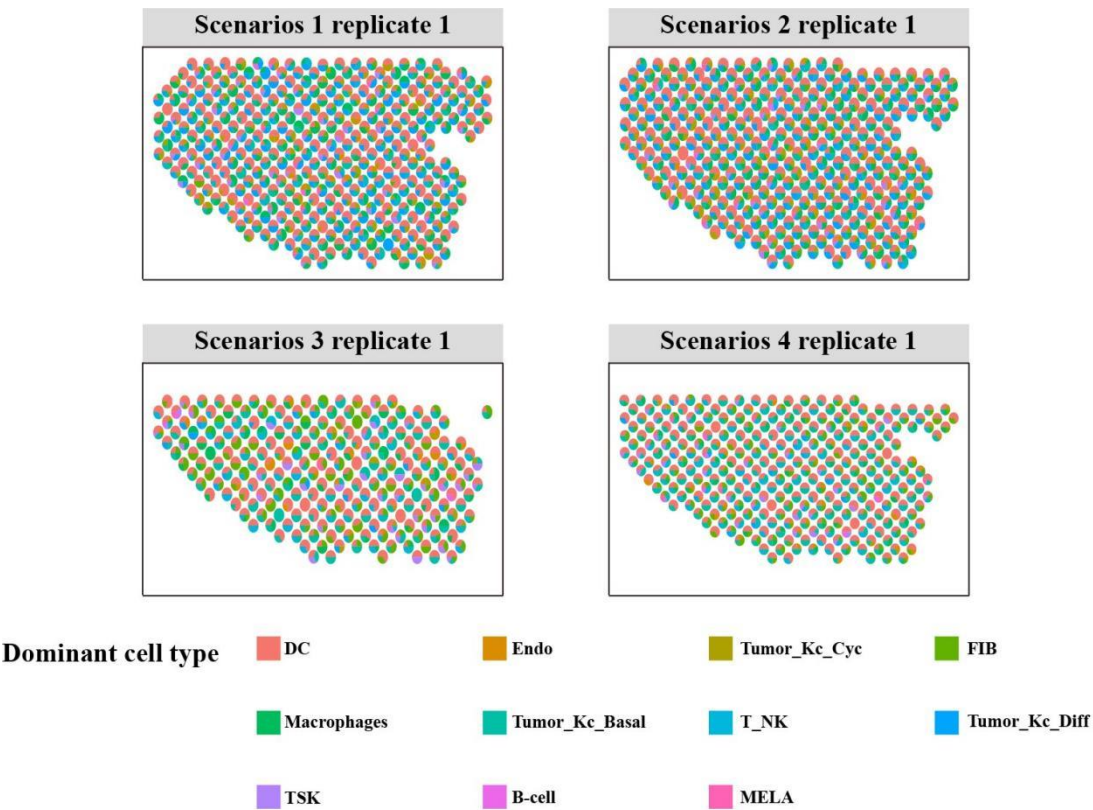


Supplementary Figures

Supplementary Fig. 1. Composition of cell types for each spot in replicate 1 of the four scenarios.

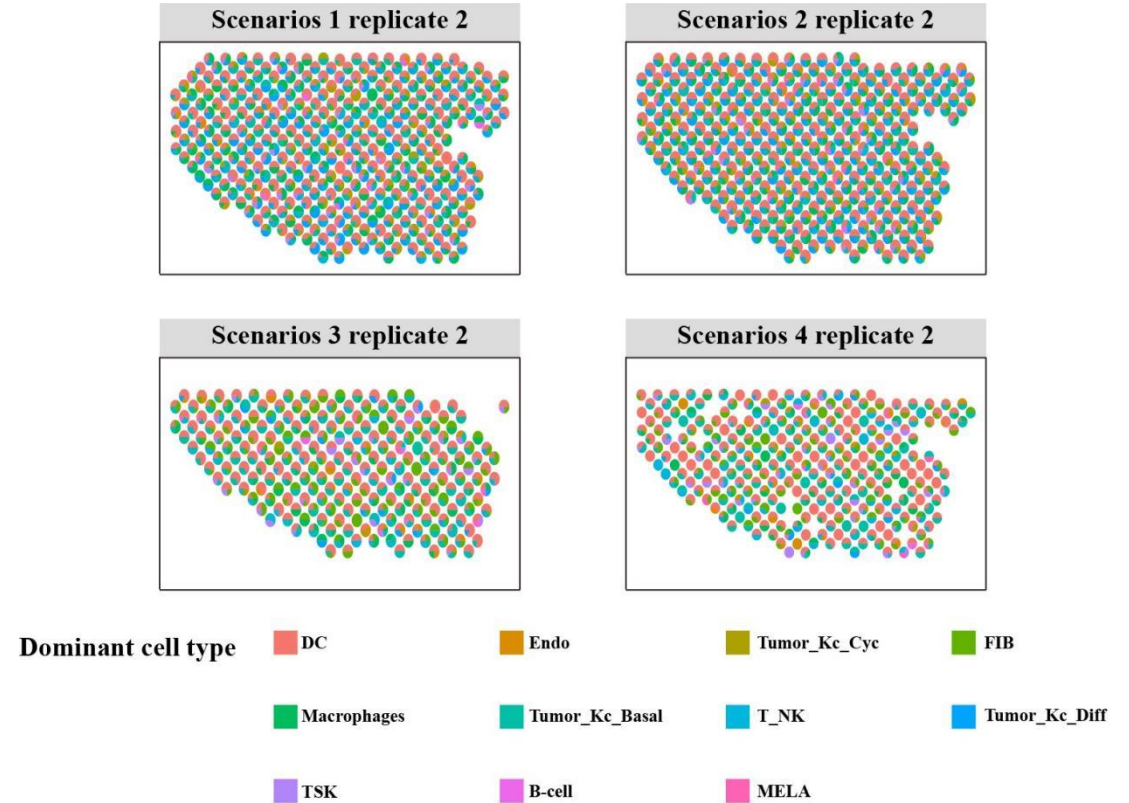
These simulated data contain 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and MELA) that were obtained from the single-cell RNA sequencing (scRNA-seq)¹ data of SCC_P2.rep2 data². We censored and merged the cell types of the single-cell data during the creation of the simulated data. All of the simulated data in the four scenarios were also generated using single-cell data from SCC_P2.rep2. There are four simulation scenarios in total. In scenario 1, we constructed spatial transcriptomics data using single-cell data and did not impose additional conditions. In scenario 2, we used single-cell data to construct spatial transcriptomics data and added different spatial pattern (details in Supplementary Note 3). In scenario 3, we used single-cell data to construct spatial transcriptomics data and added different cellular patterns (details in Supplementary Note 3). In scenario 4, we constructed spatial transcriptomics data using single-cell data and added both different cellular patterns and spatial patterns (details in Supplementary Note 3). To visualize the simulations, we plotted scatter pie charts for the four scenarios spaces of replicate 1.



24 **Supplementary Fig. 2. Composition of cell types for each spot in replicate 2 of the**
25 **four scenarios.**

26 These data are replicates of simulated data. These simulated data contain 11 cell types
27 (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK,
28 Tumor_Kc_Diff, TSK, B-cell, and MELA) that were obtained from the scRNA-seq¹
29 data of SCC_P2.rep2 data². We censored and merged the cell types of the single-cell
30 data during the creation of the simulated data. All of the simulated data in the four
31 scenarios were also generated using single-cell data from SCC_P2.rep2. There are
32 four simulation scenarios in total. In scenario 1, we constructed spatial
33 transcriptomics data using single-cell data and did not impose additional conditions.
34 In scenario 2, we used single-cell data to construct spatial transcriptomics data and
35 added different spatial pattern (details in Supplementary Note 3). In scenario 3, we
36 used single-cell data to construct spatial transcriptomics data and added different
37 cellular patterns (details in Supplementary Note 3). In scenario 4, we constructed
38 spatial transcriptomics data using single-cell data and added both different cellular
39 patterns and spatial patterns (details in Supplementary Note 3). To visualize the
40 simulations, we plotted scatter pie charts for the four scenarios spaces of replicate 2.

41



42

43

44

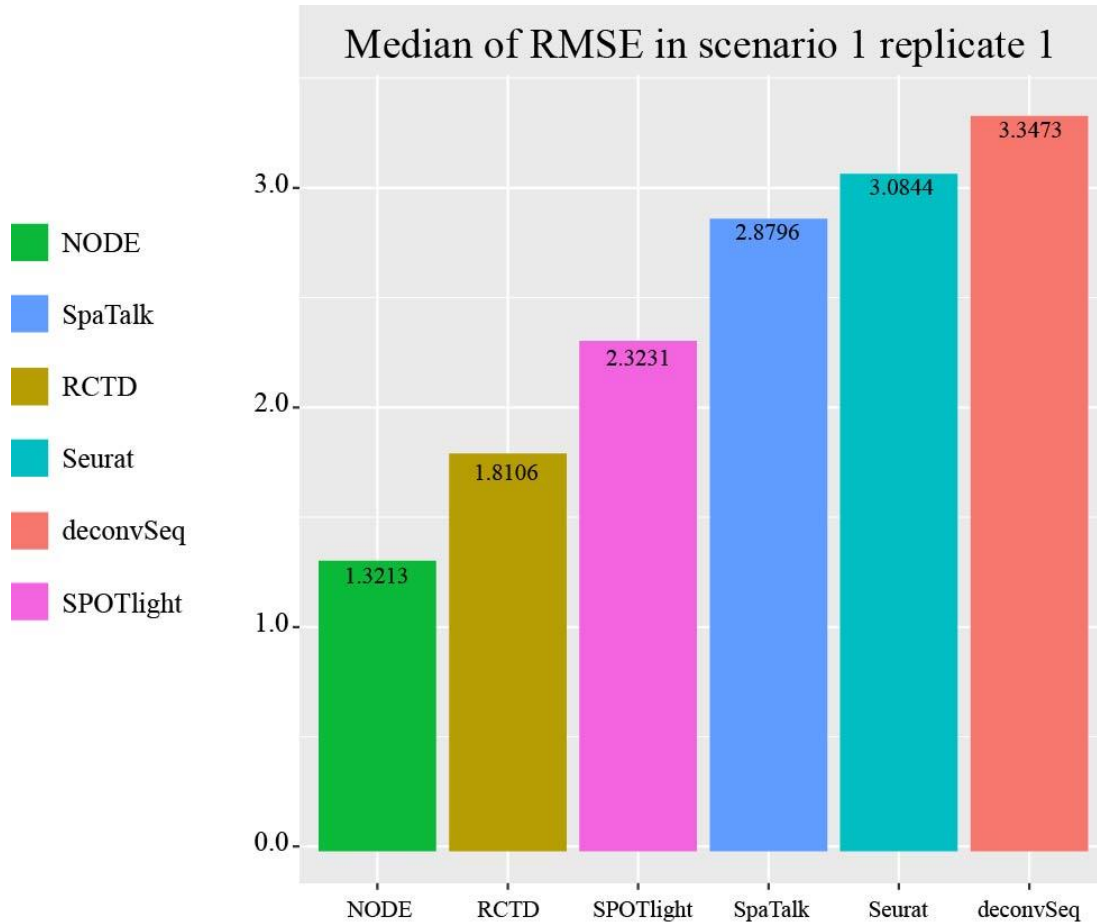
45

46

47

48 **Supplementary Fig. 3. Comparison of median RMSE for each method in**
49 **scenarios 1 replicate 1.**

50 We plotted RMSE from the six deconvolution methods in the simulated data scenarios
51 1 replicate 1. Specifically, we calculated the difference between the deconvolution
52 results and the simulated data in all spots (by RMSE), and also, since the results of the
53 RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶ methods are all calculated as the
54 proportion of cells in each category, we first used the number of cells in the results of
55 the simulated data to reduce the results of these methods into the number of cells,
56 after which we analyzed them using RMSE median. Specifically, we calculated the
57 median of all their RMSEs and plotted bar graphs. And in the results, NODE obtained
58 a lower RMSE median, responding to the accuracy of NODE in the deconvolution.

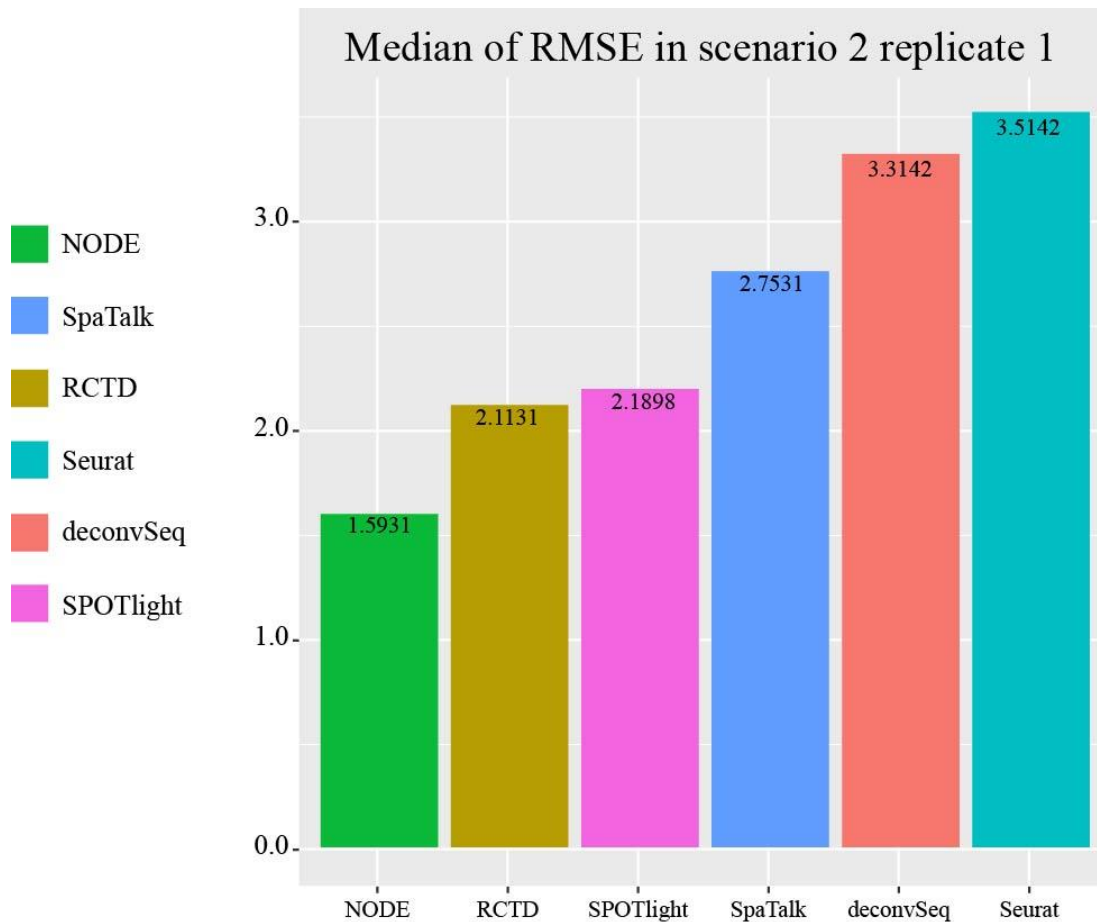


59

60

61 **Supplementary Fig. 4. Comparison of median RMSE for each method in**
62 **scenarios 2 replicate 1.**

63 We plotted RMSE from the six deconvolution methods in the simulated data scenarios
64 2 replicate 1. Specifically, we calculated the difference between the deconvolution
65 results and the simulated data in all spots (by RMSE), and also, since the results of the
66 RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶ methods are all calculated as the
67 proportion of cells in each category, we first used the number of cells in the results of
68 the simulated data to reduce the results of these methods into the number of cells,
69 after which we analyzed them using RMSE, specifically, we calculated the median of
70 all their RMSEs and plotted bar graphs. And in the results, NODE obtained a lower
71 RMSE, responding to the accuracy of NODE in the deconvolution.

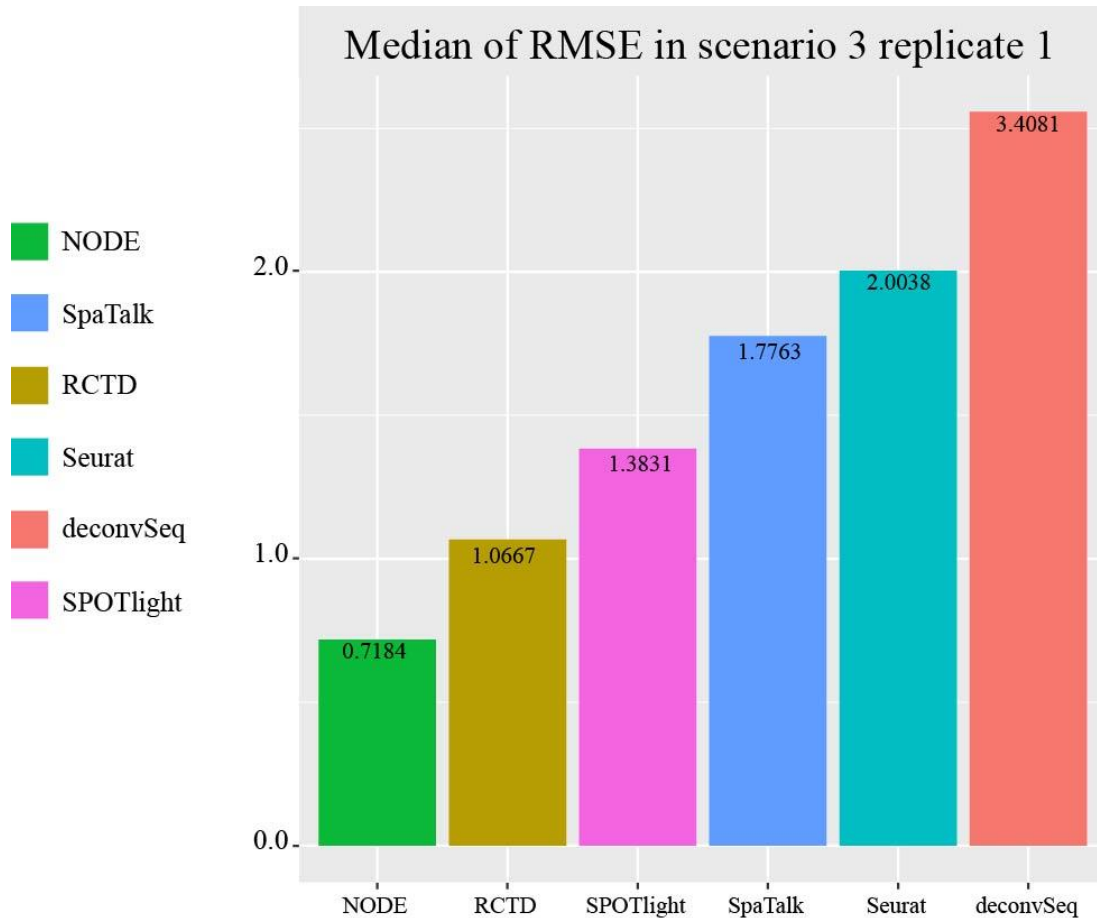


72

73

74 **Supplementary Fig. 5. Comparison of median RMSE for each method in**
75 **scenarios 3 replicate 1.**

76 We plotted RMSE from the six deconvolution methods in the simulated data scenarios
77 3 replicate 1. Specifically, we calculated the difference between the deconvolution
78 results and the simulated data in all spots (by RMSE), and also, since the results of the
79 RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶ methods are all calculated as the
80 proportion of cells in each category, we first used the number of cells in the results of
81 the simulated data to reduce the results of these methods into the number of cells,
82 after which we analyzed them using RMSE, specifically, we calculated the median of
83 all their RMSEs and plotted bar graphs. And in the results, NODE obtained a lower
84 RMSE, responding to the accuracy of NODE in the deconvolution.

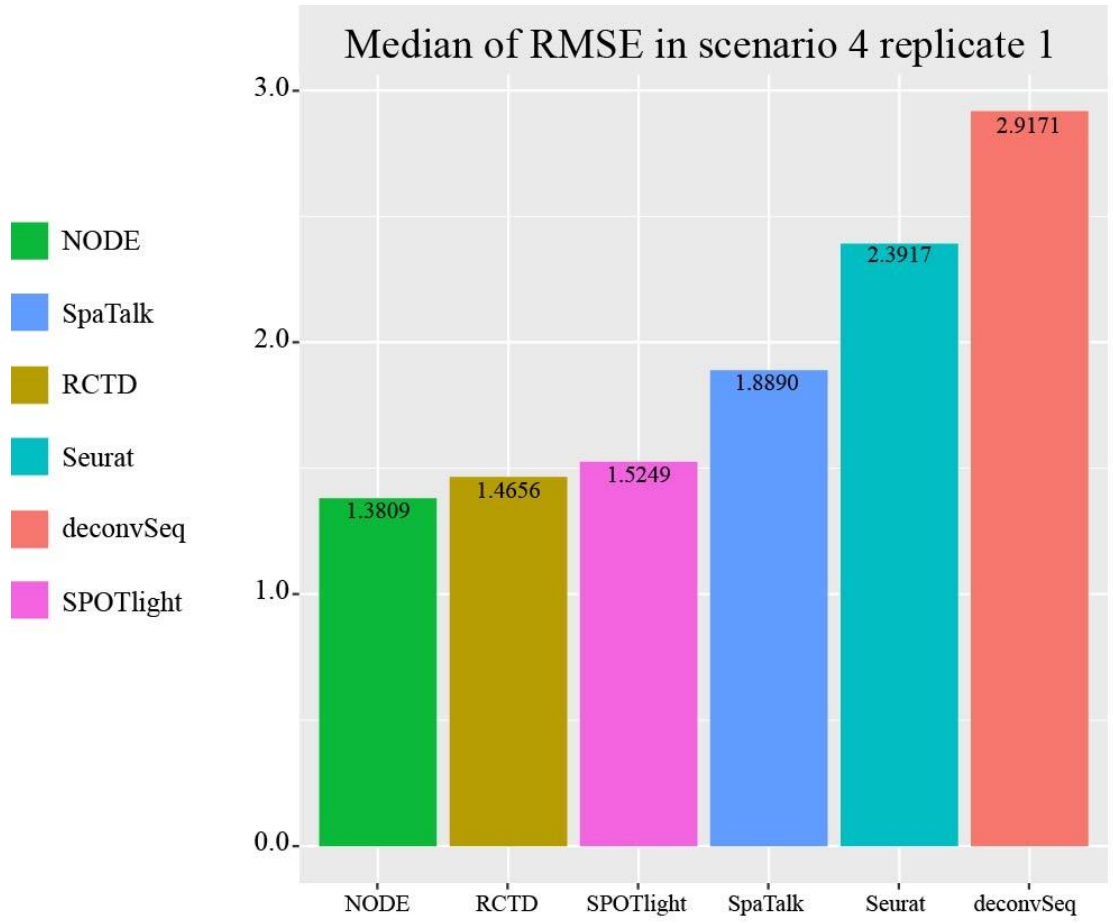


85

86

87 **Supplementary Fig. 6. Comparison of median RMSE for each method in**
88 **scenarios 4 replicate 1.**

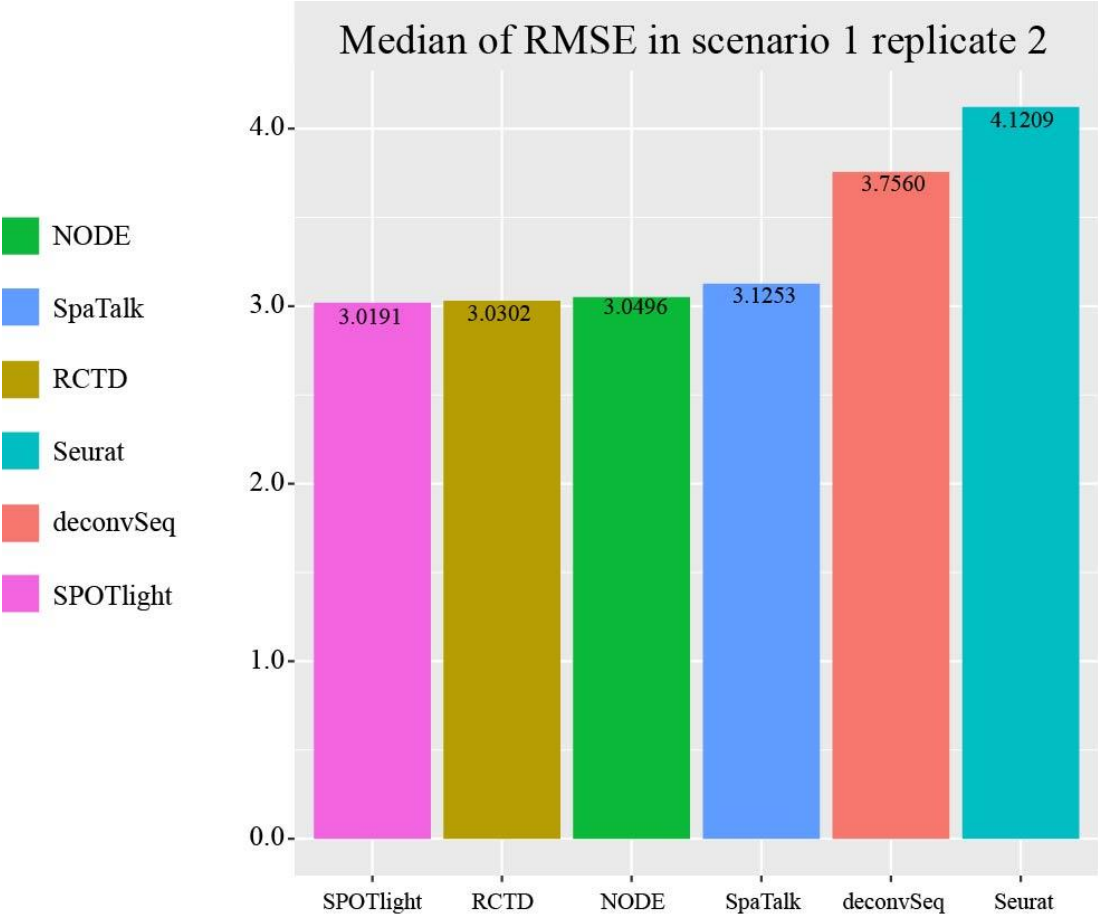
89 We plotted RMSE from the six deconvolution methods in the simulated data scenarios
90 4 replicate 1. Specifically, we calculated the difference between the deconvolution
91 results and the simulated data in all spots (by RMSE), and also, since the results of the
92 RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶ methods are all calculated as the
93 proportion of cells in each category, we first used the number of cells in the results of
94 the simulated data to reduce the results of these methods into the number of cells,
95 after which we analyzed them using RMSE, specifically, we calculated the median of
96 all their RMSEs and plotted bar graphs. And in the results, NODE obtained a lower
97 RMSE, responding to the accuracy of NODE in the deconvolution.



98
99
100

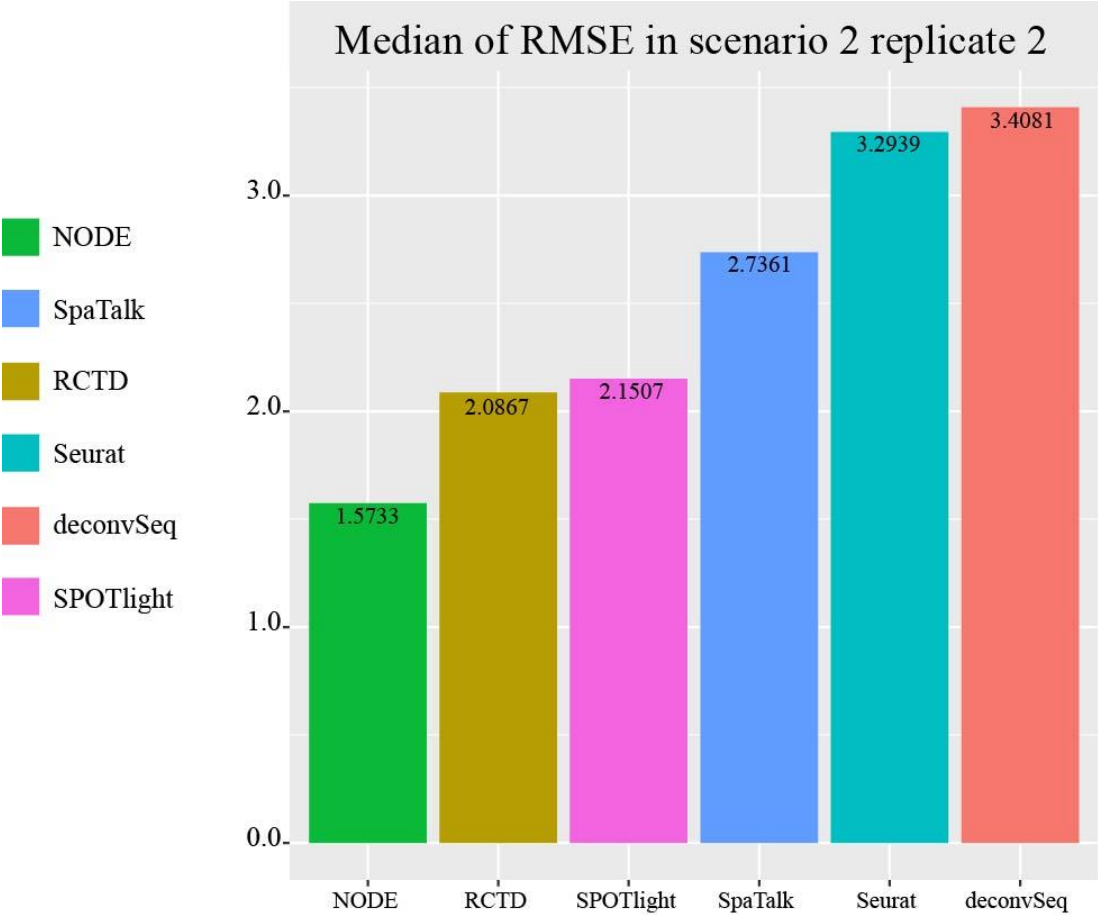
Supplementary Fig. 7. Comparison of median RMSE for each method in scenarios 1 replicate 2.

We plotted RMSE from the six deconvolution methods in the simulated data scenarios 1 replicate 2. Specifically, we calculated the difference between the deconvolution results and the simulated data in all spots (by RMSE), and also, since the results of the RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶ methods are all calculated as the proportion of cells in each category, we first used the number of cells in the results of the simulated data to reduce the results of these methods into the number of cells, after which we analyzed them using RMSE, specifically, we calculated the median of all their RMSEs and plotted bar graphs. And in the results, NODE obtained a lower RMSE, responding to the accuracy of NODE in the deconvolution.



