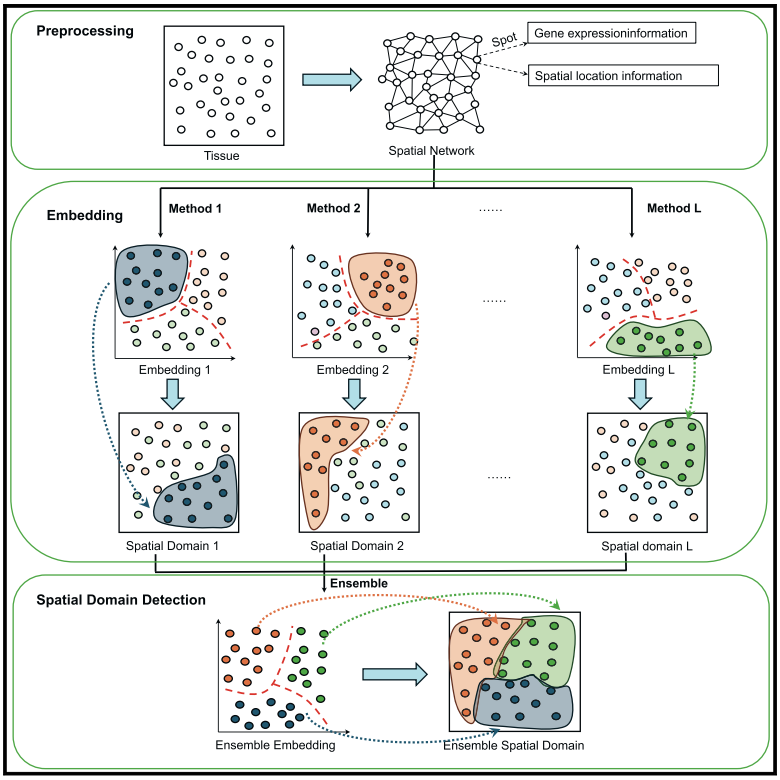
Cell Reports
Methods

Article

WEST is an ensemble method for spatial transcriptomics analysis

Graphical abstract



Authors

Jiazhang Cai, Huimin Cheng, Shushan Wu, Wenxuan Zhong, Guo-Cheng Yuan, Ping Ma

Correspondence

wenxuan@uga.edu (W.Z.), guo-cheng.yuan@mssm.edu (G.-C.Y.), pingma@uga.edu (P.M.)

In brief

Cai et al. present a method, WEST, for detecting the spatial domain with spatial transcriptomics data. WEST exploits the idea of ensemble learning to integrate the embeddings from multiple deep learning-based methods to improve performance. WEST provides a new perspective for improving the performance of the existing methods.

Highlights

- WEST identifies spatial domains in spatial transcriptomics datasets
- By incorporating ensemble learning, WEST offers robust performance and generalizability
- WEST outperforms existing methods in direct benchmarking
- We provide open-source Python code and tutorials

Cai et al., 2024, Cell Reports Methods *4*, 100886
November 18, 2024
https://doi.org/10.1016/j.crmeth.2024.100886

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Article

WEST is an ensemble method for spatial transcriptomics analysis

Jiazhang Cai,[1] Huimin Cheng,[2] Shushan Wu,[1] Wenxuan Zhong,[1,*] Guo-Cheng Yuan,[3,*] and Ping Ma[1,4,*]
[1]Department of Statistics, University of Georgia, 310 Herty Drive, Athens, GA 30602, USA
[2]Department of Biostatistics, Boston University, 801 Massachusetts Avenue, Crosstown Center, Boston, MA 02118, USA
[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 1468 Madison Avenue, New York, NY 10029, USA
[4]Lead contact
*Correspondence: wenxuan@uga.edu (W.Z.), guo-cheng.yuan@mssm.edu (G.-C.Y.), pingma@uga.edu (P.M.)
https://doi.org/10.1016/j.crmeth.2024.100886

MOTIVATION Spatial transcriptomics has spurred the development of many analytical approaches for integrating spatial and expression information. These methods often process the data differently, yield different results when applied to the same datasets, and may not be appropriate to use in every scenario. In order to develop a more generalizable approach, we developed WEST, an ensemble method for identifying spatial domains in spatial transcriptomics datasets that integrates the strengths of multiple approaches in order to provide robust, generalizable performance.

SUMMARY

Spatial transcriptomics is a groundbreaking technology, enabling simultaneous profiling of gene expression and spatial orientation within biological tissues. Yet when analyzing spatial transcriptomics data, effective integration of expression and spatial information poses considerable analytical challenges. Although many methods have been developed to address this issue, many are platform specific and lack the general applicability to analyze diverse datasets. In this article, we propose a method called the weighted ensemble method for spatial transcriptomics (WEST) that utilizes ensemble techniques to improve the performance and robustness of spatial transcriptomics data analytics. We compare the performance of WEST with six methods on both synthetic and real-world datasets. WEST represents a significant advance in detecting spatial domains, offering improved accuracy and flexibility compared to existing methods, making it a valuable tool for spatial transcriptomics data analytics.

INTRODUCTION

Spatial transcriptomics technologies have witnessed remarkable advancements in recent years, enabling the joint profiling of gene expression and precise location at single-cell resolution.[1–4] The increasing availability of commercial platforms underscores the growing significance of spatial transcriptomics in biological studies.[5,6] Spatial transcriptomics offers both unique opportunities and challenges.[7] In contrast to single-cell or bulk sequencing techniques, which may compromise the spatial structure of tissues during preparation, spatial transcriptomics stands out by reconstructing spatial domains, providing valuable biological insights.[7,8] These spatial domains prove instrumental in unraveling interactions among various organs or understanding how distinct tissue components respond to diverse microenvironments, particularly in the examination of samples with multiple organs.[8–10] Despite its potential, effective integration of spatial information with gene expression data remains a complex and ongoing challenge.

Recently, several studies have utilized unsupervised deep learning-based methods to overcome this challenge. Hu et al. introduced SpaGCN,[11] which implements a graph convolutional network as the main structure of the model and uses unsupervised deep embedding[12] to get a lower dimensional embedding that integrates the spatial and gene expression information and get the clustering results simultaneously. Ren et al. introduced SpaceFlow,[13] which uses Deep Graph Infomax[14] to encode the spatial and gene expression information. It exploits a contrastive learning strategy to learn the encoder, where the original graph is considered the positive sample, and the graph with random node permutation is considered the negative sample. Xu et al. developed Spatial Embedded Deep Representation (SEDR),[15] which implements a variational graph autoencoder[16] as the main structure, along with a masked self-supervised learning framework to integrate spatial and gene expression

