A

	BayesSpace	SpaGCN	STAGATE	stLearn	BASS	SEDR	SpatialPCA
Mouse Brain	√	√	√	√	√	√	√
Human Breast Cancer	√	√	√	√	√	√	√
DLPFC	√	√	√	√	√	√	√
Mouse Olfactory Bulb	√	√	√	×	√	√	√
Mouse Cortex (1 slice)	√	√	√	×	√	√	√
Mouse Cortex (3 slices)	√	√	√	×	√	√	√
Human SCC	√	√	√	×	√	√	√

B

Simulation	Dataset resource	Technology	# spots	# genes	# noise (%)	# cell types
SimData1	Mouse Embryo	Stereo-seq	1,000 2,000 3,000 4,000 5,000 6,000 7,000 8,000 9,000 10,000 11,000 12,000 13,000 14,000 15,000	2,000 (HVGs)	no	17
SimData2	Mouse Brain	10x Visium	2,688	300 600 900 1,200 1,500 1,800 2,100 2,400 2,700 3,000 HVGs	no	20
SimData3	Mouse Brain	10x Visium	2,688	300 600 900 1,200 1,500 1,800 2,100 2,400 2,700 3,000 SVGs	no	20
SimData4	Mouse Brain	10x Visium	2,688	3,000 (SVGs)	10 20 30 40 50 60 70 80 90 100	20
SimData5	Mouse Brain	10x Visium	2,688	3,000 (SVGs)	no	5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

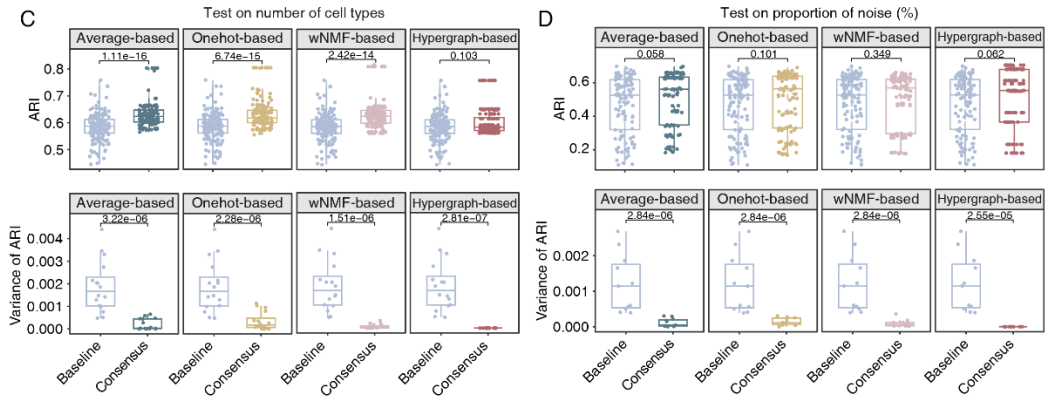

Figure S1. Overview of the simulated datasets. (*A*) Presentation of the clustering algorithm executed on each dataset, with Mouse Olfactory Bulb and STARmap data excluded from stLearn due to lack of image information. (*B*) Visual display of the simulated datasets. We simulated 5 datasets SimData1-SimData5, showing the data resource, technology, number of spots, HVGs/SVGs, and cell types in sequence. All datasets utilize SpaGCN as the baseline algorithm. (*C,D*) Comparison of ARI scores and variances between consensus methods and baseline methods when testing on different numbers of cell types (*C*) and different noise proportions (*D*) (with statistical testing).