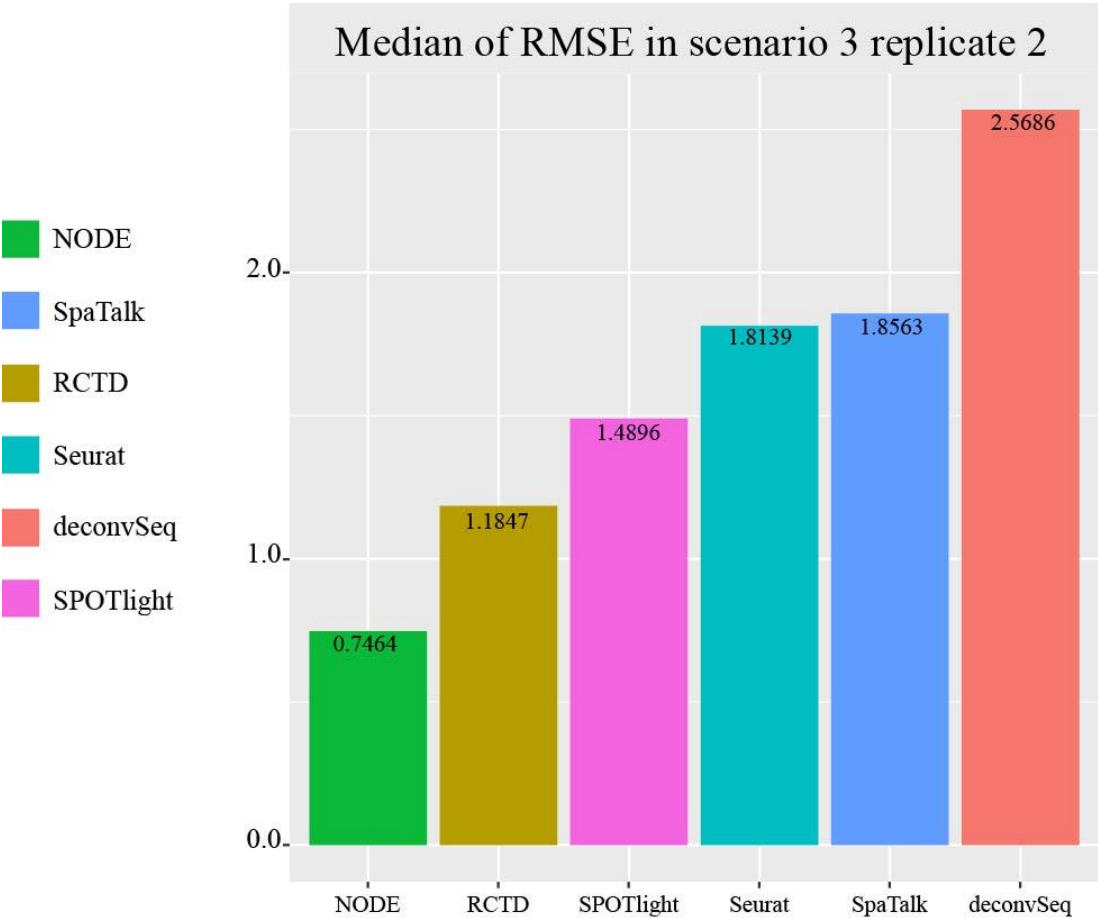
Supplementary Fig. 8. Comparison of median RMSE for each method in scenarios 2 replicate 2.

We plotted RMSE from the six deconvolution methods in the simulated data scenarios 2 replicate 2. Specifically, we calculated the difference between the deconvolution results and the simulated data in all spots (by RMSE), and also, since the results of the RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶ methods are all calculated as the proportion of cells in each category, we first used the number of cells in the results of the simulated data to reduce the results of these methods into the number of cells, after which we analyzed them using RMSE, specifically, we calculated the median of all their RMSEs and plotted bar graphs. And in the results, NODE obtained a lower RMSE, responding to the accuracy of NODE in the deconvolution.



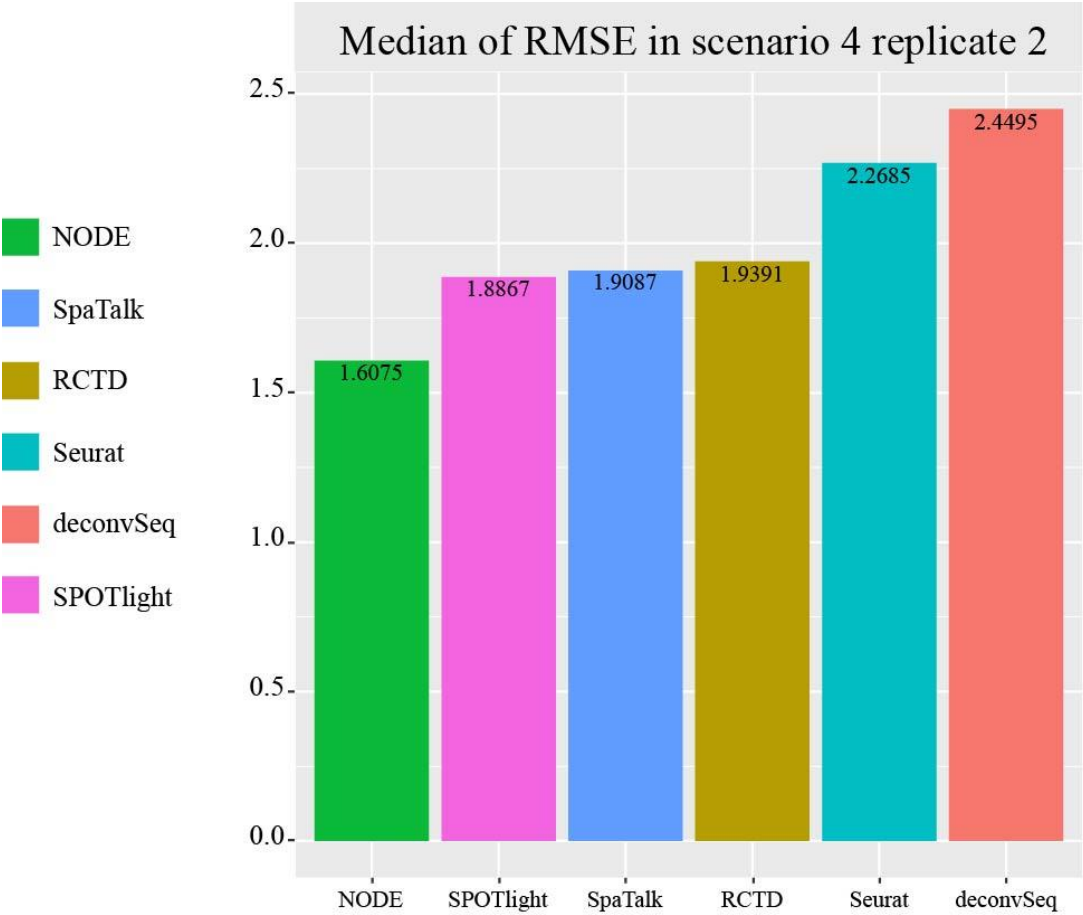
Supplementary Fig. 9. Comparison of median RMSE for each method in scenarios 3 replicate 2.

We plotted RMSE from the six deconvolution methods in the simulated data scenarios 3 replicate 2. Specifically, we calculated the difference between the deconvolution results and the simulated data in all spots (by RMSE), and also, since the results of the RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶ methods are all calculated as the proportion of cells in each category, we first used the number of cells in the results of the simulated data to reduce the results of these methods into the number of cells, after which we analyzed them using RMSE, specifically, we calculated the median of all their RMSEs and plotted bar graphs. And in the results, NODE obtained a lower RMSE, responding to the accuracy of NODE in the deconvolution.



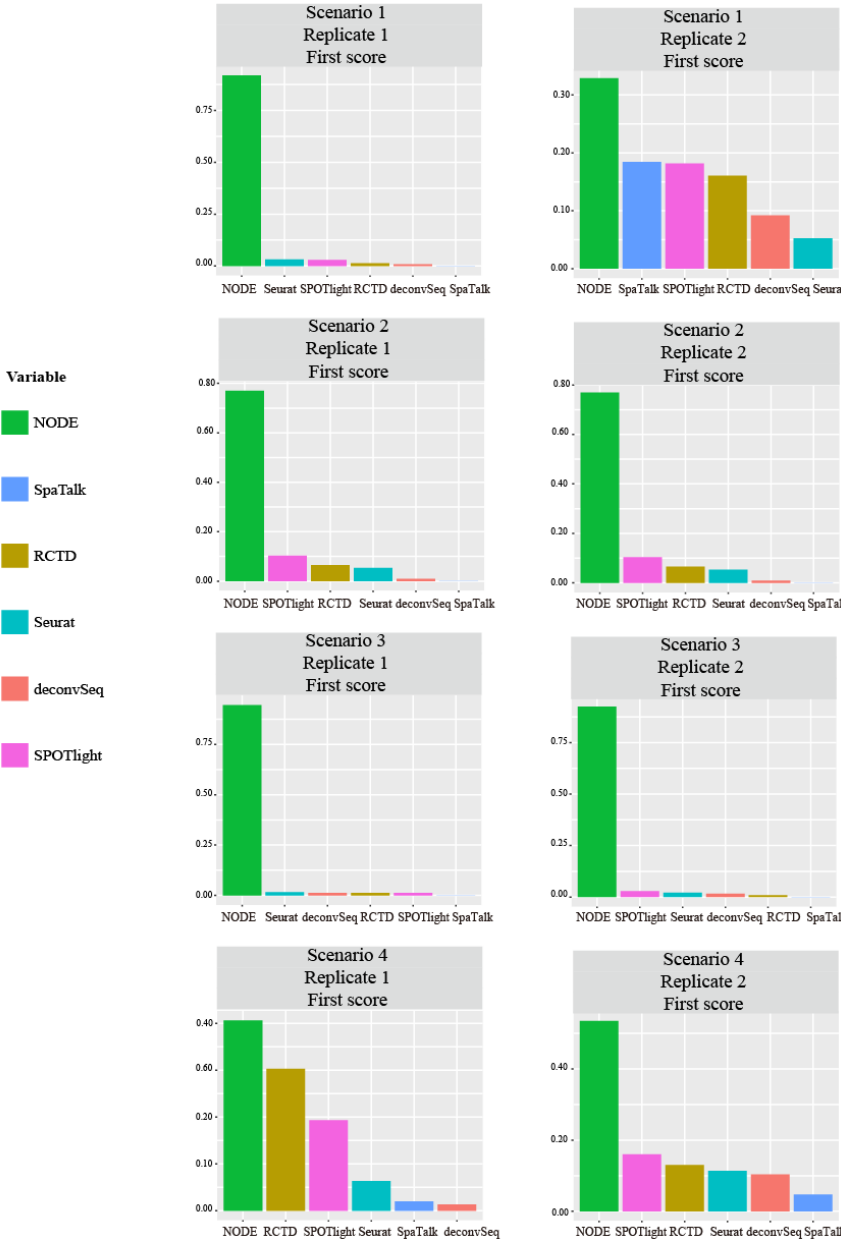
Supplementary Fig. 10. Comparison of median RMSE for each method in scenarios 4 replicate 2.

We plotted RMSE from the six deconvolution methods in the simulated data scenarios 4 replicate 2. Specifically, we calculated the difference between the deconvolution results and the simulated data in all spots (by RMSE), and also, since the results of the RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶ methods are all calculated as the proportion of cells in each category, we first used the number of cells in the results of the simulated data to reduce the results of these methods into the number of cells, after which we analyzed them using RMSE, specifically, we calculated the median of all their RMSEs and plotted bar graphs. And in the results, NODE obtained a lower RMSE, responding to the accuracy of NODE in the deconvolution.



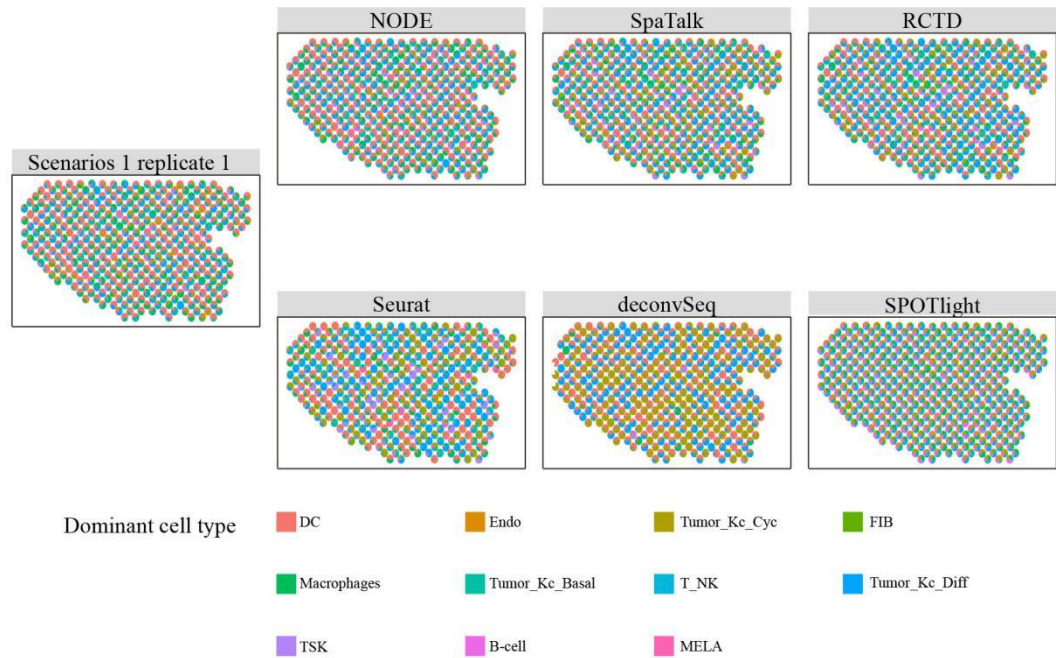
Supplementary Fig. 11. Best performance scores of different deconvolution methods in simulated data (First score).

Frequency of the best performance of all the deconvolution methods in the two replicates of the four scenarios. Specifically, we applied the different deconvolution methods to the simulated data, and calculated the root mean square error (RMSE) for every methods in each spot, and then we calculated the percentage of the number of spots in which each method obtained the smallest RMSE, and used this frequency value as the best performance scores.



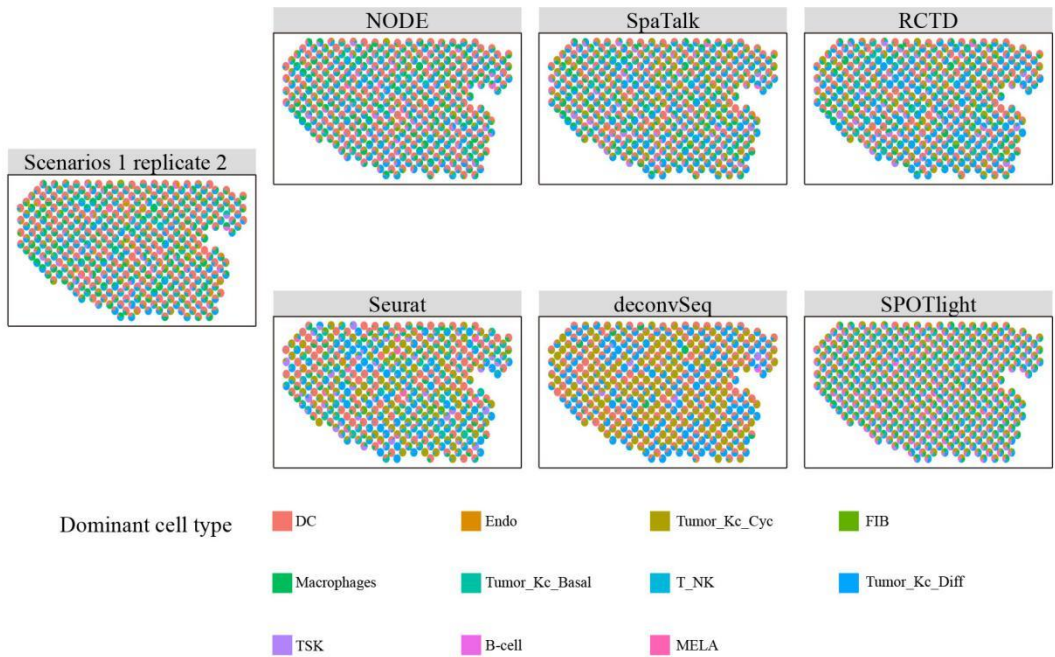
Supplementary Fig. 12. Deconvolution comparison of NODE with other deconvolution methods on simulated data scenario 1 replicate 1.

In this simulated data, we constructed spatial transcriptomics data directly from single-cell data. We censored and merged the cell types of the single-cell data during the creation of the simulated data. The simulated data contained 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and MEA), and the single-cell data used to construct the spatial transcriptomics data were from SCC_P2.rep2. We applied NODE with other deconvolution methods (SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵, and SPOTlight⁶) to this simulated data set, and then we plotted the percentage of each cell type on each spot based on the cellular inference results of the deconvolution and plotted the spatial scatter pie charts. Finally, we compared all the deconvolution results with the simulated data and found that in the color distribution of the pie charts, the results of NODE were consistent with the distribution of the cells in the simulated data, whereas the other methods appeared to infer results that were inconsistent with the simulated data.



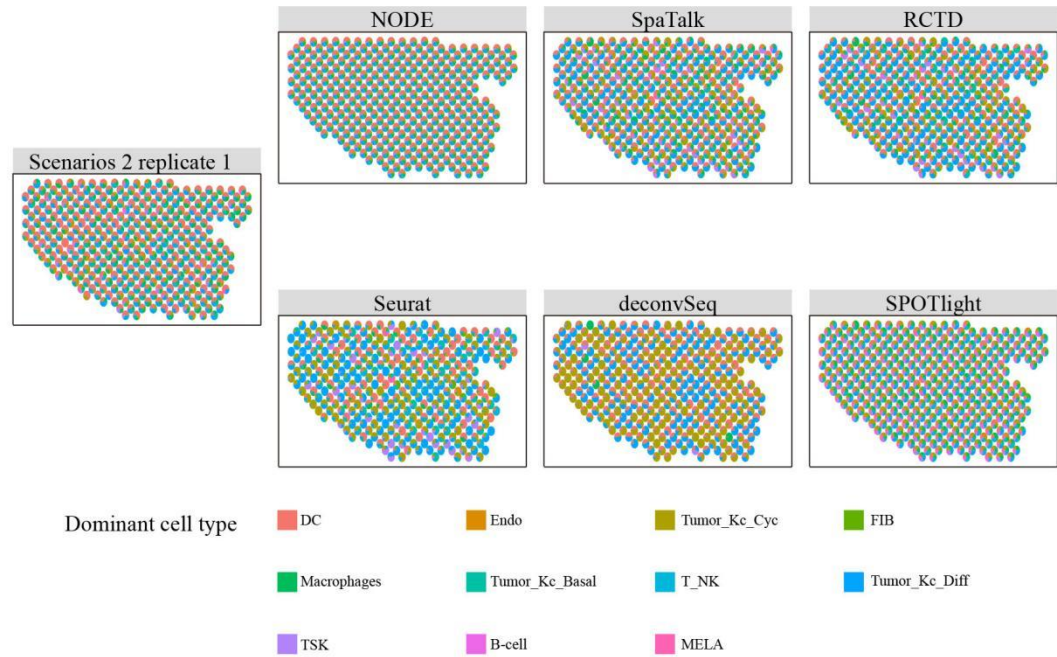
Supplementary Fig. 13. Deconvolution comparison of NODE with other deconvolution methods on simulated data scenario 1 replicate 2.

In this simulated data, we constructed spatial transcriptomics data directly from single-cell data. We censored and merged the cell types of the single-cell data during the creation of the simulated data. To avoid accidental data, we performed a duplicate. The simulated data contained 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and ME LA), and the single-cell data used to construct the spatial transcriptomics data were from SCC_P2.rep2. We applied NODE with other deconvolution methods (SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵, and SPOTlight⁶) to this simulated data set, and then we plotted the percentage of each cell type on each spot based on the cellular inference results of the deconvolution and plotted the spatial scatter pie charts. Finally, we compared all the deconvolution results with the simulated data and found that in the color distribution of the pie charts, the results of NODE were consistent with the distribution of the cells in the simulated data, whereas the other methods appeared to infer results that were inconsistent with the simulated data.



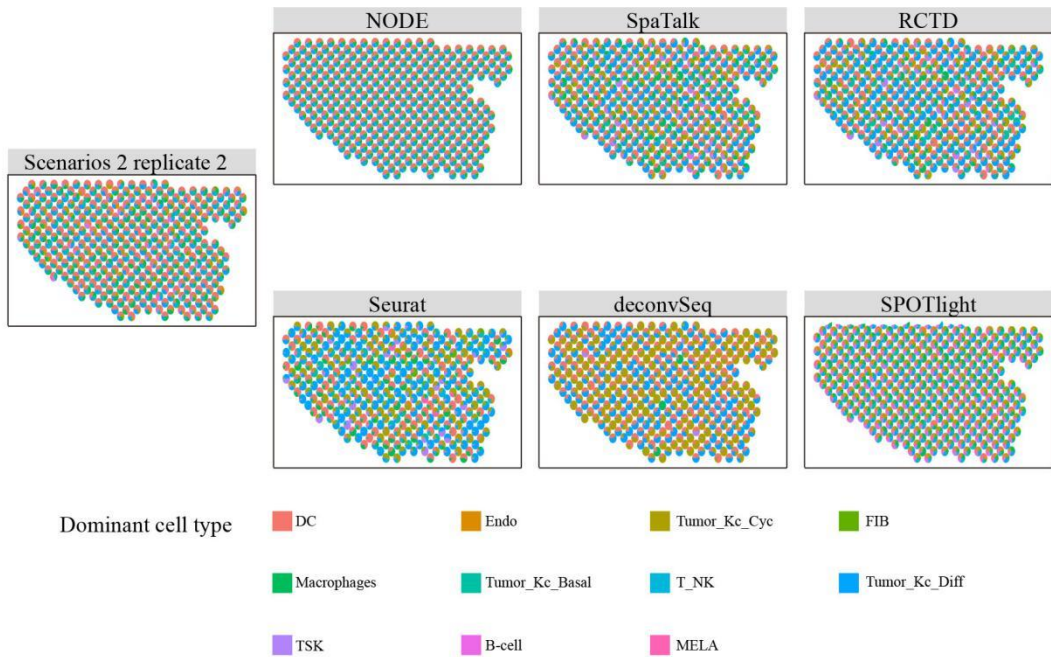
Supplementary Fig. 14. Deconvolution comparison of NODE with other deconvolution methods on simulated data scenario 2 replicate 1.

In this simulated data, after we constructed the spatial transcriptomics data with single-cell data, we incorporated different spatial patterns (details in Supplementary Note 3). We censored and merged the cell types of the single-cell data during the creation of the simulated data. The simulated data contained 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and MELA), and the single-cell data used to construct the spatial transcriptomics data were from SCC_P2.rep2. We applied NODE with other deconvolution methods (SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶) to this simulated data set, and then we plotted the percentage of each cell type on each spot based on the cellular inference results of the deconvolution and plotted the spatial scatter pie charts. Finally, we compared all the deconvolution results with the simulated data and found that in the color distribution of the pie charts, the results of NODE were consistent with the distribution of the cells in the simulated data, whereas the other methods appeared to infer results that were inconsistent with the simulated data.



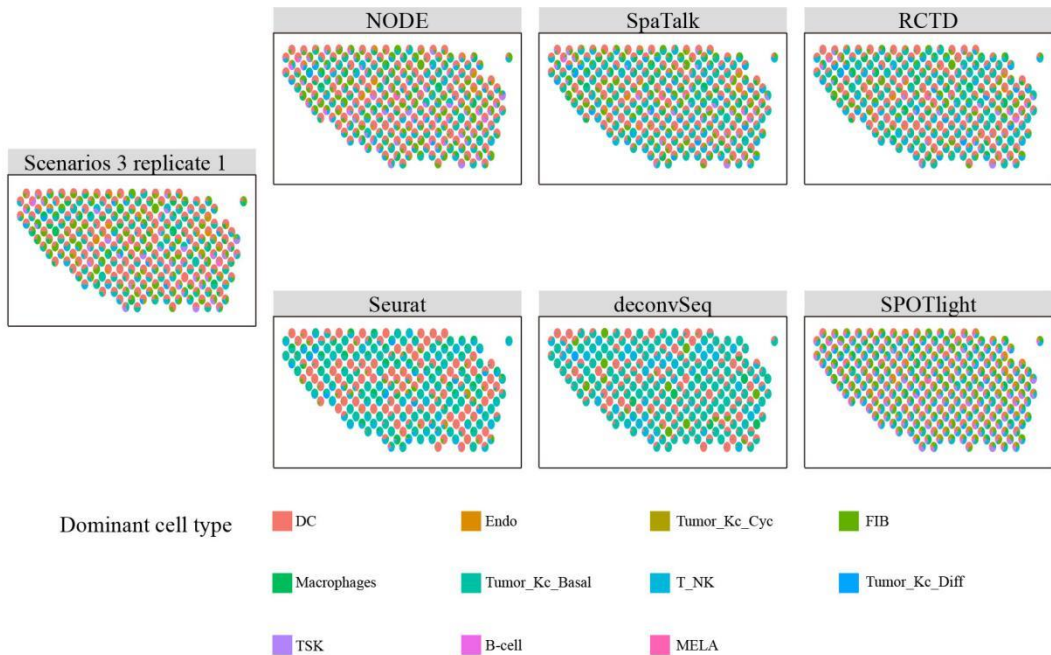
Supplementary Fig. 15. Deconvolution comparison of NODE with other deconvolution methods on simulated data scenario 2 replicate 2.

In this simulated data, after we constructed the spatial transcriptomics data with single-cell data, we incorporated different spatial patterns (details in Supplementary Note 3). To avoid accidental data, we performed a duplicate. We censored and merged the cell types of the single-cell data during the creation of the simulated data. The simulated data contained 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and MELA), and the single-cell data used to construct the spatial transcriptomics data were from SCC_P2.rep2. We applied NODE with other deconvolution methods (SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶) to this simulated data set, and then we plotted the percentage of each cell type on each spot based on the cellular inference results of the deconvolution and plotted the spatial scatter pie charts. Finally, we compared all the deconvolution results with the simulated data and found that in the color distribution of the pie charts, the results of NODE were consistent with the distribution of the cells in the simulated data, whereas the other methods appeared to infer results that were inconsistent with the simulated data.



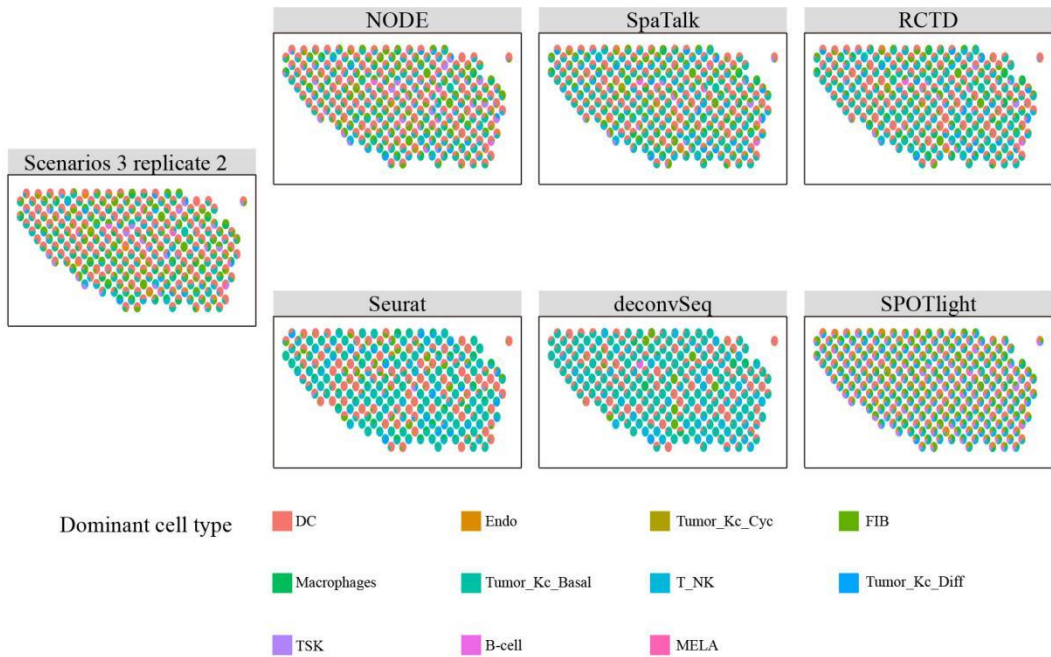
Supplementary Fig. 16. Deconvolution comparison of NODE with other deconvolution methods on simulated data scenario 3 replicate 1.

In this simulated data, after we constructed the spatial transcriptomics data with single-cell data, we incorporated different cellular patterns (details in Supplementary Note 3). We censored and merged the cell types of the single-cell data during the creation of the simulated data. The simulated data contained 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and MELA), and the single-cell data used to construct the spatial transcriptomics data were from SCC_P2.rep2. We applied NODE with other deconvolution methods (SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶) to this simulated dataset, and then we plotted the percentage of each cell type on each spot based on the cellular inference results of the deconvolution and plotted the spatial scatter pie charts. Finally, we compared all the deconvolution results with the simulated data and found that in the color distribution of the pie charts, the results of NODE were consistent with the distribution of the cells in the simulated data, whereas the other methods appeared to infer results that were inconsistent with the simulated data.