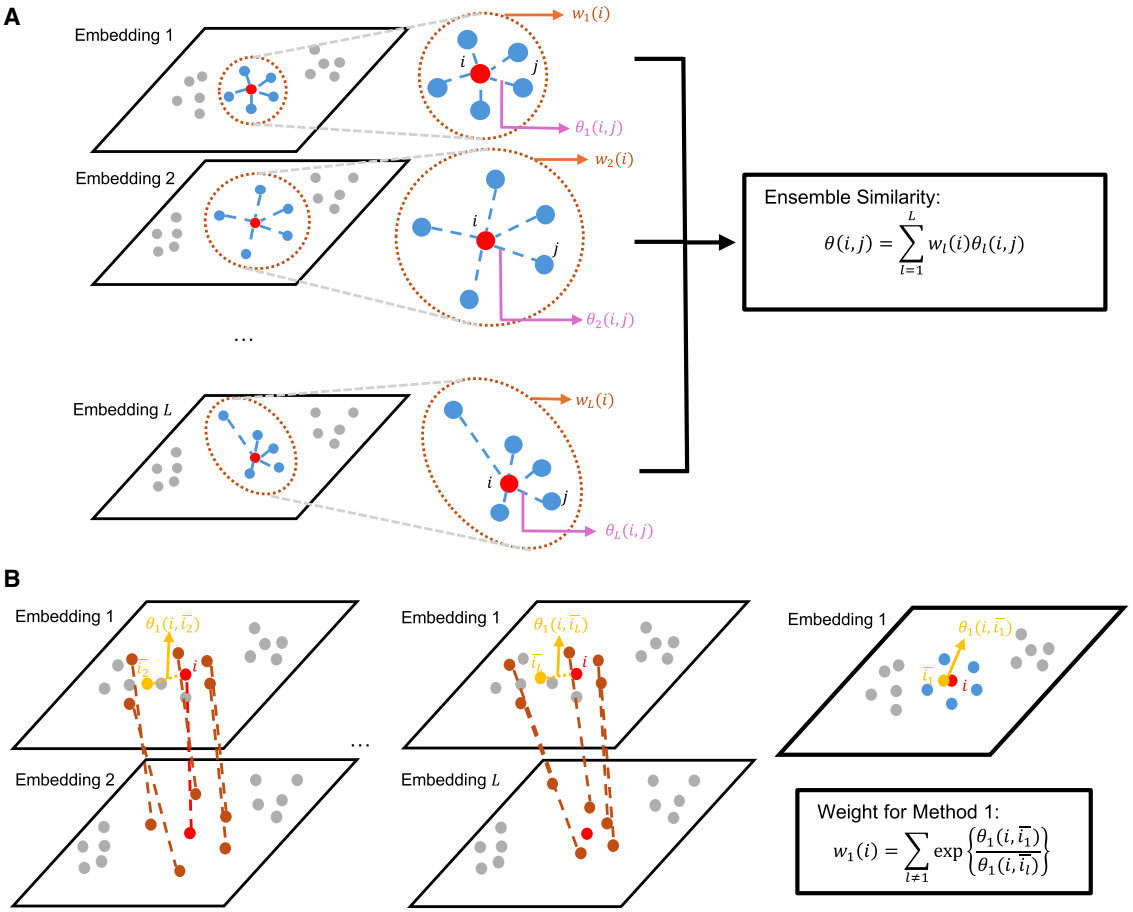


Figure 1. A summarized workflow of WEST

(A) WEST ensembles the embedding from multiple methods by calculating the ensemble similarity.

(B) A toy example of how to calculate the weight for one method.

information on a low-dimensional embedding space. Adaptive Graph Attention Auto-encoder for Spatial Transcriptomics, (STAGATE),[17] developed by Dong et al., exploits graph attention autoencoder[18] to learn a latent representation. Xu et al. introduced a customizable deep learning framework for spatial transcriptomics (DeepST),[19] which uses a complex autoencoder structure. It inserts a graph autoencoder into a denoising autoencoder to combine both the processed gene expression information and the spatial information.

While these methods have impressive performance in various datasets, it is also clear that further improvements can be made. Due to the difference in the structure of each algorithm, they integrate the gene expression and spatial information in different ways. During our experiment, we find that each method has its own strengths and weaknesses in different situations. To this end, we introduce a method called the weighted ensemble method for spatial transcriptomics (WEST), which tries to incorporate the strengths of different methods to improve model performance by integrating the embeddings from each method. By testing our method on both synthetic data and real data generated from different platforms, including 10X Genomics Visium, spatial transcriptomics (ST), and sequential fluorescence *in situ* hybridization (seqFISH), we show that our method is more robust and accurate compared with other existing benchmark methods.

RESULTS

Overview of WEST

WEST utilizes the idea of weighted nearest neighbors (WNNs)[20] to integrate the embeddings from various existing deep learning-based methods to enhance performance. The workflow of WEST is illustrated in Figure 1A. The main idea of WEST is to construct a similarity matrix that measures the similarity between each pair of two spots using the embeddings from multiple methods. Initially, WEST calculates the similarity score for the embeddings of each method, denoted by $\theta_l(i,j)$ for method l, where i and j represent two arbitrary spots on the tissue. Then, for spot i, WEST calculates a weight for method l, denoted by $w_l(i)$. Figure 1B presents a toy example to illustrate the calculation of the weight for one method at one spot. For method 1 and spot i, denote the average of its K nearest neighbors in method 1's

embedding space (marked in blue) as $\tilde{l}_1$ (marked in yellow). For spot i's K nearest neighbors in other embedding spaces (marked in brown), consider their corresponding embedding learned by method 1 (marked as brown in embedding 1). Then calculate the similarity score between the spot i and the average of its corresponding neighbors, denoted by $\theta_l(i,\tilde{l}_l)$. Since $\theta_l(i,\tilde{l}_l)$ measures the similarity between spot i and its neighbors, the ratio $\frac{\theta_l(i,\tilde{l}_1)}{\theta_l(i,\tilde{l}_l)}$ can be used to compare the embedding quality between method 1 and method l. The weight for method 1 at spot i is calculated as the sum of the aforementioned ratios to each method. The final similarity matrix is calculated using a weighted summation of the similarity of each individual method. The similarity matrix can be used for multiple subsequent analyses, like spatial domain detection, spatial network construction, and pseudo-trajectory inference. The similarity matrix can also be converted to new embedding using multidimensional scaling for further diverse analyses.

Due to the generally high cost of deep learning-based methods, ensembling excessive methods dramatically increases the computational burden. Besides, from our experiments, we find that the performance of the spatial domain identification shows a very limited improvement after combining more than two methods. Hence, in the real data analyses, we only test the performance of WEST, which integrates two methods, and present the best performance results. The individual methods are implemented in parallel to save computation time if the computation resource allows. In contrast, the computation time for the ensemble step is negligible compared with the ensembled methods. Since the input of WEST is the embeddings from different methods, whose dimension is usually under 100, the computation time is not much affected by the dimension of the data.

Application on the simulated data

To assess the performance of WEST, we applied it to several simulated datasets featuring various shapes. For the simulated dataset with a square shape, we consider an idealized situation where the tissue has a square shape containing six distinct spatial domains (Figure 2A). For the simulated dataset with an annular shape, we consider an idealized situation where the tissue has six distinct annular spatial domains (Figure 2B). The spots within each domain are randomly sampled from the tissue following a uniform distribution. As a further simplification, we assume that each spatial domain is filled with only one cell type, which is the source of variation of gene expression patterns. While these assumptions cannot fully describe the complexity of biological variations, the simplicity of the resulting spatial structure is ideal for evaluating spatial domain detection methods. The expression level of every gene is generated from a Negative Binomial distribution with different means in different domains. Based on that, we vary the dispersion of the distribution and generate 10 replications for each dispersion. More details about the simulation setting can be found in the STAR Methods.

We compare the performance of WEST with SpaceFlow,[13] SpaGCN,[11] SEDR,[15] DeepST,[19] STAGATE,[17] and Leiden.[21] Among the aforementioned methods, Leiden exploits only the gene expression level of each spot and can be considered the baseline of the subsequent comparisons. We utilize the adjusted rand index (ARI)[22] to evaluate each method's clustering performance. To make a fair comparison, in all subsequent analyses, the input data of all the methods come from the same preprocessing, and the parameters of every method are set to be the default value. Figure 2 shows the visualized result of each method on one replication. The results indicate that WEST achieves the highest ARI score and performs the best in detecting the spatial domains among all methods tested. Notably, due to the similar gene expression means of domains B and D (4 and 3, respectively), STAGATE struggles to differentiate between these two domains and clusters them as one. Conversely, while DeepST can distinguish between domains B and D, it is unable to separate domain E from domain F, which has gene expression means of 8 and 6, respectively. By integrating the two methods, WEST leverages the strengths and mitigates the weaknesses of each, accurately identifying all spatial domains and outperforming all individual methods.

The WEST results shown in Figure 2 and all subsequent analyses are the results that ensemble the two best-performing methods due to the high computation cost and limited improvement with more than two methods. To further test the performance of WEST, we also implement WEST with different combinations of methods and different numbers of methods. The detailed results can be found in Figures S4 and S5. Since the performance of WEST highly depends on the methods it ensembles, in some cases, the performance of WEST is not as good as that of other methods. However, we found that the performance is always better than that of the ensembled methods, which indicates that WEST can be considered an extra step that is able to improve the current performance. The detailed results of other combinations are presented in the supplemental information.

Application on human dorsolateral prefrontal cortex tissue

To evaluate the performance of WEST in real datasets, we analyze a publicly available 10X Genomics Visium dataset obtained from the human dorsolateral prefrontal cortex (DLPFC) study.[5] The data contain the expression levels of 33,538 genes on 12 DLPFC slices from three neurotypical adult donors at $\sim$55 μm resolution. In the original study, each sample is manually segmented into seven parts based on cytoarchitecture and selected gene markers, including six cortical layers from layer 1 (L1) to layer 6 (L6) of the human DLPFC and white matter (WM). The rest of the spots are annotated as "not known." Figure 3A shows the H&E image of one slice as an example, and the manually annotated spatial domains are shown in Figure 3B. The annotation reveals a laminar pattern for all the parts on all slices, following the order of WM to L6 to L1. We use the manual annotation as the ground-truth spatial domain to evaluate the performance of our method and compare it with six existing methods, including Leiden, SpaceFlow, SpaGCN, SEDR, DeepST, and STAGATE. As before, we use the ARI to quantify the accuracy of each method.

