

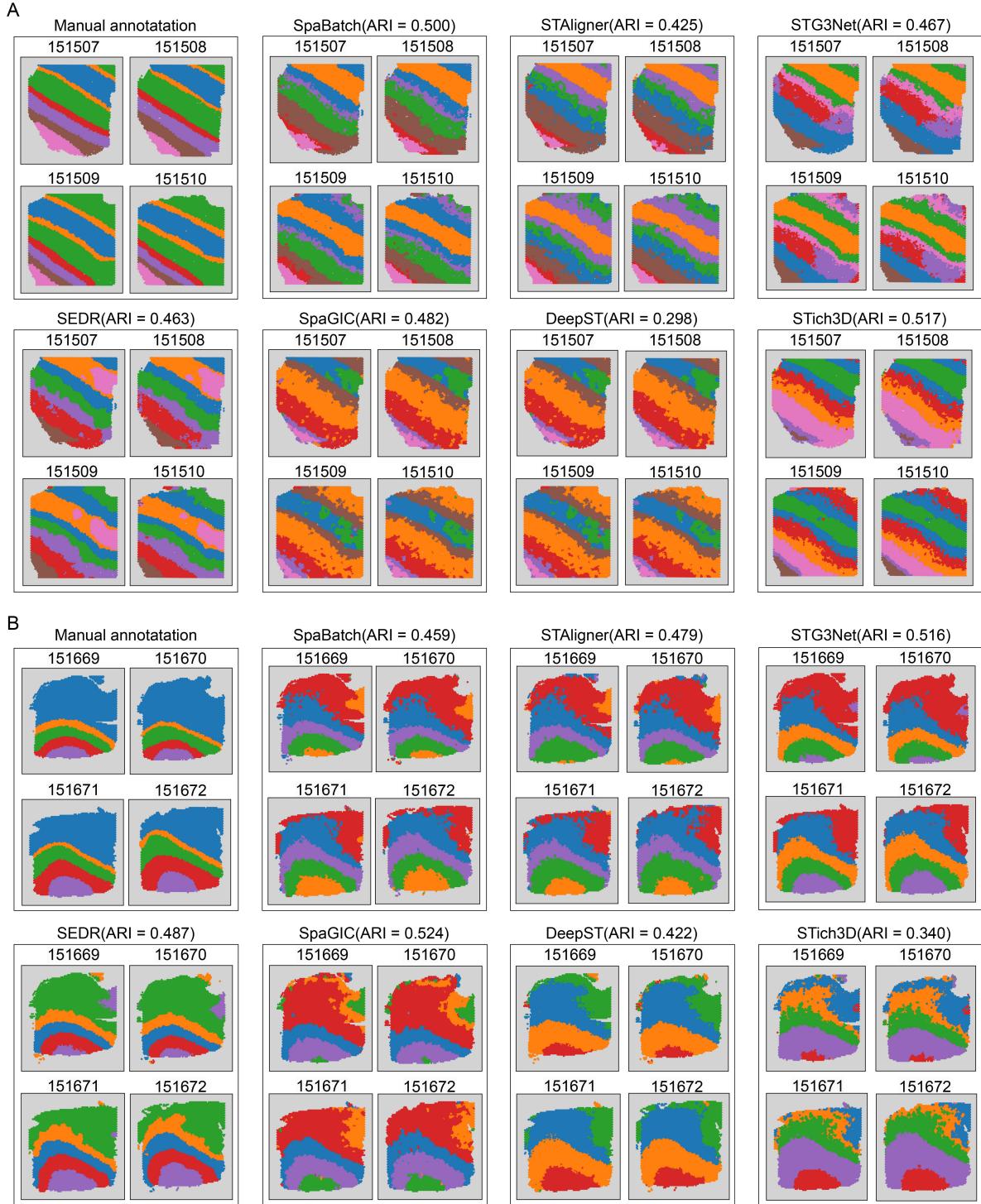
**Supplementary Materials for  
SpaBatch: Batch alignment of spatial transcriptomics data  
using deep learning**

Jinyun Niu<sup>✉</sup>, Donghai Fang<sup>✉</sup>, and Wenwen Min<sup>✉</sup>

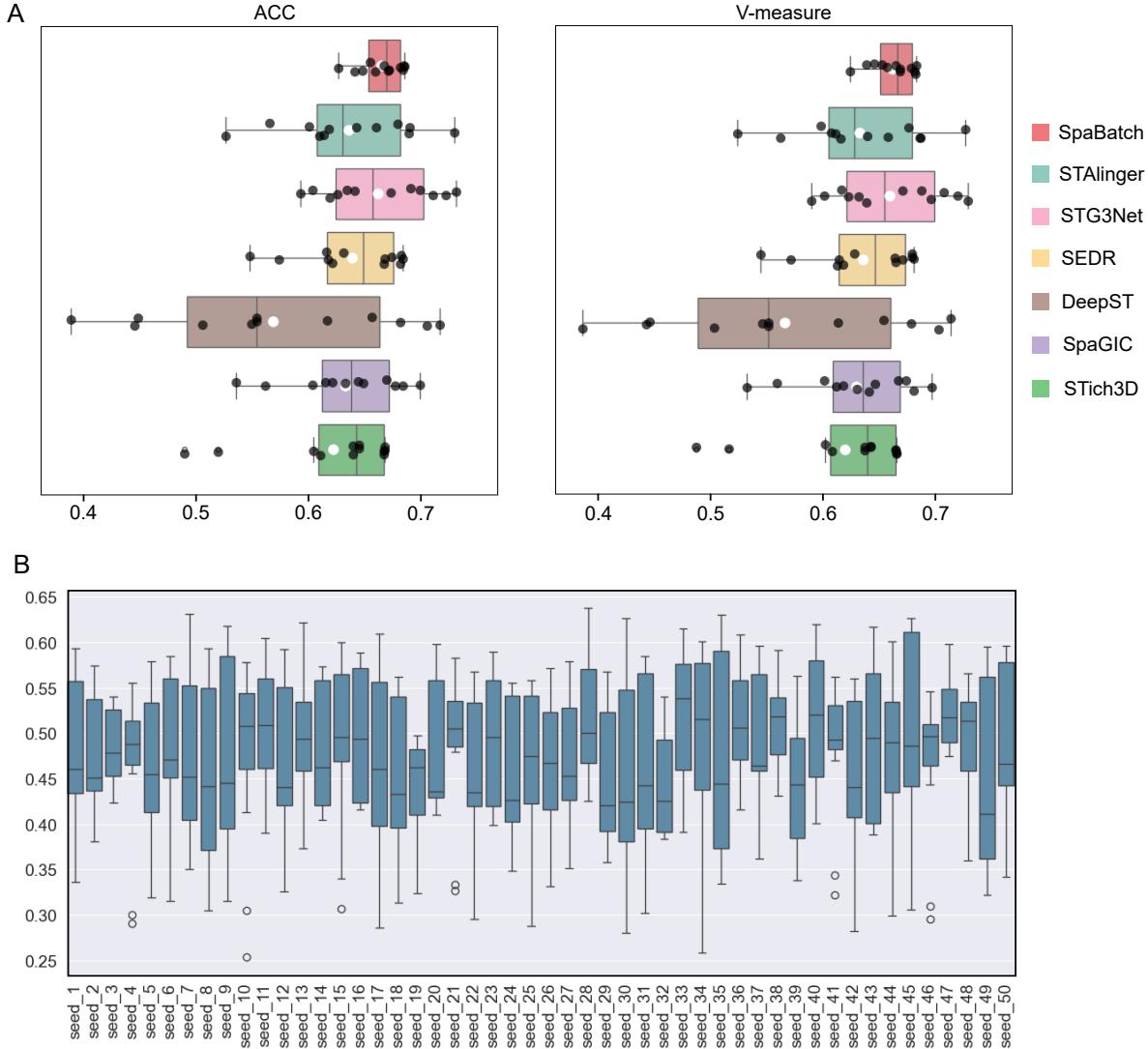
School of Information Science and Engineering, Yunnan University, 650091, Yunnan, China