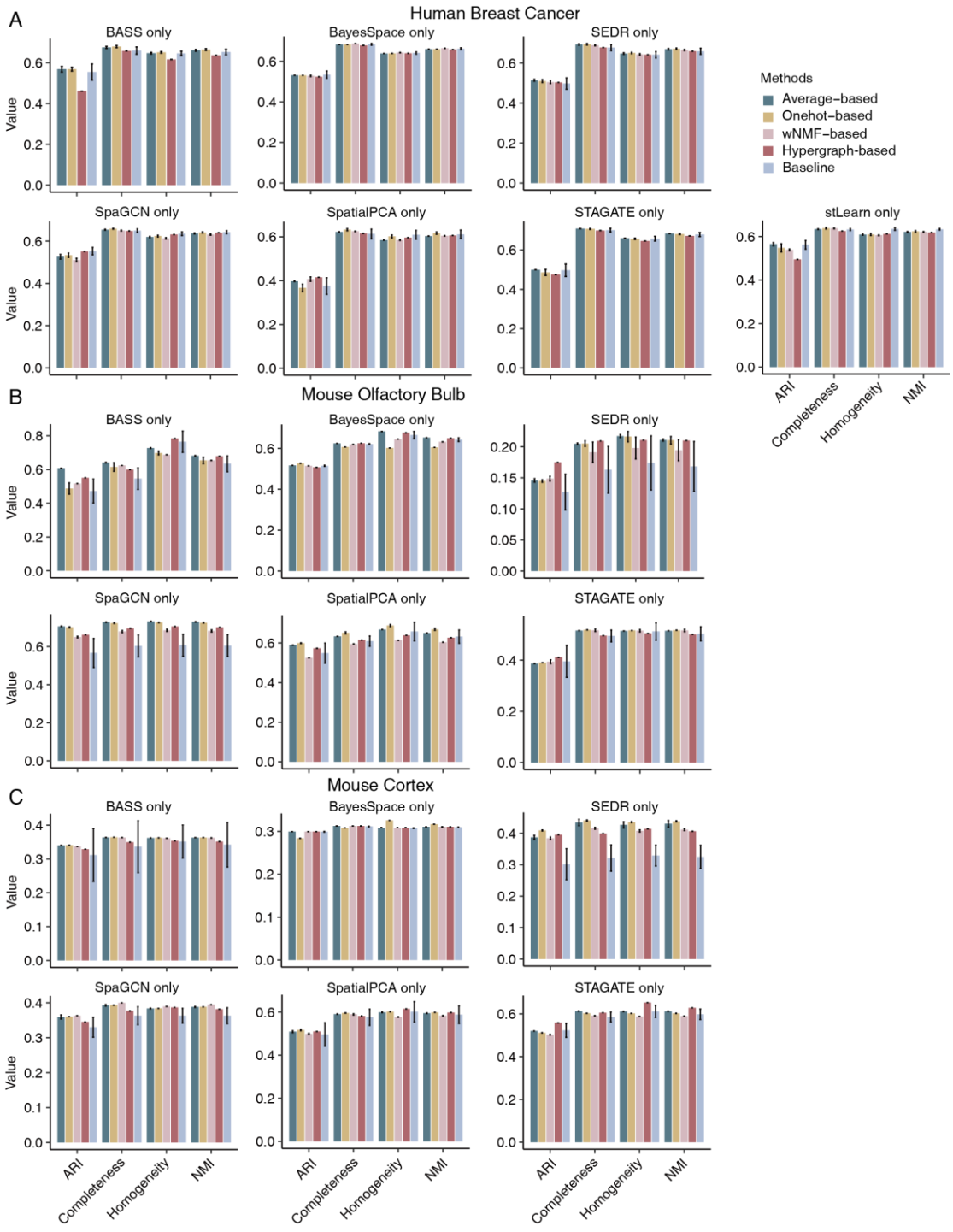

Figure S2. Performance of consensus strategies on individual clustering algorithm. (*A*-*C*) Consensus clustering accuracies based on individual baseline algorithms for Human Breast Cancer data (*A*), Mouse Olfactory Bulb data (*B*), and Mouse Cortex data (*C*), along with the calculation of 4 indicators (ARI, NMI, Completeness score and Homogeneity score).

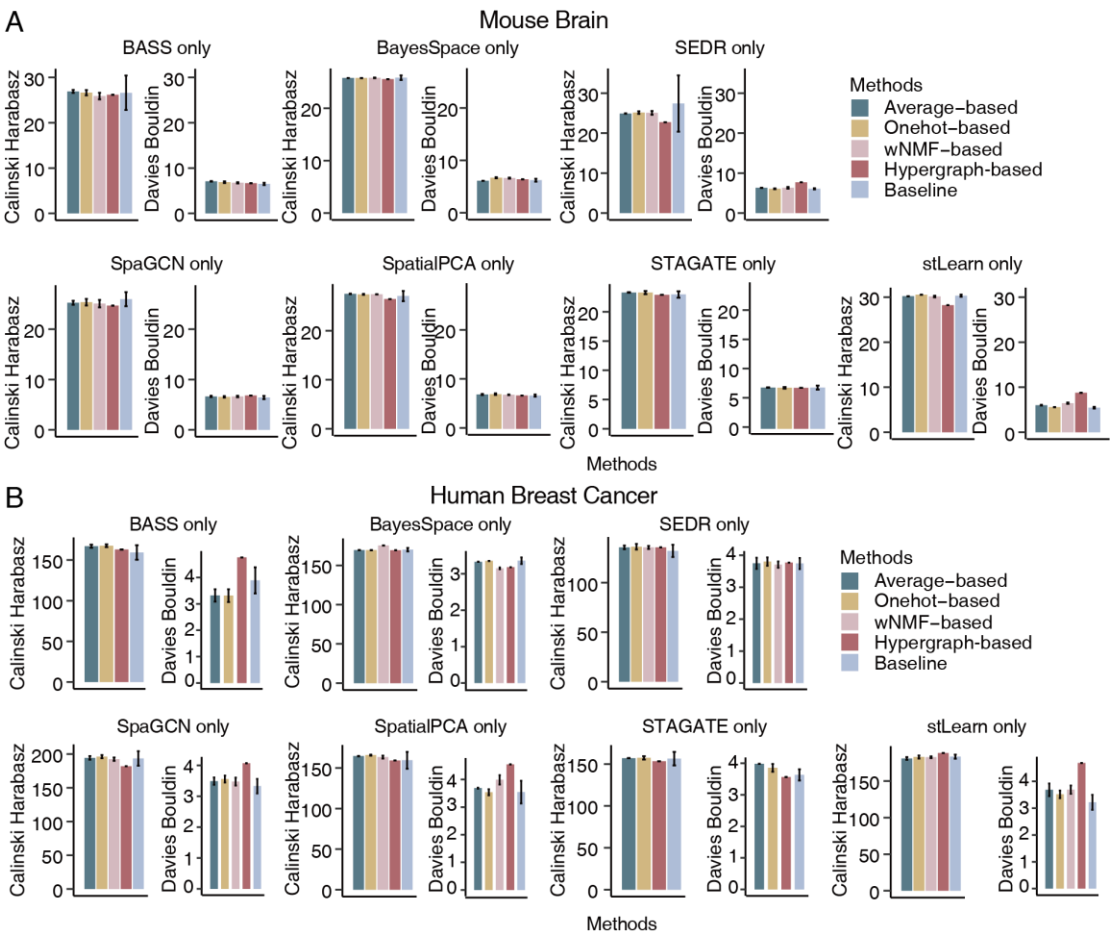

Figure S3. Performance of consensus strategies on individual clustering algorithm. (*A*,*B*) Consensus clustering accuracies based on individual baseline algorithms for Mouse Brain data (*A*), and Human Breast Cancer data (*B*), along with the calculation of 2 indicators (Calinsky Harabasz score and Davies Bouldin score).