Figure 3C shows the boxplot of the ARI for each method on all the slices, and WEST has a higher ARI than the rest of the methods on the average of all samples. The visualization of the comparison of the performance of spatial domain detection of

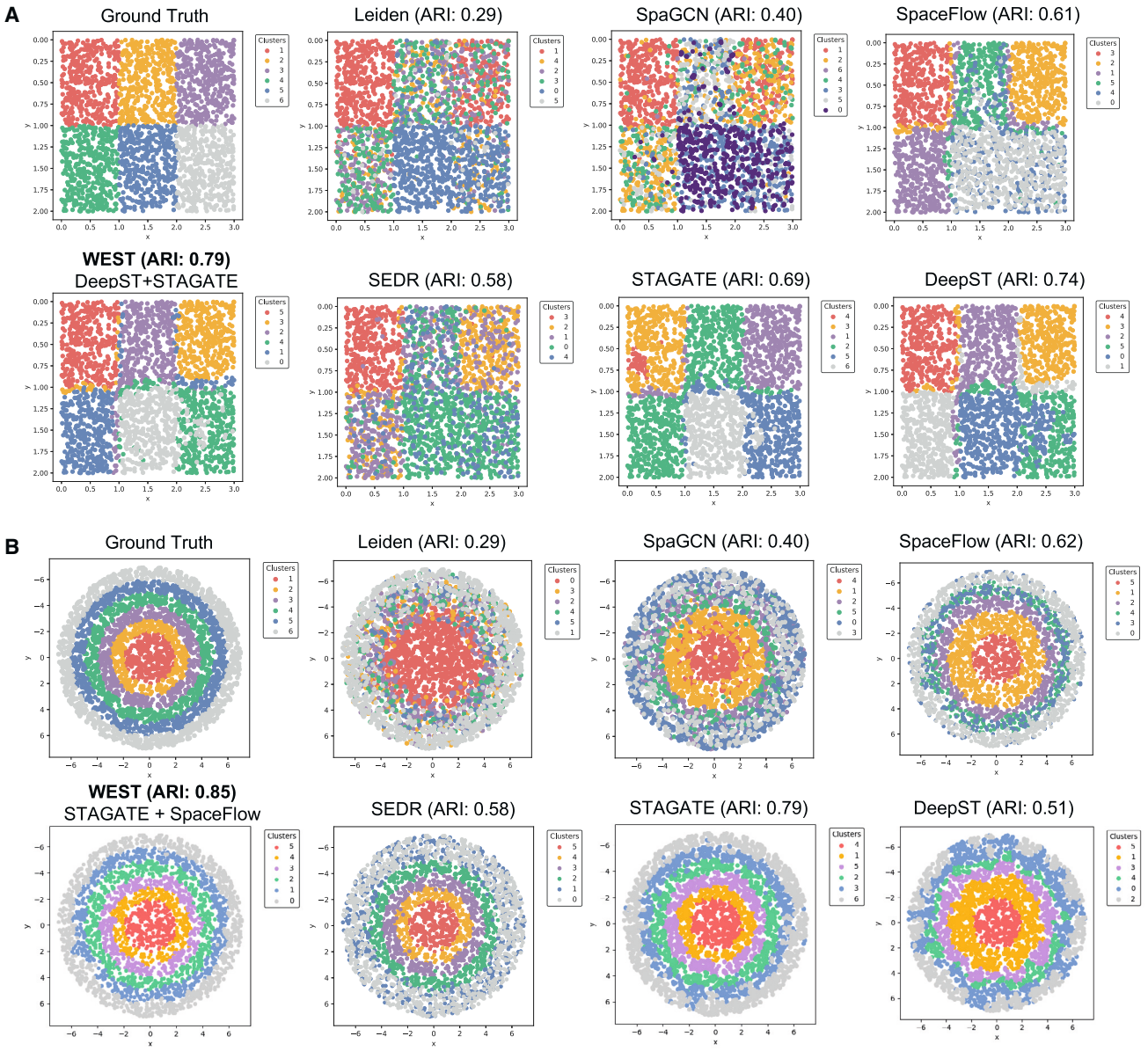


Figure 2. Performance on the simulated data

(A) Comparison of the performance of spatial domain identification between WEST and other methods on the squared simulation dataset.

(B) Comparison of the performance of spatial domain identification between WEST and other methods on the annular simulation dataset.

one sample is shown in Figure 3D. The detailed performance of WEST and other methods on the rest of the slices can be found in Figure S6 and Table S1. We present the performance of WEST that ensembles different methods in two ways of combination. Both of them achieve a higher ARI compared with all the individual methods. Specifically, the selected combinations show similar properties. Both SpaGCN and SEDR are able to capture the laminar spatial pattern but with a relatively fuzzy boundary. On the other hand, SpaceFlow and STAGATE have a much smoother boundary, but the detected domains are not consistent with the annotated domain, especially for L4 (purple)–L6 (green). After ensembling them using WEST, the result absorbs the strength from each ensembled method, which inherits the global pattern from SpaGCN and SEDR as well as the sharp boundaries from SpaceFlow and STAGATE, thereby outperforming both individual methods. Similar results can be found in other samples as well, which are presented in the supplemental information.

Application on mouse brain data

We also implement WEST on two mouse brain Visium datasets[23] with different scales that use different imaging methods. One