2 Jinyun Niu<sup>④</sup>, Donghai Fang<sup>⑤</sup>, and Wenwen Min<sup>⑥</sup>

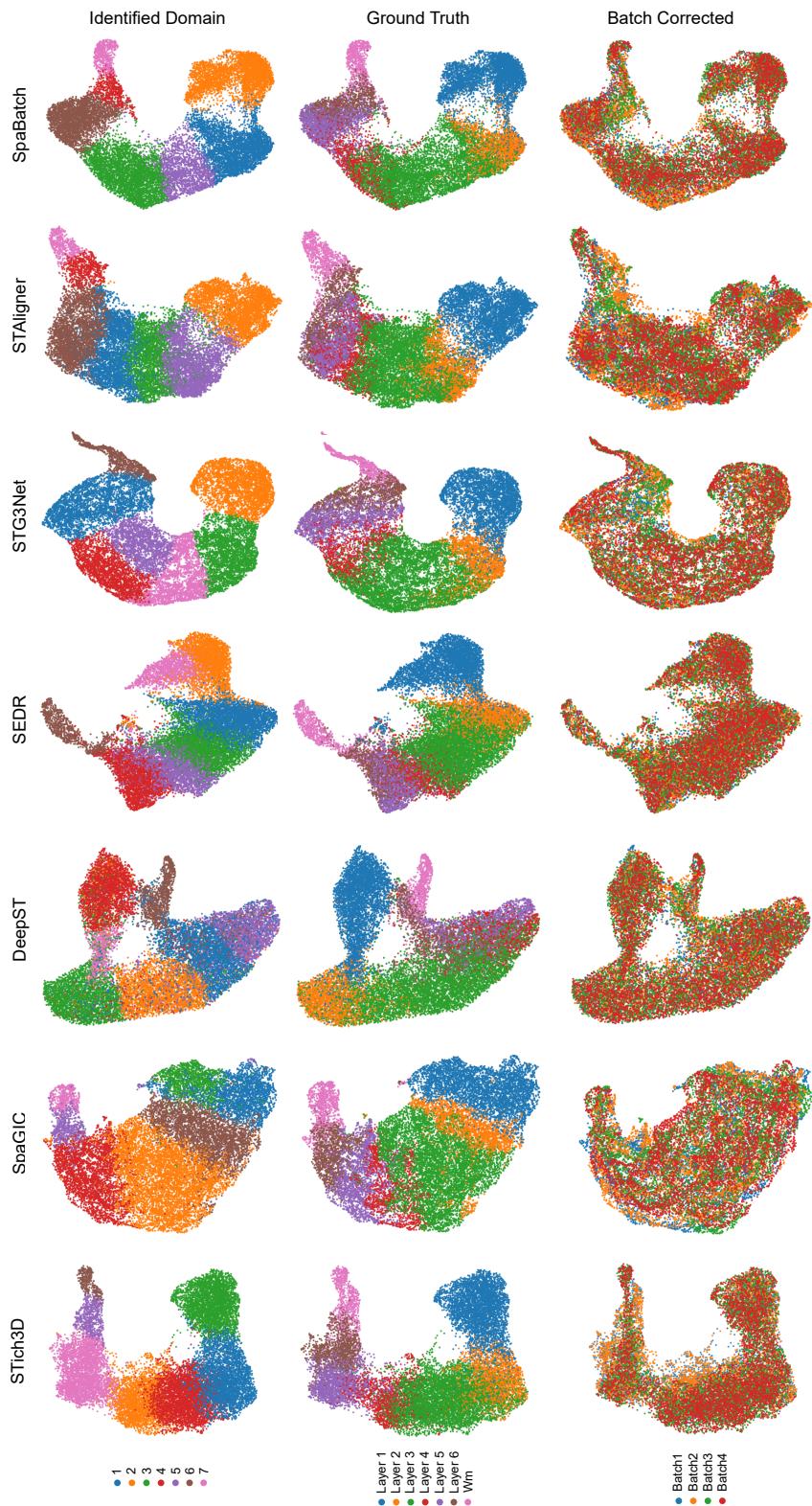
## 1 Supplementary Figures



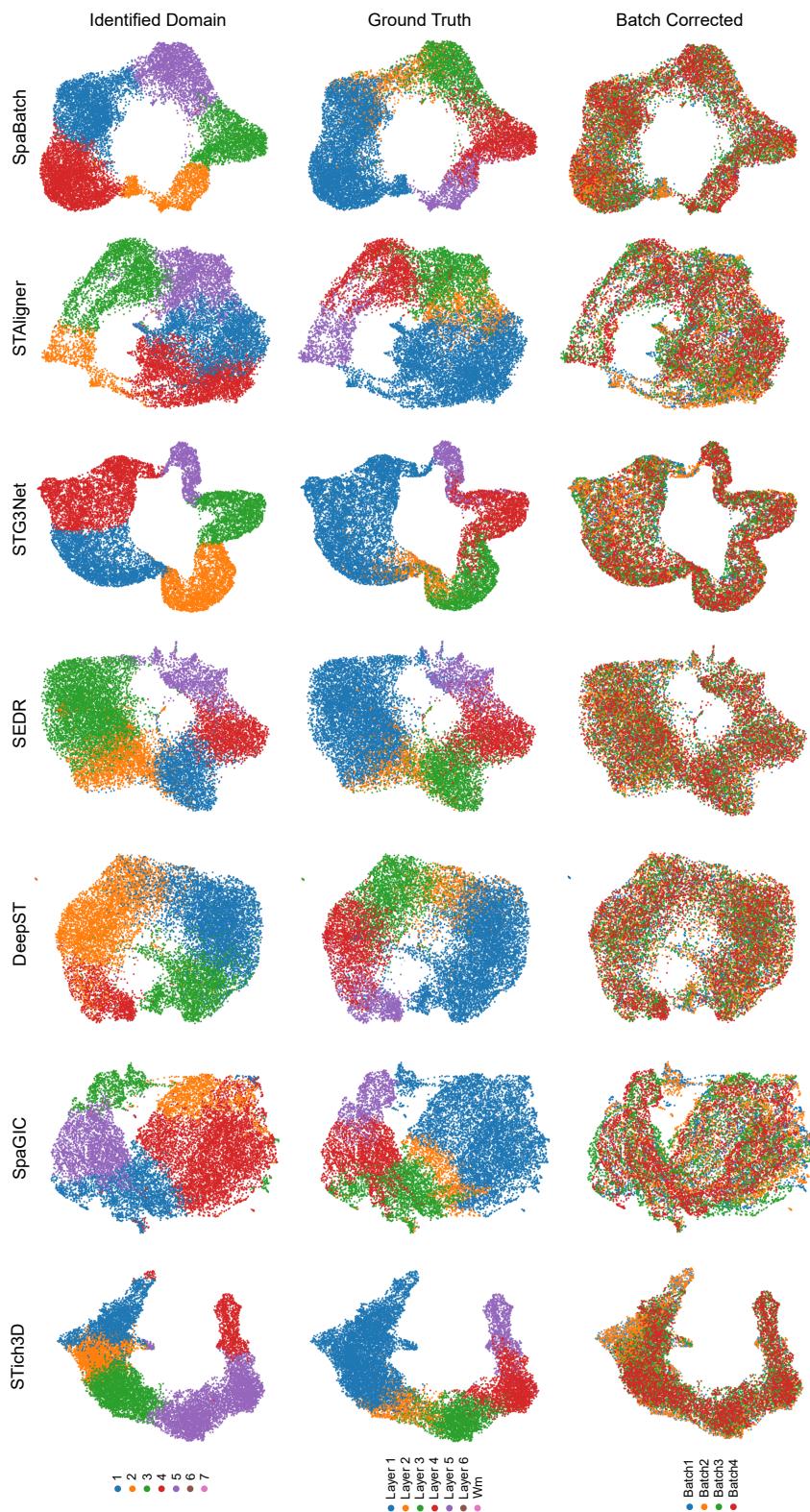
**Fig. S1** Spatial domain identification results for four different slices from (A) Donor 1 and (B) Donor 2 using various methods.