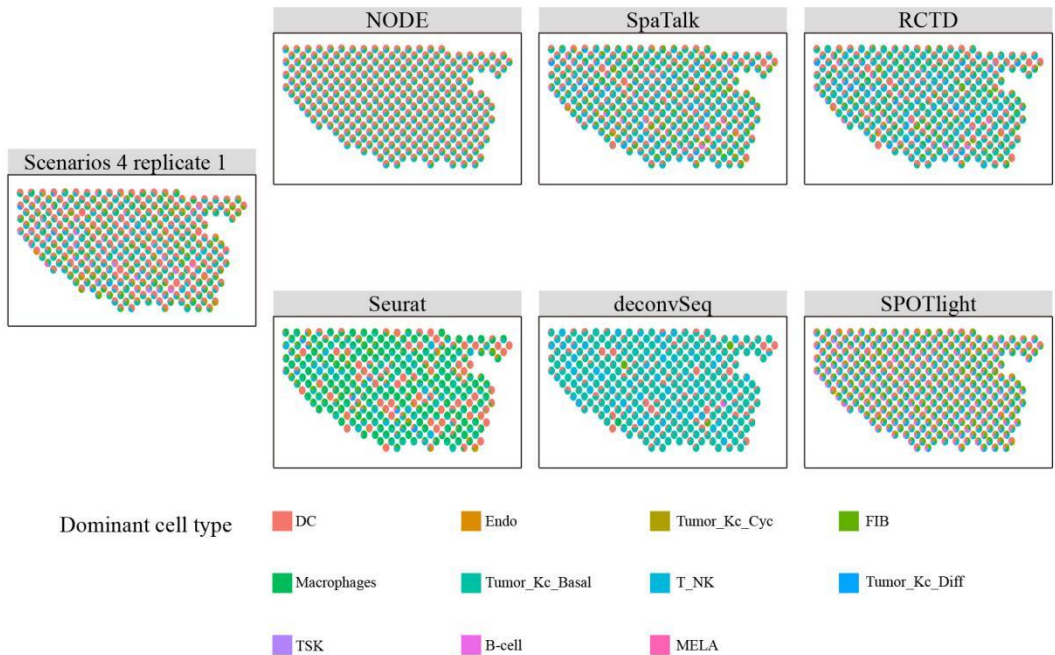
Supplementary Fig. 17. Deconvolution comparison of NODE with other deconvolution methods on simulated data scenario 3 replicate 2.

In this simulated data, after we constructed the spatial transcriptomics data with single-cell data, we incorporated different cellular patterns (details in Supplementary Note 3). To avoid accidental data, we performed a duplicate. We censored and merged the cell types of the single-cell data during the creation of the simulated data. The simulated data contained 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and MELA), and the single-cell data used to construct the spatial transcriptomics data were from SCC_P2.rep2. We applied NODE with other deconvolution methods (SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶) to this simulated data set, and then we plotted the percentage of each cell type on each spot based on the cellular inference results of the deconvolution and plotted the spatial scatter pie charts. Finally, we compared all the deconvolution results with the simulated data and found that in the color distribution of the pie charts, the results of NODE were consistent with the distribution of the cells in the simulated data, whereas the other methods appeared to infer results that were inconsistent with the simulated data.



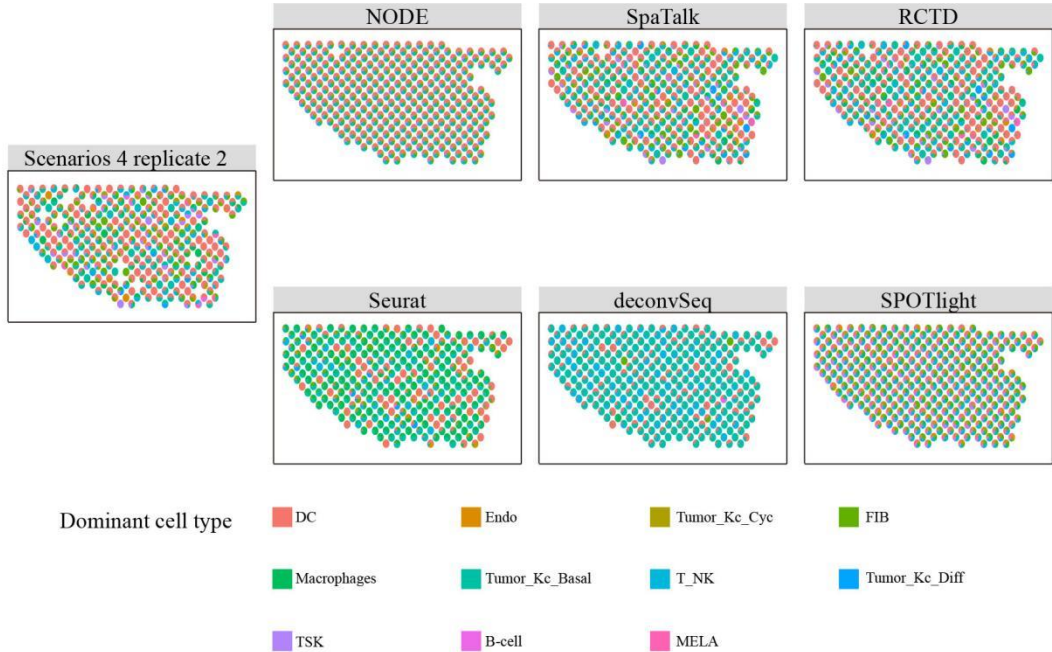
Supplementary Fig. 18. Deconvolution comparison of NODE with other deconvolution methods on simulated data scenario 4 replicate 1.

In this simulated data, after we constructed the spatial transcriptomics data with single-cell data, we incorporated different cellular patterns and spatial patterns (details in Supplementary Note 3). We censored and merged the cell types of the single-cell data during the creation of the simulated data. The simulated data contained 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and MELA), and the single-cell data used to construct the spatial transcriptomics data were from SCC_P2.rep2. We applied NODE with other deconvolution methods (SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶) to this simulated data set, and then we plotted the percentage of each cell type on each spot based on the cellular inference results of the deconvolution and plotted the spatial scatter pie charts. Finally, we compared all the deconvolution results with the simulated data and found that in the color distribution of the pie charts, the results of NODE were consistent with the distribution of the cells in the simulated data, whereas the other methods appeared to infer results that were inconsistent with the simulated data.



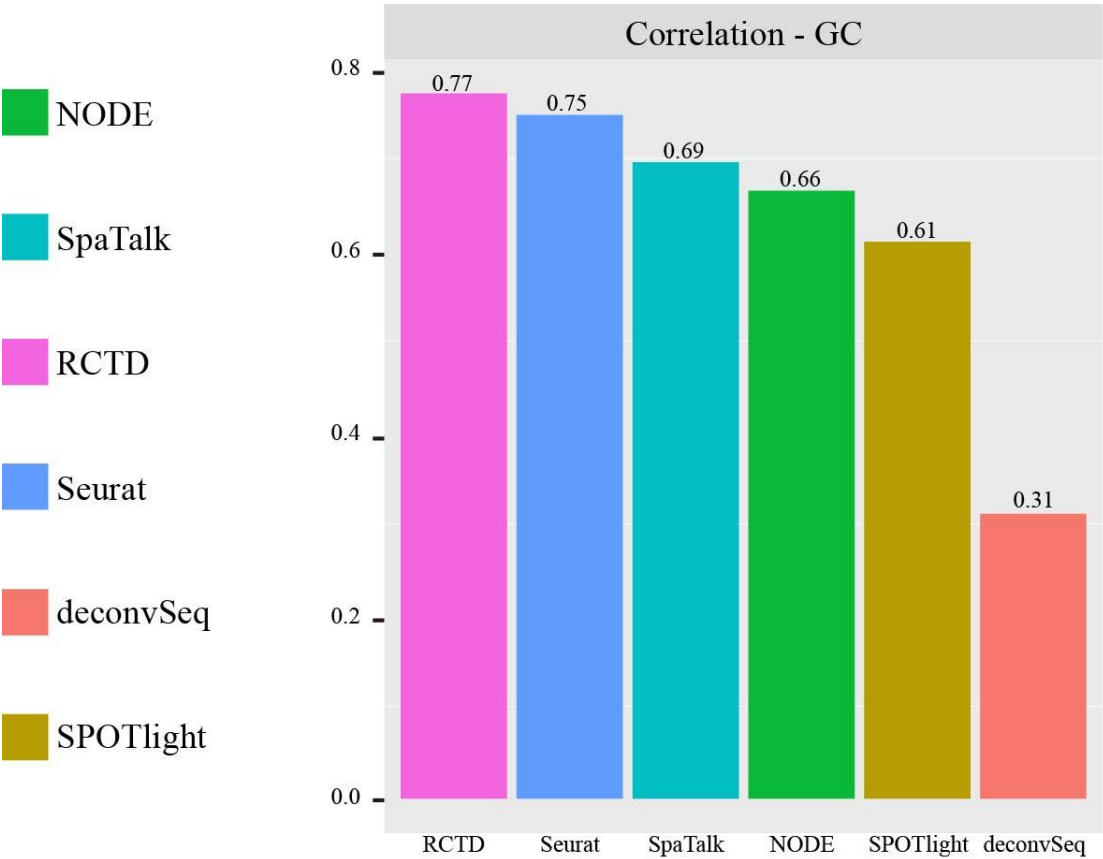
Supplementary Fig. 19. Deconvolution comparison of NODE with other deconvolution methods on simulated data scenario 4 replicate 2.

In this simulated data, after we constructed the spatial transcriptomics data with single-cell data, we incorporated different cellular patterns and spatial patterns (details in Supplementary Note 3). To avoid accidental data, we performed a duplicate. We censored and merged the cell types of the single-cell data during the creation of the simulated data. The simulated data contained 11 cell types (DC, Endo, Tumor_Kc_Cyc, FIB, Macrophages, Tumor_Kc_Basal, T_NK, Tumor_Kc_Diff, TSK, B-cell, and MELA), and the single-cell data used to construct the spatial transcriptomics data were from SCC_P2.rep2. We applied NODE with other deconvolution methods (SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵ and SPOTlight⁶) to this simulated data set, and then we plotted the percentage of each cell type on each spot based on the cellular inference results of the deconvolution and plotted the spatial scatter pie charts. Finally, we compared all the deconvolution results with the simulated data and found that in the color distribution of the pie charts, the results of NODE were consistent with the distribution of the cells in the simulated data, whereas the other methods appeared to infer results that were inconsistent with the simulated data.



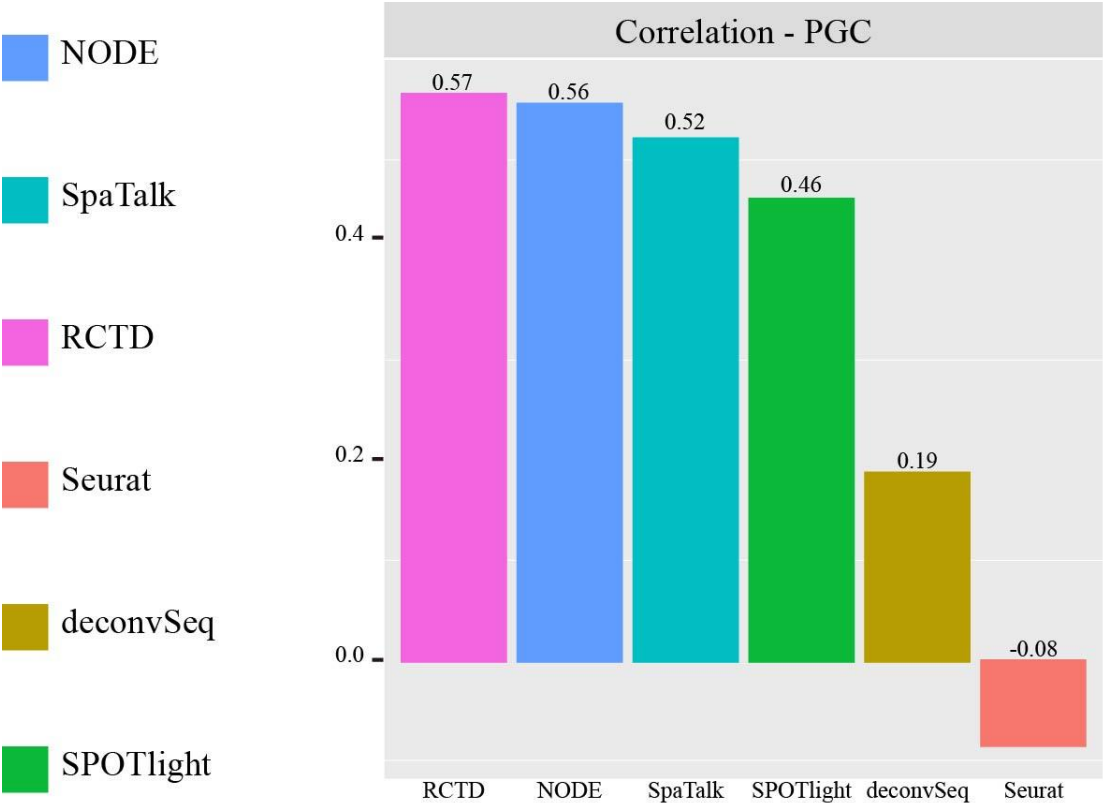
Supplementary Fig. 20. Comparison of the correlation coefficients of GC cell distribution vectors with the expression vectors of marker genes (*Penk*) corresponding to GC by different deconvolution methods in mouse olfactory bulb (MOB) data⁸.

We calculated the correlation coefficients in the MOB data and compared the results of the methods. Specifically, after applying all the deconvolution methods to the MOB data, we extracted the distribution of GC cells in all the spots based on different methods. After that, we extracted the expression of marker genes (*Penk*)⁸ corresponding to GC cells in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of GC cells and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution methods. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene. NODE ranked fourth in this set of results but NODE was very close to other results that outperform NODE methods.



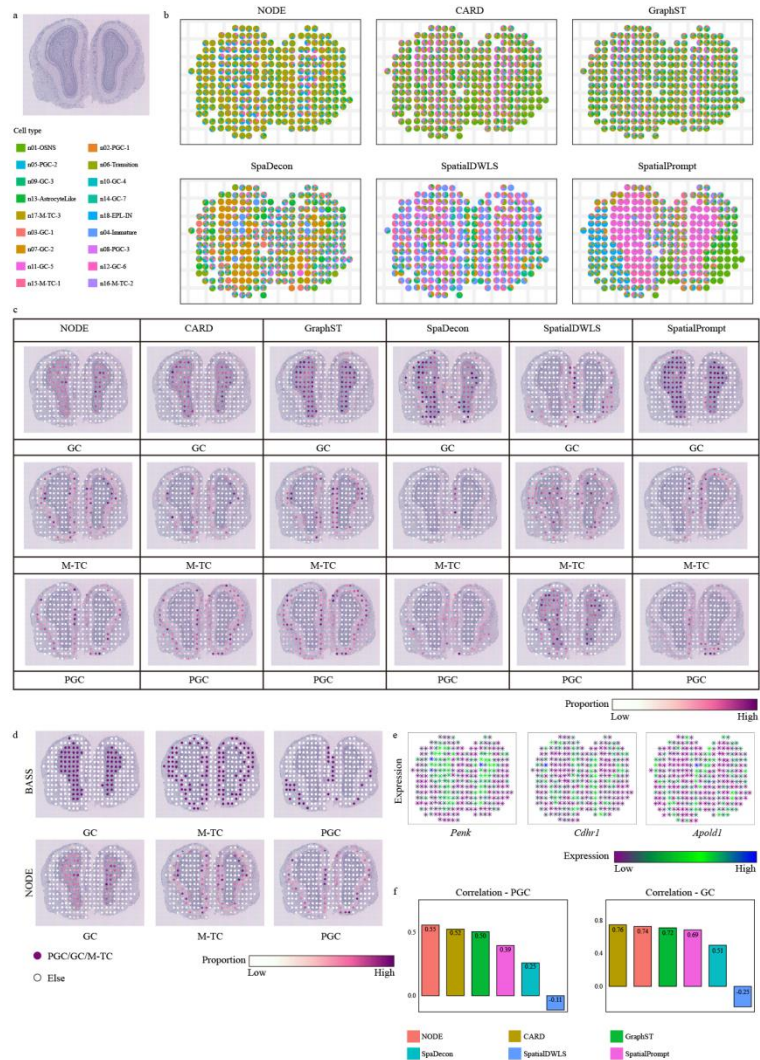
Supplementary Fig. 21. Comparison of the correlation coefficients of PGC cells distribution vectors with the expression vectors of marker genes (*Apold1*) corresponding to PGC cell by different deconvolution methods in mouse olfactory bulb (MOB) data.

We calculated the correlation coefficients in the MOB data and compared the results of the methods. Specifically, after applying all the deconvolution methods to the MOB data, we extracted the distribution of PGC cells in all the spots based on different methods. After that, we extracted the expression of marker genes (*Apold1*)⁸ corresponding to PGC cells in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of PGC cells and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution methods. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene. In this set of results, NODE ranked second and differed from the optimal result by only 0.01.



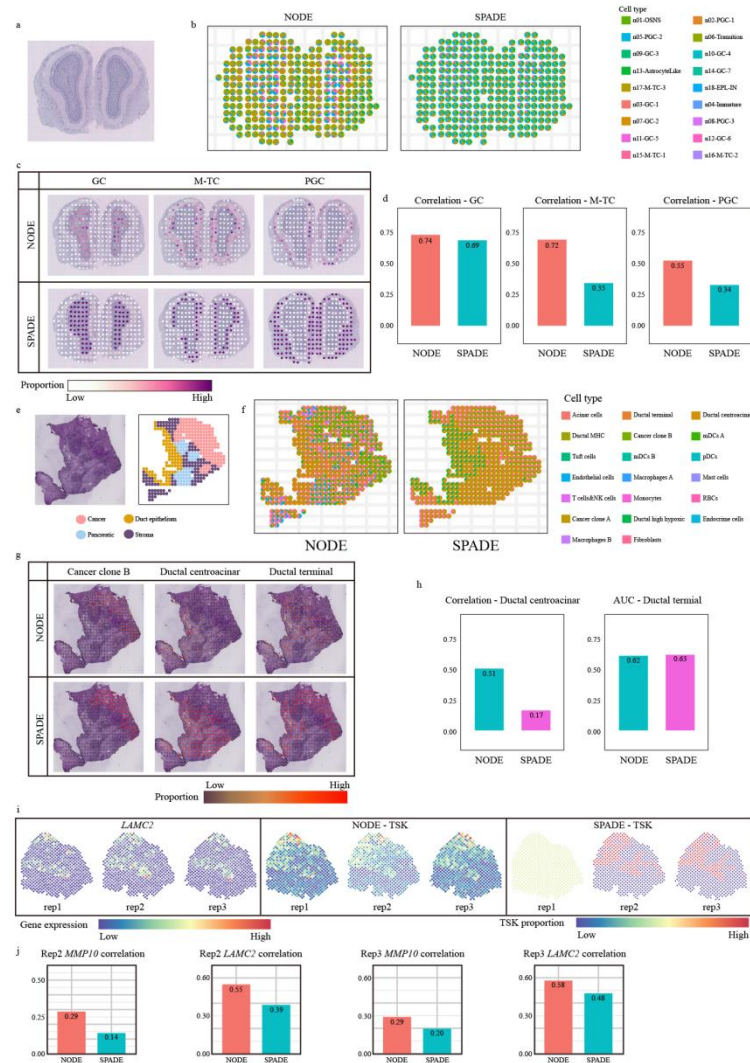
Supplementary Fig. 22. Comparison of NODE's performance with recently developed methods (CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², SpatialPrompt¹³, and BASS¹⁴) on mouse olfactory bulb (MOB) data.

a. H&E stained image of MOB. **b.** The spatial scatter pie chart displays the proportion and composition of cells at each spatial location in the inferred results of different deconvolution methods (CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³). **c.** The distribution of granule cells (GC), mitral/tufted cells (M-TC), and periglomerular cells (PGC) in the inferred results of different deconvolution methods. **d.** Aggregation locations of GC, M-TC, and PGC cells identified by BASS and aggregation locations and cellular compositions of GC, M-TC, and PGC cells identified by NODE. **e.** Expression of marker genes for GC, M-TC, and PGC cells: *Penk*, *Cdhr1*, and *Apold1*. **f.** Correlation coefficients between PGC cells and their marker genes across all deconvolution methods and AUC values for GC cells and their marker genes across all deconvolution methods.



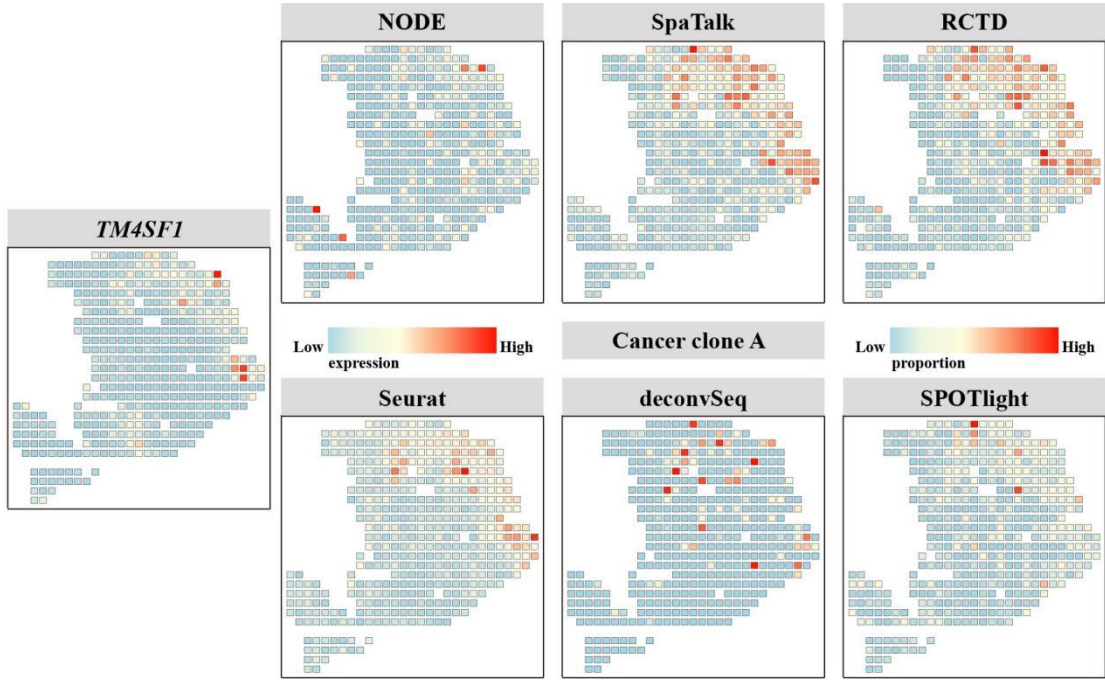
Supplementary Fig. 23. Comparison of NODE's performance with recently developed method SPADE¹⁵.

a, H&E stained image of MOB. **b**, the spatial scatter pie chart displays the proportion and composition of cells at each spatial location in the inferred results of different deconvolution methods (NODE and SPADE¹⁵) **c**, the distribution of granule cells (GC), mitral/tufted cells (M-TC), and periglomerular cells (PGC) in the inferred results. **d**, Pearson correlation coefficients between GC, M-TC, and PGC and their marker genes in NODE and SPADE results. **e**, H&E stained image of PDAC and manual annotation of PDAC. **f**, the spatial scatter pie chart displays the proportion and composition of cells at each spatial location in the inferred results of different deconvolution methods (NODE and SPADE) in PDAC data. **g**, the distribution of cancer clone B, ductal centroacinar, and ductal terminal cells in the inferred results. **h**, Pearson correlation coefficients between ductal centroacinar cells and their marker gene and AUC values for ductal terminal cells and their marker gene. **i**, *LAMC2* gene expression in SCC data and TSK identified by NODE and SPADE. **j**, comparison of correlation coefficients and AUC values for NODE and SPADE in SCC data.



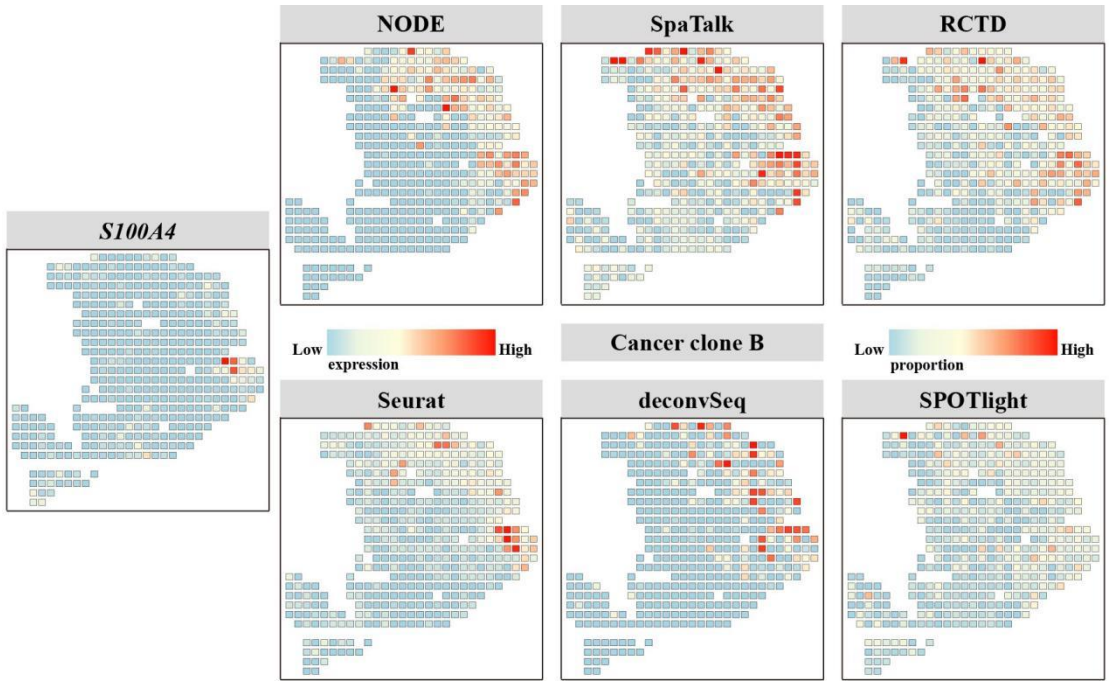
Supplementary Fig. 24. Comparison of the distribution and occupancy of cancer clone A in space inferred by each of the deconvolution methods and the expression of Cancer clone A's corresponding marker genes (*TM4SF1*)¹⁶⁻¹⁸ in space in human pancreatic ductal adenocarcinoma (PDAC) data¹⁹.

The proportion of Cancer clone A inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*TM4SF1*)¹⁶⁻¹⁸. We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression. In the results, the inferred results of NODE showed a distribution consistent with the marker genes.



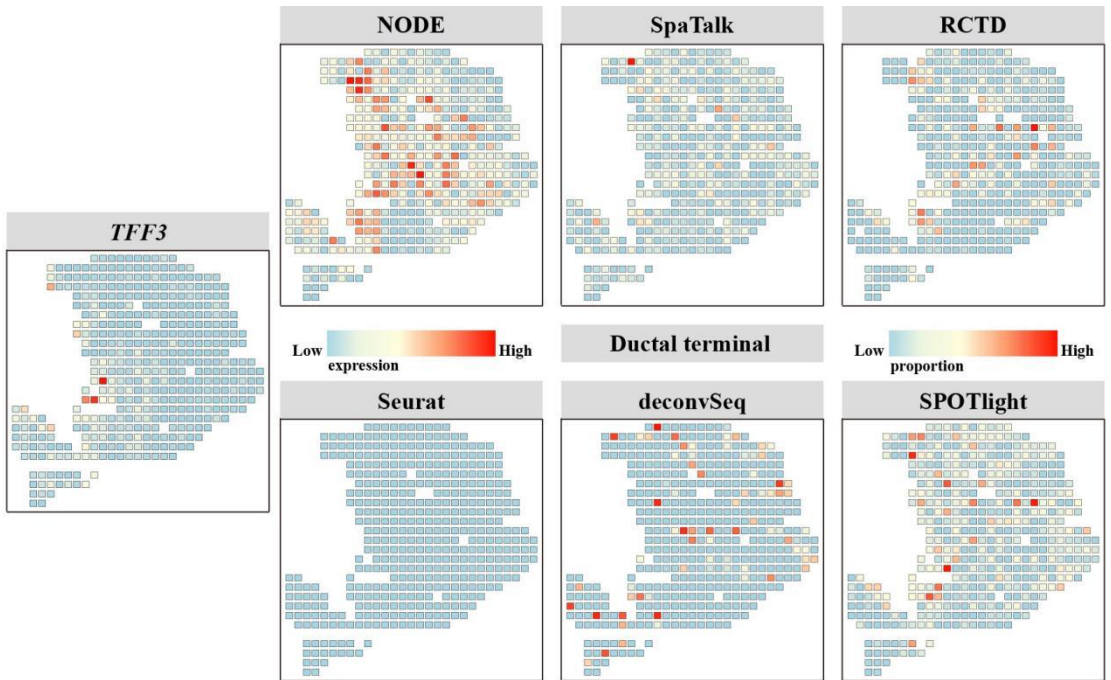
Supplementary Fig. 25. Comparison of the distribution and occupancy of cancer clone B in space inferred by each of the deconvolution methods and the expression of Cancer clone B's corresponding marker genes (*S100A4*)¹⁶⁻¹⁸ in space in human pancreatic ductal adenocarcinoma (PDAC) data.

The proportion of Cancer clone B inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*S100A4*)¹⁶⁻¹⁸. We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression. In the results, the inferred results of NODE showed a distribution consistent with the marker genes. And NODE precisely localized Cancer clone B to biologically significant tissue regions. NODE localized Cancer clone B to the cancer region and in the results of NODE, this type of cells did not infiltrate into other regions, and NODE found a clear demarcation, which is consistent with the available histological annotations. In contrast, other deconvolution methods were unable to accurately localize the region where the cells were located and did not find a clear demarcation.



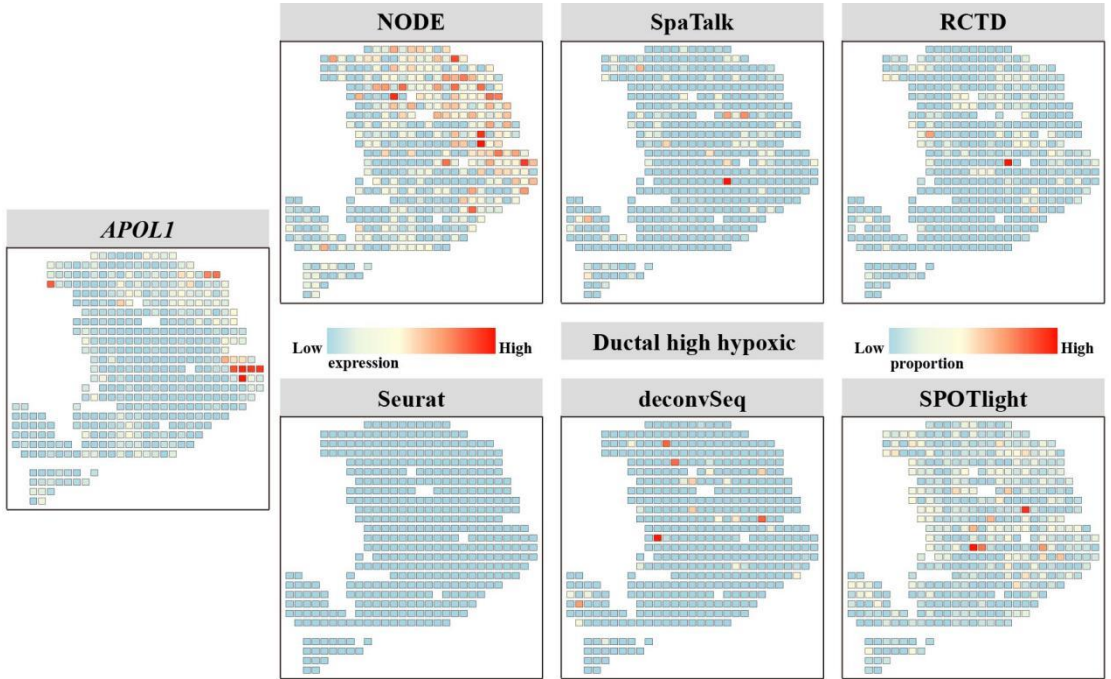
Supplementary Fig. 26. Comparison of the distribution and occupancy of ductal terminal in space inferred by each of the deconvolution methods and the expression of ductal terminal's corresponding marker genes (*TFF3*)¹⁶⁻¹⁸ in space in human pancreatic ductal adenocarcinoma (PDAC) data.