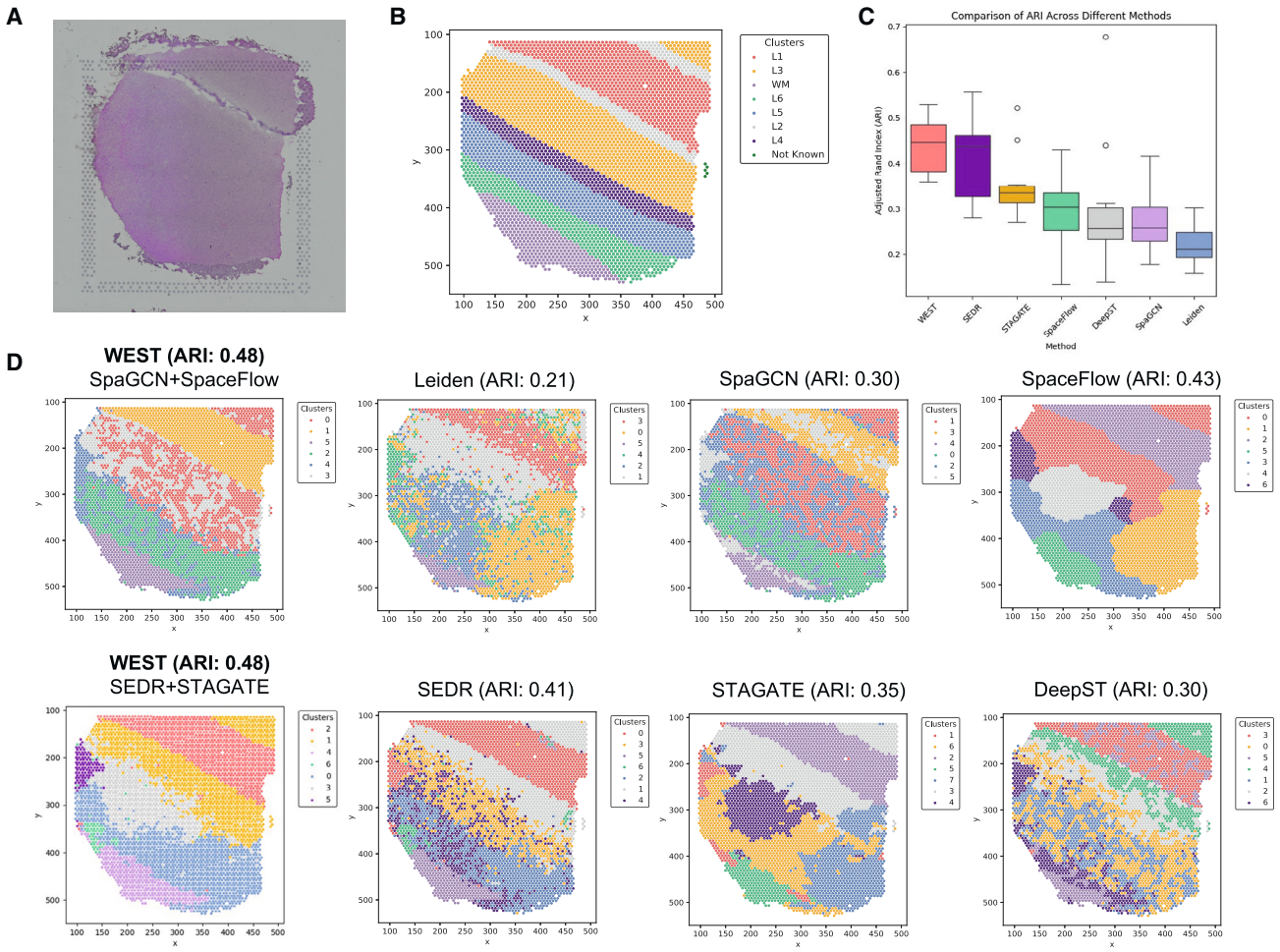


Figure 3. Performance on the DLPFC data

(A) The H&E image of one DLPFC slice.

(B) The manually annotated spatial domains.

(C) Comparison of ARI for each method on all DLPFC slices.

(D) Comparison of the performance of the spatial domain identification between WEST and other methods.

uses the fluorescence image, the other uses the H&E stain image. The two datasets measure the gene expression of different parts of the mouse brain. The detailed manually annotated spatial domain is provided in Figures 4A and 4B. Due to the differences in the techniques used in the two datasets, the size of the data is also different. The fluorescence image data contain the gene expression information of 16,562 genes on 704 spots, and there are a total of 11 annotated spatial domains. The H&E image data contain 18,078 genes' expression on 2,688 spots with 17 manually annotated spatial domains.

Figures 4C and 4D present the comparison of the performance of the spatial domain identification between WEST and the other methods for the fluorescence image and H&E image data, respectively. WEST ensembles SpaGCN and STAGATE for the fluorescence image data and ensembles Leiden and SEDR for the H&E image data. The ARI of WEST has an improvement compared to the ensembled methods and the other individual methods. Due to the integration of spatial information, some small spatial domains would be overwhelmed during the spatial smoothing process; e.g., the fiber tract (colored light gray) in the middle of the right side of the H&E image is not detected in SEDR. However, after ensembling the embedding of Leiden, which does not exploit spatial information, WEST could keep the aforementioned domain. Besides, WEST can well identify complex spatial patterns while keeping the boundary smooth, which makes the identified spatial domains more consistent with the annotated domains.

Application on the mouse embryo data

To test the performance of WEST on single-cell resolution ST data, we apply our method to a published seqFISH mouse embryo dataset.[24] The experiment detects 351 barcoded genes at 19,416 cells, which are assigned to 24 distinct cell types based on single-cell references, as shown in Figure 5A. Based on the anatomic structures and gastrulation atlas[25] as the reference,

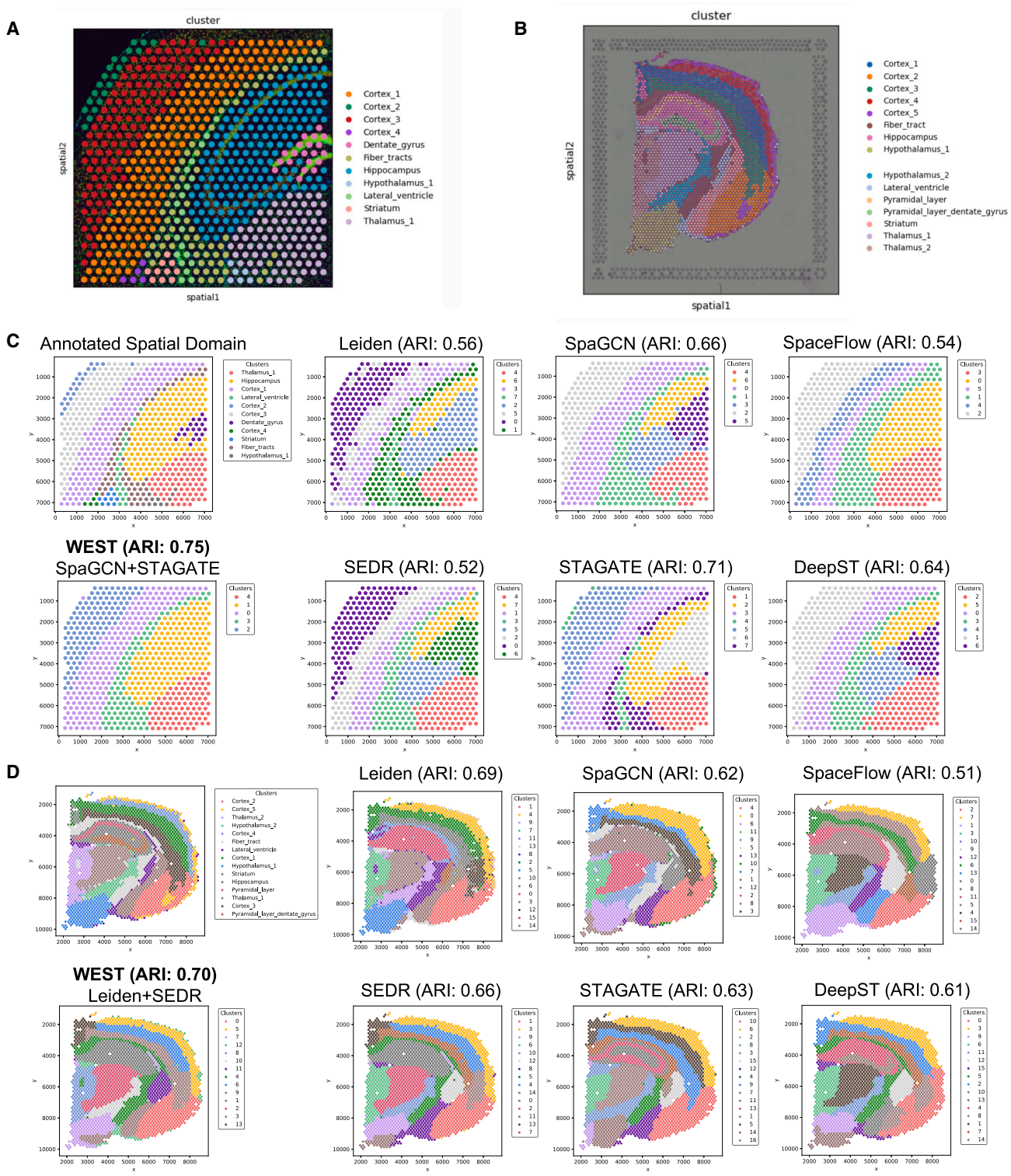


Figure 4. Performance on the Visium mouse brain data

Two Visium datasets of mouse brains were used to test the performance of WEST.

(A) The manually annotated domains of the mouse brain tissue with a fluorescence image.

(B) The manually annotated domains of the mouse brain tissue with an H&E image.

(C) Comparison of the performance between WEST and the other methods on the fluorescence data.

(D) Comparison of the performance between WEST and the other methods on the H&E data.