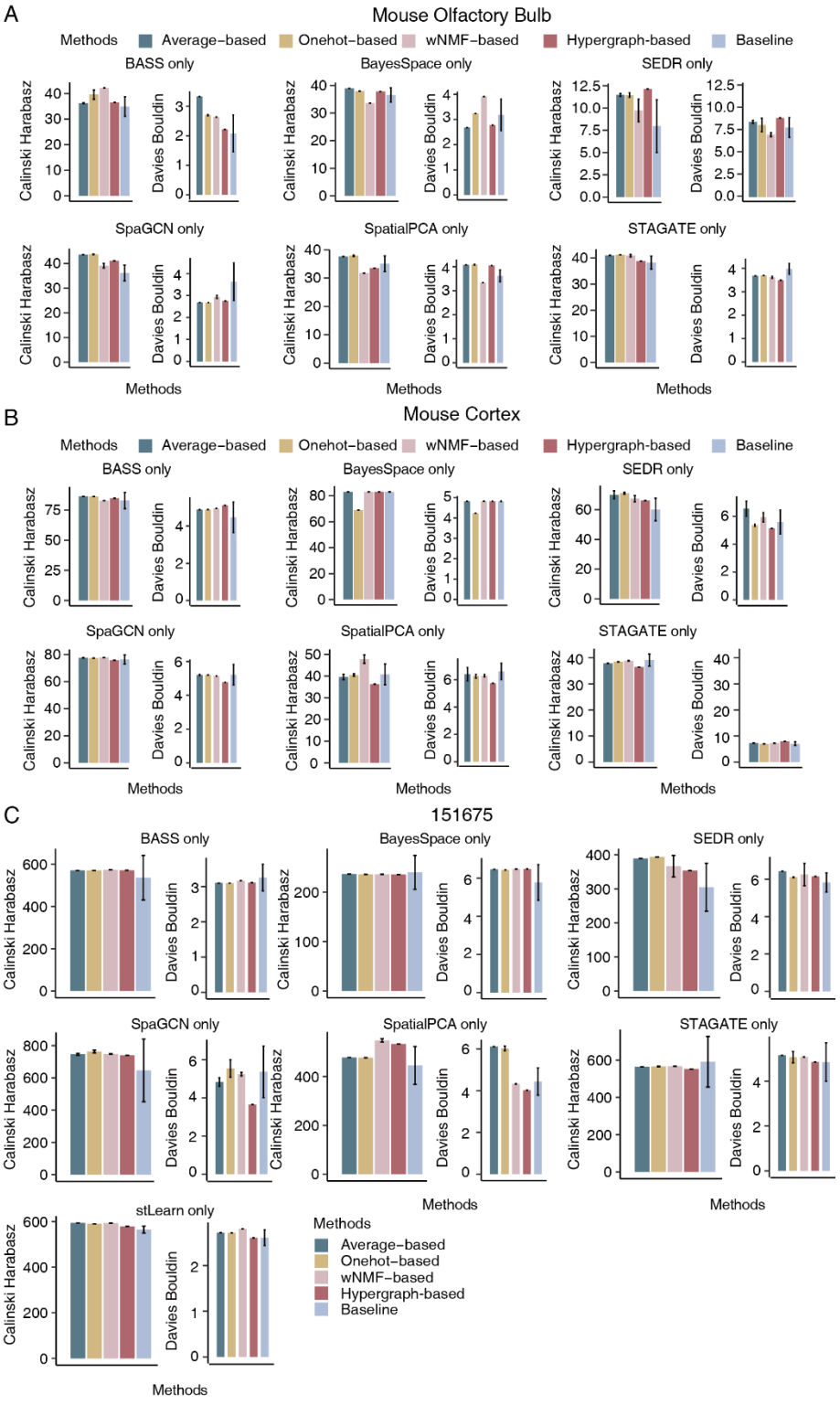

Figure S4. Performance of consensus strategies on individual clustering algorithm. (*A*-*C*) Consensus clustering accuracies based on individual baseline algorithms for Mouse Olfactory Bulb data (*A*), Mouse Cortex data (*B*), and DLPFC sample 151675 data (*C*), along with the calculation of 2 indicators (Calinsky Harabasz score and Davies Bouldin score).