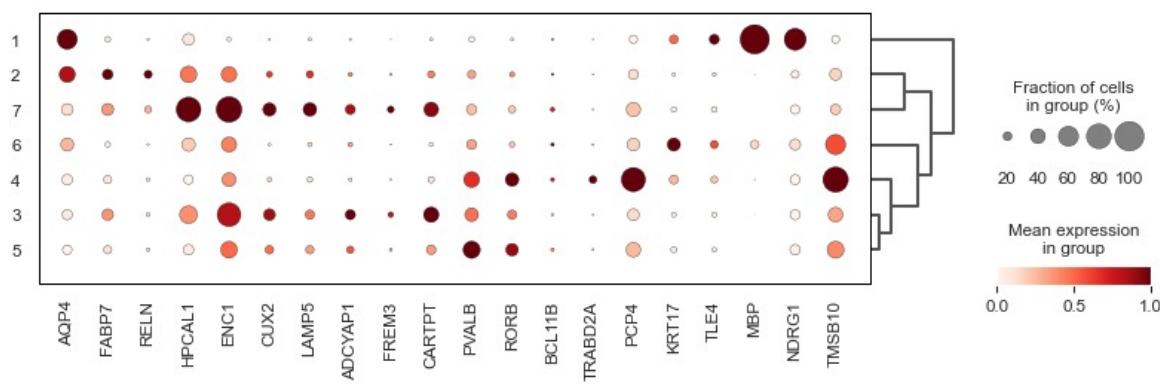
**Fig. S2** A. Boxplot of clustering accuracy across all the 12 slices grouped by Donor of the DLPFC dataset in terms of Average Clustering Consistency (ACC) and V-measure. B. The clustering ARI of SpaBatch in all 12 slices grouped by Donor under the default hyperparameters with different random seeds.



**Fig. S3** UMAP visualization of Donor1 embeddings colored by identified domains (left), ground truth (middle), and batch corrected (right).

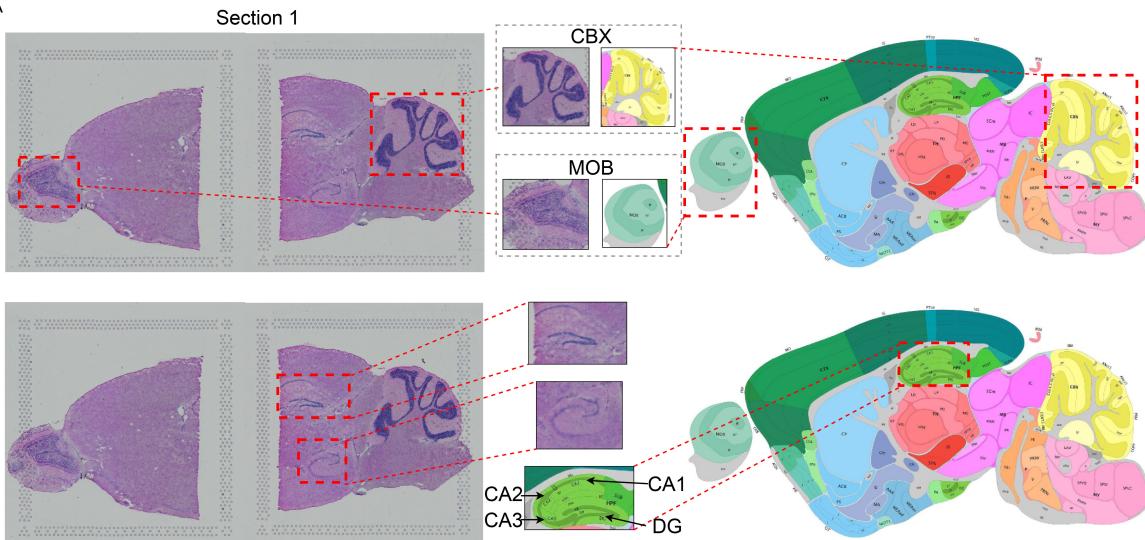


**Fig. S4** UMAP visualization of Donor2 embeddings colored by identified domains (left), ground truth (middle), and batch corrected (right).

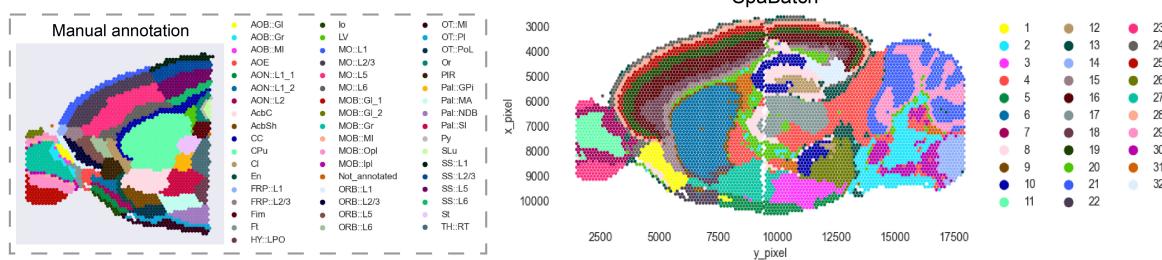


**Fig. S5** Dot plot of layer marker genes defined by SpaBatch, visualizing the expression patterns of marker genes across the seven layers (layer 1–6 and WM) identified in sample 3 of the DLPFC dataset.