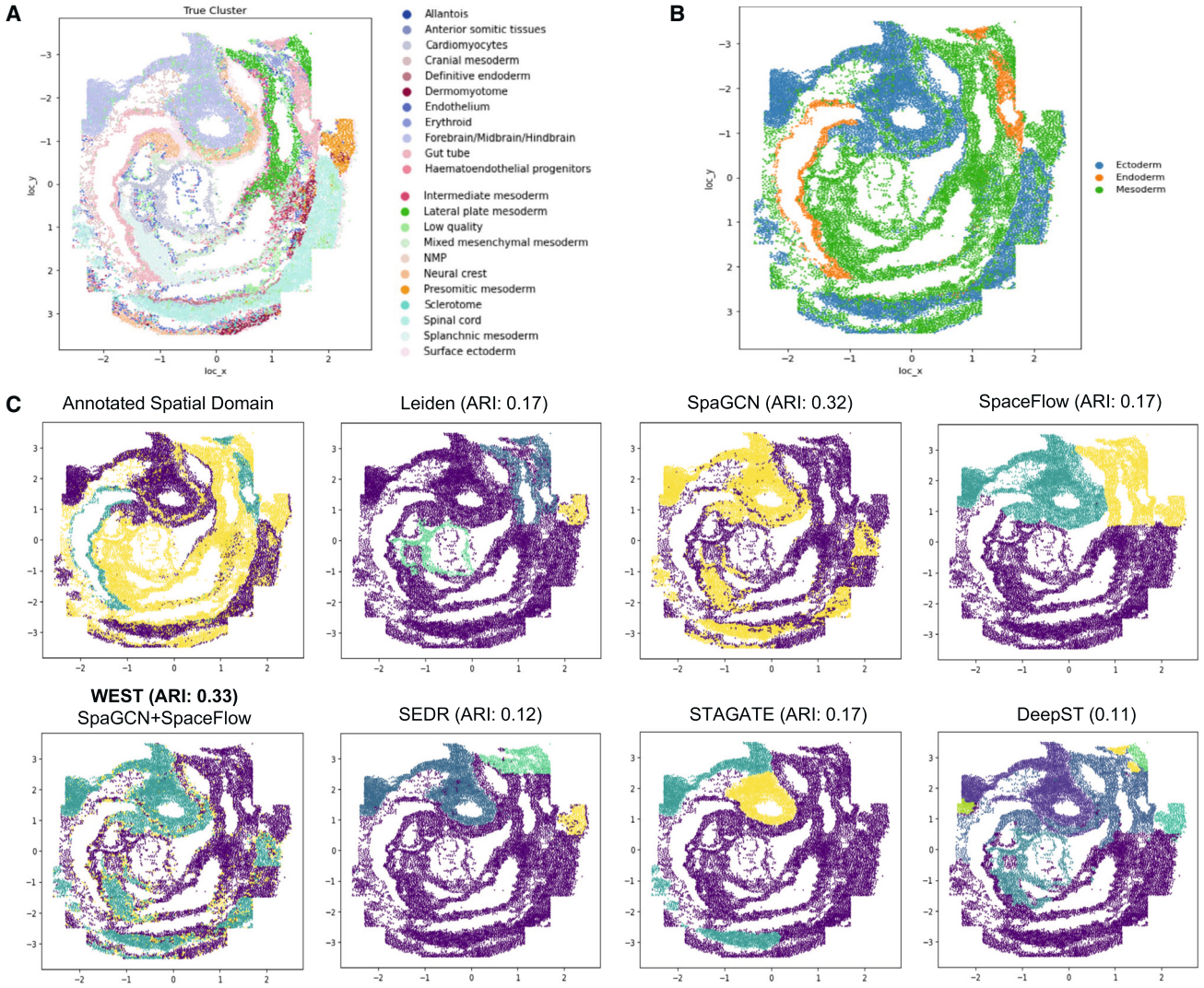


Figure 5. Analysis of the seqFISH mouse embryo data

(A) The manually annotated cell types for the mouse embryo.

(B) The annotated three germ layers (ectoderm, mesoderm, and endoderm).

(C) The comparison of the performance of the spatial domain identification between WEST and the other methods.

the embryo is manually segmented into three distinct germ layers: ectoderm, endoderm, and mesoderm (Figure 5B).

We consider the three germ layers as the ground truth spatial domains and use ARI to evaluate the performance of WEST and other existing methods. As shown in Figure 5C, WEST has the highest ARI compared to other methods. Since the seqFISH data are at single-cell resolution, and the spatial domain is annotated based on the cell type, which does not exploit the spatial information, the annotated domains are not continuous, and parts of the different domains overlap each other. Hence, the performance of the existing methods has an unsatisfactory performance in most areas. Compared with them, WEST has the highest ARI and identifies the spatial domains that are most consistent with the annotated domains. It proves the generalizability of WEST for different types of datasets.

Application on human pancreatic cancer data

To evaluate the performance of WEST on the low-resolution dataset, we performed experiments on the ST dataset for the human pancreatic ductal adenocarinoma sample.[26] The dataset provides a comprehensive view of gene expression within the tissue, captured at a spatial resolution of 100 μm per spot. The study detects 3,913 barcoded genes at 428 spots. Based on the H&E image, four clusters are defined—cancer region, nonmalignant duct epithelium, stroma, and normal pancreatic tissue—based on distinct histological features, which are presented in Figure 6A.

We regard the four histologically distinct regions as ground-truth spatial domains and use ARI to evaluate the performance of WEST and six other methods, as shown in Figures 6B and 6C. Due to the size of the spot in ST, the boundary of the

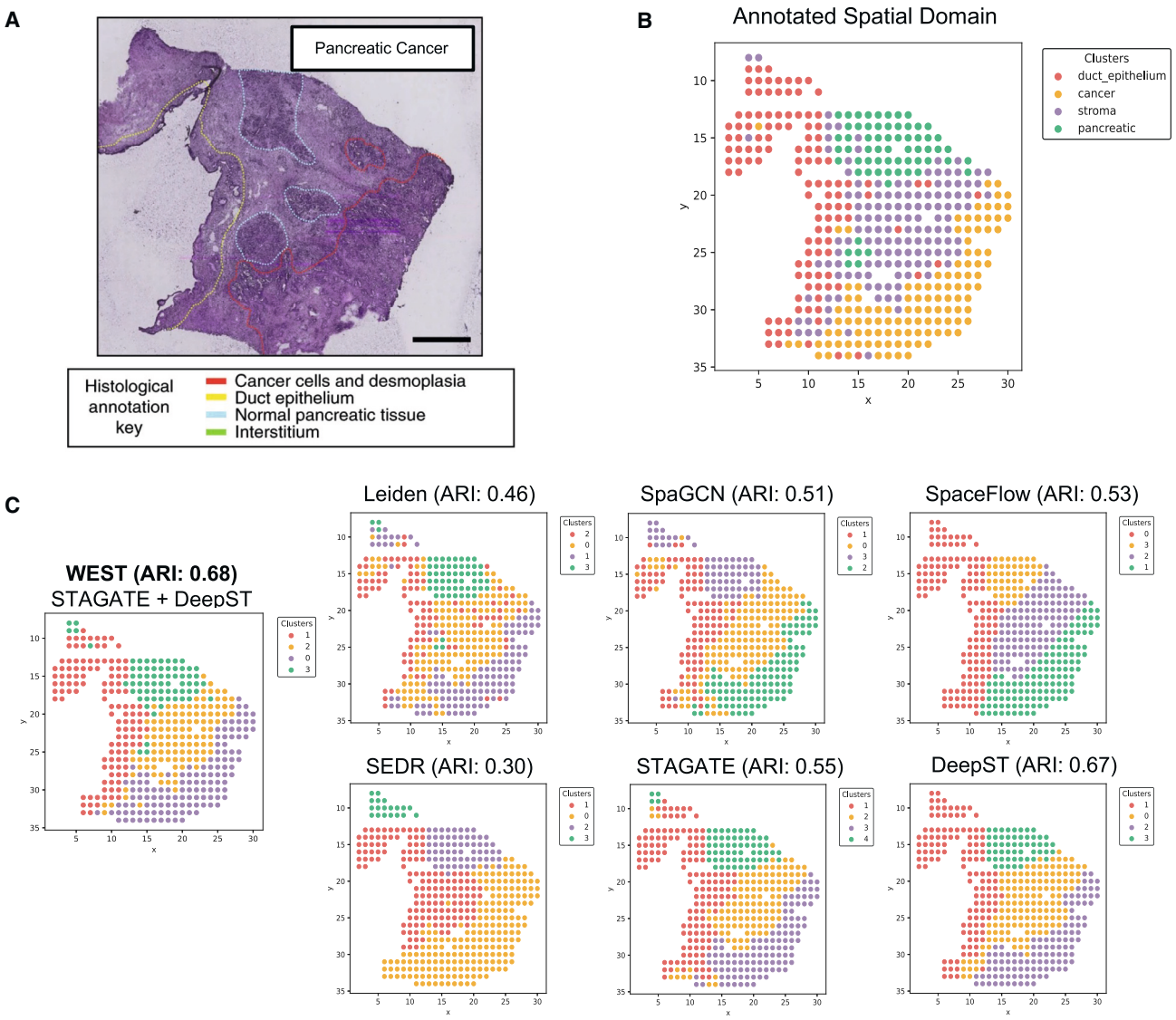


Figure 6. Performance on the human pancreatic cancer data

(A) The H&E image of the human pancreatic cancer tissue and histological annotation keys.