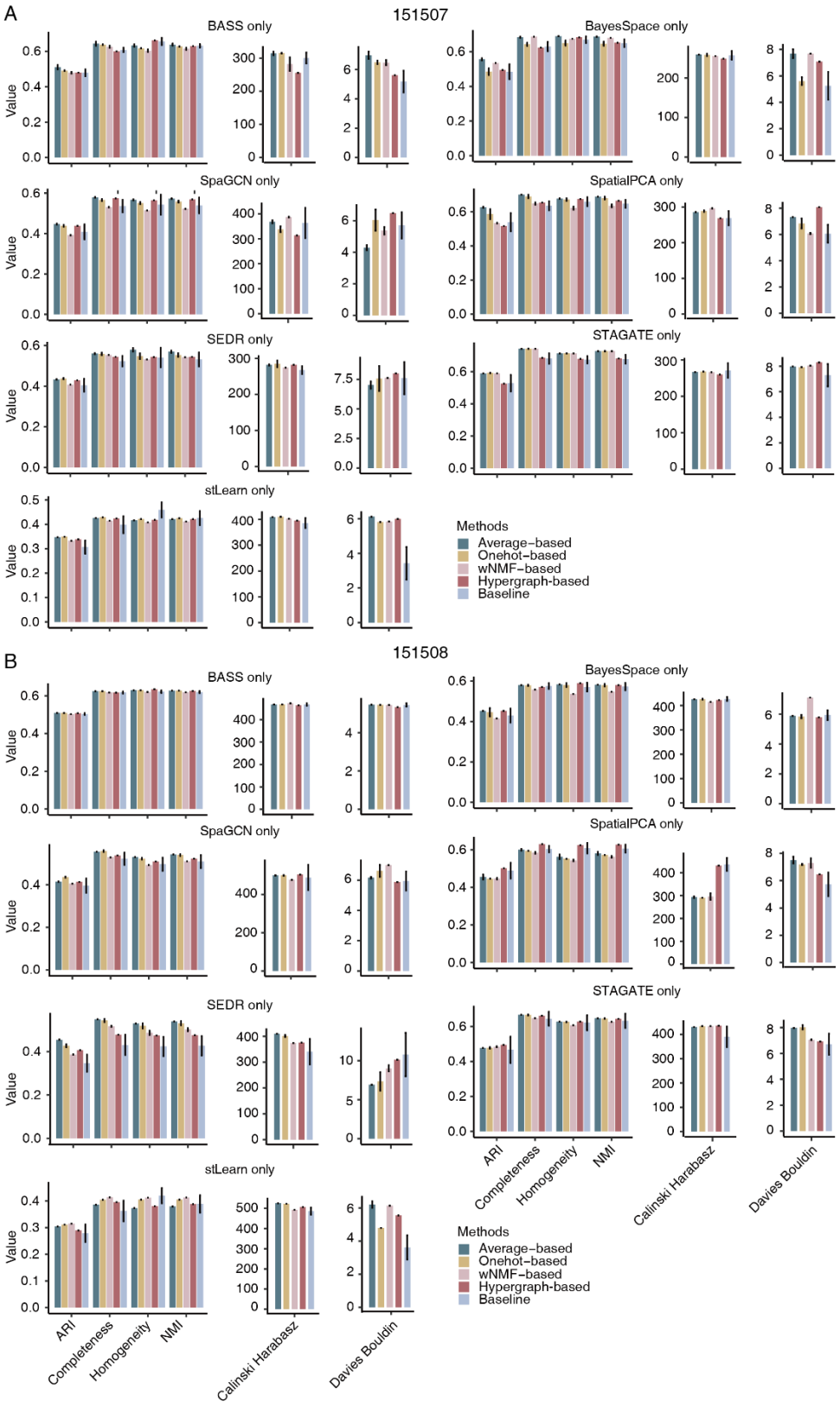

Figure S5. Performance of consensus strategies on individual clustering algorithm. (*A*,*B*) Consensus clustering accuracies based on individual baseline algorithms for DLPFC sample 151607 data (*A*), and sample 151608 data (*B*), along with the calculation of 6 indicators (ARI, NMI, Completeness score, Homogeneity score, Calinsky Harabasz score and Davies Bouldin score).