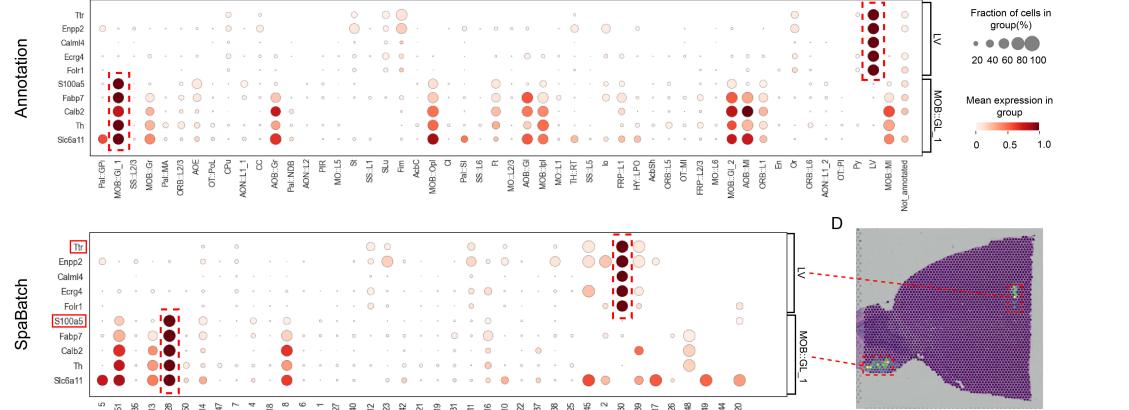
A



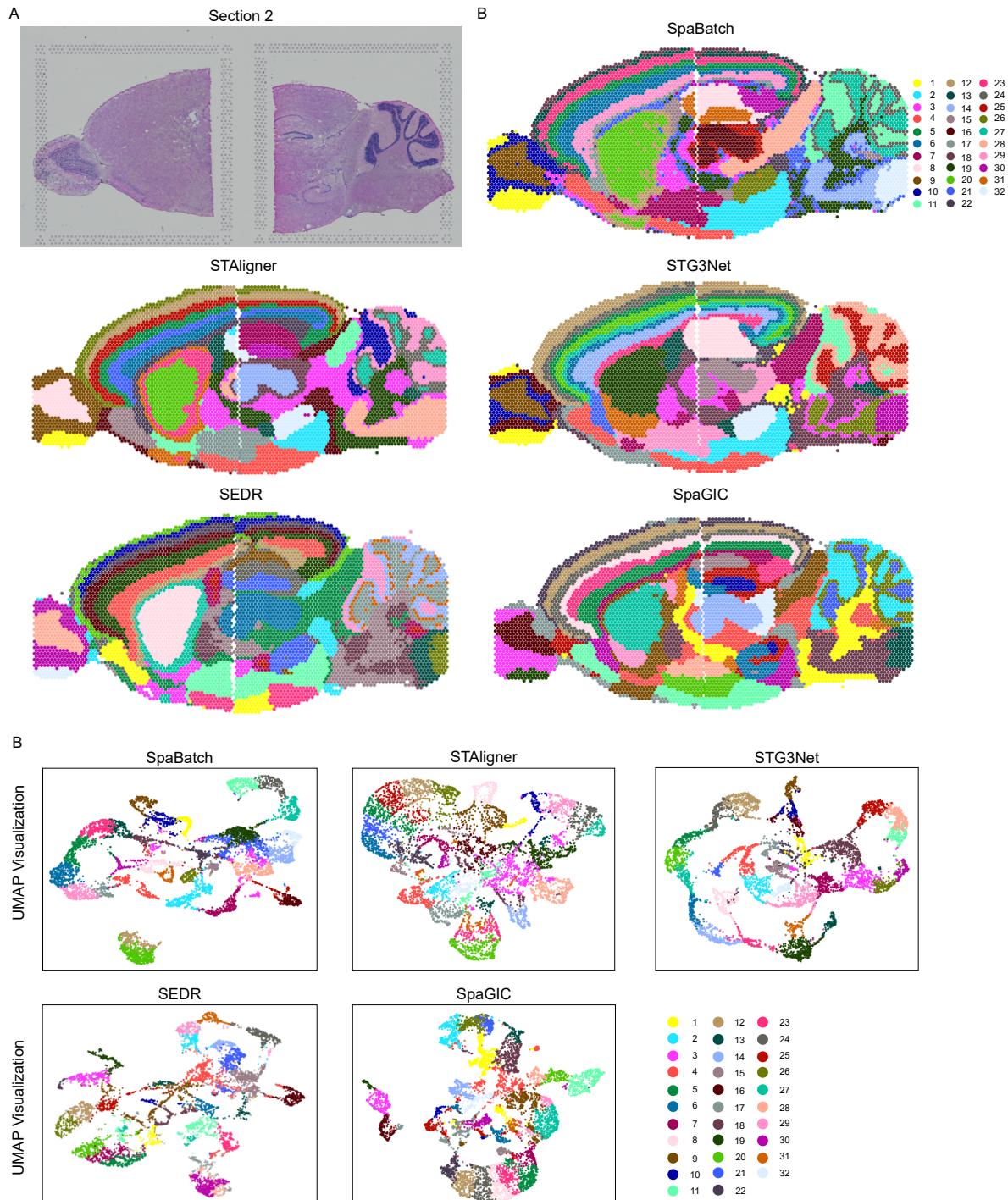
B



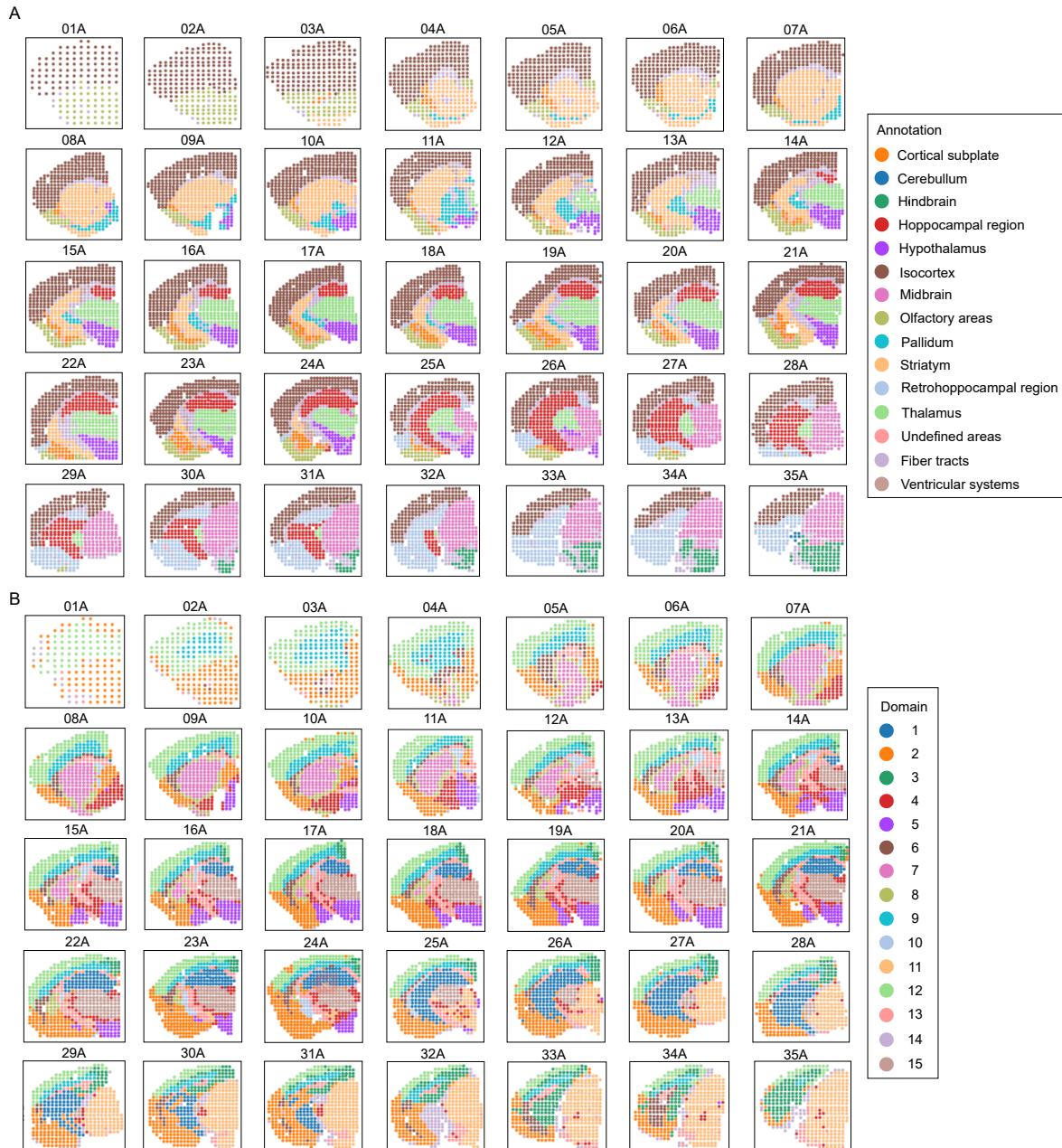
C



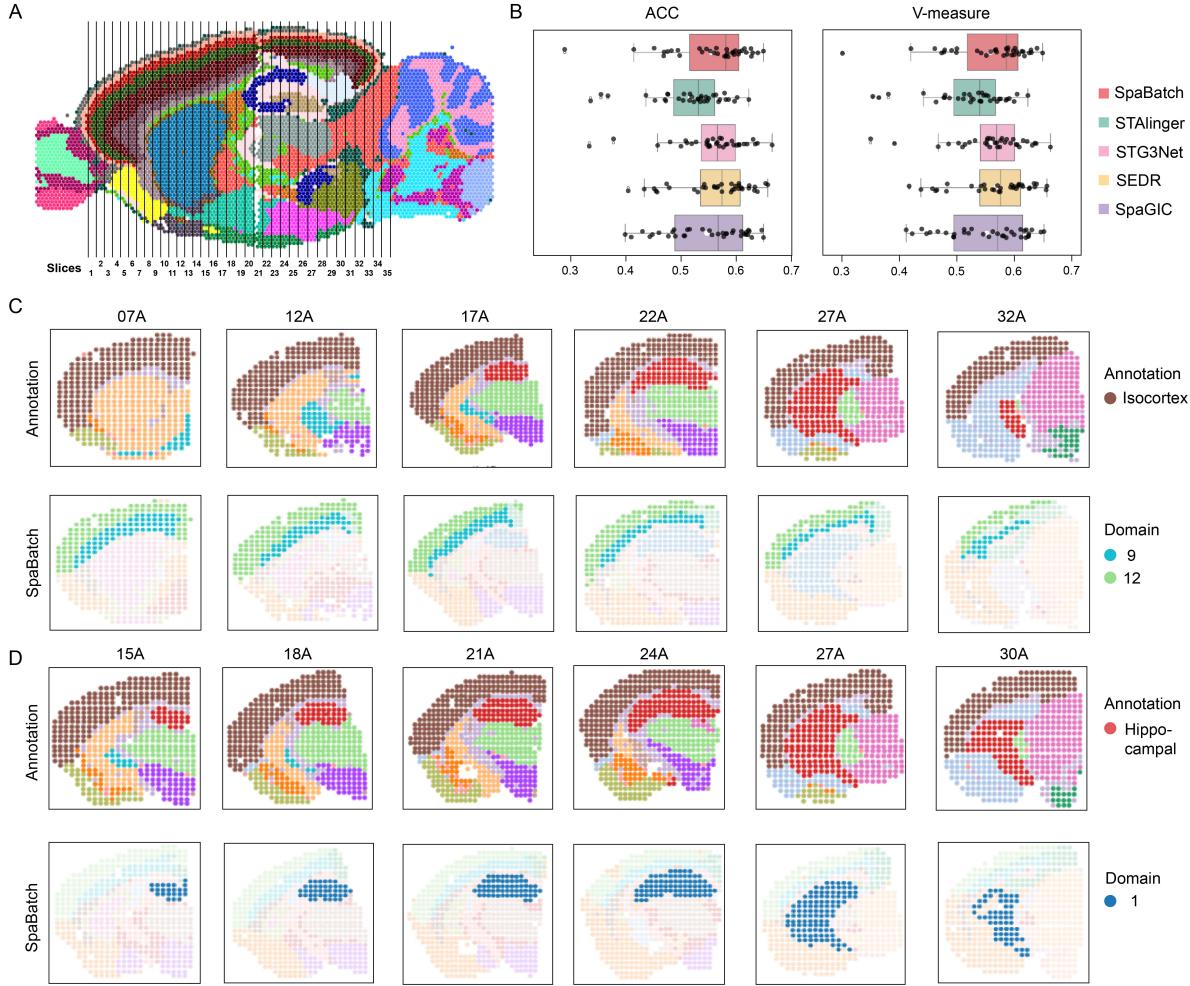
**Fig. S6** SpaBatch enables the stitching of adjacent sagittal anterior and posterior sections (Section 1) of the mouse brain to create a large composite section while accurately identifying specific spatial subdomains. A. H&E images of sagittal anterior and posterior sections of Section 1, the mouse cortex reference atlas provided by the Allen Brain Institute, and the corresponding specific spatial subdomains. B. Spatial domain identification results of sagittal anterior and posterior sections of Section 1 using SpaBatch. C. Differential analysis results of different spatial domains (Domain 