The proportion of ductal terminal inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*TFF3*)¹⁶⁻¹⁸. We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression. In the results, NODE localized the vast majority of cells to ductal regions such as the left and middle of the tissue, which is consistent with previous histological annotations¹⁹.



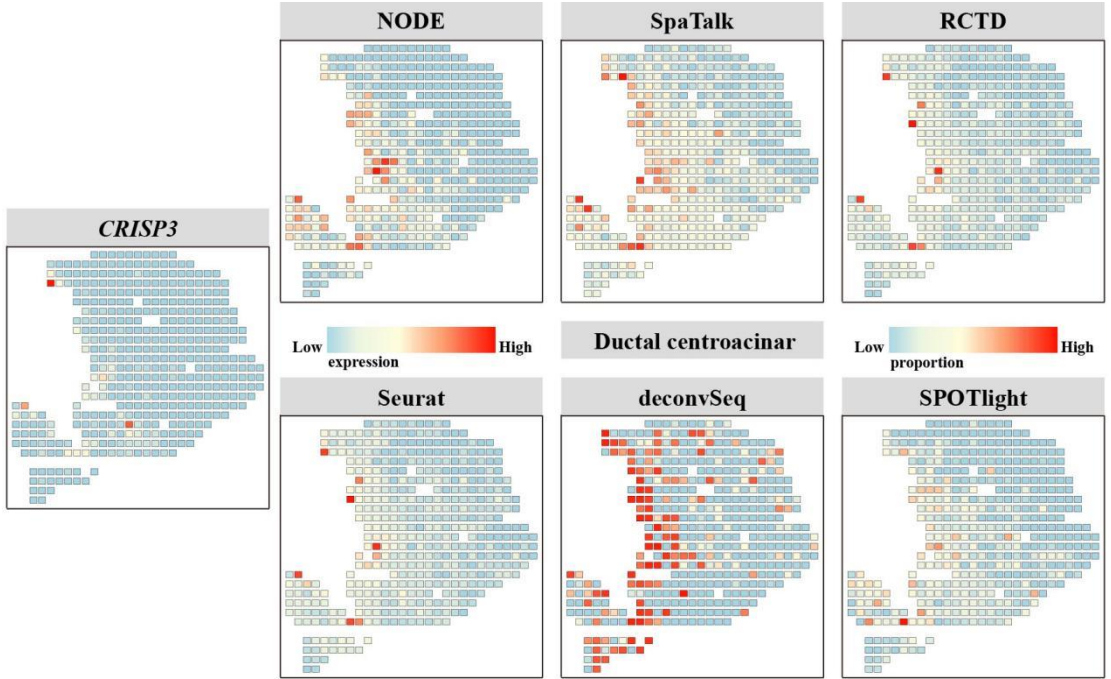
Supplementary Fig. 27. Comparison of the distribution and occupancy of ductal high hypoxic in space inferred by each of the deconvolution methods and the expression of ductal high hypoxic's corresponding marker genes (*APOL1*)¹⁶⁻¹⁸ in space in human pancreatic ductal adenocarcinoma (PDAC) data.

The proportion of ductal high hypoxic inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*APOL1*)¹⁶⁻¹⁸. We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression. In the results, the inferred results of NODE showed a distribution consistent with the marker genes. In the results, NODE localized ductal high hypoxic predominantly to the cancerous region. This is consistent with previous studies that such cells contribute to the development of tumor cells²⁰. Also, the distribution of ductal high hypoxic cells is consistent with the distribution of their marker genes. This demonstrates the performance of NODE. In contrast, other deconvolution methods did not show such localization and other deconvolution mis-localized ductal high hypoxic, and their results were also not consistent with the distribution of marker genes.



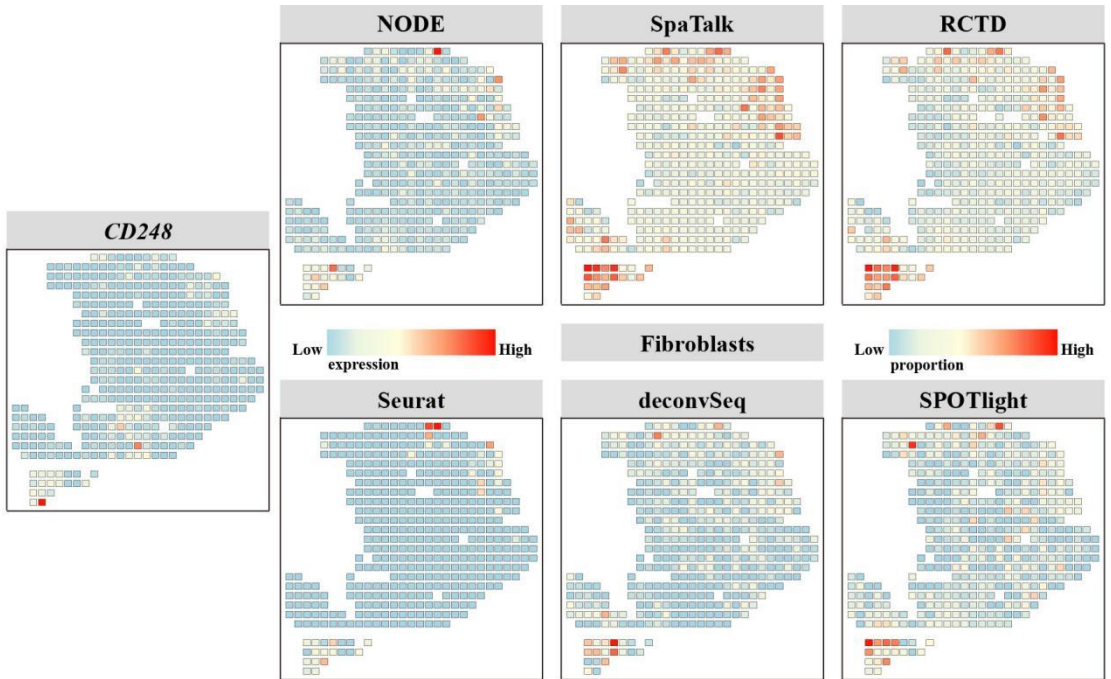
Supplementary Fig. 28. Comparison of the distribution and occupancy of ductal centroacinar in space inferred by each of the deconvolution methods and the expression of ductal centroacinar's corresponding marker genes (*CRISP3*) in space in human pancreatic ductal adenocarcinoma (PDAC) data.

The proportion of ductal centroacinar inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*CRISP3*). We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression. Ductal centroacinar cells inferred by NODE are enriched in the ductal epithelium region and partially in the normal pancreatic tissue region, consistent with the spatial expression pattern of the marker gene *CRISP3* and early literature¹⁹. In contrast, this feature was not found by other methods, and none of their deconvolution results revealed the characteristic cellular distribution of ductal centroacinar aggregates toward normal tissue.



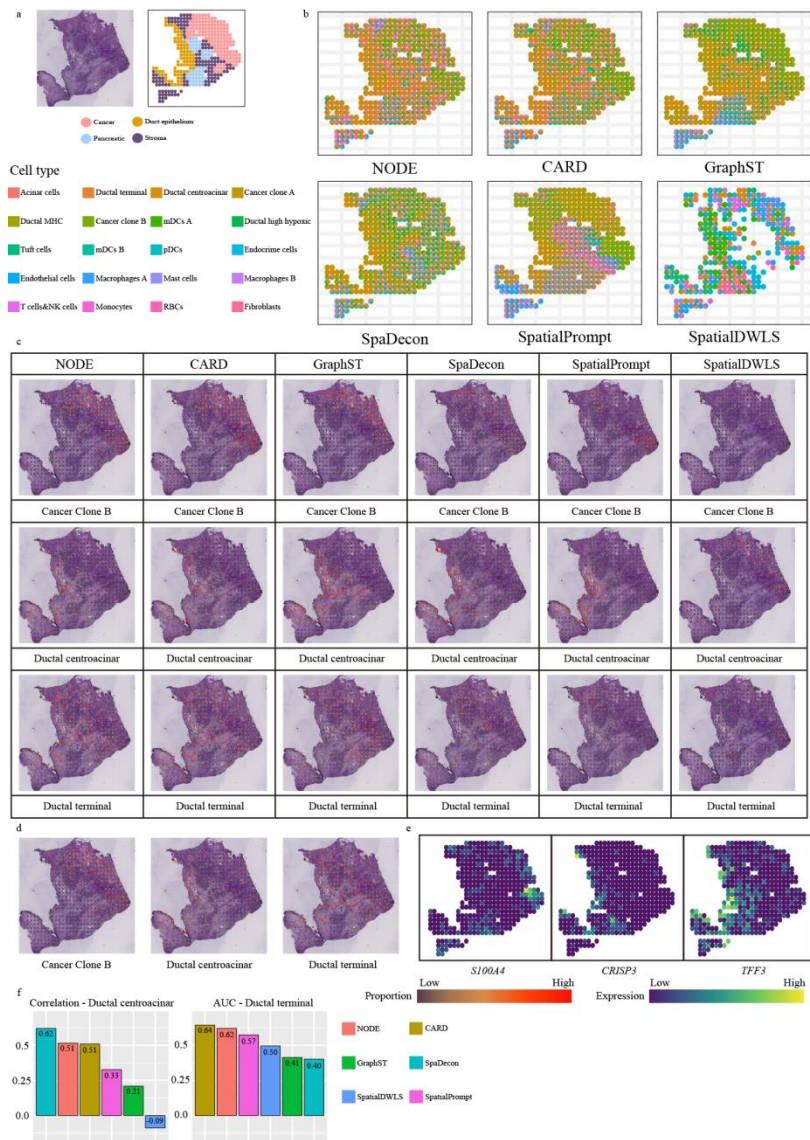
Supplementary Fig. 29. Comparison of the distribution and occupancy of Fibroblasts in space inferred by each of the deconvolution methods and the expression of Fibroblasts' corresponding marker genes (*CD248*) in space in human pancreatic ductal adenocarcinoma (PDAC) data.

The proportion of Fibroblasts inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*CD248*). We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression. In this result, the expression of the marker gene (*CD248*) was mainly concentrated in the lower left and upper right portions of the section, while the expression was smaller in all other positions. The deconvolution results of NODE coincided with this distribution feature. Although other results also found this feature, there were more false positives that appeared in other reverse convolution results in the distribution of other positions except for the lower left and upper right part of the section.



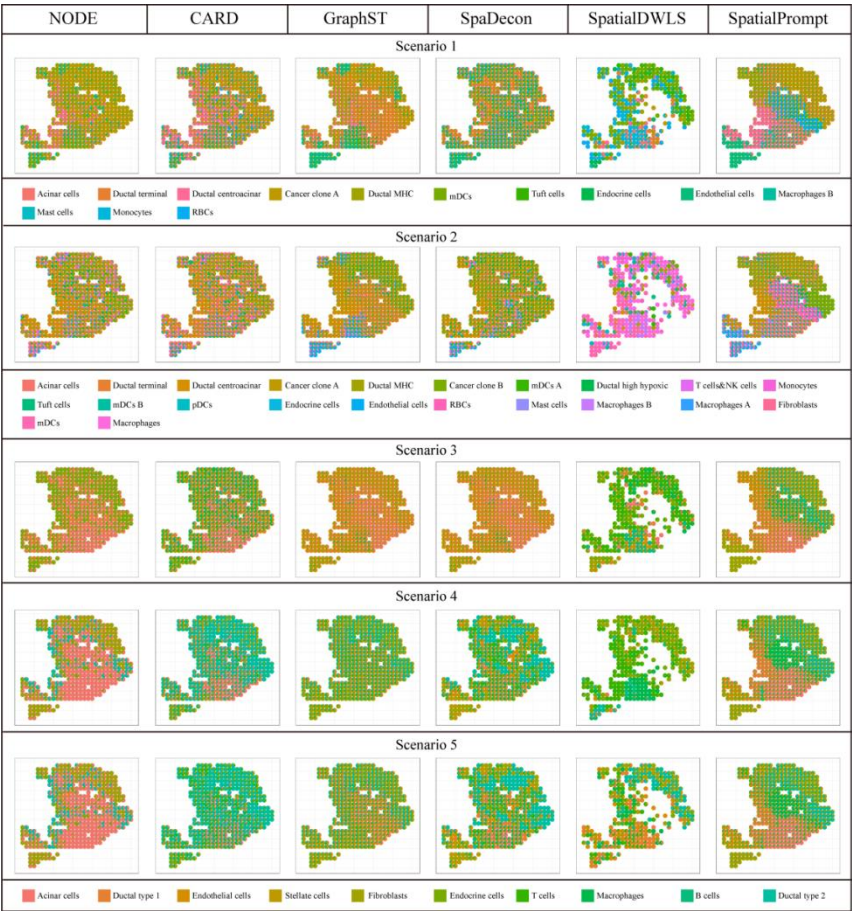
Supplementary Fig. 30. Comparison of NODE's performance with recently developed methods (CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³) on human pancreatic ductal adenocarcinoma (PDAC) data.

a. H&E stained image of PDAC and manual annotation of PDAC. **b.** The spatial scatter pie chart displays the proportion and composition of cells at each spatial location in the inferred results of different deconvolution methods(CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³). **c.** The distribution of cancer clone B, ductal centroacinar, and ductal terminal in the inferred results of different deconvolution methods. **d.** Aggregation locations of cancer clone B, ductal centroacinar, and ductal terminal cells identified by NODE. **e.** Expression of marker genes for cancer clone B, ductal centroacinar, and ductal terminal cells: *S100A4*, *CRISP3*, and *TFF3*. **f.** Correlation coefficients between ductal centroacinar cells and their marker genes across all deconvolution methods and AUC values for ductal terminal cells and their marker genes across all deconvolution methods.



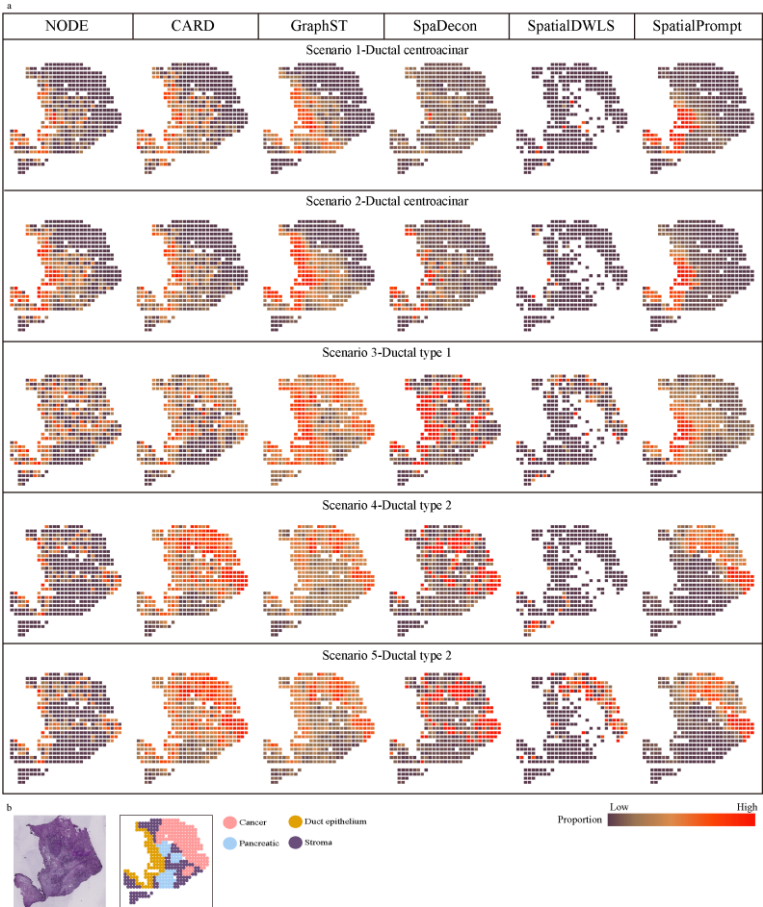
Supplementary Fig. 31. Comparison of NODE's performance with recently developed methods (CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³) on human pancreatic ductal adenocarcinoma (PDAC) data with unmatched scRNA-seq.

Performance of CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³ on PDAC data analysis with unmatched scRNA-seq data. In scenario 1, we used the scRNA-seq dataset from another patient (PDAC-B) for deconvolution (ductal centroacinar). In scenario 2 We used scRNA-seq data fusion of matched scRNA-seq with PDAC-B (ductal centroacinar). In addition to using the scRNA-seq data from the same study, we also obtained one external scRNA-seq data (10x Chromium) from Peng et al (Genome Sequence Archive under project PRJCA001063).²¹ This data was collected from 11 control pancreas and 24 PDAC tumors and was sequenced using the 10x Chromium platform. The data contains 24,005 genes and 57,530 cells from 10 cell types. We split the data into three subsets: the PengNormal data that consists of the 11 control samples; the PengTumor data that consists of the 24 PDAC tumors; and the Peng data that consists of all 35 samples. Subsequently, we recorded the deconvolution of PDAC-A using these three sets of scRNA-seq data as reference data in scenario 3 (ductal type 1), scenario 4 (ductal type 2) and scenario 5 (ductal type 2) respectively.



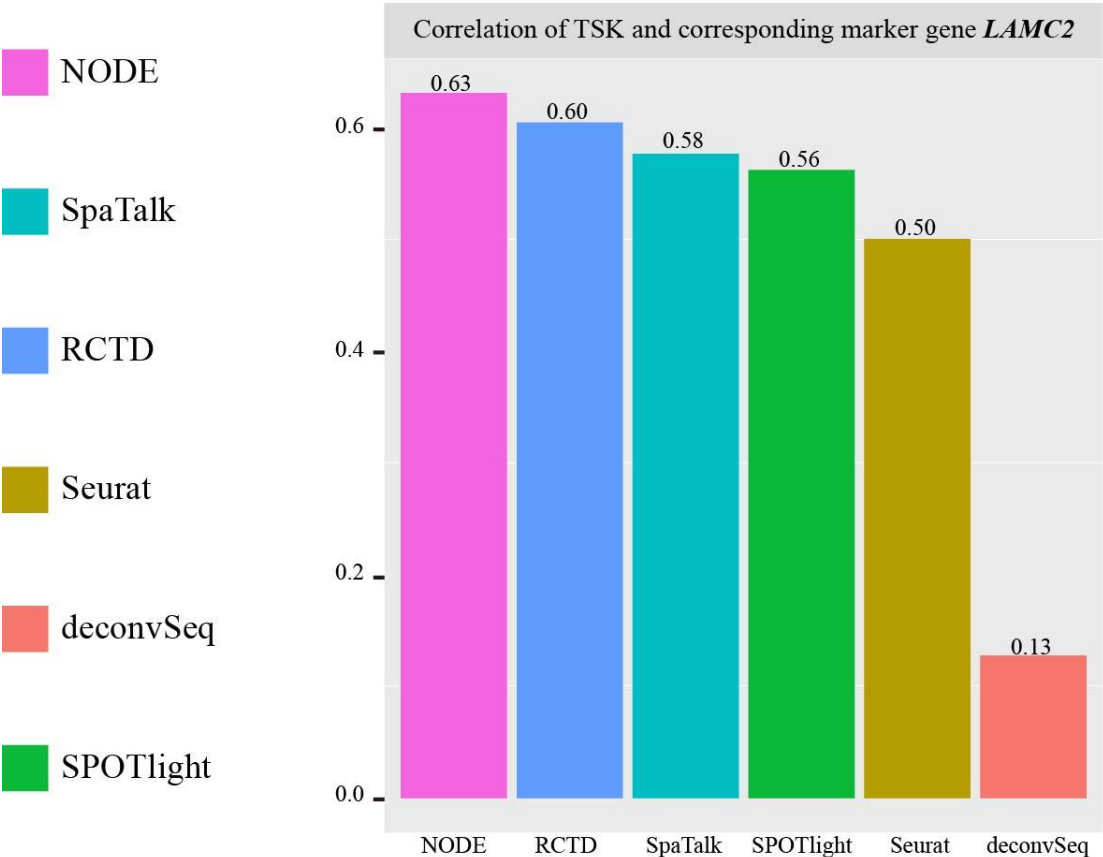
Supplementary Fig. 32. Comparison of NODE's performance with recently developed methods (CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³) on human pancreatic ductal adenocarcinoma (PDAC) data with unmatched scRNA-seq.

a. Performance of CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³ on PDAC data analysis with unmatched scRNA-seq data. In scenario 1, we used the scRNA-seq dataset from another patient (PDAC-B) for deconvolution (ductal centroacinar). In scenario 2 We used scRNA-seq data fusion of matched scRNA-seq with PDAC-B (ductal centroacinar). In addition to using the scRNA-seq data from the same study, we also obtained one external scRNA-seq data (10x Chromium) from Peng et al (Genome Sequence Archive under project PRJCA001063).²¹ This data was collected from 11 control pancreas and 24 PDAC tumors and was sequenced using the 10x Chromium platform. The data contains 24,005 genes and 57,530 cells from 10 cell types. We split the data into three subsets: the PengNormal data that consists of the 11 control samples; the PengTumor data that consists of the 24 PDAC tumors; and the Peng data that consists of all 35 samples. Subsequently, we recorded the deconvolution of PDAC-A using these three sets of scRNA-seq data as reference data in scenario 3 (ductal type 1), scenario 4 (ductal type 2) and scenario 5 (ductal type 2) respectively. **b.** H&E stained image of PDAC and manual annotation of PDAC.



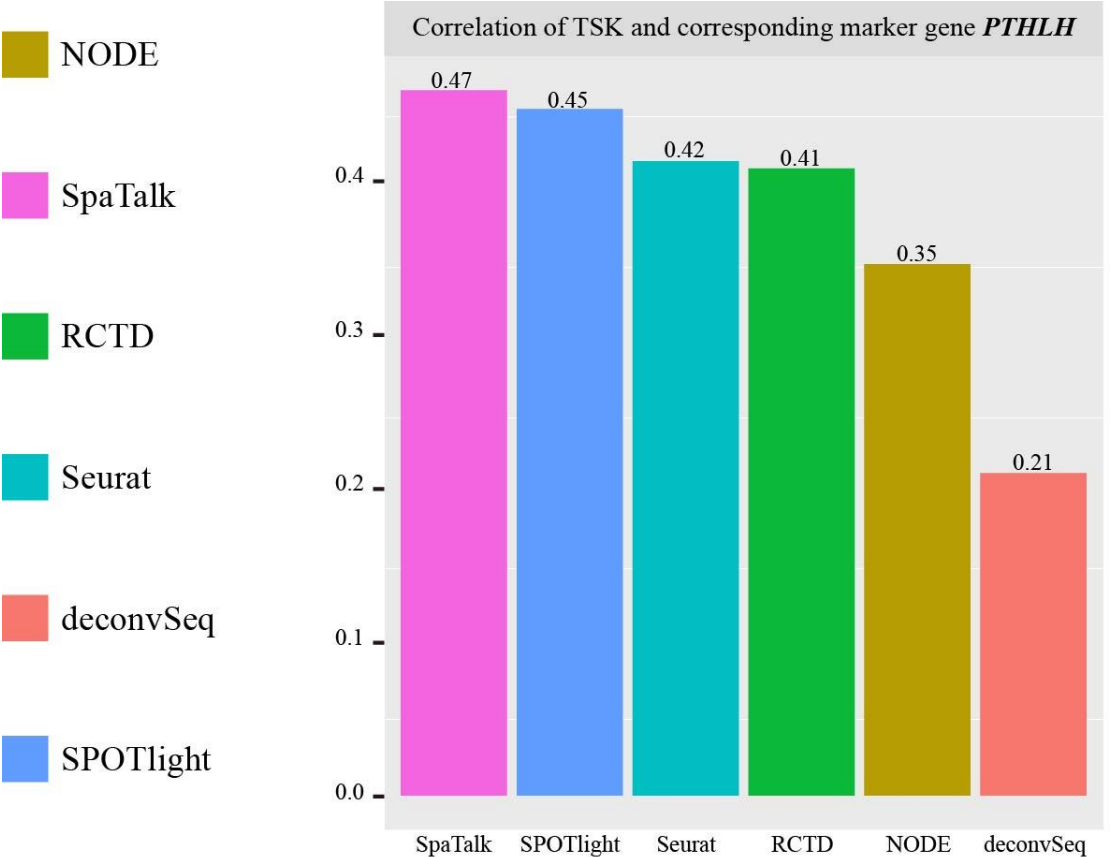
Supplementary Fig. 33. Comparison of the correlation coefficients between TSK cell distribution vectors and the expression vectors of marker genes (*LAMC2*)² corresponding to GC cell by different deconvolution methods in human skin squamous cell carcinoma (SCC) ST dataset².

In the slice one data of the SCC data, we compared the correlation coefficients of each of the deconvolution methods. Specifically, after applying all the deconvolution methods to the SCC data (SCC_P2.rep1), we extracted the distribution of TSK in all the spots based on different methods. After that, we extracted the expression of marker genes (*LAMC2*) corresponding to TSK in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of TSK and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution method. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene. In this set of correlation coefficient results, NODE ranks first, having the highest correlation coefficient. In other words, the cellular distribution inferred by NODE was most similar to the distribution of marker genes.



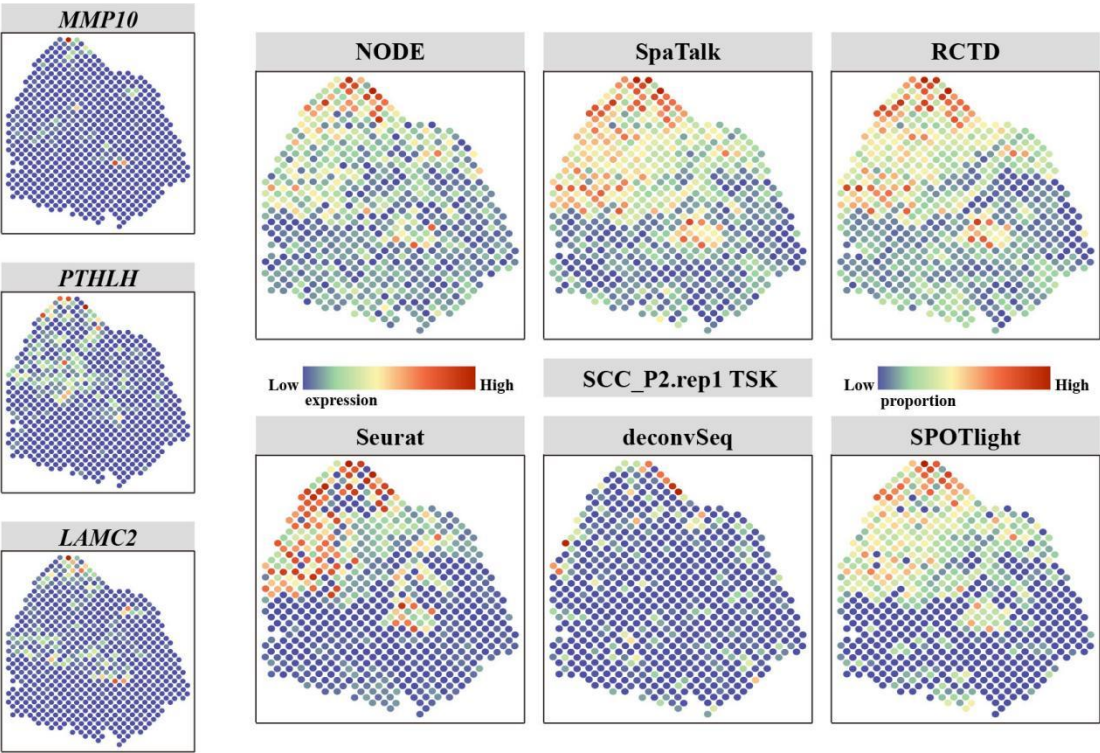
Supplementary Fig. 34. Comparison of the correlation coefficients between TSK cell distribution vectors and the expression vectors of marker genes (*PTHLH*) corresponding to TSK cell by different deconvolution methods in human skin squamous cell carcinoma (SCC) ST dataset.

In the slice one data of the SCC data, we compared the correlation coefficients of each of the deconvolution methods. Specifically, after applying all the deconvolution methods to the SCC data (SCC_P2.rep1), we extracted the distribution of TSK in all the spots based on different methods. After that, we extracted the expression of marker genes (*PTHLH*) corresponding to TSK in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of TSK and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution method. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene.