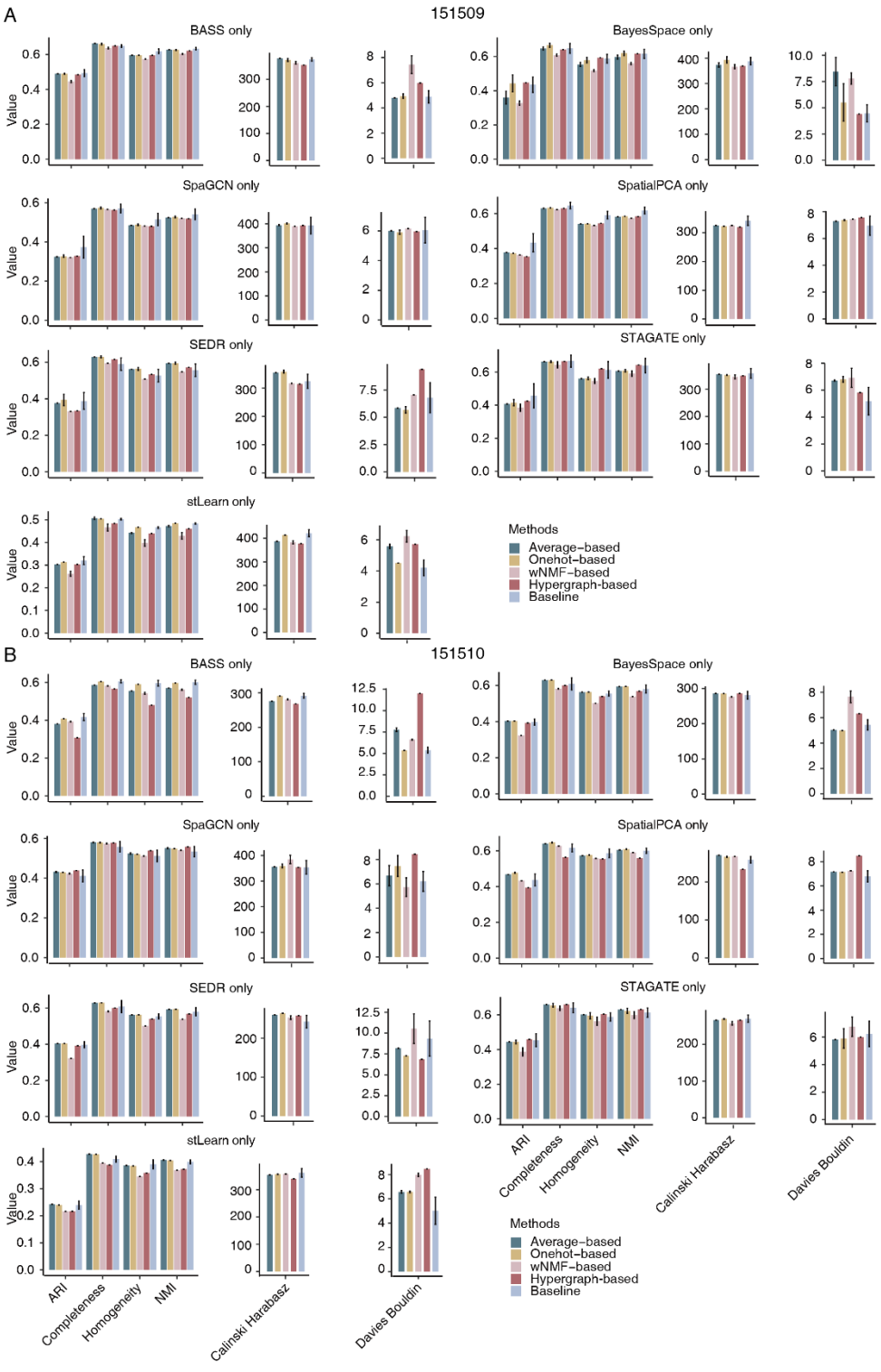

Figure S6. Performance of consensus strategies on individual clustering algorithm. (*A*,*B*) Consensus clustering accuracies based on individual baseline algorithms for DLPFC sample 151609 data (*A*), and sample 151610 data (*B*), along with the calculation of 6 indicators (ARI, NMI, Completeness score, Homogeneity score, Calinsky Harabasz score and Davies Bouldin score).

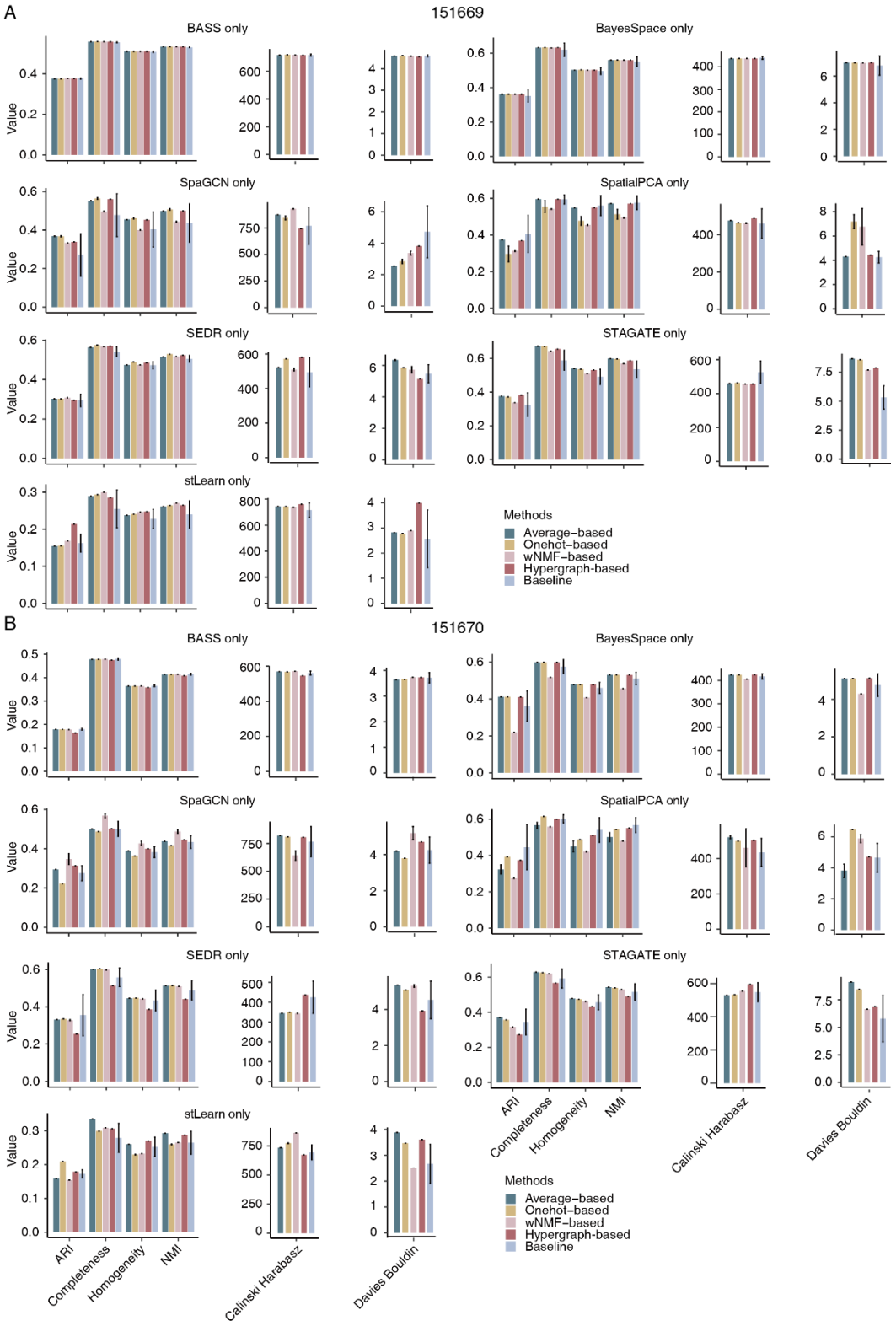

Figure S7. Performance of consensus strategies on individual clustering algorithm. (*A*,*B*) Consensus clustering accuracies based on individual baseline algorithms for DLPFC sample 151669 data (*A*), and sample 151670 data (*B*), along with the calculation of 6 indicators (ARI, NMI, Completeness score, Homogeneity score, Calinsky Harabasz score and Davies Bouldin score).