28 and 30) in sagittal anterior and posterior sections of Section 1 using SpaBatch, which align well with the manually annotated spatial domains (MOB::Gl and LV). D. Visualization of marker genes (*S100a5* and *Ttr*) for spatial domains 28 and 30.



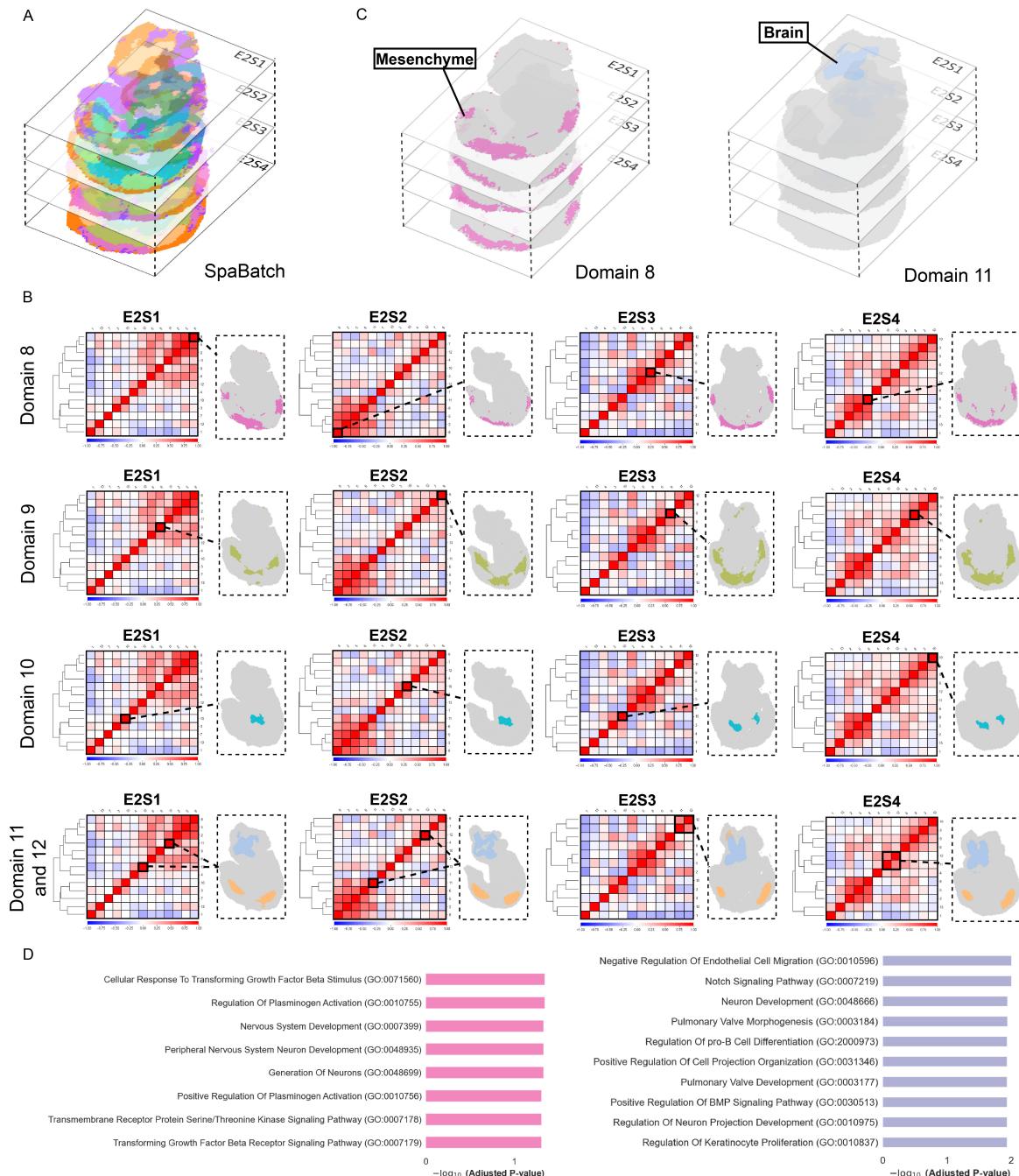
**Fig. S7** SpaBatch can stitch adjacent sagittal anterior and posterior sections of the mouse brain (section 2) to generate a large composite slice. A. The H&E images of mouse sagittal anterior and posterior Section 2. B. Results of spatial domain identification using SpaBatch and other methods on the sagittal anterior and posterior Section 2 of mouse. C. Results of UMAP visualization using SpaBatch and other methods on the sagittal anterior and posterior Section 2 of mouse.