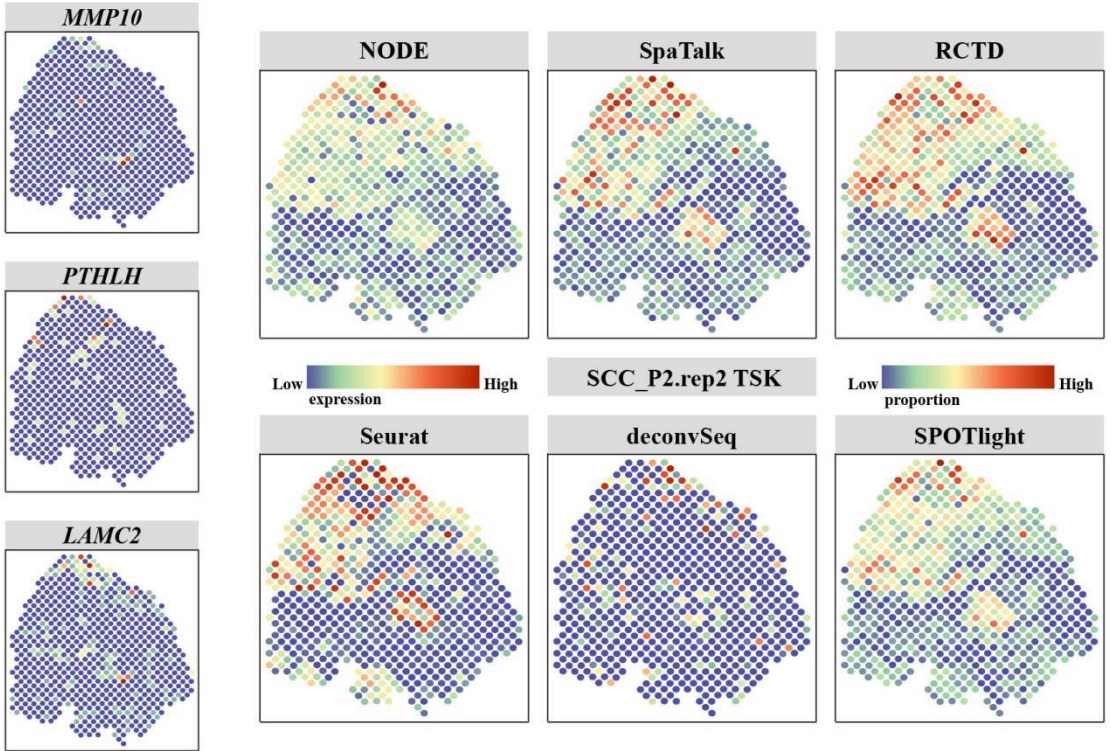
Supplementary Fig. 35. Comparison of the distribution and occupancy of TSK in space inferred by each of the deconvolution methods and the expression of TSK's corresponding marker genes (*MMP10*, *PTHLH*, *LAMC2*) in space in human skin squamous cell carcinoma (SCC) data (SCC_P2.rep1).

We analyzed slice 1 of the SCC data (SCC_P2.rep1). Here, we demonstrate the proportion of TSK inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*MMP10*, *PTHLH*, *LAMC2*). We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression. In this result, the expression of the marker gene (*MMP10*, *PTHLH*, *LAMC2*) was mainly concentrated in the upper right portions of the section, while the expression was smaller in all other positions. The deconvolution results of NODE coincided with this distribution feature. Although other results also found this feature, there were more false positives that appeared in other deconvolution results in the distribution of other positions except for the upper right part of the section. In particular, in the central region, the marker genes were not expressed in the spatial transcriptomics data, and other methods have judged this part of the region as a location of cell aggregation.



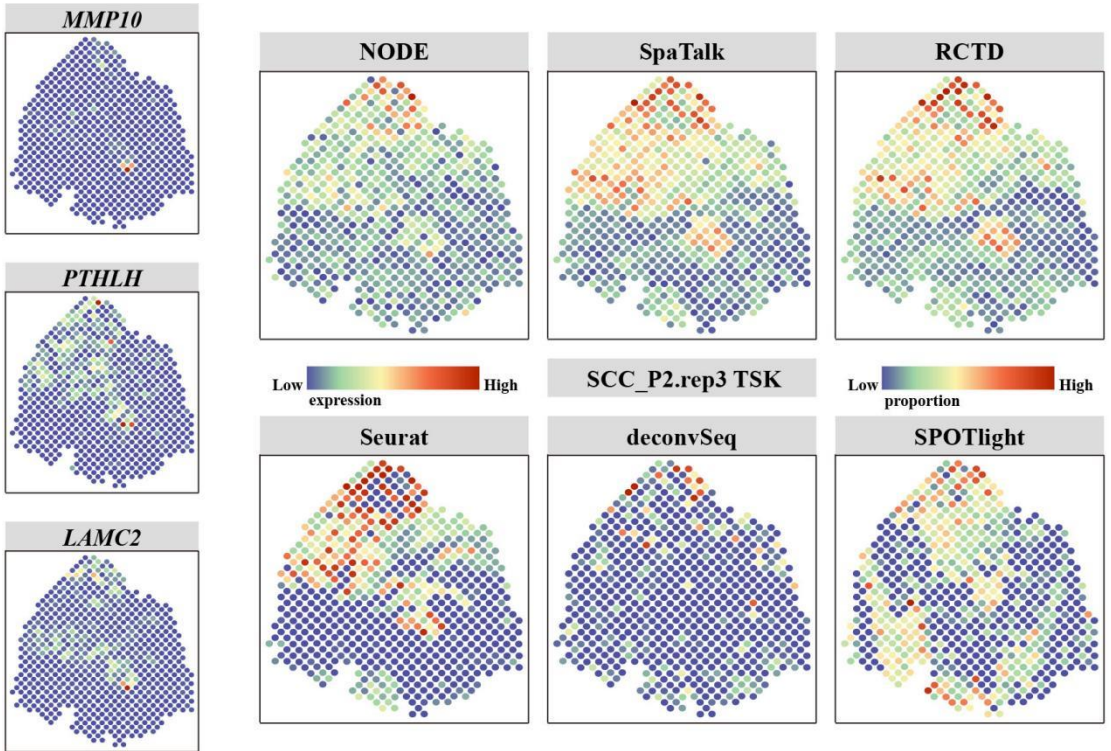
Supplementary Fig. 36. Comparison of the distribution and occupancy of TSK in space inferred by each of the deconvolution methods and the expression of TSK's corresponding marker genes (*MMP10*, *PTHLH*, *LAMC2*) in space in human skin squamous cell carcinoma (SCC) data (SCC_P2.rep2).

We analyzed slice 2 of the SCC data (SCC_P2.rep2). Here, we demonstrate the proportion of TSK inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*MMP10*, *PTHLH*, *LAMC2*). We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression.



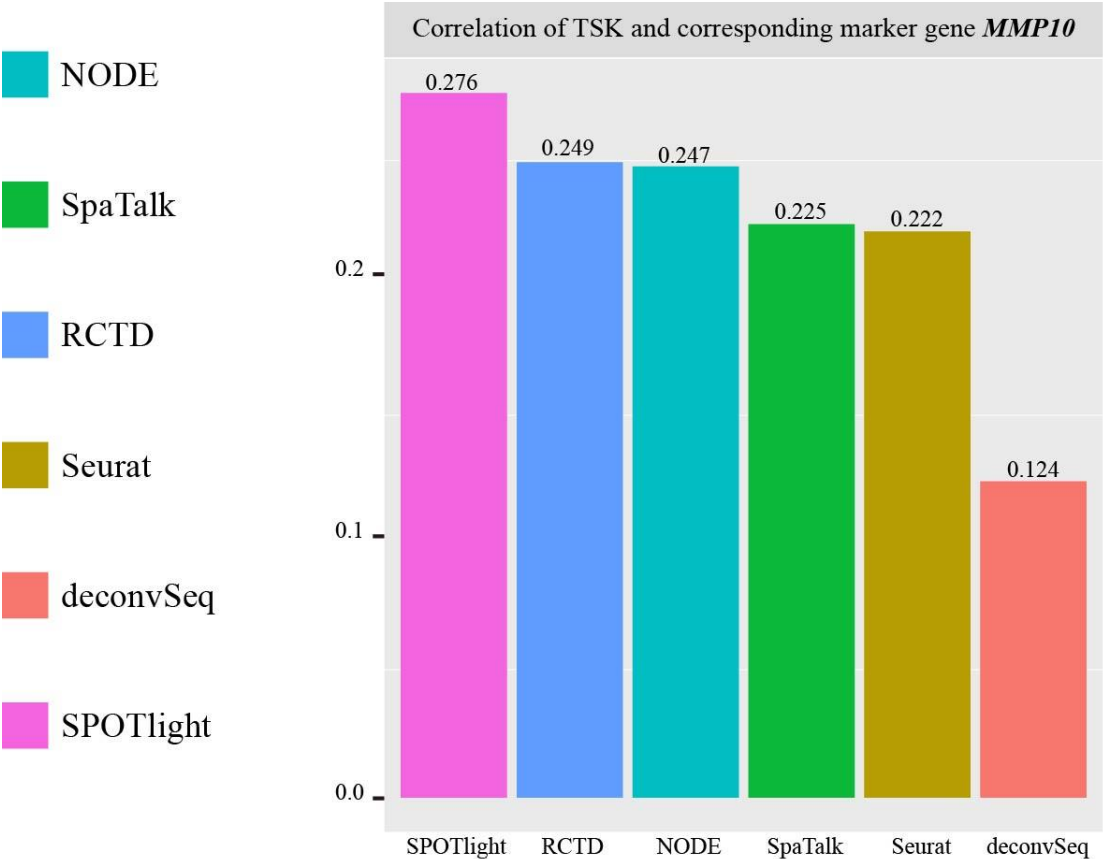
Supplementary Fig. 37. Comparison of the distribution and occupancy of TSK in space inferred by each of the deconvolution methods and the expression of TSK's corresponding marker genes (*MMP10*, *PTHLH*, *LAMC2*) in space in human skin squamous cell carcinoma (SCC) data (SCC_P2.rep3).

We analyzed slice 3 of the SCC data (SCC_P2.rep3). Here, we demonstrate the proportion of TSK inferred by NODE and other deconvolution methods on each spatial location and the expression levels of the corresponding cell-type-specific marker genes (*MMP10*, *PTHLH*, *LAMC2*). We plotted the distribution of cells with the expression of the corresponding marker genes according to the magnitude of their values in a scatter heat map. High proportion or expression represent more clustered cell distribution or gene expression.



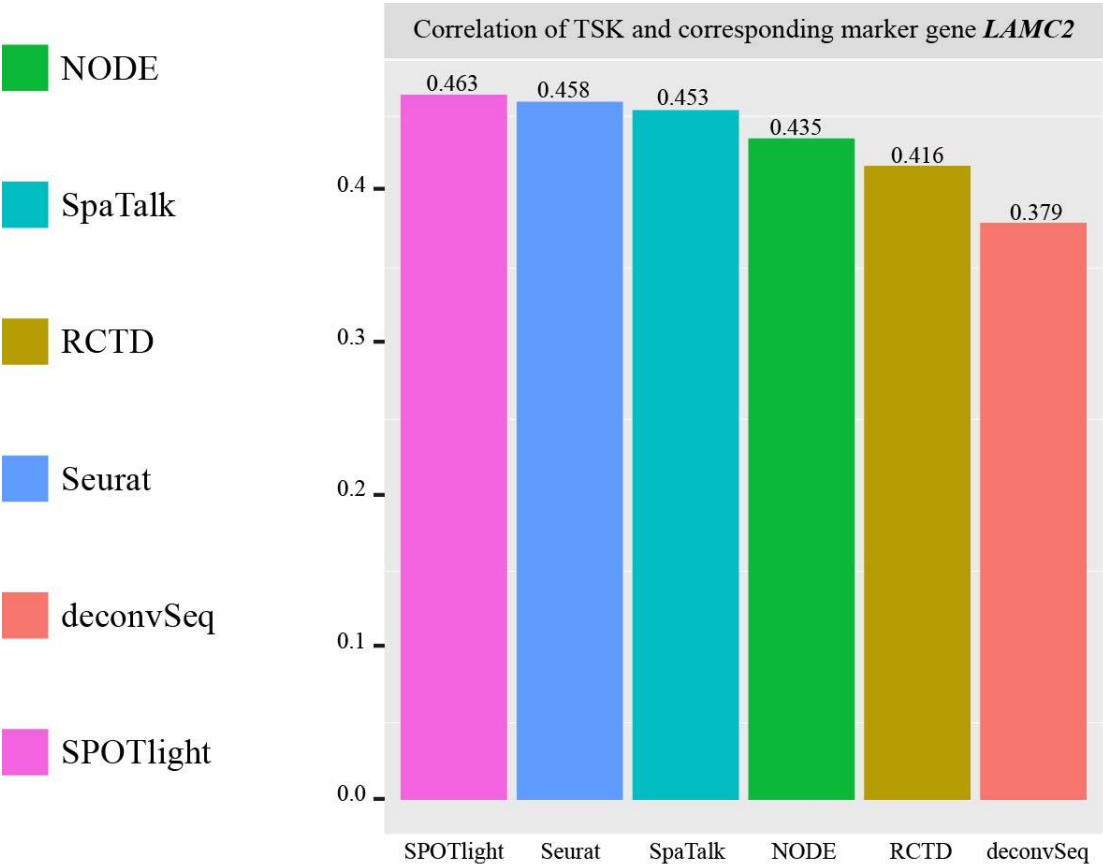
Supplementary Fig. 38. Comparison of the correlation coefficients between TSK cell distribution vectors and the expression vectors of marker genes (*MMP10*) corresponding to TSK cell by different deconvolution methods in human skin squamous cell carcinoma (SCC) ST dataset (SCC_P2.rep2).

In the slice two data of the SCC data, we compared the correlation coefficients of each of the deconvolution methods. Specifically, after applying all the deconvolution methods to the SCC data (SCC_P2.rep2), we extracted the distribution of TSK in all the spots based on different methods. After that, we extracted the expression of marker genes (*MMP10*) corresponding to TSK in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of TSK and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution method. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene. In this result, NODE ranked third in the correlation coefficient, differing from the second by only 0.002.



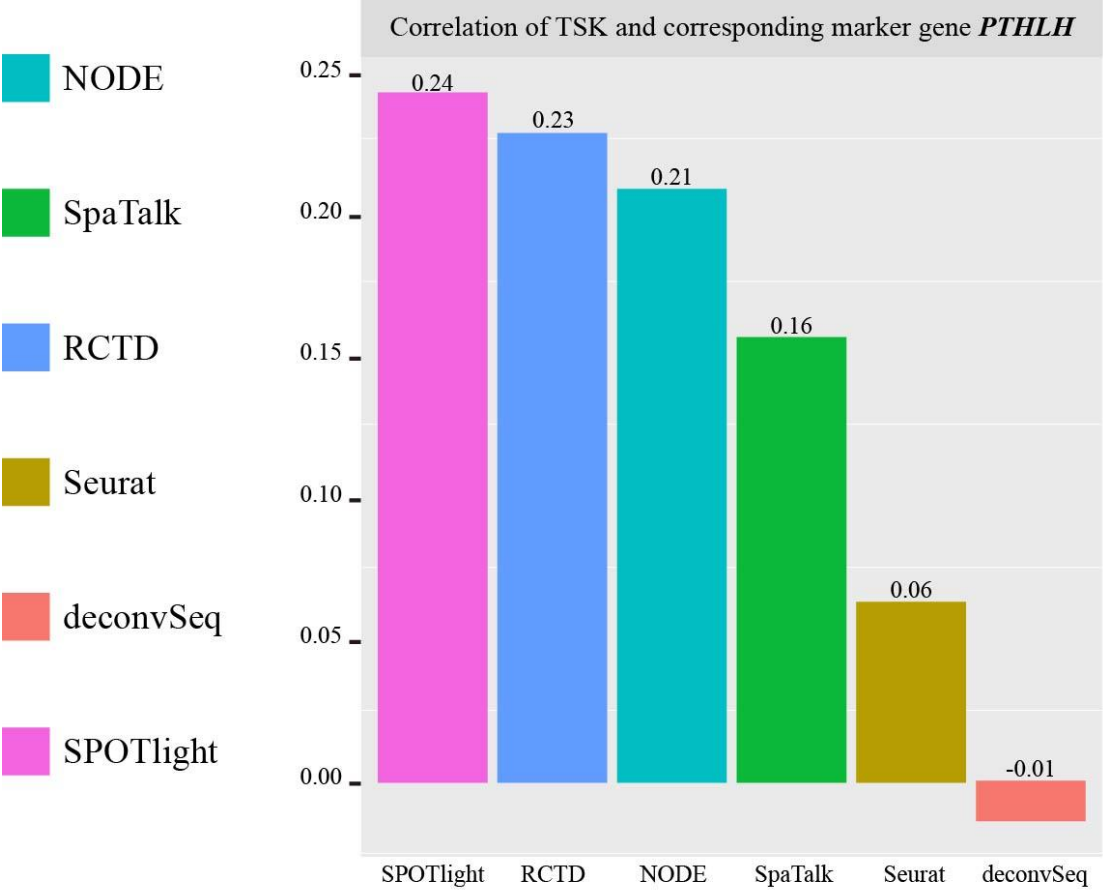
Supplementary Fig. 39. Comparison of the correlation coefficients between TSK cell distribution vectors and the expression vectors of marker genes (*LAMC2*) corresponding to TSK cell by different deconvolution methods in human skin squamous cell carcinoma (SCC) ST dataset (SCC_P2.rep2).

In the slice two data of the SCC data, we compared the correlation coefficients of each of the deconvolution methods. Specifically, after applying all the deconvolution methods to the SCC data (SCC_P2.rep2), we extracted the distribution of TSK in all the spots based on different methods. After that, we extracted the expression of marker genes (*LAMC2*) corresponding to TSK in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of TSK and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution method. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene.



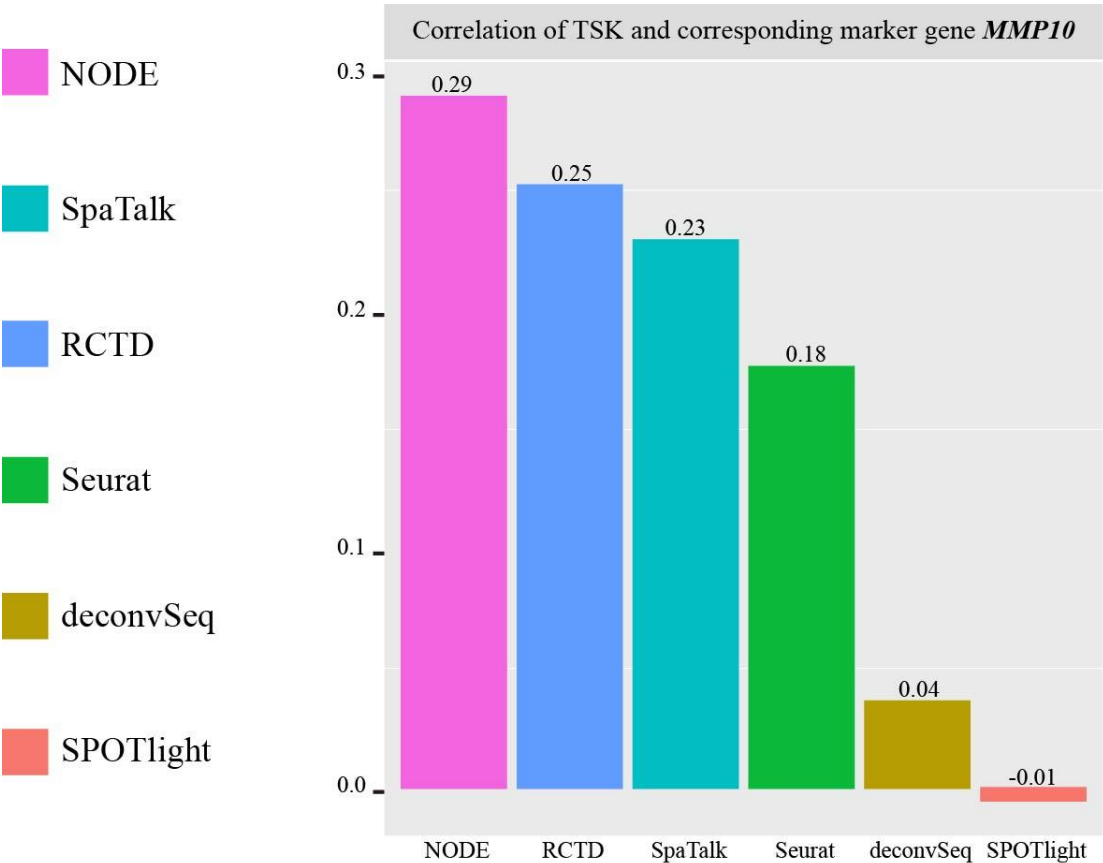
Supplementary Fig. 40. Comparison of the correlation coefficients between TSK cell distribution vectors and the expression vectors of marker genes (*PTHLH*) corresponding to TSK cell by different deconvolution methods in human skin squamous cell carcinoma (SCC) ST dataset (SCC_P2.rep2).

In the slice two data of the SCC data, we compared the correlation coefficients of each of the deconvolution methods. Specifically, after applying all the deconvolution methods to the SCC data (SCC_P2.rep2), we extracted the distribution of TSK in all the spots based on different methods. After that, we extracted the expression of marker genes (*PTHLH*) corresponding to TSK in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of TSK and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution method. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene. In this result, NODE ranked third in the correlation coefficient, differing from the first only by 0.03.



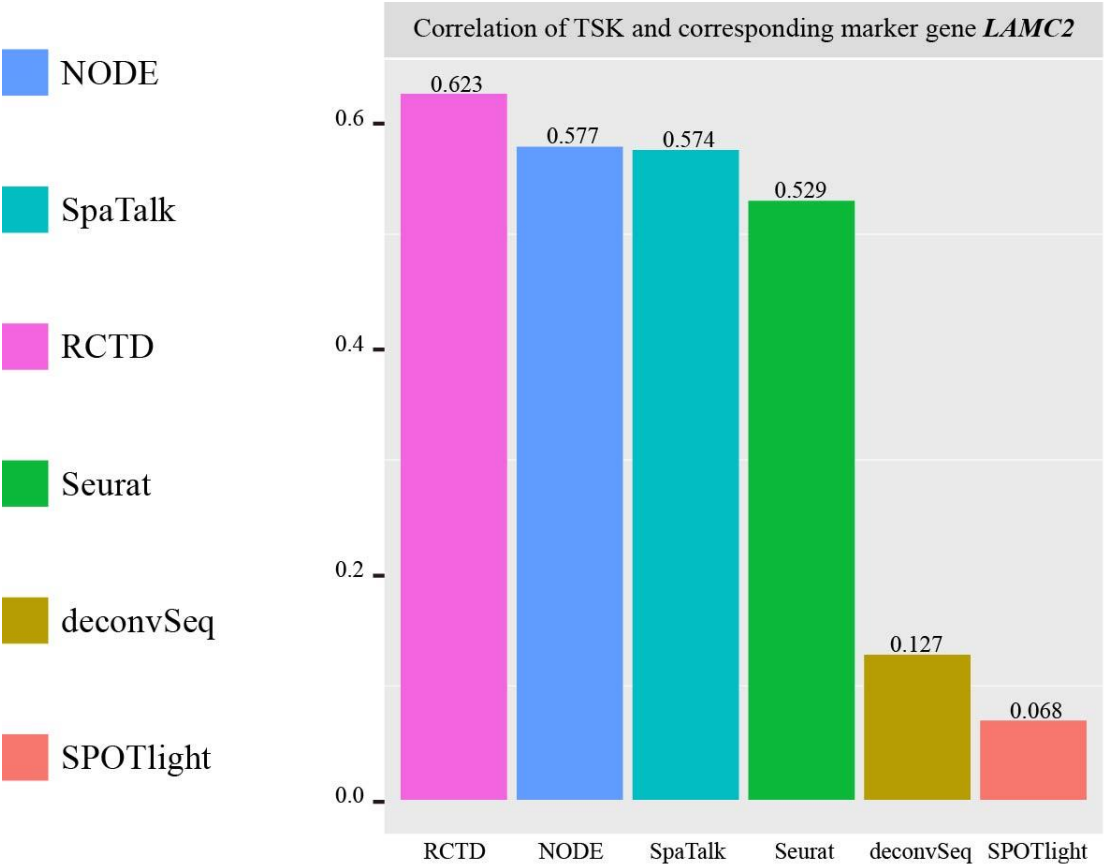
Supplementary Fig. 41. Comparison of the correlation coefficients between TSK cell distribution vectors and the expression vectors of marker genes (*MMP10*) corresponding to TSK cell by different deconvolution methods in human skin squamous cell carcinoma (SCC) ST dataset (SCC_P2.rep3).

In the slice three data of the SCC data, we compared the correlation coefficients of each of the deconvolution methods. Specifically, after applying all the deconvolution methods to the SCC data (SCC_P2.rep3), we extracted the distribution of TSK in all the spots based on different methods. After that, we extracted the expression of marker genes (*MMP10*) corresponding to TSK in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of TSK and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution method. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene. In this result, NODE ranked first in the correlation coefficient and possessed the highest correlation coefficient, in other words, the cellular distribution results inferred by NODE were the most similar to the distribution of marker genes.



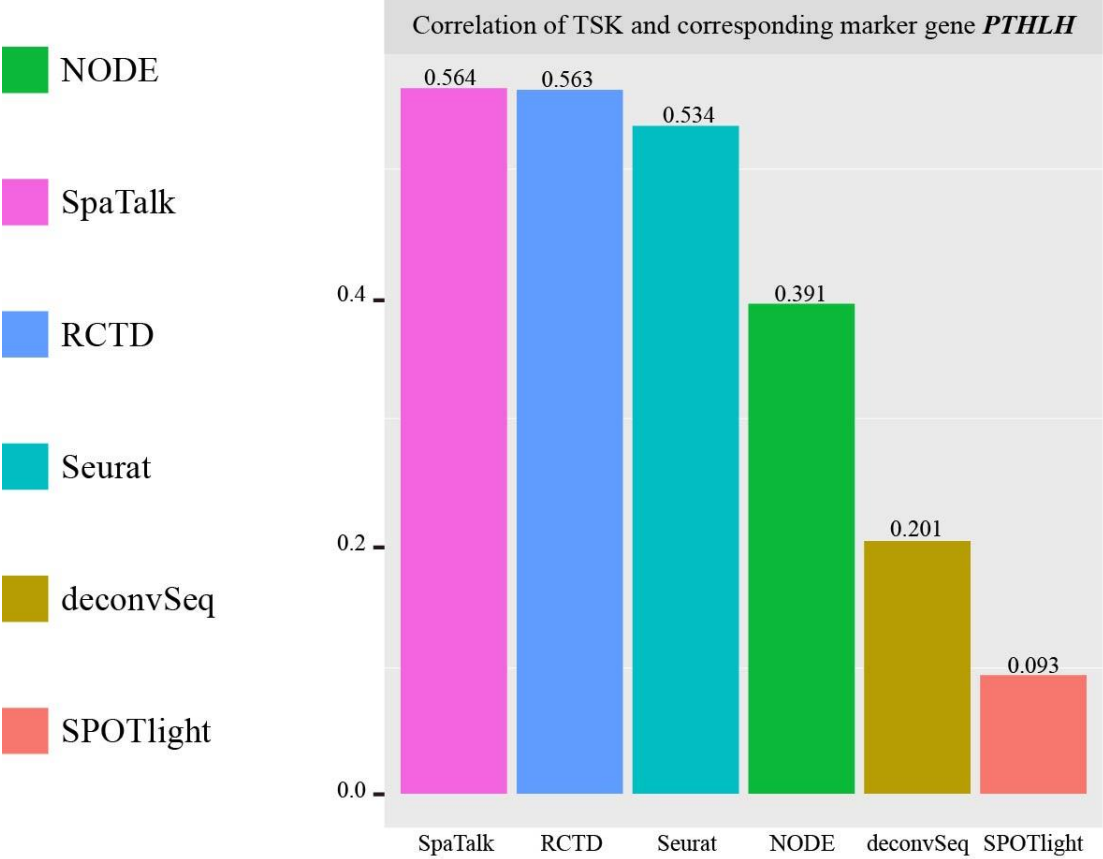
Supplementary Fig. 42. Comparison of the correlation coefficients between TSK cell distribution vectors and the expression vectors of marker genes (*LAMC2*) corresponding to TSK cell by different deconvolution methods in human skin squamous cell carcinoma (SCC) ST dataset (SCC_P2.rep3).

In the slice three data of the SCC data, we compared the correlation coefficients of each of the deconvolution methods. Specifically, after applying all the deconvolution methods to the SCC data (SCC_P2.rep3), we extracted the distribution of TSK in all the spots based on different methods. After that, we extracted the expression of marker genes (*LAMC2*) corresponding to TSK in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of TSK and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution method. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene. In this result, NODE ranked second in the correlation coefficient.



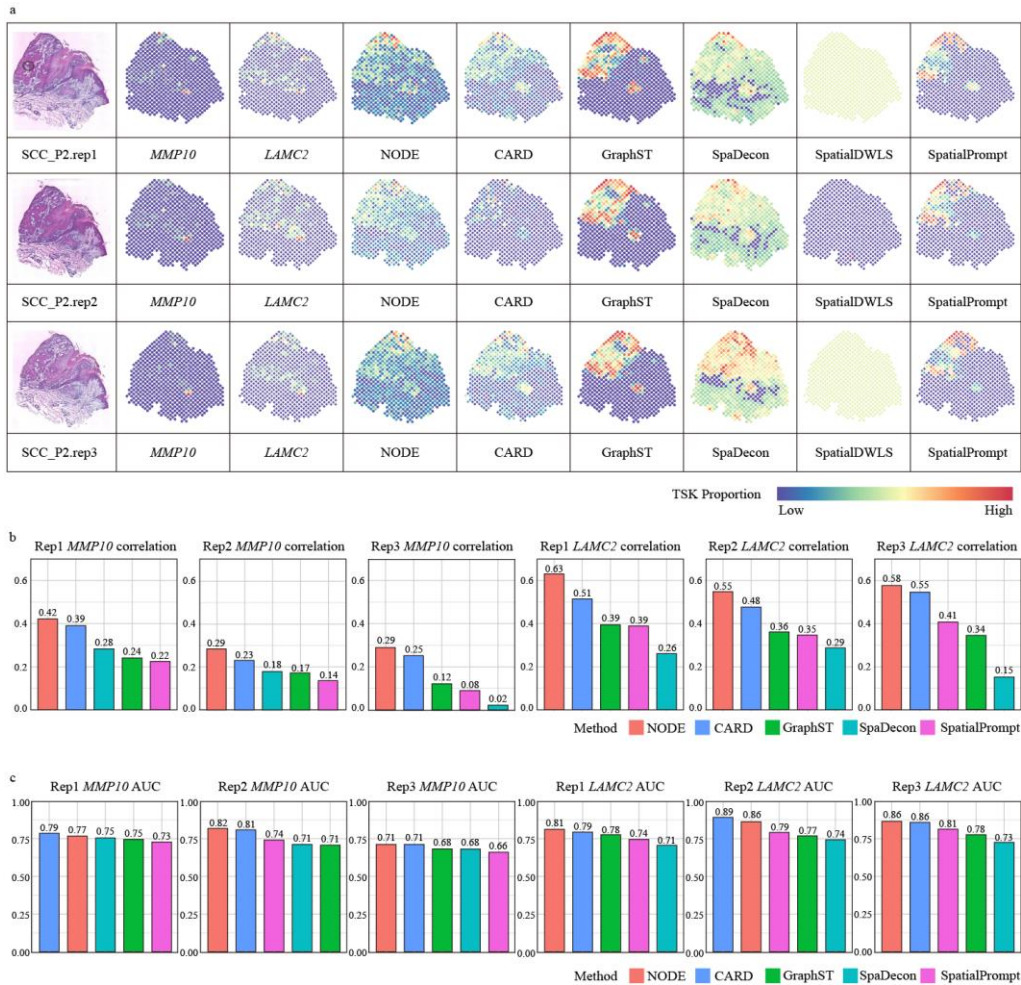
Supplementary Fig. 43. Comparison of the correlation coefficients between TSK cell distribution vectors and the expression vectors of marker genes (*PTHLH*) corresponding to TSK cell by different deconvolution methods in human skin squamous cell carcinoma (SCC) ST dataset (SCC_P2.rep3).