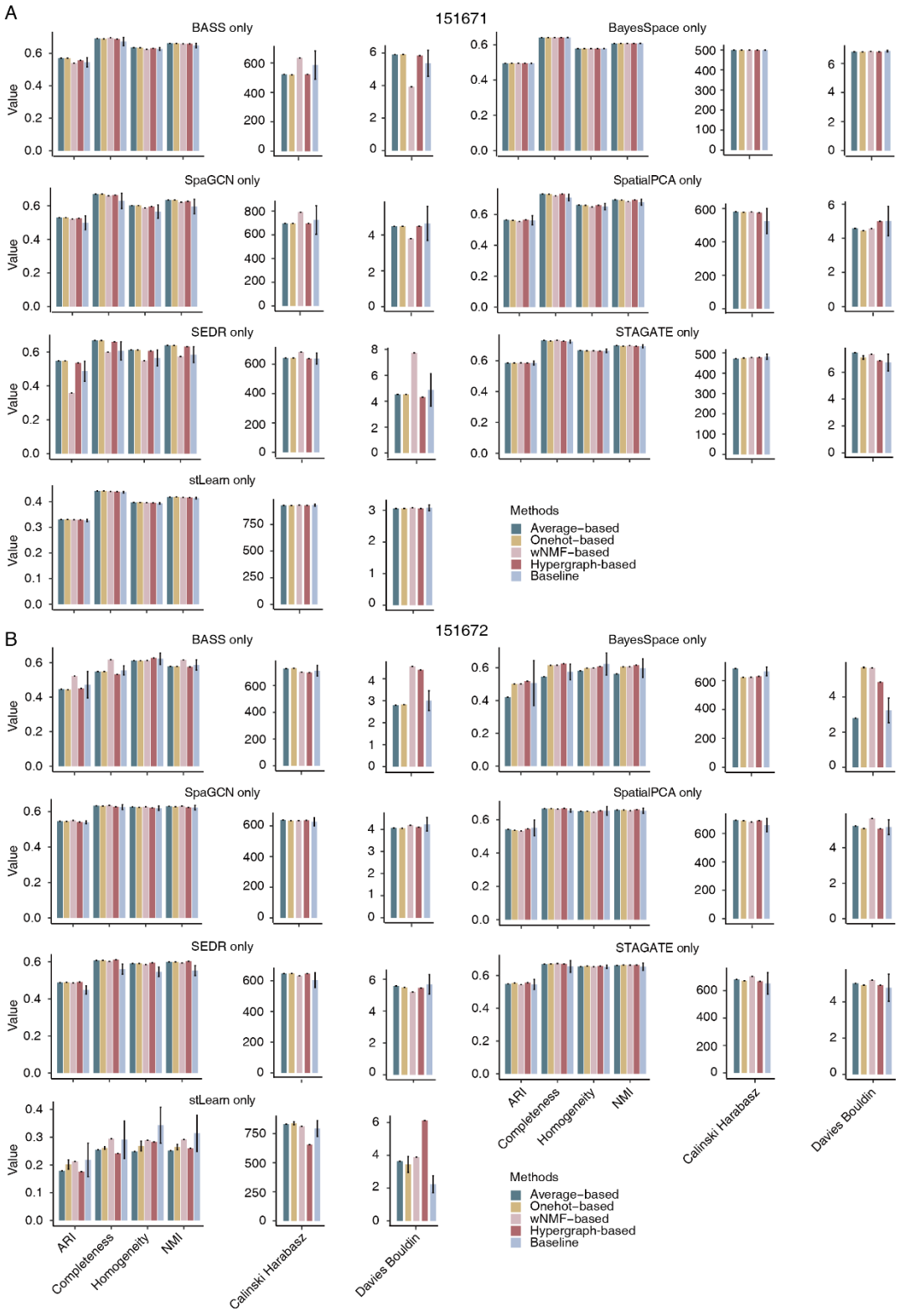

Figure S8. Performance of consensus strategies on individual clustering algorithm. (*A*,*B*) Consensus clustering accuracies based on individual baseline algorithms for DLPFC sample 151671 data (*A*), and sample 151672 data (*B*), along with the calculation of 6 indicators (ARI, NMI, Completeness score, Homogeneity score, Calinsky Harabasz score and Davies Bouldin score).