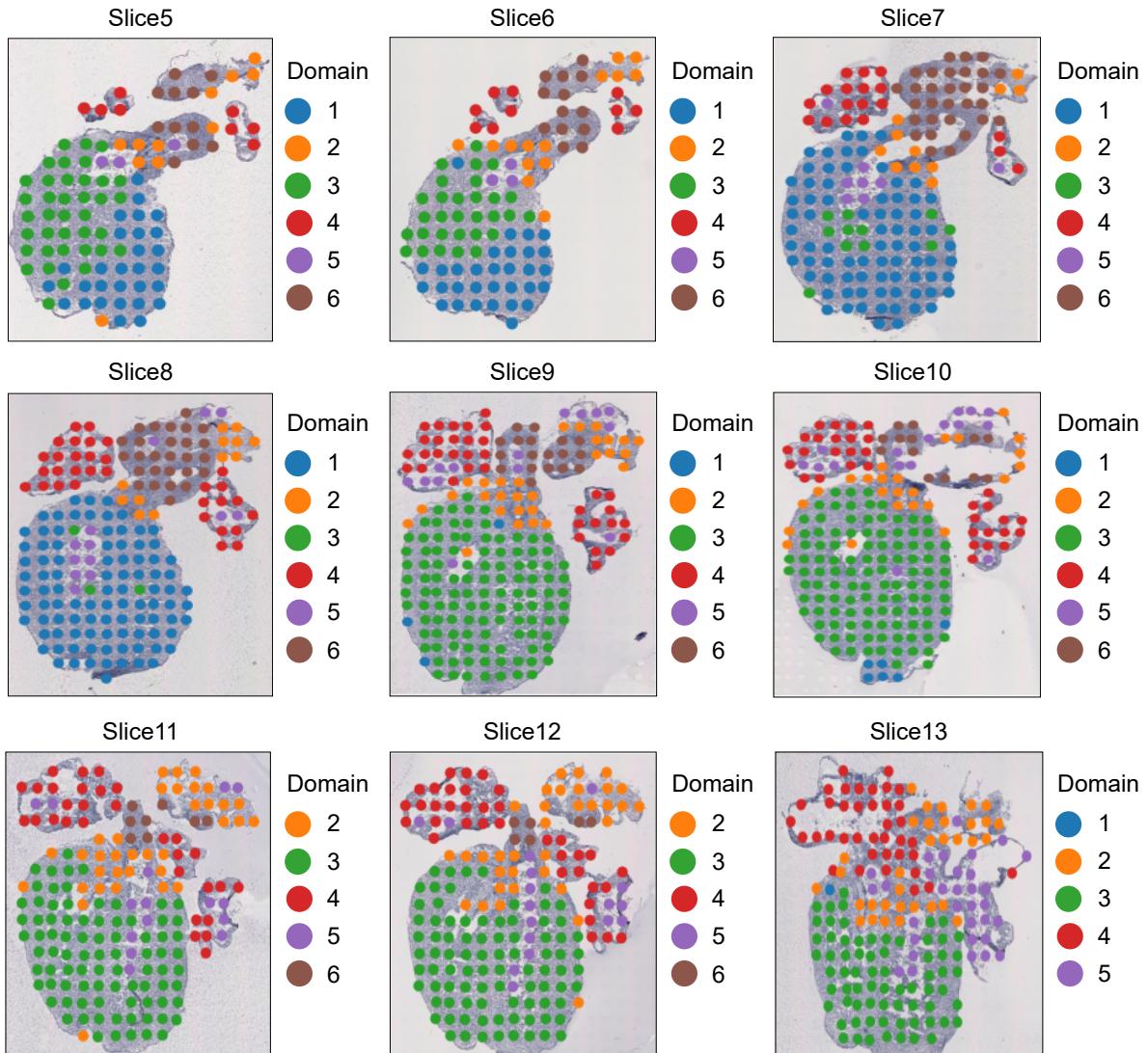
**Fig. S8** A. Manual annotation of mouse brain data from 35 slices and B. SpaBatch spatial domain identification results.



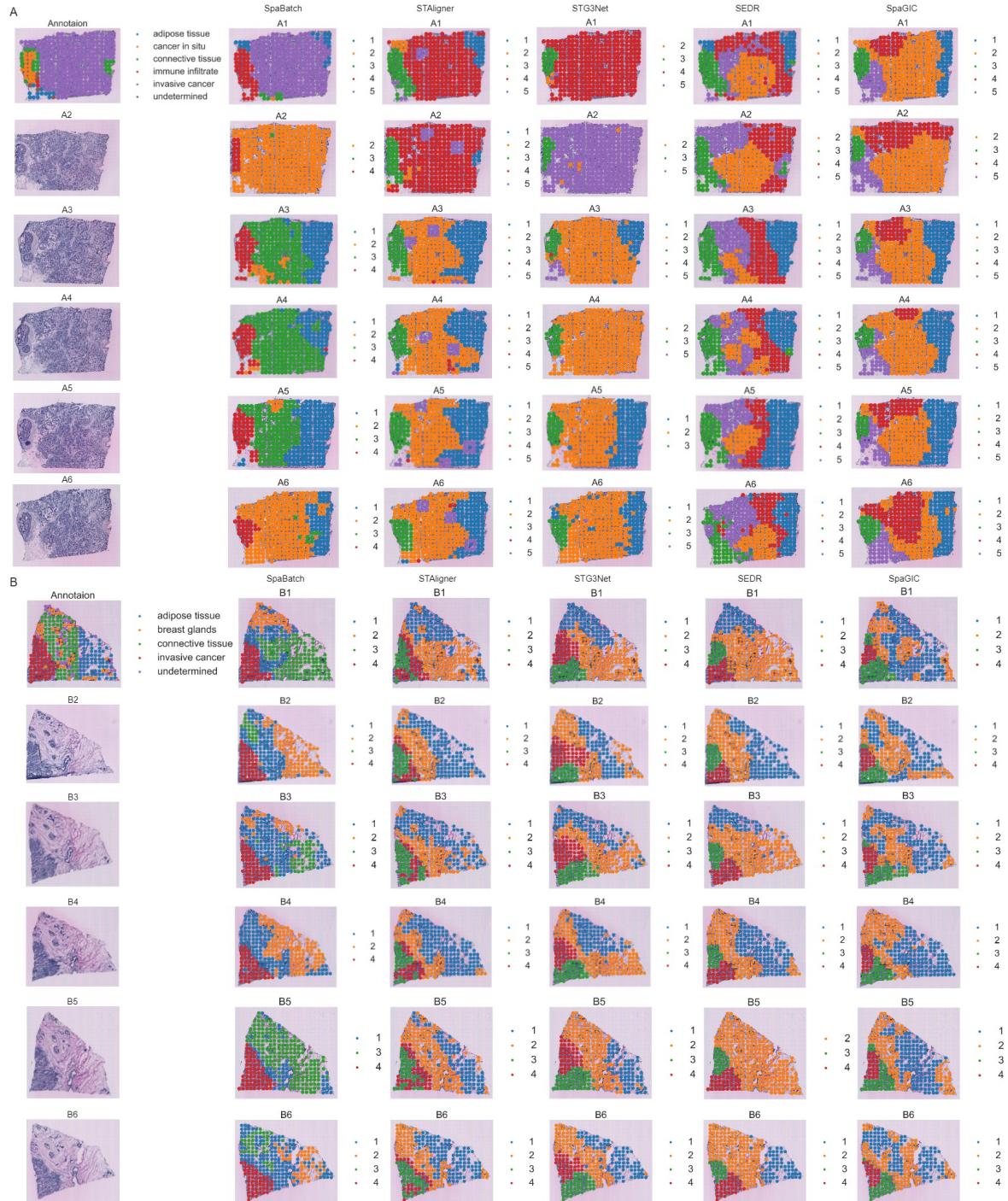
**Fig. S9** A. Spatial domain identification results using SpaBatch on sagittal adult mouse brain, with the 35 slices mapped onto the sagittal mouse brain. B. Box plots of ACC and V-measure values for clustering the 35 adult mouse brain slices using SpaBatch and other methods. C. Spatial domains 9 and 12 identified by SpaBatch correspond with the manually annotated Isocortex in cross-slice spatial domain identification. D. Spatial domain 1 identified by SpaBatch corresponds with the manually annotated Hippocampal region in cross-slice spatial domain identification.