In the slice three data of the SCC data, we compared the correlation coefficients of each of the deconvolution methods. Specifically, after applying all the deconvolution methods to the SCC data (SCC_P2.rep3), we extracted the distribution of TSK in all the spots based on different methods. After that, we extracted the expression of marker genes (*PTHLH*) corresponding to TSK in the spatial transcriptomics data and calculated the correlation coefficients between the distribution of TSK and the expression of marker genes. Finally, we plotted the correlation coefficients as bar graphs based on the deconvolution method and marked the magnitude of the correlation coefficients on the bars. The y-axis (vertical axis) represents the correlation coefficient, and the x-axis (horizontal axis) represents the deconvolution method. A larger value on the y-axis indicates that the distribution of cells in the deconvolution result is closer to that of the marker gene.



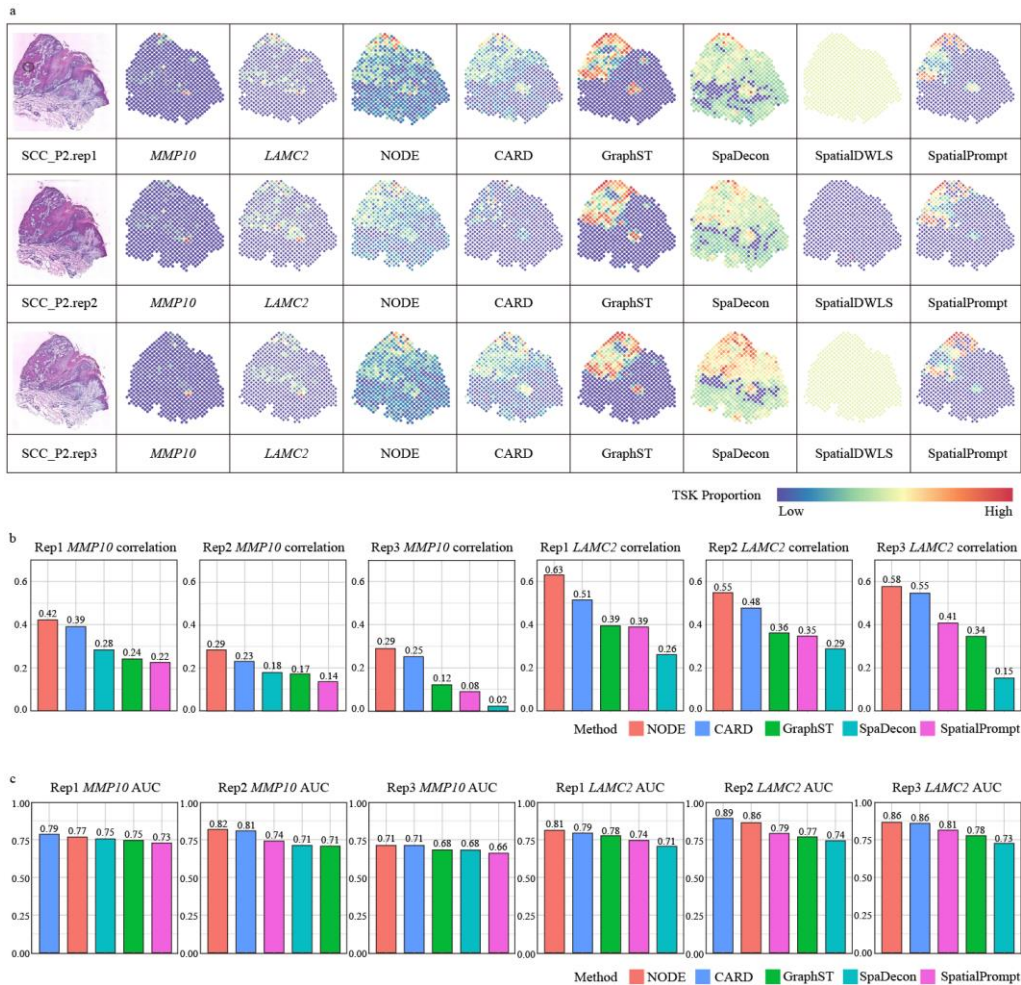
Supplementary Fig. 44. Comparison of NODE's performance with recently developed methods (CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³) on human skin squamous cell carcinoma(SCC) data.

a. H&E stained image of PDAC, clustering locations and expression of marker genes(*MMP10*,*LAMC2*), and the Spatial scatter heat map showing the proportion and distribution of TSK cells in the inferred results of different deconvolution methods(CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹² and SpatialPrompt¹³). **b.** Correlation coefficients for TSK cells and their marker genes *MMP10* and *LAMC2* in the deconvolution results of all methods across replicates 1-3 (rep1-rep3). **c.** AUC values for TSK cells and their marker genes *MMP10* and *LAMC2* in the deconvolution results of all methods across replicates 1-3 (rep1-rep3).



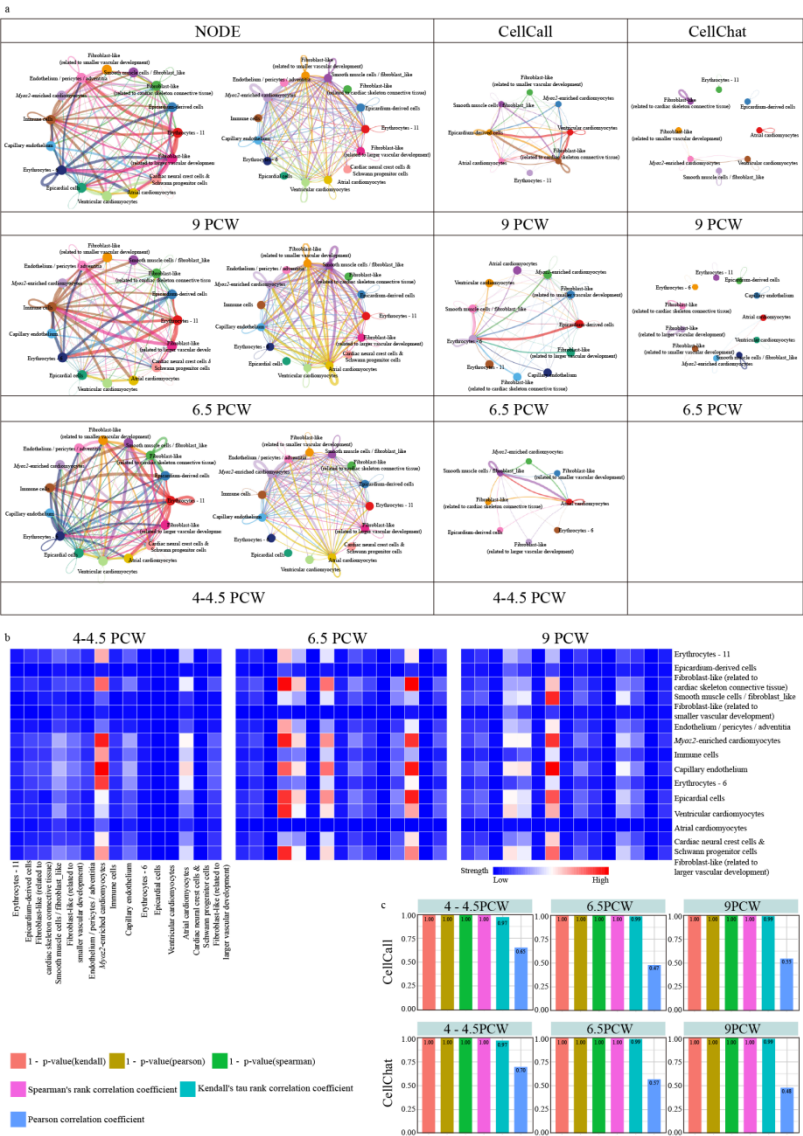
Supplementary Fig. 45. Comparison of NODE's ability to resist spatial noise and batch effects with recently developed methods (CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³) on human skin squamous cell carcinoma(SCC) data.

a. H&E stained image of PDAC, clustering locations and expression of marker genes(*MMP10*,*LAMC2*), and the Spatial scatter heat map showing the proportion and distribution of TSK cells in the inferred results of different deconvolution methods(CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³). **b.** Correlation coefficients for TSK cells and their marker genes *MMP10* and *LAMC2* in the deconvolution results of all methods across replicates 1-3 (rep1-rep3). **c.** AUC values for TSK cells and their marker genes *MMP10* and *LAMC2* in the deconvolution results of all methods across replicates 1-3 (rep1-rep3).



Supplementary Fig. 46. Communication performance of CellCall, CellChat, and NODE in human cardiac datasets.

a. Three stages of inferred communications by NODE, CellCall, and CellChat. For ease of visualization, NODE's inferred communications are divided into two parts. **b.** Heatmap of NODE-inferred communications across the three periods. The cell order on the horizontal axis for 6.5 PCW and 9 PCW aligns with that of 4-4.5 PCW. **c.** Correlation coefficients of CellChat and CellCall with NODE. To calculate these coefficients, it is necessary to extend CellCall and CellChat to the same resolution as NODE for cellular communication analysis after scRNA-seq.



Supplementary Table

Supplementary Table 1. List of cells and their corresponding marker genes in the three spatially resolved transcriptomics datasets.

Cell-Type Name	Corresponding Marker Gene Name	Belonging to Data Set	Cell layers/ Tissue domains of marker gene distribution
GC	Penk	Mouse Olfactory Bulb (MOB)	GCL
M-TC	Cdhr1	Mouse Olfactory Bulb (MOB)	MCL
PGC	Apold	Mouse Olfactory Bulb (MOB)	GL
Cancer clone A	TM4SF1	Human pancreatic ductal adenocarcinoma (PDAC)	Cancerous region
Cancer clone B	S100A4	Human pancreatic ductal adenocarcinoma (PDAC)	Cancerous region
Ductal terminal	TFF3	Human pancreatic ductal adenocarcinoma (PDAC)	Duct epithelium region
Ductal high hypoxic	APOL1	Human pancreatic ductal adenocarcinoma (PDAC)	Cancerous region
Ductal centroacinar	CRISP3	Human pancreatic ductal adenocarcinoma (PDAC)	Duct epithelium region
Fibroblasts	CD248	Human pancreatic ductal adenocarcinoma (PDAC)	Cancerous region and Stroma
Tumor specific Keratinocyte (TSK)	MMP10	Human skin squamous cell carcinoma (SCC)	
Tumor specific Keratinocyte (TSK)	PTHLH	Human skin squamous cell carcinoma (SCC)	
Tumor specific Keratinocyte (TSK)	LAMC2	Human skin squamous cell carcinoma (SCC)	

The table contains cell-type name (1st column), corresponding marker gene name (2nd column), belonging to data set (3rd column), cell layers/tissue domains of marker gene distribution (4th column). We colored cell types from the same dataset with marker genes in the same colors.

Supplementary Table 2. List of 5 spatially resolved transcriptomics datasets and corresponding scRNA-seq datasets we used in our analysis. The scRNA-seq data sets were used to construct cell-type-specific profiles for each deconvolution method.

Data Set	Protocol	Year	#Gene	#Spatial locations	scRNAseq /Spatial	Data Type
Mouse Olfactory Bulb (Replicate12)	Spatial Transcriptomics	2016	16034	282	GSE121891	Spatial
Human pancreatic ductal adenocarcinoma (PDAC-A)	Spatial Transcriptomics	2020	25753	428	GSM3036911 (GSE111672)	Spatial
Human pancreatic ductal adenocarcinoma (PDAC-B)	Spatial Transcriptomics	2020	19739	225	GSM3405534 (GSE111672)	Spatial
Human skin squamous cell carcinoma (SCC_P2.rep1)	Spatial Transcriptomics	2020	17138	666	GSM4284316 (GSE144240)	Spatial
Human skin squamous cell carcinoma (SCC_P2.rep2)	Spatial Transcriptomics	2020	17344	649	GSM4284317 (GSE144240)	Spatial
Human skin squamous cell carcinoma (SCC_P2.rep3)	Spatial Transcriptomics	2020	17883	638	GSM4284318 (GSE144240)	Spatial
Developing human heart data (4-4.5PCW;3)	Spatial Transcriptomics	2019	39739	69	Developing human heart	Spatial
Developing human heart data (6.5PCW;8)	Spatial Transcriptomics	2019	39739	186	Developing human heart	Spatial
Developing human heart data	Spatial Transcriptomics	2019	39739	243	Developing human heart	Spatial

(9PCW;16)						
GSE121891	10x Chromium	2018	18560	21746	Mouse Olfactory Bulb	scRNA seq
GSE111672 (PDAC-A)	inDrop	2020	19736	1926	Human PDAC	scRNA seq
GSE111672 (PDAC-B)	inDrop	2020	19736	1733	Human PDAC	scRNA seq
GSM4284227 (SCC-P2)	10x Chromium	2020	32738	2736	Human SCC	scRNA seq
Developing human heart data	10x Chromium	2019	15323	3777	Developing human heart	scRNA seq

The table contains data set data set name (**1st** column), experimental platform (**2nd** column), year of publication (**3rd** column), number of genes (**4th** column), number of cells or spatial locations (**5th** column), scRNA-seq reference datasets used or used the scRNA-seq dataset to deconvolute the spatial dataset (**6th** column), data type to identify whether it is a spatial dataset or single cell RNA-seq dataset (**7th** column), H&E staining image used in this study (obtained from the website link, see the bottom of the table). We colored the pair of spatial datasets and scRNA-seq dataset reference as the same color to represent that deconvolution was performed using these two datasets. GSEXXX represents the GEO accession number that the corresponding dataset can be downloaded from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>.

MOB: <https://www.spatialresearch.org/>

PDAC-A: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111672>

PDAC-B: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111672>

SCC: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144240>

Developing human heart data: <https://www.spatialresearch.org/>

Supplementary Table 3. The running time (s) and computational efficiency of NODE and other deconvolution methods.

	Computation time (Seconds)					
Method	Human heart 4-4.5 PCW (genes = 13815 spots = 55)	Human heart 6.5 PCW (genes = 13815 spots = 103)	Human heart 9 PCW (genes = 13815 spots = 242)	MOB (genes = 14453 spots = 282)	simulation 2 (genes = 7515 spots = 339)	simulation 4 (genes = 7515 spots = 300)
NODE (CPU)	366.7	1042.7	8850.8	19041	16124	10173
SpaTalk (CPU)	431.5	744.03	1417.34	4034.2	480.4	380.8
RCTD (CPU)	171	251.63	460.4	1085.8	402.2	270.2
Seurat (CPU)	379.4	413.9	416.8	934.7	368.6	151.4
deconvSeq (CPU)	7364.8	7499.3	7496.7	35639.9	1441.1	609
SPOTlight (CPU)	645.2	806.5	807.34	844.9	252.1	173.9
CARD (CPU)	13.8	14.6	21.7	99.6	7.5	4.7
GraphST* (GPU)	35.1	39.4	43.4	337.2	38.5	32.8
SpaDecon (CPU)	77.6	69.6	73.8	420.2	32.5	19.3
SpatialDWLS (CPU)	43.8	69.0	120.6	410.1	92.6	44.3
SpatialPrompt (CPU)	34.4	33.4	36.8	264	29.9	23.6

***The GraphST require a GPU environment, so its runtime had been accelerated by GPU. And the runtime of other methods is the only runtime on CPU.**

Supplementary Table 4. The peak memory usage (MB) and computational efficiency of NODE and other deconvolution methods.

	Peak memory usage (MB)					
Method	Human heart 4-4.5 PCW (genes = 13815 spots = 55)	Human heart 6.5 PCW (genes = 13815 spots = 103)	Human heart 9 PCW (genes = 13815 spots = 242)	MOB (genes = 14453 spots = 282)	simulation 2 (genes = 7515 spots = 339)	simulation 4 (genes = 7515 spots = 300)
NODE	1245.3	1250.3	1265.0	7273.3	1200.3	897.8
SpaTalk	958.5	1138.7	3443	5856.2	3716.7	2797.4
RCTD	2098.1	2538.5	2361.1	5827	3695.3	2774.9
Seurat	2571.1	3129.9	2903.7	14531.5	3307.1	2799.7
deconvSeq	5856	5446	6251.5	33726.8	3331.6	2527.2
SPOTlight	2231	2199.7	3081	12189.1	1071.9	836.6
CARD	1214.1	1274.8	1310.7	6993.8	598.6	330.9
GraphST*	1881.8	1889.4	1911.4	10172.8	925.3	590.2
SpaDecon	1368.3	1373.4	1388.1	7396.1	864.9	529.8
SpatialDWLS	2776.9	2780.4	2791.1	15903.4	1313.3	719.9
SpatialPrompt	2558.0	2568.1	2597.4	7332.1	1539.7	1448.6

***The GraphST require a GPU environment, so GraphST not only uses general RAM but also spends a large of video memory. But in this table, we only calculated the usage of general RAM, and were not involved in the usage of video memory.**

To demonstrate the computational performance, we performed a uniform computational test for NODE and other 10 methods (GraphST¹⁰, SpaDecon¹¹, SpatialPrompt¹³, SpaTalk⁷, RCTD³, Seurat⁴, donvSeq⁵, SPOTlight⁶, CARD⁹ and SpatialDWLS¹²) on 6 datasets. And we used 4 real datasets and 2 simulated datasets to test the performance of these methods in speed (Table 3) and running peak memory (Table 4).

These methods were run on 12th Gen Intel(R) Core (TM) i9-12900H (20 cores), ~2.5GHZ (CPU). GraphST¹⁰ require an extra GPU, NVIDIA GeForce RTX 3060 Laptop GPUs. We can see that NODE rank 7th~10th in the 6 datasets for speed test (Table 3), and it rank 1st~6th in the 6 datasets for memory test (Table 4).

In brief, NODE is not a fast method (Table 3), but NODE is a good method in terms of accuracy as well as memory usage (Table 4). When it comes to large data,

994 NODE will perform well in terms of memory usage.
995

	Datasets				
Method	Simulated dataset	MOB dataset	PDAC dataset	SCC dataset	Developing human heart dataset
NODE	✓	✓	✓	✓	✓
SpaTalk	✓	✓	✓	✓	×
RCTD	✓	✓	✓	✓	×
Seurat	✓	✓	✓	✓	×
deconvSeq	✓	✓	✓	✓	×
SPOTlight	✓	✓	✓	✓	×
CARD	×	✓	✓	✓	×
GraphST	×	✓	✓	✓	×
SpaDecon	×	✓	✓	✓	×
SpatialDWLS	×	✓	✓	✓	×
SpatialPrompt	×	✓	✓	✓	×
CellChat	×	×	×	×	✓
CellCall	×	×	×	×	✓

997

998 NODE is a method for spatial transcriptomics deconvolution with the consideration of
999 spot-to-spot communications. So we primarily compared NODE against ten recently
1000 developed deconvolution algorithms (GraphST, SpaDecon, SpatialPrompt, SpaTalk,
1001 RCTD, Seurat, deconvSeq, SPOTlight, CARD and SpatialDWLS) across spatial
1002 transcriptomics datasets (MOB, PDAC and SCC). The CellChat and CellCall are
1003 cellular communication algorithms, and were used to compare the communication
1004 between spots with NODE based on a developing human heart data. The primary
1005 function of the BASS algorithm is the identification of spatial domains and BASS is
1006 not ideally suited for spatial transcriptomics deconvolution, so we only compared
1007 BASS with NODE in one dataset (MOB).

1008 Then, we add a new Supplementary Table 5 to summarize the information of
1009 methods applied datasets.

Supplementary Notes

Supplementary Note 1. Methodological details of NODE

1.1. Optimization model building and deconvolution results

Here, we provided details on how we construct the optimization model and deconvolute the result. Also, we describe how to do multiple batch iterations. In spatial transcriptomics data, each capture position is not the gene expression of a single cell but the mixed expression information of a group of cells, therefore, if we can know the gene expression of each single cell, we can multiply the cell expression with the number of cells to get the gene expression of each capture position. The converse is also possible, only we currently know the mixed expression of the cells and need to solve for the number of cells, so we use the optimization concept to decompose the spatial transcriptomics data as follows. Here we assumed that we have obtained the matrix Y and B as described in the main text.

$$Y = X_1 B^T + \epsilon \quad (s1)$$

where ϵ is the residual difference between the estimated value and the true value. In other words, the goal of optimization is to minimize ϵ . Before optimization, X_1 is the decision variable, it is the initial value of the cellular composition and distribution matrix. We first optimize the initial value according to the equation (s1), and subsequently we quantify the spatial communication matrix W_0 according to the initial value X_1 . It is worth noting here that we only take into account the effect of the neighboring spots and not those that are far away from each other. So, we optimize W_0 by putting a restriction on that each spot is only affected around the four closest spots. In other words, each row of W_0 has only four locations that do not have a value of 0. And it is important to note that W_0 is only a former solution before optimization. This optimization is formalized as follows:

$$Y = W_0 Y + X_1 B^T + \epsilon \quad (s2)$$

In equation (s2), we can obtain the initial communication strength W_0 , This is the initial solution state of the spatial communication matrix. Subsequently, we bring back W_0 to solve for a more precise cellular composition and distribution X_2 as follows:

$$Y = W_0 Y + X_2 B^T + \epsilon \quad (s3)$$

In the optimization of equation (s3), we will get the result of the deconvolution based on the communications state W_0 . And in this step, we will get our final cellular composition and distribution X_2 .

1.2. Solving for spatial communications

Here, we provided details on how we solve the spatial communications and interactions. After the initial solving of the optimization model, we will get the optimized solution for the communication matrix and the cell distribution matrix. However, in the last step, we did not continue to use optimization to get the final spatial communication matrix W_1 , but performed an inverse solution to seek the optimal W_1 . Here, this decision is from a limitation during optimization that NODE will only consider local communication at each location when performing

optimization. For avoiding this local-only optimization, we used an inverse matrix solution that enables the communication matrix to operate on a global scale. After obtaining the inverse matrix solution, we imposed local constraints. Although these constraints still limited communication to the local area, we considered the global impact during the inverse matrix solution process to derive the final local communication. Next, we assume that we had obtained the matrix Y and B as described in the main text. After obtaining the cellular composition and distribution matrix X_2 , we consider the global communications roles so we bring the X_2 back to solve for the final communications. From equation (s3) the specific details are as follows:

$$W_1 Y = Y - X_2 B^T \quad (s4)$$

Here W_1 denotes the communication matrix, it is only a formal solution and we still compute for final communication matrix.

$$Y Y^{-1} Y = E \quad (s5)$$

$$W_1 = (Y - X B^T) * Y^{-1} \quad (s6)$$

After the equation (s4), we solve for Y based on the generalized inverse matrix in equation (s5)^{22,23}, and subsequently solve for W_1 in equation (s6). In an organization, the communication strength should be distributed over the full space. But here we only consider the communication in the neighborhood. We do not consider the communication between spot pairs that are far away from each other. Therefore, after solving, we put a restriction on W_1 that each spot is only affected around its four closest neighbors, that is, the communication exists only between the spot and the four spots closest to it.

In summary, we integrated the models of NODE and finally summarize the main models of NODE as follows:

$$Y = X_1 B^T + \epsilon_1 \quad (s7)$$

$$Y = W_1 Y + X_2 B^T + \epsilon_2 \quad (s8)$$

The final model is the optimization formula in equation (s7) and (s8) for optimizing the W_1 and X_2 , by minimizing ϵ_1 and ϵ_2 .

1.3. Optimize model details:

NODE contains a total of two optimization algorithms, the Sequential Least Squares Programming (SLSQP) algorithm²⁴ and the Differential Evolution (DE) algorithm²⁵, and provides a parameter, named 'method_optimize', to the user. The parameter allows the user to choose the optimization method according to their own situation. In the optimization problem, NODE will call SLSQP or DE to deconvolute the tissue and infer the cellular communication according to the parameter (i.e., 'method_optimize') set by the user. During the optimization of NODE, NODE will set cellular composition and distribution matrix and spatial communication matrix as decision variables, respectively, to complete the deconvolution and spatial communication inference. NODE uses two convergence criteria: reaching the maximum number of iterations or achieving the optimization goal. Due to the characteristics of spatial transcriptomic data, NODE is less affected by local optimal solutions when optimizing using the SLSQP algorithm. In addition, NODE provides a

heuristic algorithm like DE algorithm. The heuristic algorithm can well help NODE to get rid of the local optimal solution, thus being able to perform supplementary analysis when the SLSQP algorithm is not satisfied. In the specific implementation of NODE, the SLSQP algorithm is implemented by the minimize python function in the scipy library (<https://scipy.org/>), while the DE algorithm is implemented through the DE python function in scikit-opt library (<https://github.com/guofei9987/scikit-opt>).

1.4. Model details:

NODE provides a parameter, named 'method_optimize', and this parameter allows the user to freely choose the specific optimization model. There are two values for 'method_optimize', 1 and 2, when the value is 1 NODE selects the Sequential Least Squares Programming (SLSQP) algorithm²⁴ for optimization and when the value is 2 NODE selects the Differential Evolution (DE) algorithm²⁵ for solving the problem. To better illustrate the specific optimization algorithm for NODE. Let us reformulate the optimization problem with the notation. In our previous account, the core optimization problem for NODE is established by the following equation:

$$Y = W_1Y + X_2B^T + \epsilon_2 \quad (s9)$$

To illustrate the principle, we omit the subscripts here and notate the equation as:

$$Y = WY + XB^T + \epsilon \quad (s10)$$

The Y denotes the expression matrix of the spatial transcriptomics data with a form spatial location (spot) multiplying the gene expression. Each element in Y represents a gene's expression value in a spot. The X denotes the cellular composition and distribution matrix, where each row represents a spot and each column represents a cell type. Each value in X represents the proportion of a cell type in a spot. The B matrix is the cell-type-specific reference basis matrix from scRNA-Seq data. The row of the B matrix represents gene, and B matrix's column represents cell type. The W denotes the spatial communication matrix between spatial locations (spots). The value on its element represents the communication strength. The ϵ is the residual between the estimated value and the true value. In the optimal solution of this model, W and X are our decision variables that need be solved in iterations that minimize the residuals ϵ . Therefore, we construct the optimization problem shown below:

$$\begin{aligned} \min_{x,w} \quad & F(X, W) = \|Y - (WY + XB^T)\|_2 \\ \text{subject to} \quad & X \geq 0, \\ & W + 1 \geq 0, \\ & -W + 1 \geq 0 \end{aligned} \quad (s11)$$

Here, since X characterizes the distribution of cells and is therefore a non-negative matrix, and W takes values between -1 and 1 as the communication strength in our hypothesis will not be stronger than its own expression.

To facilitate formal derivation, we combine the decision variables X and W into a single decision variable X^* and denote the objective function as $F(X^*)$. It should be noted that combining the decision variables X and W into X^* is solely for the purpose of facilitating formal derivation. Subsequently, we transform equation (s11) to the standard form by denoting the conditions of X and W as $G(X^*)$, and by introducing a negative factor in the conditions of the inequality (multiplying both

sides of the equation by minus one at the same time). After the transformation of the above equation, we obtain a standard optimization problem as follows:

$$\begin{aligned} & \min_{X^*} F(X^*) \\ & \text{subject to} \quad G(X^*) \leq 0 \end{aligned} \quad (s12)$$

In equation (s12), X^* denotes the merged decision variable, $G(X^*)$ denotes all the constraints on the decision variable, and $F(X^*)$ denotes the objective function on the merged decision variable, which is the same as the objective function in equation (s11), only in a different form. At this point, NODE established a least squares optimization problem, and then NODE optimally solved the original problem using the SLSQP and DE algorithms.

1.5. Sequential Least Squares Programming (SLSQP):

SLSQP is an algorithm for solving constrained optimization problems, especially for problems with nonlinear objective functions and constraints. It solves optimization problems by iteratively transforming them into linear or quadratic programming problems step by step. SLSQP in NODE consists of the following five main steps.

Step1: Initialize the initial solution X_0^* .

Step2: Establish the quadratic approximation subproblem of the Taylor expansion. Simultaneously start an iterative search. For the k th solution of this iteration X_k^* . NODE will turn it into a subproblem with approximation constraints.

$$\begin{aligned} & \min_{X^*} F(X_k^* + \Delta x^*) \approx F(X_k^*) + \nabla F(X_k^*)^T \Delta x^* + \frac{1}{2} \Delta x^{*T} H_k \Delta x^* \\ & \text{subject to} \quad G(X_k^* + \Delta x^*) \approx G(X_k^*) + \nabla G(X_k^*)^T \Delta x^* \leq 0 \end{aligned} \quad (s13)$$

Here X_k^* represents the solution at the k th iteration. $\nabla F(X_k^*)$ is the gradient of the objective function at X_k^* . $\nabla G(X_k^*)$ is the gradient of the constraint function at X_k^* . H_k is the Hessian matrix or approximate Hessian matrix at X_k^* , and Δx^* is the next iteration step and the decision variable for this subproblem. At this point the original problem is transformed into a local quadratic programming problem. Also, the local quadratic programming problem is a convex optimization problem, so the Karush-Kuhn-Tucker (KKT) conditions is a sufficient condition for this optimization problem, so according to the KKT condition, NODE can find the new optimal change Δx^* and get a solution for the next step X_{k+1}^* .

Step3: Construct the Lagrangian function, solve the quadratic programming subproblem, and establish the KKT conditions based on the subproblem.

$$L(\Delta x^*, \lambda) = F(X_k^*) + \nabla F(X_k^*)^T \Delta x^* + \frac{1}{2} \Delta x^{*T} H_k \Delta x^* + \lambda G(X_k^* + \Delta x^*) \quad (s14)$$

Karush-Kuhn-Tucker (KKT) conditions:

$$\begin{aligned} L(\Delta x^*, \lambda) &= F(X_k^*) + \nabla F(X_k^*)^T \Delta x^* + \frac{1}{2} \Delta x^{*T} H_k \Delta x^* + \lambda G(X_k^* + \Delta x^*) \\ G(X_k^*) + \nabla G(X_k^*)^T \Delta x^* &\leq 0 \end{aligned} \quad (s15)$$

$$\frac{\partial L(\Delta x^*, \lambda)}{\partial \Delta x^*} = 0, \quad \frac{\partial L(\Delta x^*, \lambda)}{\partial \lambda} = 0 \quad (s16)$$

$$\lambda G(X_k^* + \Delta x^*) = 0 \quad (s17)$$

$$\lambda \geq 0 \quad (s18)$$

Equation (s14) is the Lagrange function for the quadratic subproblem, where λ is the Lagrange multiplier, equation (s15) is a first-order approximation of the constraints in the KKT condition, equation (s16) represents the stationarity condition within the KKT conditions, equation (s17) represents the complementary slackness condition, and equation (s18) represents the dual feasibility condition. Through the KKT conditions, we can solve for the updated value of the current iteration point $X_{k+1}^* = X_k^* + \Delta x^*$ and the new Lagrange multiplier λ . Then, we proceed to check whether the termination criteria are met.