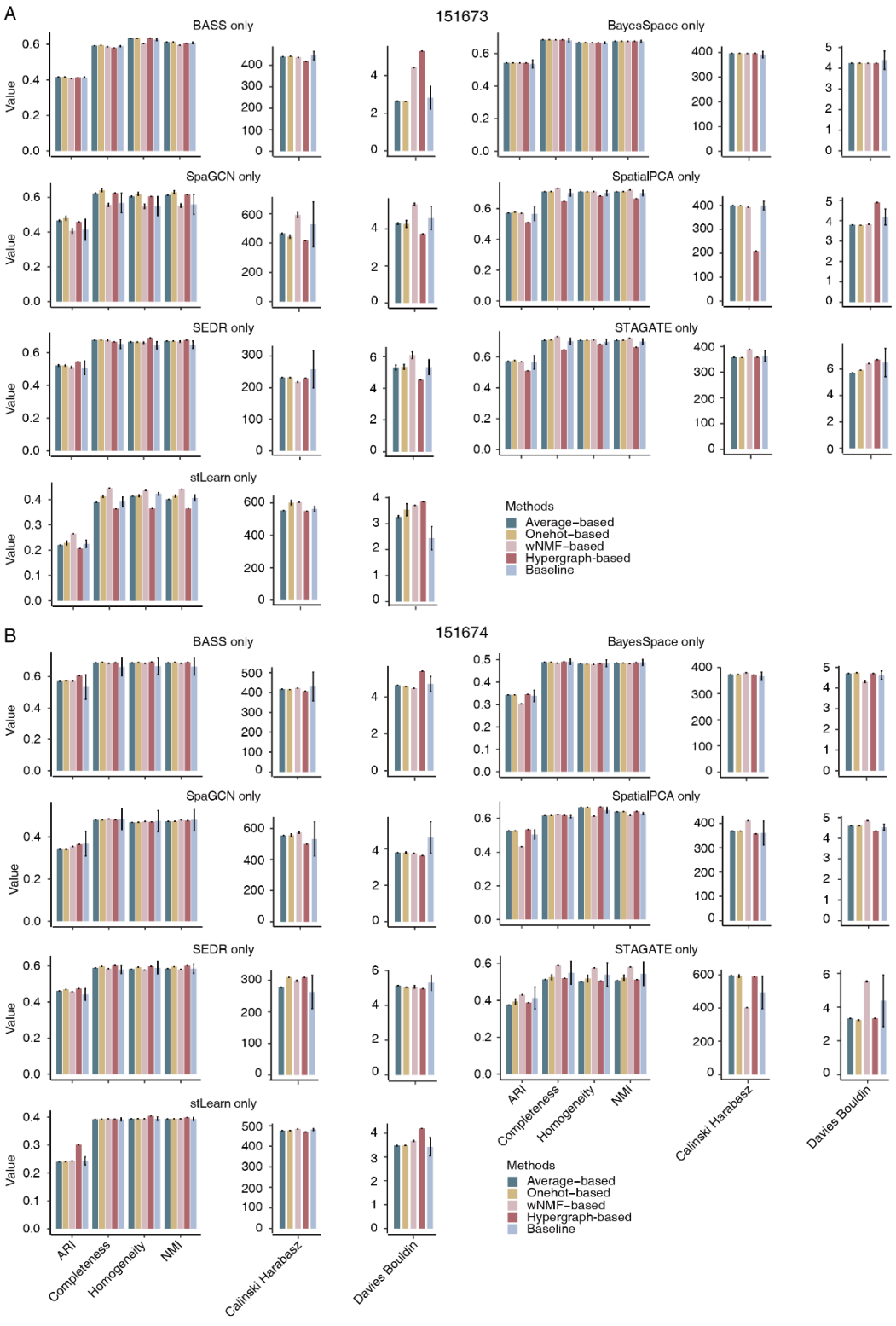

Figure S9. Performance of consensus strategies on individual clustering algorithm. (*A*,*B*) Consensus clustering accuracies based on individual baseline algorithms for DLPFC sample 151673 data (*A*), and sample 151674 data (*B*), along with the calculation of 6 indicators (ARI, NMI, Completeness score, Homogeneity score, Calinsky Harabasz score and Davies Bouldin score).