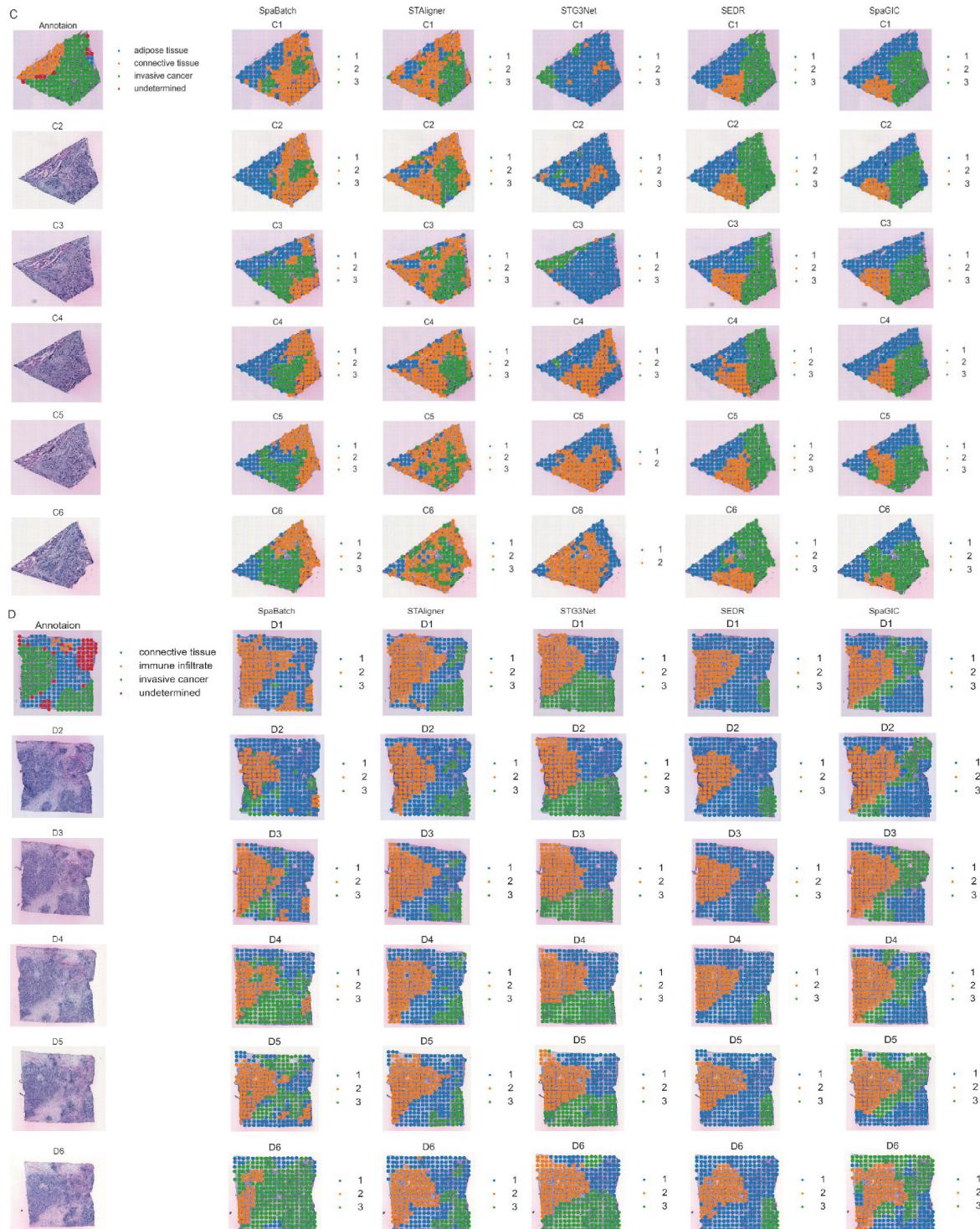
**Fig. S10** A. The spatial domains identified by SpaBatch on four sections of E2S1, E2S2, E2S3, and E2S4 in E9.5 embryos. B. The correlation matrix of clusters identified by SpaBatch highlights a group of highly correlated clusters, marked by the black box, concentrated in the mesenchyme, sclerotome, primitive gut tube, and brain regions. C. The regions exhibiting correlation in the mesenchyme and brain were consistently identified by SpaBatch across all four sections. D. GO enrichment pathway of domain 11 and 12.



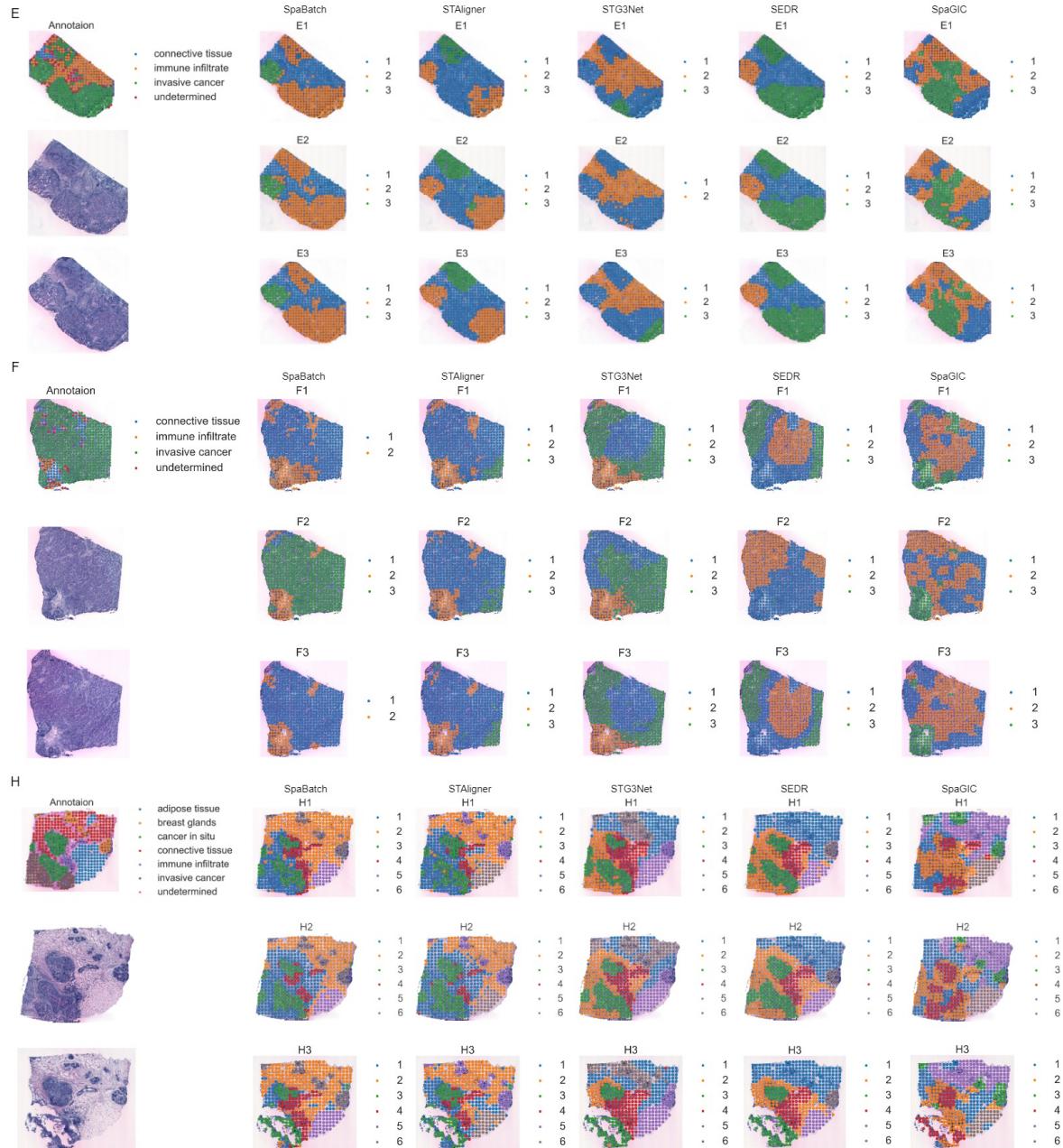
**Fig. S11** SpaBatch performs spatial domain identification across all sections of the human heart at 6.5 PCW.



