison the performance between SpaGene with unequal weights and SpaGene with equal weights. A) AUC plots of SpaGene with equal weights (SpaGene_e,red) and unequal weights (Spagene_u,blue) in simulated datasets with a hotspot pattern containing 10,000 genes and 3,000 locations across different effect sizes (x axis) and pattern sizes (point shapes); B) AUC plots of SpaGene with equal weights (SpaGene_e,red) and unequal weights (Spagene_u,blue) in simulated datasets with a bi-circularity pattern containing 10,000 genes and 3,000 locations across different effect sizes (x axis) and pattern sizes (point shapes); B) The same with A) but simulated datasets with a bi-quarter circularity pattern; C) Visualization of spatial patterns by *Nppa* and *Gm45716* and their significances from SpaGene_e and SpaGene_u; D) Visualization of spatial patterns by *Wfdc2* and *Zfp235* and their significances from SpaGene_e and SpaGene_u; Noted: In the unequal weights setting : W_i=0.5+i*0.5/(2*k), i=0,1,2…2*k; In the equal weights setting: W_i=1, i=0,1,2…2*k, where i is the degree, and k is the parameter to build the nearest graph.



Supplemental Fig. S35. The effect of the number of k nearest neighbors on the performance of SpaGene. The pairwise correlation between results with different k-values (4,8, 24, and 48). A) MERFISH data; B) 10X Visium mouse brain; C) Slideseq V2 mouse cerebellum data.