Step4: Check whether the maximum number of iterations has been reached or if the change in function value $|F(X_k^*) - F(X_{k+1}^*)|_2$ is less than a specified threshold (Default is 10^{-9}). Here, the Euclidean norm is used. Or the variable change value Δx^* is less than a specified threshold (Default is 10^{-9}). If either condition is met, terminate the iteration; otherwise, return to Step 2 and continue the iterative process.

Since the cell distribution matrix to be solved in the spatial transcriptomic data often does not have a high cellular content, the problem of local optimal solution of SLPQP is not significant, and therefore we set the initial solution at 0, which is a reasonable treatment. And there are a few important points is that, in the optimization calculation of NODE if faced with huge high-dimensional data, the Hessian matrix calculation is very huge, the user here can do the calculation of Hessian matrix approximation, at the same time in practice the Lagrange multiplier can also be achieved through specific iteration, such changes will not affect the performance of the NODE. In the specific implementation of NODE, we directly used the scipy library, and the specific details of the algorithm can be found in the scipy document and official website (<https://scipy.org/>).

1.6. Differential Evolution (DE):

Although the local optimal solution problem of NODE in SLSQP is almost negligible, we still take this into account, so we add a heuristic optimization algorithm to NODE: Differential Evolution. Users can call it by setting the parameter 'method_optimize' to 2. The core idea of differential evolutionary algorithms is to generate new candidate solutions by “differencing” between individuals within a population. This method does not depend on the gradient information of the problem, and therefore can be used for discontinuous, non-differentiable and complex objective function optimization. In this problem, the utilization of heuristic optimization algorithms can well eliminate the problem of local optimal solutions. We restate our optimization problem again:

$$\begin{aligned} \min_{X^*} \quad & F(X^*) \\ \text{subject to} \quad & G(X^*) \leq 0 \end{aligned} \quad (\text{s12})$$

Step1: The algorithm randomly generates all individuals within the population, each representing a possible solution. Assuming a population size of N (NODE defaults to 50), the DE generates N possible solutions X^* . This N individuals we call the population. Since these individuals are first-generation individuals, we denote them as X_{1i}^* . Here i denotes the i th individual in the population and 1 denotes the first generation of the population.

Step2: Differential variation. For each individual X_{ki}^* in the population (The i th

individual of the k th generation of the population), the DE algorithm generates a “variation vector” (i.e., a candidate for a new solution) from the differences between the other individuals in the population. Here i denotes the i th individual in the population.

$$V_{(k+1)i}^* = X_{kj}^* + F_v * (X_{km}^* - X_{kn}^*) \quad (j, m, n \neq i) \quad (s19)$$

Here F_v is the coefficient of variation. It tends to have a value between [0,2] (NODE defaults to 0.5). X_{kj}^* , X_{km}^* , and X_{kn}^* are three individuals of the k th generation.

X_{kj}^* , X_{km}^* , and X_{kn}^* must satisfy $X_{kj}^* \neq X_{ki}^*$, $X_{km}^* \neq X_{ki}^*$, and $X_{kn}^* \neq X_{ki}^*$. Specifically, here the i th individual of the k th generation of the population is mutated from three different individuals from previous generation.

Step3: Intersections (combinations). In order to increase the diversity of the population, each variation vector $V_{(k+1)i}^*$ is crossed with the current individual X_{ki}^* to produce a test vector $U_{(k+1)i}^*$:

$$U_{(k+1)i}^* = \begin{cases} V_{(k+1)i}^* & \text{if } \text{rand}(0,1) \leq C_r \\ X_{ki}^* & \text{otherwise} \end{cases} \quad (s20)$$

In equation (s20) C_r is the crossover probability (NODE defaults to 0.3), which indicates how much information the child inherits from the parent during the crossover process. $V_{(k+1)i}^*$ is mutated individual and X_{ki}^* is individual from previous generation. $U_{(k+1)i}^*$ is the i th potential individual of next generation.

Step4: Natural selection. Comparisons between offspring and parents are performed to optimize the whole population. At this step, NODE will actually select a new generation of population.

$$X_{(k+1)i}^* = \begin{cases} U_{(k+1)i}^* & \text{if } F(U_{(k+1)i}^*) < F(X_{ki}^*) \\ X_{ki}^* & \text{otherwise} \end{cases} \quad (s21)$$

Step5: Iterative cycle. Repeat steps 2-4 until the preset termination conditions are met (maximum number of iterations reached or the fitness function reaches a threshold). In this process, NODE records the best individuals in the population and returns after terminating the iteration.

In the DE implementation of NODE, in each generation of mutation, to prevent the population from not iterating, NODE forces one individual to mutate in order to prevent the entire population from not evolving.

1.7. Convergence criteria

Finally, we will present the convergence criteria for two optimization methods. In SLSQP, there are a total of three criteria for terminating the iteration, 1. the

maximum number of convergence is reached (NODE defaults to 150) 2. the change in the function value is less than the threshold (NODE defaults to 10^{-9}), 3. the change in the parameters is less than the threshold (NODE defaults to 10^{-9}). In DE, since it is a heuristic algorithm, NODE forces it to execute to the maximum number of iterations (NODE defaults to 700). In the specific implementation of NODE, we rely on the scikit-opt library, and the specific details of the algorithm can refer to its document and official website (<https://scikit-opt.github.io/scikit-opt/#/en/README> and <https://github.com/guofei9987/scikit-opt>).

1.8. Reference basis matrix and its relationship to spatial transcriptomics

Here, we provided details on how we construct the reference basis matrix from the scRNA-seq data, and how the reference basis matrix is related to the spatial transcriptomics data. First, we described how we constructed the reference basis matrix. After we get the single-cell data, we compared the single-cell data with the spatial transcriptomics data to find their common genes. Then, we deleted the non-common genes in the single-cell data and the spatial transcriptomics data after finding the common genes. Subsequently, we counted the number of different cell types in the single-cell data based on the clustered cell labeling information. After that, we summed up the gene expressions of the cells belonging to the same type and divided them by the total number of such cell types, which in turn led to the average expression of all genes in each cell type. Then, these averaged expressions, placed together in columns form the reference basis matrix. The rows of the reference basis matrix represent genes and the columns represent cell types.

Next, we provided details on how the reference basis matrix is related to the spatial transcriptomics data. We assumed that there are g genes in the reference matrix with k cell types. We denote B as the $g \times k$ reference matrix. The reference basis matrix is solved to represent the expression levels of different types of individual cells. Thus, by the number of these cells, and the communications with the different spots, we can get the expression of the spatial transcriptomics, further we denoted the communication between spots by $W_1 Y$, which leads to the decomposition equation, i.e.

$$Y = W_1 Y + X_2 B^T + \epsilon_2 \quad (s22)$$

In the above equation, $W_1 Y$ represents the communications network formed by spot pairs communicating with each other, and $X_2 B^T$ represents the gene expression profile formed by multiplying the number of cells by the expression profile of the cells. We denoted Y as the $n \times g$ gene expression matrix of the same set of informative genes measured at n spatial locations. The n is the number of spots. The rows of the Y matrix represent spots and the columns represent genes. We denoted X_2 as the $n \times k$ cellular composition and distribution matrix, where each row of X_2 represents the proportions of the k cell types on each spatial location. We denoted W_1 as the $n \times n$ communication matrix. In this decomposition, our goal is to use X_2 as the decision variable and minimize ϵ_2 . This in turn yields the cellular composition of each spot.

1.9. Results solving and constraints

Here, we provided details on how we constructed the constraints and solve the results. We restricted a total of two sets of conditions. First, the X_1 and X_2 to be optimized represents the meaning of the number of cells and contain a non-negative limit, while considering the pore size of the sequencing, we added an upper limit on the number of cells, i.e., we added an upper limit of 8 (which can be modified by the user according to their data) for each cell type at the time of optimization. Second, we also added constraints to the communication matrix. Specifically, the effect of a spot on itself we regard as maximal, i.e., the effect of a spot on the expression of other spots does not exceed its effect on itself, so we restrict the communications coefficients (W_0) to be between -1 and 1 in optimization search (which can be modified by the user according to their data).

In the solving, we used Python programming language for the code, which provides two solver libraries. They are based on scipy and scikit-opt, respectively. Users are free to choose the solution library according to their needs. The package scipy is faster, but sometimes it cannot strictly satisfy the iteration conditions, so it needs to impose an external condition to filter the selection. The package scikit-opt is strict, and user can freely set the optimization parameters, but it takes a longer time, and sometimes it may be trapped in local optimal solutions. Overall, the user is free to choose the solution library and set the parameters of the library according to their own needs, which makes NODE very flexible to use.

1.10. Calculation of AUC values and setting of thresholds

(1) Setting of gene and cell thresholds and calculation of AUC values:

Here is the detail how AUC values are calculated in the comparison of cells and their corresponding marker genes. Specifically, we determined thresholds based on the R software that maps gene heat maps of spatial transcriptomics data, and the purpose of determining thresholds was to find the presence of spots that clustered high gene expression. Subsequently, we compared the gene expression value of each spot with the threshold value, if the expression value is less than the threshold value the spot value was set to 0 and vice versa to 1, based on this we constructed the reference data. Subsequently, according to this method, we also did a 0 or 1 selection of cell types corresponding to the deconvolution results, and subsequently, the AUC values were calculated based on these two sets of 0-1 vectors.

(2) Setting of thresholds and calculation of AUC values in validation of cellular communications in simulated data:

Here, we detail how AUC values are calculated when comparing simulated data cell communications with inferred communications from NODE. Specifically, we selected the maximum three signal strengths from each spot in the simulated data and set them to 1, and other signals are considered as 0. At the same time, we did the same for NODE, and subsequently we placed columns in relative positions based on the two 0-1 matrices obtained, and subsequently carried out the computation of the AUC value.

Supplementary Note 2. Description on compared deconvolution methods

We compared NODE with five recently developed deconvolution methods such as SpaTalk⁷, RCTD³, Seurat⁴, deconvSeq⁵, and SPOTlight⁶. Specifically, SpaTalk relies on a graph network and knowledge graph to model and score the ligand-receptor-target signaling network between spatially proximal cells by dissecting cell-type composition. For spot-based ST data, the cell types with different weights were used as the reference to project the cells from scRNA-seq data onto the spatial spot. Through random sampling and deep iteration processes, the optimal cellular combination that most resembled the spatial spot was refined to reconstruct the single-cell ST data for spot-based ST data. RCTD directly models expression count data from spatial transcriptomics based on a Poisson factor analysis model, which extends the linear factor analysis models commonly used for bulk RNA-seq deconvolution. In the factor analysis model, RCTD introduces additional variance parameters to account for the different platform effects between scRNA-seq and spatial transcriptomics. Seurat deconvolution using finding anchors between reference single-cell data and spatial transcriptomics data. DeconvSeq utilizes a generalized linear model to model effects of tissue type on feature quantification, which is specific to the data structure of the sequencing type used. Estimated model coefficients can then be used to predict the cell type mixture within a tissue. SPOTlight takes low-dimensional components from both scRNA-seq and spatial transcriptomics as input. The low-dimensional components of scRNA-seq are referred to as cell-type specific topic profiles in SPOTlight and are obtained through non-negative matrix factorization and aggregated across cells within each cell type. With the input of the low-dimensional components, SPOTlight relies on a nonnegative least squares estimation procedure commonly used in bulk RNA-seq deconvolution for spatial transcriptomics deconvolution.

After that we added the method of comparing: CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², SpatialPrompt¹³, BASS¹⁴ and SPADE¹⁵. CARD combines cell-type-specific expression information from single-cell RNA sequencing (scRNA-seq) with correlation in cell-type composition across tissue locations. GraphST, a graph self-supervised contrastive learning method. For the cell-type deconvolution task, GraphST will train an auto-encoder to learn informative cellular features from the scRNA-seq data in an unsupervised way. SpaDecon is a semi-supervised learning approach that incorporates gene expression, spatial location, and histology information for cell-type deconvolution. SpatialDWLS is an extension of dampened weighted least squares (DWLS). SpatialDWLS performs deconvolution through this modeling. In addition, by incorporating the spatial information from the real spatial data, SpatialPrompt uses an iterative approach inspired from the message passing layer in the graph neural net to compute weighted mean expression of the local microenvironment for each spatial spot. BASS, on the other hand, utilizes bayesian models for a posteriori deconvolution and domain identification. SPADE emphasize preserving spot-to-spot relationships in its deconvolution.

Supplementary Note 3. Simulation design.

3.1. Model-free simulation design.

We used the scikit-opt module in NODE for deconvolution in all simulated data. We performed extensive simulations to compare NODE with other existing deconvolution methods, and used single-cell data from SCC data² for the construction of simulated data. In the original data, there were 32738 genes and 2736 cells. In the construction of the simulated data, we performed preliminary processing of the single-cell data. First, we deleted cells that expressed the gene number less than 5% of all genes, and subsequently deleted genes that expressed in cells less than 5% of all cells. After deleting these cells and genes, we constructed the reference basis matrix based on the remaining cells and genes. After that, we also deleted the cells without distributional features and individual labeling according to the literature. Finally, we got the filtered scRNA-seq data. The filtered scRNA-seq data contains 1972 cells, 7515 genes with 11 cell types. These cell types include DC (n = 872), Endo (n = 73), Tumor_KC_Cyc (n = 140), FIB (n = 22), Macrophages (n = 255), Tumor_KC_Basal (n = 149), T_NK (n = 85), Tumor_KC_Diff (n = 319), TSK (n = 32), B-cell (n = 5), and MELA (n = 20)². Next, we constructed spatial transcriptomics data. The spatial coordinates of 649 spots were provided in the spatial coordinate's information. First, we randomly perturbed the number of each cell type, so that the number of each cell type was randomly generated, and the number of each cell type after randomly increasing or decreasing is regarded as the number of each type of cell contained in the spatial transcriptomics data. After obtaining the total number of cells, we appropriately reduced the dimension of the spatial transcriptomics data according to the total number of cells, so that the average number of cells in each spot will not be too large or too small. After setting the dimension of the spatial transcriptomics data, we randomly placed each cell type. Specifically for each cell type, one cell at a time was taken and placed randomly into a spot until all cells were placed. In this way, we obtained the cellular distribution of the spatial transcriptomics data. Subsequently, we performed the construction of spatial transcriptomics expression data based on the cell type information of single cells. Specifically, for each spot in the cellular distribution of the spatial transcriptomics data, we performed a randomized construction based on the cell type. For example, in a certain spot, it contains 3 A cells and 3 B cells. Then when we constructed its gene expression data, took three cells of cell type A and three cells of cell type B randomly in the single cell data, summed up all the gene expressions of cells, and regarded their gene co-expression as the gene expression in that spot. Following the above process, we generated the expression profiles of all the spots and completed the construction of spatial transcriptomics data.

3.2. Simulation analysis scenarios

In order to better simulate the real situation, we added different spatial and cellular modes to construct different spatial transcriptomics data.

(1) Scenarios 1: Spatial transcriptomics data were generated as model-free simulation design, without additional modes addition. Scenario 1 is a simulation of the base case.

(2) Scenarios 2: The different spatial patterns were added. Specifically, after generating the simulated data according to the baseline simulation process, we added different spot-to-spot communications to the spatial transcriptomics data, added the communications among spots on each other, to generate new spatial transcriptomics data, and applied the new data for the deconvolution.

(3) Scenarios 3: We censored the single-cell data to delete the cells with low-quality expression and maintain cells with higher-quality expression, then used the new single-cell data to construct the simulated data. Other construction methods, same as scenario 1. The single cell data used in this scenario contains 11 cell types, 7515 genes, 1005 cells. These cell types include DC (n = 438), Endo (n = 38), Tumor_KC_Cyc (n = 72), FIB (n = 13), Macrophages (n = 129), Tumor_KC_Basal (n = 76), T_NK (n = 44), Tumor_KC_Diff (n = 161), TSK (n = 18), B-cell (n = 4), and MELA (n = 12).

(4) Scenarios 4: Both different cellular patterns and spatial patterns were added. Specifically, we added spatial communications, i.e., different cellular communications, on the spatial transcriptomics data in scenario 3 to generate new spatial transcriptomics expression data for the deconvolution. In this scenario, the single-cell data used are the same as in scenario 3.

3.3. Repeated experiments and analysis of results.

After we constructed the simulated data for the four scenarios, we repeated the experiment to avoid experimental bias and obtained a total of eight sets of simulated data. First, we plotted pie charts for eight datasets based on repetition (Supplementary Fig. 1-2). After constructing the spatial transcriptomics data, we used the single-cell data that generates this spatial transcriptomics data as its reference data to perform the deconvolution of various methods. After applying all the deconvolution methods to the simulated data, we transformed all the deconvolution results into the inferred number of distinct cells for the spatial transcriptomics data. Subsequently, we analyzed and compared the results with the true distribution of cells in the simulated data and calculated the root mean square error (RMSE). The performance of different deconvolution methods is reflected by comparing the median RMSE calculated by each method (Supplementary Fig. 3-10). We then added the percentage of spots with the smallest RMSE values calculated by each method to calculate their first scores (Supplementary Fig. 11). Then we ranked all the RMSEs and calculated the frequency at which each method performed best. Finally, we plotted the pie charts based on the number of cells of each type and compared them to the real situation (Supplementary Fig. 12-19).

Supplementary Note 4. Details on preprocessing spatial transcriptomics and single cell.

4.1. Mouse olfactory spatial transcriptomics data.

We downloaded the mouse olfactory bulb spatial transcriptomics data from the spatial transcriptomics research website (<https://www.spatialresearch.org/>). This data consists of gene expression measurements in the form of read counts that are collected on a number of spatial locations known as spots. We focused our analysis on MOB section #12 which contains 16,034 genes and 282 spatial locations. For deconvolution, we obtained the scRNA-seq data from Gene Expression Omnibus (GEO) database, (<https://www.ncbi.nlm.nih.gov/geo/>) with accession number GSE121891 to serve as the reference. This scRNA-seq data was collected from the mouse olfactory bulb and contains 18,560 genes and 21746 cells. The cells have already been clustered into the following main cell types: n03-GC-1 (n = 1914), n02-PGC-1 (n=1037), n07-GC-2 (n=3608), n17-M/TC-3 (n=927), n06-Transition (n = 2085), n01-OSNs (n = 1200), n13-AstrocyteLike (n = 2950), n09-GC-3 (n = 273), n14-GC-7 (n = 679), n10-GC-4 (n = 234), n05-PGC-2 (n = 377), n18-EPL-IN (n = 161), n04-Immature (n = 3910), n16-M/TC-2 (n = 116), n08-PGC-3 (n = 279), n11-GC-5 (n = 1669), n12-GC-6 (n = 237), and n15-M/TC-1 (n = 90). Subsequently, we merged these cell types to obtain the main ones: GC, PGC, M-TC, Transition, OSNs, AstrocyteLike, EPL-IN, Immature. We used gene-wide information and intersected only the genes from the spatial transcriptomics data with the single-cell data, yielding 14,454 public genes. Then, we deconvolved the spatial transcriptomics data with single-cell data.

4.2. Human pancreatic ductal adenocarcinoma data.

We downloaded the human pancreatic ductal adenocarcinoma (PDAC) data from GEO database with accession number GSE111672. This dataset consists of both spatial transcriptomics data and scRNA-seq data collected on the same tissue through inDrop. We focused our analysis on PDAC for the patient ID GSM3036911 (PDAC-A). The spatial transcriptomics data of PDAC contains 25753 genes and 428 spatial locations. The scRNA-seq data are available in the same study, PDAC-A-inDrop, which is a matched data collected from the same patient GSM3036911. PDAC-A-inDrop consists of 19,736 genes and 1,926 cells. The cells have already been clustered into the following main 20 cell types: Acinar cells (n = 13), Ductal terminal (n = 350), Ductal centroacinar (n = 529), Cancer clone A (n = 126), Ductal MHC Class II (n = 287), Cancer clone B (n = 170), mDCs A (n = 12), Ductal high hypoxic (n = 215), Tuft cells (n = 32), mDCs B (n = 33), pDCs (n = 13), Endocrine cells (n = 3), Endothelial cells (n = 11), Macrophages A (n = 21), Mast cells (n = 14), Macrophages B (n = 19), T cells&NK cells (n = 40), Monocytes (n = 18), RBCs (n = 15), and Fibroblasts (n = 5). We used gene-wide information and intersected only the genes between the spatial transcriptomics data and the single-cell data, yielding 8781 public genes. Then, we deconvolved the spatial transcriptomics data with single-cell data.

When using data from non-matching scRNA-seq, we acquired a total of two additional scRNA-seq data: the PDAC-B-inDrop as well as an external dataset²¹.

PDAC-B-inDrop consists of 19,736 genes and 1,733 cells. The cells have already been clustered into the following main 13 cell types: Acinar cells (n = 6), Ductal terminal (n = 736), Ductal centroacinar (n = 152), Cancer clone A (n = 339), Ductal MHC Class II (n = 211), mDCs (n = 35), Tuft cells (n = 37), Endothelial cells (n = 159), Macrophages (n = 9), Mast cells (n = 13), Monocytes (n = 20), Endocrine cells (n = 13), and RBCs (n = 3).

This external dataset was collected from 11 control pancreas and 24 PDAC tumors and was sequenced using the 10x Chromium platform. The data contains 24,005 genes and 57,530 cells from 10 cell types. We split the data into three subsets: the PengNormal data that consists of the 11 control samples; the PengTumor data that consists of the 24 PDAC tumors; and the Peng data that consists of all 35 samples. PengNormal have already been clustered into the following main 9 cell types: acinar cells (n = 1423), B cells (n = 31), Ductal type 1 cells (n = 7671), endocrine cells (n = 270), endothelial cells (n = 3983), fibroblast cells (n = 940), macrophage cells (n = 559), stellate cells (n = 623), and T cells (n = 44). PengTumor have already been clustered into the following main 10 cell types: acinar cells (n = 512), B cells (n = 2416), Ductal type 1 cells (n = 2446), Ductal type 2 cells (n = 11315), endocrine cells (n = 459), endothelial cells (n = 5134), fibroblast cells (n = 5802), macrophage cells (n = 4802), stellate cells (n = 5284), and T cells (n = 3616). Peng data have already been clustered into the following main 10 cell types: acinar cells (n = 1935), B cells (n = 2447), Ductal type 1 cells (n = 10317), Ductal type 2 cells (n = 11315), endocrine cells (n = 729), endothelial cells (n = 9117), fibroblast cells (n = 6742), macrophage cells (n = 5361), stellate cells (n = 5907), T cells (n = 3660). For these supplemental scRNA-seq data, we did the same operation as PDAC-A-indrop to intersect these scRNA-seq data with spatial transcriptomic data.

4.3. Human skin squamous cell carcinoma ST dataset.

We downloaded the human skin squamous cell carcinoma (SCC) ST dataset from GEO database with accession number GSE144240. This dataset consists of both spatial transcriptomics data and scRNA-seq data collected on the same tissue using the 10x Chromium technology. We focused our analysis on SCC_P2 which had three slices for the patient ID GSM4284316 (denoted as rep1), GSM4284317 (denoted as rep2), and GSM4284318 (denoted as rep3). The spatial transcriptomics data of GSM4284316 contains 17138 genes and 666 spatial locations. The spatial transcriptomics data of GSM4284317 contains 17344 genes and 649 spatial locations. The spatial transcriptomics data of GSM4284318 contains 17883 genes and 638 spatial locations. The scRNA data matched to the spatial transcriptomics data contained 32738 genes and 2736 cells. The cells have already been clustered into the following main 21 cell types: Tumor_KC_Diff (n = 319), Keratinocyte (n = 250), Tumor_KC_Basal (n = 149), Tumor_KC_Cyc (n = 140), Endothelial-Cell (n = 73), Pilosebaceous (n = 342), Fibroblast (n = 22), Melanocyte (n = 20), TSK (n = 32), Multiplet (n = 48), Eccrine (n = 5), LC (n = 332), PDC (n = 106), CLEC9A (n = 61), CD1C (n = 410), Mac (n = 255), ASDC (n = 69), T-cell (n = 82), MDSC (n = 13),

B-Cell (n = 5), and NK (n = 3). We used gene-wide information and intersected only the genes both the spatial transcriptomics data and the single-cell data. We solved the spatial transcriptomics data for each of the three slices with the single-cell data to obtain the common set of genes, leaving 16400 genes (rep1), 16613 genes (rep2), and 17045 genes (rep3), respectively.

4.4. Developing human heart data

We downloaded the spatial transcriptomics data and scRNA-seq data of developing human heart data from <https://www.spatialresearch.org>. This data contains information on gene expression in the developing human heart providing sliced data for three time periods which are 4-4.5, 6.5, and 9 post-conception weeks (PCW), respectively. There are nineteen data slices spanning three periods: the 4-4.5 PCW period comprises four slices, the 6.5 PCW period comprises nine slices, and the 9 PCW period comprises six slices. Each of the nineteen slices in the original document is labeled with an ordinal number from 1 to 19. The spatial transcriptomics data of all slices contains 39739 genes and 3111 spatial locations. The scRNA-seq data matched to the spatial transcriptomics data contained 15323 genes and 3777 cells. The cells have already been clustered into the following main 15 cell types and the erythrocytes were divided into two classes, which are distinguished in this paper by the original clustering labels: Capillary endothelium (n = 673), Ventricular cardiomyocytes (n = 508), Fibroblast-like (related to cardiac skeleton connective tissue) (n = 466), Epicardium-derived cells (n = 395), Fibroblast-like (related to smaller vascular development) (n = 344), Smooth muscle cells / fibroblast-like) (n = 265), Erythrocytes-6 (n = 190), Atrial cardiomyocytes (n = 157), Fibroblast-like (related to larger vascular development) (n = 151), Epicardial cells (n = 137), Endothelium / pericytes / adventitia (n = 131), Myoz2-enriched cardiomyocytes (n = 97), Immune cells (n = 75), Cardiac neural crest cells & Schwann progenitor cells (n = 75)²⁶. We used gene-wide information and intersected only the genes both the spatial transcriptomics data and the single-cell data, yielding 13841 public genes. Then, we deconvolved the spatial transcriptomics data with single-cell data.

Supplementary Note 5. Co-localization in deconvolutional results

5.1. MOB colocalization mode:

NODE infers those mitral/tufted cells (M-TC) and periglomerular cells (PGC) are spatially colocalized, offering spatial localization evidence for their reciprocal synapses with mitral/tufted cells dendritic tufts in the glomerulus.²⁷.

5.2. PDAC colocalization mode:

NODE also found the co-location condition in PDAC, such as ductal high-hypoxic cells and cancer cells, endothelial cells and fibroblast cells. The role of co-localization of high-hypoxic cells and cancer cells, on the other hand, is to support the formation of the hypoxic and nutrient-poor tumor microenvironment and the role of colocalization of endothelial cells and fibroblast cells is in pancreatic cancer stroma interaction of the tumor microenvironment^{20,28}.

5.3. SCC colocalization mode:

Infiltration of immune cells with tumor cells is a great indication of co-localization. In the extrapolated results of NODE, Tumor_KC_Cyc, Tumor_KC_Diff, and TSK showed a pattern of co-localization with B-cell, and it is consistent with immune infiltration².

Supplementary Note 6. Comparison of NODE and recent development deconvolution methods.

Here, we evaluated the performance of NODE with five recent deconvolution methods (CARD⁹, GraphST¹⁰, SpaDecon¹¹, SpatialDWLS¹², and SpatialPrompt¹³) in mouse olfactory bulb (MOB) data⁸, human pancreatic ductal adenocarcinoma (PDAC) data¹⁹, and human skin squamous cell carcinoma (SCC) data². Here, we have chosen these methods because they are recently developed deconvolution methods with higher accuracy. Thus, these methods represent the current development of the deconvolution task well. So, we chose these methods to compare with NODE. In addition, we also have chosen an extra method BASS¹⁴. The BASS¹⁴ method was not designed for the deconvolution of spatial transcriptomic data but for domain detection and spatial clustering of cell type. We compared the ability of NODE and BASS to recognize and localize tissue layer. However, considering that BASS is not designed for deconvoluting cell composition in spots, we only compared the performance both the NODE and BASS methods using MOB data.

By the MOB and PDAC data, NODE recognized the hierarchical structure of tissue (Supplementary Fig. 22 and 30). Meanwhile NODE outperformed all other methods to achieve the best deconvolution result in SCC data (Supplementary Fig. 44). Overall, NODE demonstrated a clear advantage in these recent deconvolution results.