See next page



See next page



**Fig. S12** Spatial domain detection results of SpaBatch and other methods across all slice groups (A-H) in the HER2-positive breast cancer dataset.

## 2 Supplementary Tables

**Table S1** Overview of comparative spatial domain identification methods.

Method	Methodology	Input Data	Downstream tasks	Link
STAGATE [Dong and Zhang, 2022]	Graph attention autoencoders	Spatial location data Gene expression data	Spatial domain identification Visualization Trajectory inference Denosing	<a href="https://github.com/zhanglabtools/STAGATE/">https://github.com/zhanglabtools/STAGATE/</a>
STAligner [Zhou et al., 2023]	Graph attention autoencoders	Spatial location data Gene expression data	Spatial domain identification Visualization Trajectory inference GO enrichment analysis	<a href="https://github.com/zhoux85/STAligner/">https://github.com/zhoux85/STAligner/</a>
STG3Net [Fang et al., 2024]	Graph neural autoencoder	Spatial location data Gene expression data	Spatial domain identification Visualization Spatial domain detect Denosing	<a href="https://github.com/wenwenmin/STG3Net/">https://github.com/wenwenmin/STG3Net/</a>
DeepST [Xu et al., 2022]	Variational graph autoencoders	Spatial location data Gene expression data Histology information	Spatial domain identification Visualization Trajectory inference	<a href="https://github.com/JiangBioLab/DeepST/">https://github.com/JiangBioLab/DeepST/</a>
SEDR [Xu et al., 2024]	Variational graph autoencoders	Spatial location data Gene expression data	Spatial domain identification Visualization Trajectory inference Denosing	<a href="https://github.com/HzFu/SEDR/">https://github.com/HzFu/SEDR/</a>
SpaGIC [Liu et al., 2024]	Graph neural autoencoder	Spatial location data Gene expression data	Spatial domain identification Visualization Trajectory inference Denosing	<a href="https://github.com/Liuwei-CS/SpaGIC">https://github.com/Liuwei-CS/SpaGIC</a>
STitch3D [Wang et al., 2023]	Graph neural autoencoder	Spatial location data Gene expression data sc-RNA data	Spatial domain identification Cell-type deconvolution Visualization Trajectory inference Denosing	<a href="https://github.com/YangLabHKUST/STitch3D">https://github.com/YangLabHKUST/STitch3D</a>

**Table S2** Summary of the ST data used in this study.

Platform	Tissue	Section	Number of domains	Spots	Related figures
10x Visium	Human dorsolateral prefrontal cortex (DLPFC)	151507	7	4226	
		151508	7	4384	
		151509	7	4789	
		151510	7	4634	
		151669	5	3661	
		151670	5	3498	Figure 2-3
		151671	5	4110	Figure S1-S5
		151672	5	4015	
		151673	7	3639	
		151674	7	3673	
		151675	7	3592	
		151676	7	3460	
10x Visium	Sagittal mouse brain	Section 1 anterior	32	2695	
		Section 1 posterior	32	3355	Figure 4
		Section 2 anterior	32	2825	Figure S6-S7
		Section 2 posterior	32	3289	
ST	Coronal mouse brain	01A	4	152	
		02A	6	240	
		03A	8	269	
		04A	8	326	
		05A	9	361	
		06A	11	403	
		07A	11	488	
		08A	12	470	
		09A	10	491	
		10A	12	522	
		11A	12	494	
		12A	13	506	
		13A	12	509	
		14A	14	580	
		15A	13	556	
		16A	14	560	
		17A	14	580	
		18A	14	577	Figure 5
		19A	13	591	Figure S8-S9
		20A	13	536	
		21A	14	620	
		22A	14	589	
		23A	14	574	
		24A	13	617	
		25A	13	508	
		26A	12	548	
		27A	11	508	
		28A	11	460	
		29A	11	462	
		30A	11	527	
		31A	11	548	
		32A	9	524	
		33A	9	484	
		34A	10	478	
		35A	8	428	
Stereo-seq	Mouse embryos	E2S1	14	5913	
		E2S2	13	5292	Figure 6
		E2S3	13	4356	Figure S10
		E2S4	13	5059	

Platform	Tissue	Section	Number of domains	Spots	Related figures
10x Visium	Human heart	PCW4.5-5 01	2	54	
		PCW4.5-5 02	2	57	
		PCW4.5-5 03	2	66	
		PCW4.5-5 04	2	55	
		PCW6.5 01	6	100	
		PCW6.5 02	6	101	
		PCW6.5 03	6	155	Figure 7
		PCW6.5 04	6	182	Figure S11
		PCW6.5 05	6	212	
		PCW6.5 06	6	203	
		PCW6.5 07	6	173	
		PCW6.5 08	6	180	
		PCW6.5 09	6	174	
		Group A 01	5	346	
10x Visium	HER2-positive breast cancer	Group A 02	5	325	
		Group A 03	5	359	
		Group A 04	5	343	
		Group A 05	5	332	
		Group A 06	5	360	
		Group B 01	4	295	
		Group B 02	4	270	
		Group B 03	4	298	
		Group B 04	4	283	
		Group B 05	4	289	
		Group B 06	4	277	
		Group C 01	3	176	
		Group C 02	3	187	
		Group C 03	3	180	
		Group C 04	3	184	Figure 8
		Group C 05	3	181	Figure S12-S14
		Group C 06	3	178	
		Group D 01	3	306	
		Group D 02	3	303	
		Group D 03	3	301	
		Group D 04	3	302	
		Group D 05	3	306	
		Group D 06	3	315	
		Group E 01	3	587	
		Group E 02	3	572	
		Group E 03	3	570	
		Group F 01	3	691	
		Group F 02	3	695	
		Group F 03	3	712	
		Group H 01	6	613	
		Group H 02	6	603	
		Group H 03	6	510	

## References

- Dong and Zhang, 2022. Dong, K. and Zhang, S. (2022). STAGATE: Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications*, 13(1):1739.
- Fang et al., 2024. Fang, D., Zhu, F., et al. (2024). Multi-slice spatial transcriptomics data integration analysis with stg3net. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 509–514. IEEE.
- Liu et al., 2024. Liu, W., Wang, B., Bai, Y., Liang, X., Xue, L., and Luo, J. (2024). Spagic: graph-informed clustering in spatial transcriptomics via self-supervised contrastive learning. *Briefings in Bioinformatics*, 25(6):bbae578.
- Wang et al., 2023. Wang, G., Zhao, J., Yan, Y., et al. (2023). Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. *Nature Machine Intelligence*, 5(11):1200–1213.
- Xu et al., 2022. Xu, C., Jin, X., et al. (2022). DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Research*, 50(22):e131–e131.
- Xu et al., 2024. Xu, H., Fu, H., et al. (2024). SEDR: Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Medicine*, 16(1):1–15.
- Zhou et al., 2023. Zhou, X., Dong, K., et al. (2023). Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nature Computational Science*, 3(10):894–906.