(B) The annotated spatial domains.

(C) Comparison of the performance of the spatial domain identification between WEST and other methods.

annotated spatial domains is not smooth and not even continuous at some point. Therefore, the boundary of the spatial domains detected by the existing methods is usually oversmoothed. WEST solves the problem by integrating the spatial pattern from the result of STAGATE and DeepST, providing the most consistent spatial domains with the manually annotated domains with the highest ARI.

DISCUSSION

As the field of ST progresses, tools designed to analyze ST data are rapidly evolving. The methodologies behind these tools span a spectrum from time-tested statistical techniques to cutting-edge deep-learning approaches.[27–30] Given the inherent differences in their underlying structures, different methods can yield different outcomes even when applied to the same dataset. Thus, we introduce WEST, which integrates the embeddings from multiple methods, harnessing their collective strengths while offsetting their individual weaknesses. The idea of the ensemble method can effectively reduce the bias and variance in each individual method and improve the generalization, robustness, and accuracy.[31–33] The results indicate that WEST is more generally applicable and outperforms individual methods when applied to varied datasets. Furthermore, WEST is able to integrate embeddings from spaces of any dimensionality and map the weighted similarity matrix to a target space, ensuring

the preservation of most of the information. This adaptability guarantees that WEST can process a wide array of ST data, capturing the complex structures of tissues with precision.

WEST has demonstrated superior results on both sequence-based data (DLPFC from Visium) and image-based data (mouse embryo from seqFISH). The structures reconstructed by WEST align closely with annotated counterparts, and the performance from the ensembled embedding is much better than each individual method. The foundational concept of WEST holds great potential for integration with more embedding methods or for exploration in diverse ST tasks. As the landscape of data analysis methods evolves, we anticipate growing relevance of our approach.

Limitations of the study

The concept behind WEST is akin to ensemble learning, which leverages the strengths of each individual algorithm. While WEST can enhance performance by borrowing information from different individual algorithms, it does not directly process the data, which means its improvement over individual algorithms depends on the performance of the ensembled methods. Additionally, if certain methods in the ensemble perform poorly, then the overall performance may be unsatisfactory. Essentially, the protocol can be viewed as a weighted average of the individual algorithms in the ensemble, making the results more robust but still susceptible to outliers.

RESOURCE AVAILABILITY

Lead contact

Requests for further information, resources, and reagents should be directed to and will be fulfilled by the lead contact, Ping Ma (pingma@uga.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This paper analyzes existing, publicly available data. The accession numbers for the datasets are listed in the key resources table. All original code has been deposited on GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table.

ACKNOWLEDGMENTS

The authors would like to acknowledge support from the US National Science Foundation under grants DMS-1925066, DMS-1903226, DMS-2124493, DMS-2311297, DMS-2319279, and DMS-2318809 and the US National Institute of Health under grants R01GM152814, RF1MH128970, and RF1MH133703.

AUTHOR CONTRIBUTIONS

Conceptualization, G.-C.Y., P.M., W.Z., and H.C.; methodology, P.M., H.C., and J.C.; coding, H.C., S.W., and J.C.; formal analysis, H.C., J.C, S.W., G.-C.Y., and P.M.; writing – original draft, J.C.; writing – review & editing, J.C., G.-C.Y., S.W., P.M, H.C., and W.Z.; supervision, G.-C.Y., P.M., and W.Z.; funding acquisition, G.-C.Y., P.M., and W.Z.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Data preprocessing
  - Data embedding
  - Ensemble embeddings
  - Subsequent analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Simulation setting
  - Methods comparison

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2024.100886.

Received: January 2, 2024
Revised: June 29, 2024
Accepted: October 15, 2024
Published: November 7, 2024

REFERENCES

1. Moses, L., and Pachter, L. (2022). Museum of spatial transcriptomics. Nat. Methods *19*, 534–546. https://doi.org/10.1038/s41592-022-01409-2.
2. Du, J., Yang, Y.-C., An, Z.-J., Zhang, M.-H., Fu, X.-H., Huang, Z.-F., Yuan, Y., and Hou, J. (2023). Advances in spatial transcriptomics and related data analysis strategies. J. Transl. Med. *21*, 330. https://doi.org/10.1186/s12967-023-04150-2.
3. Rao, A., Barkley, D., França, G.S., and Yanai, I. (2021). Exploring tissue architecture using spatial transcriptomics. Nature *596*, 211–220. https://doi.org/10.1038/s41586-021-03634-9.
4. Duan, H., Cheng, T., and Cheng, H. (2023). Spatially resolved transcriptomics: advances and applications. Blood Sci. *5*, 1–14. https://doi.org/10.1097/BS9.0000000000000141.
5. Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippani, M., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat. Neurosci. *24*, 425–436. https://doi.org/10.1038/s41593-020-00787-0.
6. Janesick, A., Shelansky, R., Gottscho, A.D., Wagner, F., Rouault, M., Beliakoff, G., de Oliveira, M.F., Kohlway, A., Abousoud, J., Morrison, C.A., et al. (2022). High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.06.510405.
7. Dries, R., Chen, J., del Rossi, N., Khan, M.M., Sistig, A., and Yuan, G.-C. (2021). Advances in spatial transcriptomic data analysis. Genome Res. *31*, 1706–1718. https://doi.org/10.1101/gr.275224.121.
8. Zhu, Q., Shah, S., Dries, R., Cai, L., and Yuan, G.-C. (2018). Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. Nat. Biotechnol. https://doi.org/10.1038/nbt.4260.
9. Edsgärd, D., Johnsson, P., and Sandberg, R. (2018). Identification of spatial expression trends in single-cell gene expression data. Nat. Methods *15*, 339–342. https://doi.org/10.1038/nmeth.4634.
10. Whiteside, T.L. (2008). The tumor microenvironment and its role in promoting tumor growth. Oncogene *27*, 5904–5912. https://doi.org/10.1038/onc.2008.271.
11. Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D.J., Lee, E.B., Shinohara, R.T., and Li, M. (2021). SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially

variable genes by graph convolutional network. Nat. Methods *18*, 1342–1351. https://doi.org/10.1038/s41592-021-01255-8.