6.1. Analysis of MOB:

We first applied all the deconvolution methods to the MOB data⁸. The MOB comprises three main anatomic layers organized in an inside-out fashion, annotated based on hematoxylin and eosin (H&E) staining images: the granule cell layer (GCL; main cell: granule cells denoted as GC), the mitral cell layer (MCL; main cell: mitral/tufted cells denoted as M-TC), and the glomerular layer (GL; main cell: periglomerular cells denoted as PGC) (Supplementary Fig. 22a). In the result, the cell type compositions inferred by NODE accurately depicted such expected layered structure (Supplementary Fig. 22b). This clear hierarchical structure was also present in the results of CARD and GraphST (Supplementary Fig. 22b). Correspondingly, SpaDecon, SpatialDWLS, and SpatialPrompt did not show such clear stratification (Supplementary Fig. 22b). While SpatialPrompt found stratification in the center of the tissue (Supplementary Fig. 22b), SpatialPrompt's performance gradually deteriorated as the tissue expands outward (Supplementary Fig. 22b). To further assess NODE's performance, the three layers of cells were individually delineated, and the inferred results from NODE and other deconvolution methods were overlaid on the tissue anatomy map to visually inspect whether the cell distribution of the inferred results matches that of the tissue (Supplementary Fig. 22c). The stratification observed in NODE's extrapolated results can precisely align with the stratification in the H&E staining images (Supplementary Fig. 22c). In contrast, CARD and GraphST were unable to accurately localize the MCL, even though they identified structures of this layer (Supplementary Fig. 22c). Additionally, the boundaries of MCL and GL were indistinct in the results inferred by both CARD and GraphST. On the other hand,

SpaDecon was unable to locate the MCL and the GCL. And in SpaDecon deconvolution result, GL was positioned in a way that did not fit the organizational structure. SpatialDWLS could not distinguish between all three major anatomical layers (Supplementary Fig. 22c). And SpatialPrompt only correctly positioned the GCL. Correspondingly, NODE not only successfully distinguished the three cell layers, but also the arrangement of the three cell layers was highly consistent with the annotation of the sections (Supplementary Fig. 22c). NODE correctly localized the major cells in the three-layer structure and clearly distinguished the boundaries of the three cell layers (Supplementary Fig. 22c). Subsequently, we performed a comparison between NODE and BASS. Importantly, while BASS can perform clustering of cell types, it does not have access to the specific composition of the spots, therefore BASS does not serve as a deconvolution method for comparison with NODE. Also, BASS does not require reference to scRNA-seq data. BASS only gets a cell type per spot. In the BASS results, BASS correctly localized GC, but when localizing M-TC and PGC, BASS did not correctly localize these two layers (Supplementary Fig. 22d). In contrast, NODE correctly localized the three cell types. Afterwards, we mapped the marker genes of the corresponding cells (GC, M-TC, and PGC) (Supplementary Fig. 22e). The results showed that the distributions of the cells obtained by NODE deconvolution showed a high degree of agreement with the distributions of their corresponding marker genes (Supplementary Fig. 22c-22e). Meanwhile, we extracted the major cell distribution vectors of the anatomical layer and the corresponding marker gene expression vectors of the spatial transcriptomics data, and calculated the Pearson correlation coefficients between them. A higher correlation coefficient means a closer distribution between the cell types and corresponding the marker genes. NODE obtained a 0.547 correlation in the GL (main cell: PGC) and is higher than the correlation from other methods (Supplementary Fig. 22f). In the GCL (main cell: GC), NODE achieved a 0.736 correlation and is the second highest performance only behind CARD 0.756 correlation (Supplementary Fig. 22f).

6.2. Deconvolution of PDAC:

We then examined the PDAC data¹⁹ for deconvolution. Following the original paper¹⁹, the tissue type of each region on the section has been labelled. There were four tissues regions, such as cancer, pancreatic, ductal and stroma regions, in PDAC (Supplementary Fig. 30a). In the H&E staining of PDAC, distinct colors in different regions of the tissue clearly indicated variations in cellular composition across these sections (Supplementary Fig. 30a)¹⁹. We then manually labeled the different regions of the organization based on the original publication (Supplementary Fig. 30a)¹⁹. After that, we calculated the deconvolution for the PDAC spatial transcriptomics data using NODE and other deconvolution methods and the results was shown in Supplementary Fig. 30b-30f. In the spatial scatter pie plot, NODE successfully identified spatial regionality similar to that observed in the case of sliced tissue (Supplementary Fig. 30a and 30b). The cell's color distribution in the NODE deconvolution result matched the color depth distribution of the sliced tissue

(Supplementary Fig. 30a and 30b). In contrast, none of the other deconvolution results, except for CARD, showed such a distribution of cells (Supplementary Fig. 30b). In the case of cell localization, only NODE and CARD showed a more satisfactory performance (Supplementary Fig. 30b). GraphST and SpatialPrompt though correctly localized the cells in the cancer and ductal epithelium (Supplementary Fig. 30b). However, they did not correctly localize other tissue cells (Supplementary Fig. 30b). On the other hand, SpaDecon was unable to distinguish between ductal epithelium and cancer areas, and there was no clear region of cell distribution in SpaDecon's deconvolution results (Supplementary Fig. 30b). In the inferred results of NODE, we could clearly see that NODE captured the stratification of different regions and revealed the fuzzy demarcation between cancer and non-cancer regions, between duct and stroma regions, and between cancer and pancreatic (Supplementary Fig. 30b). Subsequently, we labeled the aggregation states of cancer clone B, duct centroacinar cells and duct terminal cells in the deconvolution results from NODE and other deconvolution methods (Supplementary Fig. 30c). The results showed that the deconvolution results of NODE were consistent with the histologically annotated distributions (Supplementary Fig. 30c). In contrast, the methods other than CARD did not exactly find the distribution of cells in the tissue. Specifically, GraphST and SpaDecon misplaced the localization of ductal terminal cells (Supplementary Fig. 30c). The area of distribution of ductal centroacinar cells and cancer clone B in SpatialPrompt's results was much smaller than them in the original literature, and SpatialPrompt did not accurately find ductal terminal cells distribution either (Supplementary Fig. 30c). SpatialDWLS was unable to localize all three cell types (Supplementary Fig. 30c). To further characterize the performance of NODE, we also plotted the heatmaps (Supplementary Fig. 30d and 30e) of the three cells and their corresponding marker genes (cancer clone B - *SI00A4*, ductal centroacinar - *CRISP3*, ductal terminal - *TFF3*). In the results, the cell distributions in the NODE deconvolution results were consistent with their distributions relative to the marker genes and showed the same infiltration trend. Finally, in quantifying the accuracy metrics of each method, we calculated the correlation coefficients of ductal centroacinar cells and their corresponding marker genes, and the area under the curve (AUC) values of ductal terminal cells and their corresponding marker genes. Higher correlation coefficients and AUC values indicated greater accuracy of the methods. Specifically, we set the value of the gene vector higher than the upper quartile to 1 and the value lower than the upper quartile to 0. Then we calculated the AUC value of the gene vector and cell vector after the binarization. In the quantitative assessment, NODE achieved the second-highest performance in both correlation coefficient and AUC (Supplementary Fig. 30f). It ranked just below SpaDecon in correlation coefficient and just below CARD in AUC (Supplementary Fig. 30f).

6.3. Analysis of SCC ST dataset:

We finally applied NODE and other deconvolution methods to SCC data² to investigate whether all deconvolution methods can identify the focal cells. In the SCC

data, we selected tumor specific keratinocyte (TSK) as the main target for analysis. There were three repeated slices of the SCC data, and we analyzed all three repeated slices for deconvolution. Meanwhile, since the distribution of TSK in tissues was not characterized by distinct tissue images, we first localized the marker genes (*MMP10* and *LAMC2*) for TSK (Supplementary Fig. 44a). Subsequently, we applied all methods to the SCC dataset and labeled the localization of TSK in the results of each method (Supplementary Fig. 44a). Among all the results, only NODE accurately found the distribution of TSKs coincided with the distribution of their corresponding marker genes (Supplementary Fig. 44a). Notably, CARD accurately identified the distribution of TSK in rep1 and rep3, but failed to find an accurate distribution of TSK in the rep2 slice (Supplementary Fig. 44a). Although GraphST and SpatialPrompt found the aggregation locations of TSKs, the TSKs in both the deconvolution results have too many false positives (Supplementary Fig. 44a). Specifically, their distributions do not coincide with the expression of the labeled genes, and TSKs are clustered at locations with low gene expression (Supplementary Fig. 44a). Meanwhile, SpaDecon and SpatialDWLS were unable to find the accurate distribution of TSK (Supplementary Fig. 44a). Compared with all methods, NODE not only found the distribution of TSK correctly, but also had the least number of false positives in the NODE results (Supplementary Fig. 44a). Moreover, the localization of TSK by NODE was also the most suitable for the distribution of TSK marker genes (Supplementary Fig. 44a). Subsequently, we calculated the correlation coefficients and AUC values for the distribution of TSKs localized by different methods and TSK's corresponding marker genes. Here, we performed the same binarization on the marker gene vector in the AUC value calculation. The values in marker gene vector that were higher than its own upper quartile were set to 1, and vice versa to 0. Subsequently, we performed the calculation of the AUC values between TSK vector and normalization marker gene vector. A higher correlation coefficient or AUC value represented higher accuracy of the method. In the correlation coefficient calculation, NODE achieved the best performance in all replicates as well as in all genes, obtaining the highest correlation coefficient (Supplementary Fig. 44b). NODE also showed the best performance in the calculation of AUC values (only slightly inferior to CARD in the AUC of the *LAMC2* gene in the rep2 and the AUC value of the *MMP10* gene in the rep1). NODE achieved the best performance in most of the comparisons of AUC values (Supplementary Fig. 44c).

In summary, NODE demonstrated good performance in all comparative validations against the six recent methods.

6.4. Comparison of NODE's performance with recently developed method SPADE.

To further validate the accuracy of NODE, we conducted comparative analyses between NODE and SPADE¹⁵ across the mouse olfactory bulb (MOB) data, human pancreatic ductal adenocarcinoma (PDAC) data, and human skin squamous cell carcinoma (SCC) ST dataset. In MOB data, the MOB comprises three main anatomic layers organized in an inside-out fashion, annotated based on hematoxylin and eosin

(H&E) staining images: the granule cell layer (GCL; main cell: granule cells denoted as GC), the mitral cell layer (MCL; main cell: mitral/tufted cells denoted as M-TC), and the glomerular layer (GL; main cell: periglomerular cells denoted as PGC) (Supplementary Fig. 23a). In the result, the cell type compositions inferred by NODE accurately depicted such expected layered structure (Supplementary Fig. 23b). In contrast, SPADE failed to identify the stratification patterns present in the outer layers of the tissue (Supplementary Fig. 23b). Subsequently, we examined the ability of NODE and SPADE to localize the major cells of the three-layer tissue (Supplementary Fig. 23c). In GL, SPADE also identified peripheral tissues as the primary location of PGC (Supplementary Fig. 23c). Meanwhile, we extracted the major cell distribution vectors (GL, M-TC, and PGC) of the anatomical layer and the corresponding marker gene expression vectors of the spatial transcriptomics data, and calculated the pearson correlation coefficients between them. In the quantitative comparison between NODE and SPADE, NODE consistently outperformed SPADE in all correlation coefficient analyses, demonstrating higher accuracy (Supplementary Fig. 23d). Overall, NODE achieved more accurate results in the analysis of the MOB data.

Subsequently, we analyzed the PDAC dataset using NODE and SPADE. Following the original paper¹⁹, the tissue type of each region on the section has been labelled. There were four tissues regions, such as cancer, pancreatic, ductal and stroma regions, in PDAC (Supplementary Fig. 23e). In the H&E staining of PDAC, distinct colors in different regions of the tissue clearly indicated variations in cellular composition across these sections (Supplementary Fig. 23e)¹⁹. In the spatial scatter pie plot, NODE successfully identified spatial regionality similar to that observed in the case of sliced tissue (Supplementary Fig. 23e and 23f). The cell's color distribution in the NODE deconvolution result matched the color depth distribution of the sliced tissue. In contrast, SPADE only correctly identified the stratification of the ductal region in the results of the SPADE analysis (Supplementary Fig. 23f). Subsequently, we labeled the aggregation states of cancer clone B, duct centroacinar cells and duct terminal cells in the deconvolution results from NODE and SPADE. In the results, both NODE and SPADE accurately identified the localization of cells in the cancerous area (Supplementary Fig. 23g), but SPADE showed a inappropriate result in the localization of duct centroacinar cells (Supplementary Fig. 23g). SPADE's localization of duct centroacinar cells extends significantly beyond the ductal region. Finally, in quantifying the accuracy metrics of each method, we calculated the correlation coefficients of ductal centroacinar cells and their corresponding marker genes, and the area under the curve (AUC) values of ductal terminal cells and their corresponding marker genes. Higher correlation coefficients and AUC values indicated greater accuracy of the methods. Specifically, we set the value of the gene vector higher than the upper quartile to 1 and the value lower than the upper quartile to 0. Then we calculated the AUC value of the binarized gene vector and cell vector. The results showed that NODE achieved higher correlation coefficients, while SPADE obtained higher AUC values (Supplementary Fig. 23h). This indicates that NODE demonstrated greater accuracy in localizing duct

centroacinar cells (Supplementary Fig. 23h), whereas SPADE was more effective in accurately identifying ductal terminal cells (Supplementary Fig. 23h).

Finally, we finally applied NODE and SPADE to SCC data² to investigate whether they can identify the focal cells. In the SCC data, we selected tumor specific keratinocyte (TSK) as the main target for analysis. There were three repeated slices of the SCC data, and we analyzed all three repeated slices for deconvolution. Meanwhile, since the distribution of TSK in tissues was not characterized by distinct tissue images, we first localized the marker gene *LAMC2* (Supplementary Fig. 23i). Among all the results, only NODE accurately found the distribution of TSKs coincided with the distribution of their corresponding marker gene (Supplementary Fig. 23i). On the other hand, SPADE produced inconsistent results across the three replicates (Supplementary Fig. 23i). In replicate 1 (rep1), SPADE failed to localize TSK, while in the last two replicates (Supplementary Fig. 23i), the distribution of TSK identified by SPADE did not exhibit clustering patterns consistent with those of the marker gene (Supplementary Fig. 23i). Subsequently, we quantified the performance of NODE and SPADE in replicates 2 and 3 by calculating the correlation coefficients and AUC values between the distribution of TSK identified by each method and corresponding marker genes (*LAMC2* and *MMP10*) (Supplementary Fig. 23j). Here, we excluded replicate 1 (rep1) from the analysis of correlation coefficients and AUC values because SPADE failed to localize TSKs in this replicate. In all quantitative comparisons, NODE outperformed the SPADE method, achieving superior deconvolution results (Supplementary Fig. 23j).

Overall, NODE outperformed the SPADE method in the majority of cases, delivering superior results across a larger number of datasets.

Supplementary Note 7. Robustness of NODE deconvolution

For the robustness of NODE, we have done robustness tests against NODE in the following three aspects: testing the stability of NODE in using different reference scRNA-seq data, testing the ability of NODE to resist interference from spatial noise and spatial batch effects, and testing the computational performance of NODE.

7.1. Testing the stability of NODE in using different reference scRNA-seq data:

The performance of NODE remains stable even in the face of different reference scRNA-seq data. After changing the reference scRNA-seq data, NODE still recognized the cell distribution accurately, and there was no significant change in the deconvolution result of NODE. This advantage was demonstrated when we analyzed the PDAC data with different reference scRNA-seq data¹⁹. There are multiple slices of PDAC data in the original literature¹⁹, and we only used PDAC-A slice as an example in here. We did the deconvolution for the PDAC-A dataset using five different reference scRNA-seq data, categorizing into scenario 1 to scenario 5. In scenario 1, as described in the original literature, multiple slices of PDAC were analyzed¹⁹. Among these, the PDAC-B and PDAC-A slices originated from two different patients. Naturally, the scRNA-seq data matched to the PDAC-B slice is not corresponding to the PDAC-A slice¹⁹. Therefore, in scenario 1, we performed a deconvolution analysis of the PDAC-A dataset using the reference scRNA-seq data from PDAC-B dataset. Subsequently, in scenario 2, we merged the scRNA-seq datasets of PDAC-A and PDAC-B, and used the merged scRNA-seq data as the reference scRNA-seq for the deconvolution of PDAC-A. We also obtained one external scRNA-seq data (10x Chromium) from Peng et al (Genome Sequence Archive under project PRJCA001063)²¹. This data was collected from 11 control pancreas and 24 PDAC tumors, and it contains 24,005 genes and 57,530 cells from 10 cell types. We split the data into three subsets: the PengNormal data that consists of the 11 control samples, the PengTumor data that consists of the 24 PDAC tumors, and the Peng data that consists of all 35 samples. Subsequently, we recorded the deconvolution of PDAC-A using these three sets of scRNA-seq data as reference data as scenario 3, scenario 4 and scenario 5, respectively. After obtaining all unmatched scRNA-seq data, we explored the robustness of NODE and other methods to scRNA-seq datasets.

In scenario 1, we used scRNA-seq data of PDAC-B from the same study¹⁹. We used this data as a reference dataset and analyzed all methods. In the results, despite using mismatched scRNA-seq data, NODE and CARD still found layered structures that matched the original tissue images (Supplementary Fig. 31). In contrast, GraphST failed to correctly localize ductal centroacinar cells, and SpaDecon and SpatialDWLS were unable to find any stratified structures at all (Supplementary Fig. 31). SpatialPrompt correctly localized the ductal epithelium, region and the cancer region, but the previous distribution of stroma regions was no longer present (Supplementary Fig. 31). Only NODE and CARD accomplished the task of localizing the regions well and found the distribution characteristics of all cells (Supplementary Fig. 31).

In scenario 2, the reference dataset we used the mix scRNA-seq data of PDAC-A

and PDAC-B. In the results, NODE and CARD were still accurate in identifying tissues in different regions, and the change in the results of NODE and CARD was not significant compared to scenario 1 (Supplementary Fig. 31). On the contrary, the two deconvolution results (scenario 1 and scenario 2) of GraphST, SpaDecon, SpatialDWLS and SpatialPrompt changed greatly. And this phenomenon is especially noticeable in the SpatialPrompt results (Supplementary Fig. 31).

An external scRNA-seq data was used as reference for scenarios 3 to 5, and the scRNA-seq data used for scenario 3 did not contain ductal type 2 cells. After we separately analyzed the deconvolution results from all methods, NODE produced similar deconvolution results across scenarios (Supplementary Fig. 31). In contrast, the other deconvolution methods produced significant changes after replacing the reference scRNA-seq data and incorrectly localized ductal type 2 cells to the vast majority of the tissue in scenario 4 and scenario 5 (Supplementary Fig. 31). Thus, in summary, the NODE method has good robustness and shows excellent performance in referencing scRNA-seq data from different locations.

After that we also explored the distribution of specific cell types, e.g., ductal centroacinar, ductal type 1 and ductal type 2 (Supplementary Fig. 32a and 32b). In scenarios 1 and 2, NODE correctly localized the distribution of ductal centroacinar cells (Supplementary Fig. 32a). The similar distributions in the both scenarios also demonstrated the robustness of NODE (Supplementary Fig. 32a). In scenarios 3 to 5 using external reference scRNA-seq data, only NODE correctly localized ductal cells and was not affected by the difference of reference datasets, whereas the other methods incorrectly localized ductal type 1 and ductal type 2 mainly to cancerous tissues (Supplementary Fig. 32a).

7.2. The ability of NODE to resist interference from spatial noise and spatial batch effects:

Spatial noise and spatial batch effects not only affect the analysis of the data, but also impact the correctness of the algorithm in deconvolution. However, the performance of NODE in three replicates of SCC dataset² demonstrated that NODE has a good resistance to spatial noise. To better demonstrate the performance of NODE, we applied all deconvolution methods to the SCC dataset². The SCC data contains three replicate slices (rep1-rep3), each with different data characteristics. We applied all methods to the three replicate slices to analyze the impact of batch effect and spatial noise on the deconvolution algorithm. We remained focused on exploring the distribution of tumor specific keratinocyte (TSK) cells in all deconvolution results. For a more visual analysis, we first plotted the distribution of marker genes (*MMP10* and *LAMC2*) for TSK, and subsequently, we plotted the TSK aggregation status for all methods (Supplementary Fig. 45a). In the results, NODE behaved very similarly in the three replicate sections and similar cell aggregation states were found. In contrast, the distribution of CARD was not similar in the three replicates, and CARD showed a clear batch difference in rep2 (Supplementary Fig. 45a). SpaDecon also was affected by the batch effect and spatial noise, and its result in rep3 showed significantly more

aggregation than the other slices. The other methods identified too many false-positive locations (Supplementary Fig. 45a). NODE not only correctly identified the TSK distribution, but also showed excellent resistance to the batch effect and spatial noise (Supplementary Fig. 45a). To further illustrate the correctness of NODE, we calculated the correlation coefficients between the TSK distributions and the corresponding marker genes in each deconvolution results (Supplementary Fig. 45b). Higher correlation coefficient indicated higher accuracy of the method. The best performance was achieved by NODE in all replicates with all marker genes (Supplementary Fig. 45b). Subsequently, we also calculated AUC values for TSK distribution and marker gene distribution. Specifically, we calculated the upper quartile of marker gene expression. Then we set the expression of marker genes greater than the upper quartile to 1 and vice versa to 0, and subsequently performed AUC calculations with the TSK distribution. Among all marker genes in all replicates, NODE achieved the best performance in 2/3 of the cases (Supplementary Fig. 45c). A performance second only to CARD was achieved in the remaining cases (Supplementary Fig. 45c).

7.3. The computational efficiency of NODE:

We compared the computational performance of NODE with recently developed deconvolution methods, which consisted of SpaDecon¹¹, SpatialPrompt¹³, SpaTalk⁷, RCTD³, Seurat⁴, donvSeq⁵, SPOTlight⁶, GraphST¹⁰, CARD⁹ and SpatialDWLS¹². By analyzing the computation time and peak memory usage of all methods we comprehensively showed the computational efficiency of NODE. Finally, for all methods including NODE, we summarized their performance in the mouse olfactory bulb (MOB) data⁸ and human heart data²⁶, as well as two simulated data (Tables 3 and 4). Among all the data, although NODE was not a fast method, but NODE performed better in terms of memory usage. Therefore, NODE has better computational scalability. All methods ran on 12th Gen Intel(R) Core (TM) i9-12900H (20 cores), ~2.5GHZ (CPU). GraphST¹⁰ required GPU platform and ran on an extra NVIDIA GeForce RTX 3060 Laptop GPUs.

In brief, NODE is not a fast method (Table 3), but NODE is a good method in terms of accuracy as well as memory usage (Table 4). When it comes to large data, NODE will perform well in terms of memory usage.

Supplementary Note 8. Comparison of NODE with recent developed communication methods

We show the accuracy of the NODE communication. Here we compare NODE with CellChat²⁹ and CellCall³⁰. Although CellChat can be extended to spatial transcriptomics data, it still relies on the annotation of spots, and its communication results are much lower resolution than that of NODE. CellChat can only obtain cell type to cell type or domain to domain communication results, not spot to spot. But NODE can infer the flow of information to each spot in the spatial transcriptomics data. These methods are based on similar mechanism in ligand-receptor interaction. CellChat and CellCall cannot directly use the spatial transcriptomic data. So, we labeled the human heart dataset²⁶ with the dominant cell type in each spot using the deconvolution results from NODE. (We also tried annotation using other deconvolution results, but CellChat was unable to find any communication in subsequent analyses.) After annotating spatial transcriptomic data, we used the human heart dataset²⁶ to infer the communications. In results, we found that neither the CellChat nor the CellCall method identified communication information in smooth muscle cells and ventricular cardiomyocytes. The positional movement and development of smooth muscle cells is related to the development of the pulmonary veins, and the development of ventricular cardiomyocytes varies over time³¹. During development, smooth muscle cells are regulated by the cells around them³¹. But CellChat and CellCall did not identified these communication³¹. According to the literature, electrical impulse signals from the inside of the heart to the outside occur during the developmental cycle of the mouse heart³¹. However, both methods did not find the electrical pulse signals that appear in the developmental cycle of the mouse heart (Supplementary Fig. 46a)³¹. In contrast, NODE was able to find potential communication situations of smooth muscle cells and ventricular cardiomyocytes and communication messages similar to the electrical pulse signals during mouse heart development (Supplementary Fig. 46a)³¹. To better represent the cellular communication found by NODE, we also plotted the communication situation of the cell types found by NODE during the three periods (4-4.5, 6.5, and 9 post-conception weeks (PCW)) (Supplementary Fig. 46b). These communication matched the flow of cellular information described in previous studies of heart development³¹. Subsequently, in order to obtain the communications among spots in spatial transcriptome, we inferred the communications between cell types on the reference single-cell data using CellChat and CellCall, respectively. By the deconvolution results of NODE, the communication between two spots can be identified by counting the communications of cell types between the two spots. Specifically, when computing the communication between any two spots, for the spot that sent the message, we compute the sum of the communications of all cell types contained in the sender spot. Therefore, we can get the communication results among spots by CellChat and CellCall. Afterwards, we separately calculated various correlation coefficients between the communications from NODE and CellCall or CellChat (Supplementary Fig. 46c), and the higher correlation coefficients show that the communications from NODE are more similar to the communications from CellCall

or CellChat (Supplementary Fig. 46c). Finally, NODE obtained high correlation coefficients (Supplementary Fig. 46c), which proves that the spatial communication found by NODE was compatible with the communication results obtained by single cells. It shows that the spatial communication inferred by NODE is reasonable. To summarize, the communication results from NODE have achieved performance and correctness rates that exceed those of CellChat and CellCall.

Supplementary Note 9. NODE innovative features

9.1. NODE supports directly getting the number of each cell type in the deconvolution result.

NODE models the expression of genes directly in terms of the number of cells by means of matrix optimization, so that it can directly return the number of each cell type, while the cell non-occupancy ratio.

9.2. NODE supports saving at any time.

In the NODE code, we provided an incoming interface for saving paths, which allows the user to save the completed section locally for viewing during the deconvolution process. This allows the user to see the result of the deconvolution at any time, instead of having to wait for the full result to be finished

9.3. NODE supports starting the deconvolution at any position in the spatial transcriptomics data.

In the code of NODE, the parameter of the start position is provided, so that the user can change the start position of the deconvolution according to the realistic situation. NODE supports to start of the deconvolution to the last spot of the spatial transcriptome data at any spot position

References

- 1 Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine* **50**, 1-14, doi:10.1038/s12276-018-0071-8 (2018).
- 2 Ji, A. L. *et al.* Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* **182**, 497-514.e422, doi:10.1016/j.cell.2020.05.039 (2020).
- 3 Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology* **40**, 517-526, doi:10.1038/s41587-021-00830-w (2022).
- 4 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).
- 5 Andersson, A. *et al.* Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Communications Biology* **3**, 565, doi:10.1038/s42003-020-01247-y (2020).
- 6 Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Research* **49**, e50-e50, doi:10.1093/nar/gkab043 %J Nucleic Acids Research (2021).
- 7 Shao, X. *et al.* Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nature Communications* **13**, 4429, doi:10.1038/s41467-022-32111-8 (2022).
- 8 Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science (New York, N.Y.)* **353**, 78-82, doi:10.1126/science.aaf2403 (2016).
- 9 Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nature Biotechnology* **40**, 1349-1359, doi:10.1038/s41587-022-01273-7 (2022).
- 10 Long, Y. *et al.* Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nature Communications* **14**, 1155, doi:10.1038/s41467-023-36796-3 (2023).
- 11 Coleman, K., Hu, J., Schroeder, A., Lee, E. B. & Li, M. SpaDecon: cell-type deconvolution in spatial transcriptomics with semi-supervised learning. *Communications Biology* **6**, 378, doi:10.1038/s42003-023-04761-x (2023).
- 12 Dong, R. & Yuan, G.-C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biology* **22**, 145, doi:10.1186/s13059-021-02362-7 (2021).
- 13 Swain, A. K., Pandit, V., Sharma, J. & Yadav, P. SpatialPrompt: spatially aware scalable and accurate tool for spot deconvolution and domain identification in spatial transcriptomics. *Communications Biology* **7**, 639, doi:10.1038/s42003-024-06349-5 (2024).
- 14 Li, Z. & Zhou, X. BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. *Genome Biology* **23**, 168, doi:10.1186/s13059-022-02734-7 (2022).
- 15 Lu, Y., Chen, Q. M. & An, L. SPADE: spatial deconvolution for domain specific cell-type estimation. *Communications Biology* **7**, 469, doi:10.1038/s42003-024-06172-y (2024).
- 16 Zheng, B. *et al.* TM4SF1 as a prognostic marker of pancreatic ductal adenocarcinoma is involved in migration and invasion of cancer cells. *Int J Oncol* **47**, 490-498, doi:10.3892/ijo.2015.3022 (2015).

2138 17 Fu, F. *et al.* Role of Transmembrane 4 L Six Family 1 in the Development and Progression of
2139 Cancer. **7**, doi:10.3389/fmolb.2020.00202 (2020).

2140 18 Xu, D. *et al.* Lost miR-141 and upregulated TM4SF1 expressions associate with poor
2141 prognosis of pancreatic cancer: regulation of EMT and angiogenesis by miR-141 and TM4SF1
2142 via AKT. *Cancer biology & therapy* **21**, 354-363, doi:10.1080/15384047.2019.1702401
2143 (2020).

2144 19 Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell
2145 RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature*
2146 *Biotechnology* **38**, 333-342, doi:10.1038/s41587-019-0392-8 (2020).

2147 20 Morvaridi, S., Dhall, D., Greene, M. I., Pandol, S. J. & Wang, Q. Role of YAP and TAZ in
2148 pancreatic ductal adenocarcinoma and in stellate cells associated with cancer and chronic
2149 pancreatitis. *Scientific Reports* **5**, 16759, doi:10.1038/srep16759 (2015).

2150 21 Peng, J. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant
2151 progression in pancreatic ductal adenocarcinoma. *Cell research* **29**, 725-738,
2152 doi:10.1038/s41422-019-0195-y (2019).

2153 22 Ben-Israel, A. & Greville, T. N. E. *Generalized Inverses: Theory and Applications*. (Springer,
2154 2003).

2155 23 Horn, R. A. & Johnson, C. R. *Matrix Analysis*. (Cambridge University Press, 1985).

2156 24 Kraft, D. & Dfvlr, F. B. *A software package for sequential quadratic programming*. 33 s
2157 (1988).

2158 25 Ahmad, M. F., Isa, N. A. M., Lim, W. H. & Ang, K. M. Differential evolution: A recent review
2159 based on state-of-the-art works. *Alexandria Engineering Journal* **61**, 3831-3872,
2160 doi:https://doi.org/10.1016/j.aej.2021.09.013 (2022).

2161 26 Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the
2162 Developing Human Heart. *Cell* **179**, 1647-1660.e1619, doi:10.1016/j.cell.2019.11.025 (2019).

2163 27 Olude, M. A., Mustapha, O. A., Aderounmu, O. A., Olopade, J. O. & Ihunwo, A. O. Astrocyte
2164 morphology, heterogeneity, and density in the developing African giant rat (*Cricetomys*
2165 *gambianus*). *Frontiers in neuroanatomy* **9**, 67, doi:10.3389/fnana.2015.00067 (2015).

2166 28 Nielsen, M. F., Mortensen, M. B. & Detlefsen, S. Key players in pancreatic cancer-stroma
2167 interaction: Cancer-associated fibroblasts, endothelial and inflammatory cells. *World journal*
2168 *of gastroenterology* **22**, 2678-2700, doi:10.3748/wjg.v22.i9.2678 (2016).

2169 29 Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell–cell communication
2170 from single-cell transcriptomics. *Nature Protocols*, doi:10.1038/s41596-024-01045-4 (2024).

2171 30 Zhang, Y. *et al.* CellCall: integrating paired ligand–receptor and transcription factor activities
2172 for cell–cell communication. *Nucleic Acids Research* **49**, 8520-8534, doi:10.1093/nar/gkab638
2173 (2021).

2174 31 Buijtendijk, M. F. J., Barnett, P. & van den Hoff, M. J. B. Development of the human heart.
2175 *American journal of medical genetics. Part C, Seminars in medical genetics* **184**, 7-22,
2176 doi:10.1002/ajmg.c.31778 (2020).

2177