12. Xie, J., Girshick, R., and Farhadi, A. (2016). Unsupervised Deep Embedding for Clustering Analysis. In Proceedings of The 33rd International Conference on Machine Learning (PMLR)), pp. 478–487.
13. Ren, H., Walker, B.L., Cang, Z., and Nie, Q. (2022). Identifying multicellular spatiotemporal organization of cells with SpaceFlow. Nat. Commun. *13*, 4076. https://doi.org/10.1038/s41467-022-31739-w.
14. Veličković, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., and Hjelm, R.D. (2018). Deep Graph Infomax. Preprint at arXiv. https://doi.org/10.48550/arXiv.1809.10341.
15. Xu, H., Fu, H., Long, Y., Ang, K.S., Sethi, R., Chong, K., Li, M., Uddamvathanak, R., Lee, H.K., Ling, J., et al. (2024). Unsupervised spatially embedded deep representation of spatial transcriptomics. Genome Med. *16*, 12. https://doi.org/10.1186/s13073-024-01283-x.
16. Kipf, T.N., and Welling, M. (2016). Variational Graph Auto-Encoders. Preprint at arXiv. https://doi.org/10.48550/arXiv.1611.07308.
17. Dong, K., and Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nat. Commun. *13*, 1739. https://doi.org/10.1038/s41467-022-29439-6.
18. Salehi, A., and Davulcu, H. (2019). Graph Attention Auto-Encoders. Preprint at arXiv. https://doi.org/10.48550/arXiv.1905.10715.
19. Xu, C., Jin, X., Wei, S., Wang, P., Luo, M., Xu, Z., Yang, W., Cai, Y., Xiao, L., Lin, X., et al. (2022). DeepST: identifying spatial domains in spatial transcriptomics by deep learning. Nucleic Acids Res. *50*, e131. https://doi.org/10.1093/nar/gkac901.
20. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
21. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.
22. Rand, W.M. (1971). Objective Criteria for the Evaluation of Clustering Methods. J. Am. Stat. Assoc. *66*, 846–850. https://doi.org/10.1080/01621459.1971.10482356.
23. Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A.C., Kuemmerle, L.B., Rybakov, S., Ibarra, I.L., Holmberg, O., Virshup, I., et al. (2022). Squidpy: a scalable framework for spatial omics analysis. Nat. Methods *19*, 171–178. https://doi.org/10.1038/s41592-021-01358-2.
24. Lohoff, T., Ghazanfar, S., Missarova, A., Koulena, N., Pierson, N., Griffiths, J.A., Bardot, E.S., Eng, C.-H.L., Tyser, R.C.V., Argelaguet, R., et al. (2020). Highly multiplexed spatially resolved gene expression profiling of mouse organogenesis. Preprint at bioRxiv. https://doi.org/10.1101/2020.11.20.391896.
25. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490–495. https://doi.org/10.1038/s41586-019-0933-9.
26. Moncada, R., Barkley, D., Wagner, F., Chiodin, M., Devlin, J.C., Baron, M., Hajdu, C.H., Simeone, D.M., and Yanai, I. (2020). Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat. Biotechnol. *38*, 333–342. https://doi.org/10.1038/s41587-019-0392-8.
27. Liu, Y., Chen, Y., Lu, H., Zhong, W., Yuan, G.-C., and Ma, P. (2024). Orthogonal multimodality integration and clustering in single-cell data. BMC Bioinf. *25*, 164. https://doi.org/10.1186/s12859-024-05773-y.
28. Zhao, E., Stone, M.R., Ren, X., Guenthoer, J., Smythe, K.S., Pulliam, T., Williams, S.R., Uytingco, C.R., Taylor, S.E.B., Nghiem, P., et al. (2021). Spatial transcriptomics at subspot resolution with BayesSpace. Nat. Biotechnol. *39*, 1375–1384. https://doi.org/10.1038/s41587-021-00935-2.
29. Zeng, Z., Li, Y., Li, Y., and Luo, Y. (2022). Statistical and machine learning methods for spatially resolved transcriptomics data analysis. Genome Biol. *23*, 83. https://doi.org/10.1186/s13059-022-02653-7.
30. Sun, S., Zhu, J., and Zhou, X. (2020). Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. Nat. Methods *17*, 193–200. https://doi.org/10.1038/s41592-019-0701-7.
31. Zhou, Z.-H. (2012). Ensemble Methods: Foundations and Algorithms (CRC Press)).
32. Polikar, R. (2012). Ensemble Learning. In Ensemble Machine Learning: Methods and Applications, C. Zhang and Y. Ma, eds. (Springer), pp. 1–34. https://doi.org/10.1007/978-1-4419-9326-7_1.
33. Sagi, O., and Rokach, L. (2018). Ensemble learning: A survey. WIREs Data Mining and Knowledge Discovery *8*, e1249. https://doi.org/10.1002/widm.1249.
34. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.
35. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. *2008*, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.
36. McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at arXiv. https://doi.org/10.48550/arXiv.1802.03426.
37. Kruskal, J.B., and Wish, M. (1978). Multidimensional Scaling (SAGE).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
10X Genomics Visium data of human dorsolateral prefrontal cortex	Maynard et al.[5]	http://research.libd.org/spatialLIBD/
10X Genomics Visium fluorescence image data of mouse brain	Palla et al.[23]	https://support.10xgenomics.com/spatial-gene-expression/datasets
10X Genomics Visium H&E image data of mouse brain	Palla et al.[23]	https://support.10xgenomics.com/spatial-gene-expression/datasets
seqFISH data of mouse embryo	Lohoff et al.[24]	https://crukci.shinyapps.io/SpatialMouseAtlas/
Spatial Transcriptomics (ST) data of human pancreatic cancer	Moncada et al.[26]	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111672
Software and algorithms		
WEST	This paper	https://github.com/JiazhangCai/WEST/tree/main https://doi.org/10.5281/zenodo.14004971
SpaGCN	Hu et al.[11]	https://github.com/jianhuupenn/SpaGCN
SpaceFlow	Ren et al.[13]	https://github.com/hongleir/SpaceFlow
SEDR	Fu et al.[15]	https://github.com/JinmiaoChenLab/SEDR
DeepST	Xu et al.[19]	https://github.com/JiangBioLab/DeepST
STAGATE	Dong and Zhang[17]	https://github.com/QIFEIDKN/STAGATE
WNN	Hao et al.[20]	https://github.com/dylkot/pyWNN
scanpy	Wolf et al.[34]	https://github.com/scverse/scanpy
squidpy	Palla et al.[23]	https://github.com/scverse/squidpy

METHOD DETAILS

Data preprocessing

The raw data comprises two components: the gene expression level data $\boldsymbol{X} \in \mathbb{R}^{N_0 \times p_0}$, which records the counts of p_0 expressed genes in N_0 spots, and the spatial information data $\boldsymbol{S} \in \mathbb{R}^{N_0 \times 2}$, which records the 2-D coordinates of all N_0 spots. The interpretation of the term "spot" varies depending on the techniques employed. For example, in the DLPFC dataset, the "spot" corresponds to the location on the Visium panel, as used in the 10x Genomics technology. In the mouse embryo seqFISH dataset, the "spot" represents a detected single cell. To improve the quality of the data, we filter out genes that are expressed in less than 100 spots and spots detected in less than three genes. The number of genes and spots that remain after filtering are denoted by p and N, respectively.

Since the distribution of cells is not uniform across the tissue, it is challenging to relate gene expression in a spot to a known reference, such as single-cell RNA or marker gene profiles.[7] Therefore, our analysis focuses primarily on spatial domain segmentation rather than cell type identification, and there is no need to consider the deconvolution process in subsequent analysis.

Data embedding

We consider a set of N spatially distinct spots, indexed by $i = 1,\ldots,N$. Each spot i is associated with a two-dimensional spatial coordinate $\boldsymbol{s}_i = (s_{i1}, s_{i2})$ and a p-dimensional vector of filtered gene expression levels $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$. The objective of the ensemble method is to learn an embedding function $\mathcal{E} : \mathbb{R}^{N\times p} \times \mathbb{R}^{N\times 2} \to \mathbb{R}^{N\times q}$ that integrates both the spatial and gene expression information, projecting the data into a q-dimensional space. Specifically, the embedding should capture the underlying spatial structure of the tissue while preserving the differences in gene expression levels. In this paper, we mainly consider five deep learning-based methods: SpaGCN, SpaceFlow, SEDR, DeepST, and STAGATE. Besides, we include Leiden, which only exploits the gene expression information, as the benchmark. These methods output a q-dimensional embedding, which is set at 50 as the default, as the input for WEST. The detailed setting of each method is described in the Methods Comparison part.

Ensemble embeddings

Given the embeddings from L embedding methods, denoted as $\boldsymbol{e}^{(1)},\ldots,\boldsymbol{e}^{(L)}$, WEST integrates them by constructing a weighted similarity matrix that indicates the pairwise similarity between every two spots in the dataset. This weighted similarity matrix enables us to create a weighted nearest-neighbor graph for clustering or an ensemble embedding for other analyses.

To construct this matrix, we first generate the K-nearest neighbor network using the embeddings $\{\boldsymbol{e}^{(l)}\}_{l=1}^{L}$, where the value of K is predetermined. Let $\mathcal{X}^{(l)}$ denote the embedding space of method l. In $\mathcal{X}^{(l)}$, we identify the K closest neighbors of a given spot i as $knn_{i,1}^{(l)},..,knn_{i,1K}^{(l)}$, where $knn_{i,1}^{(l)}$ is the nearest neighbor of spot i, and $knn_{i,1K}^{(l)}$ is the farthest neighbor among its K neighbors. We then define the similarity score between spot i and spot j in $\mathcal{X}^{(l)}$ by:

$$\theta_l(i,j) = \exp\left\{-\frac{d\left(e_i^{(l)},e_j^{(l)}\right)-d\left(e_i^{(l)},e_{knn_{i,1}^{(l)}}^{(l)}\right)}{d\left(e_i^{(l)},e_{knn_{i,K}^{(l)}}^{(l)}\right)-d\left(e_i^{(l)},e_{knn_{i,1}^{(l)}}^{(l)}\right)}\right\}, l = 1,\ldots,L$$

where $e_i^{(l)}$ and $e_j^{(l)}$ represent for the embedding vector of spot i and j in $\mathcal{X}^{(l)}$. $d(\cdot,\cdot)$ represents a distance metric, which is the Euclidean distance as the default. The similarity score $\theta_l(i,j)$ is defined as the difference between the distance from spot i to spot j and its nearest neighbor, normalized by the difference between the distance from spot i to its nearest and furthest neighbor. Then the similarity score is scaled by the exponential kernel, which makes the score go to one when spot j is close to spot i 's nearest neighbor and goes to 0 when spot j is far away from spot i in $\mathcal{X}^{(l)}$.

To ensemble multiple methods while retaining the strengths of each, we define a weight function $w_l(i)$ for each method l at spot i on the tissue, satisfying $\sum_{l=1}^{L} w_l(i)$. WEST determines the weight function through the performance of how well the embedded data reveals significant clustering patterns. To compare the clustering performance between two embedding methods l_1 and l_2, we introduce the average neighbor vector in $\mathcal{X}^{(l_1)}$ using the neighbor set from another method l_2, denoted by $\bar{e}_{i,l_2}^{(l_1)} = \frac{1}{K}\sum_{k=1}^{K} e_{knn_{i,k}^{(l_2)}}^{(l_1)}$. We calculate the similarity score from $e_i^{(l_1)}$ to $\bar{e}_{i,l_2}^{(l_1)}$ in $\mathcal{X}^{(l_1)}$, which we denote by $\theta_{l_1}(i,\bar{i}_{l_2})$. This similarity score is small when the average distance between spot i and its K neighbors in method l_2 is small. We can use the ratio of $\theta_{l_1}(i,\bar{i}_{l_1})$ to $\theta_{l_1}(i,\bar{i}_{l_2})$ to compare the performance of method l_1 to method l_2. This ratio is large if the spot i together with its K neighbors detected in method l_1 show a stronger clustering pattern than method l_2 in $\mathcal{X}^{(l_1)}$. We define the weight function for method l at spot i as the softmax transformation of the ratios from all embedding methods:

$$w_l(i) = \frac{\sum_{l_2\neq l}\exp\left\{\frac{\theta_l(i,\bar{i}_l)}{\theta_l(i,\bar{i}_{l_2})+\epsilon}\right\}}{\sum_{l_1}\sum_{l_2\neq l_1}\exp\left\{\frac{\theta_{l_1}(i,\bar{i}_{l_1})}{\theta_{l_1}(i,\bar{i}_{l_2})+\epsilon}\right\}},$$

where ϵ is a small constant to stabilize the computation. With the weight function, we can combine the similarity score from different methods and get a weighted similarity matrix, where its (i,j)-th entry is

$$\theta(i,j) = \sum_{l-1}^{L} w_l(i)\theta_l(i,j).$$

Notably, for two different spots i and j, since the similarity score is calculated based on their neighbors respectively, $\theta(i,j)$ is usually not equal to $\theta(j,i)$. To make the similarity matrix symmetric, we use the average of the two-way similarity score as the final result: $\tilde{\theta}(i,j) = \frac{1}{2}(\theta(i,j)+\theta(j,i))$.

Subsequent analysis

The weighted similarity matrix combines information from both methods to produce a new neighbor distance. Using this matrix, we construct a weighted nearest neighbors graph in our analysis. This graph integrates the strengths of each method, uncovering clustering patterns that would have been missed if only one method had been used. The graph can be used for various bioinformatics analyses, including tissue segmentation, pseudo-trajectory inference, and more, without being limited by the dimensions of the embedding space.

For example, we apply the Leiden method[35] to the weighted graph for spatial domain segmentation. Furthermore, by replacing the original graph with the constructed weighted nearest neighbors graph, we can also generate a lower-dimensional embedding using UMAP[36] or multidimensional scaling methods.[37]

QUANTIFICATION AND STATISTICAL ANALYSIS

Simulation setting

The simulation dataset includes two types of spatial structures: squared and annular shapes. Each tissue consists of six clusters. The squared design features six square regions, while the annular design contains six annular spatial domains.

First, we generate synthetic data on a rectangular tissue with a total of 2,400 spots, randomly assigned across the entire area. This synthetic tissue is then divided into six domains of approximately equal areas (Figure 2A). For the annular tissue, we generate synthetic data on a round tissue with a total of 3,082 spots, which is then divided into six regions (Figure 2B). The first cluster is an annular region with a radius of 2 units. The subsequent clusters are annular regions, with each region's inner and outer radii progressively increasing. Specifically, the second cluster spans from an inner radius of 2 units to an outer radius of 3 units, the third cluster spans from 3 to 4 units, the fourth from 4 to 5 units, the fifth from 5 to 6 units, and the sixth from 6 to 7 units. For each cluster, the number of spots is proportional to the area of the region. In the annular region of the first cluster, spots are uniformly distributed up to the radius of 2 units. In the annular regions of the subsequent clusters, spots are uniformly distributed within their respective inner and outer radius.

Next, we generate the expression levels of 400 pseudo genes, consisting of 200 signal genes and 200 noise genes. The exact count of each signal gene at each spot is generated using a negative binomial distribution. For the six regions of the rectangular tissue, the mean values of signal gene expression are {1,4,2,3,8,6}, while for the six regions of the round tissue, the mean values are {1,2,3,5,7,9}. The dispersion values of the negative binomial distribution are kept the same across all clusters. We simulate five scenarios, each with different dispersion values for the signal genes, specifically {0.1,0.2,0.3,0.4,0.5}. The noisy gene expression is generated from a normal distribution with a mean of 0 and a standard deviation of 0.8. The visualization of the change of gene expression using UMAP is presented in Figure S1, where the first row is for the squared simulation data and the second row is for the annular simulation data. It indicates that as the dispersion increases, different clusters get closer to each other, and there is more overlap between different clusters.

We compared WEST with six other popular methods on the simulated data. The visualization of the embedding learned by each algorithm using UMAP is presented in the second and fourth rows in Figure S2. Note that Leiden's embedding is attained by the PCA results of the gene expression. We can see that WEST shows a more separate pattern of ensembled embedding, compared with STAGATE's and DeepST's embeddings. As for the annular shape shown in Figure S3, the dispersion parameter of each domain is set as 0.3. When WEST achieves the best performance (ARI 0.85), the clustering structure of the ensemble embedding is more distinct. The overlap between the two clusters is less compared with the embeddings of SpaceFlow and STAGATE. The proposed ensemble learning method can improve the methods being ensembled by leveraging the strengths and mitigating the weaknesses. In addition to ensembling the best-performed two methods, we also show the performance of WEST ensembled by all the combinations of two or three deep learning-based methods in Figures S4 and S5, and they outperformed the ensembled methods.

Methods comparison

We compare the performance of six popular spatial domain detection methods with WEST on synthetic and real datasets, including Leiden, SpaGCN, SpaceFlow, SEDR, STAGATE, and DeeST, to evaluate the quality of our proposed method. Leiden is a network community detection algorithm that clusters the data by maximizing the modularity of the constructed spatial network. SpaGCN combines spatial and gene expression information by constructing a weighted network with attributes and uses this network as the graph structure in the graph convolutional network to get a lower dimensional embedding. SpaceFlow transforms the spatial information into an unweighted network, where the edge between two spots represents whether they are spatially far away. Then, SpaceFlow uses a discriminator to train a graph convolutional network and get the low dimensional embedding. Similar to SpaceFlow, SEDR also constructs an unweighted network to present the spatial relationship of every spot to help further fit the model. The difference is that instead of using the graph convolutional network, SEDR uses the variational graph autoencoder to get the lower dimensional embedding. STAGATE utilizes a graph attention module to learn how much attention should be drawn to each location when learning the embedding. DeepST constructs a complex autoencoder structure that inserts a graph autoencoder into a denoising autoencoder to integrate the processed gene expression and spatial information.

Among the methods we introduced above, some use the H&E stain image to assist the spatial domain detection. Since the H&E stain image is not always provided with the dataset, and the quality of the image is different due to the difference in the environment of different labs, different platforms, and different techniques to be used, we do not include the H\&E image in our experiments.

During the composition, all the parameters in the algorithms are set to be the default value. To make a fair comparison, the parameters used for adjusting the number of clusters are set to get a number of clusters close to the annotation or ground truth (for simulation).