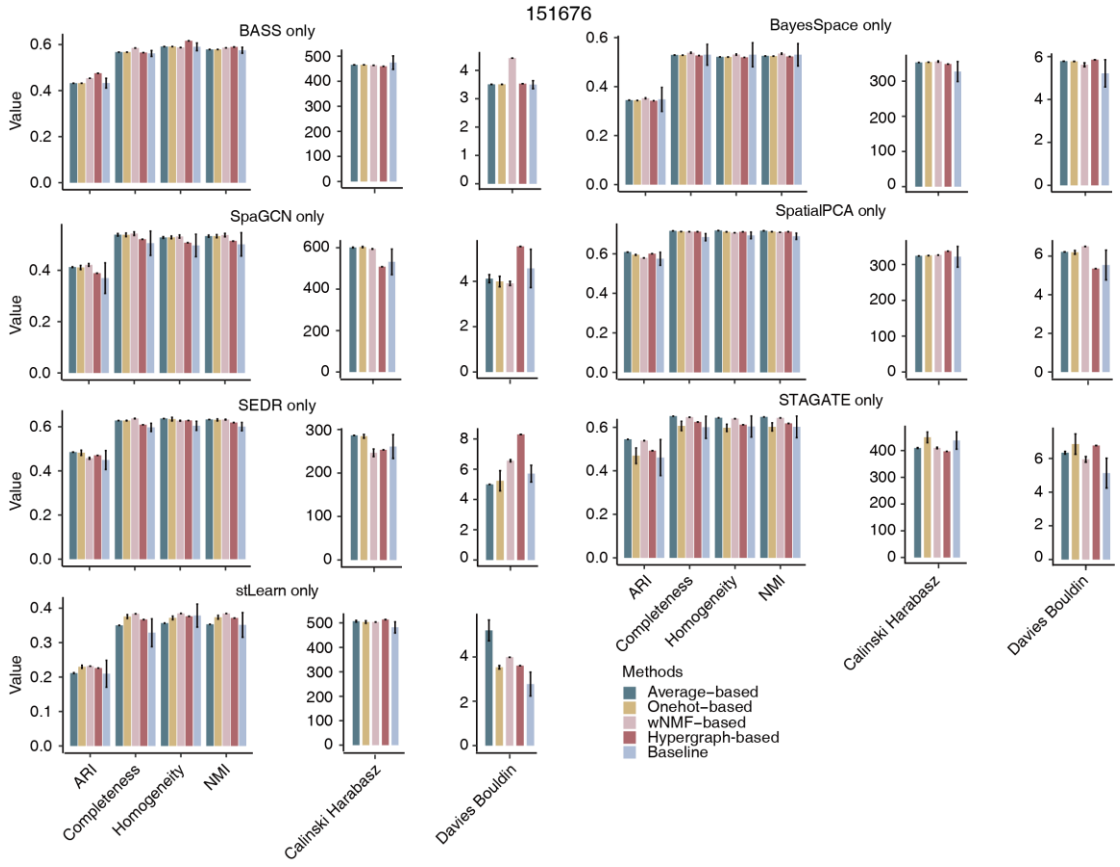

Figure S10. Performance of consensus strategies on individual clustering algorithm. Consensus clustering accuracies based on individual baseline algorithms for DLPFC sample 151676 data, along with the calculation of 6 indicators (ARI, NMI, Completeness score, Homogeneity score, Calinsky Harabasz score and Davies Bouldin score).

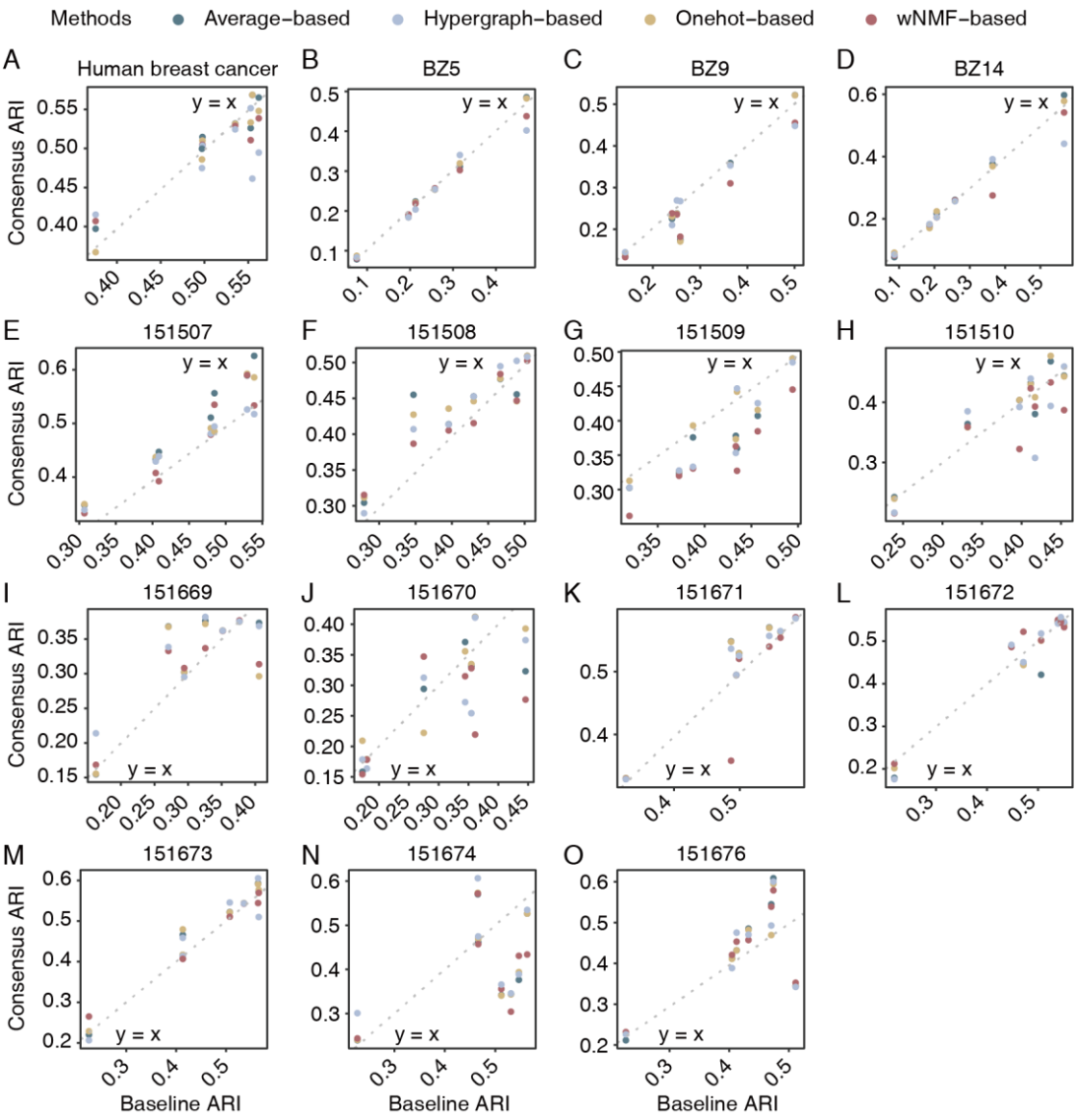

Figure S11. Performance of consensus strategies on individual clustering algorithm. (*A*-*O*) Scatter plot comparing the ARI between baseline methods (x-axis) and consensus methods (y-axis) for Human Breast Cancer data (*A*), Mouse Cortex data (*B*-*D*) and DLPFC data (*E*-*O*), with the y=x line indicating equal performance.

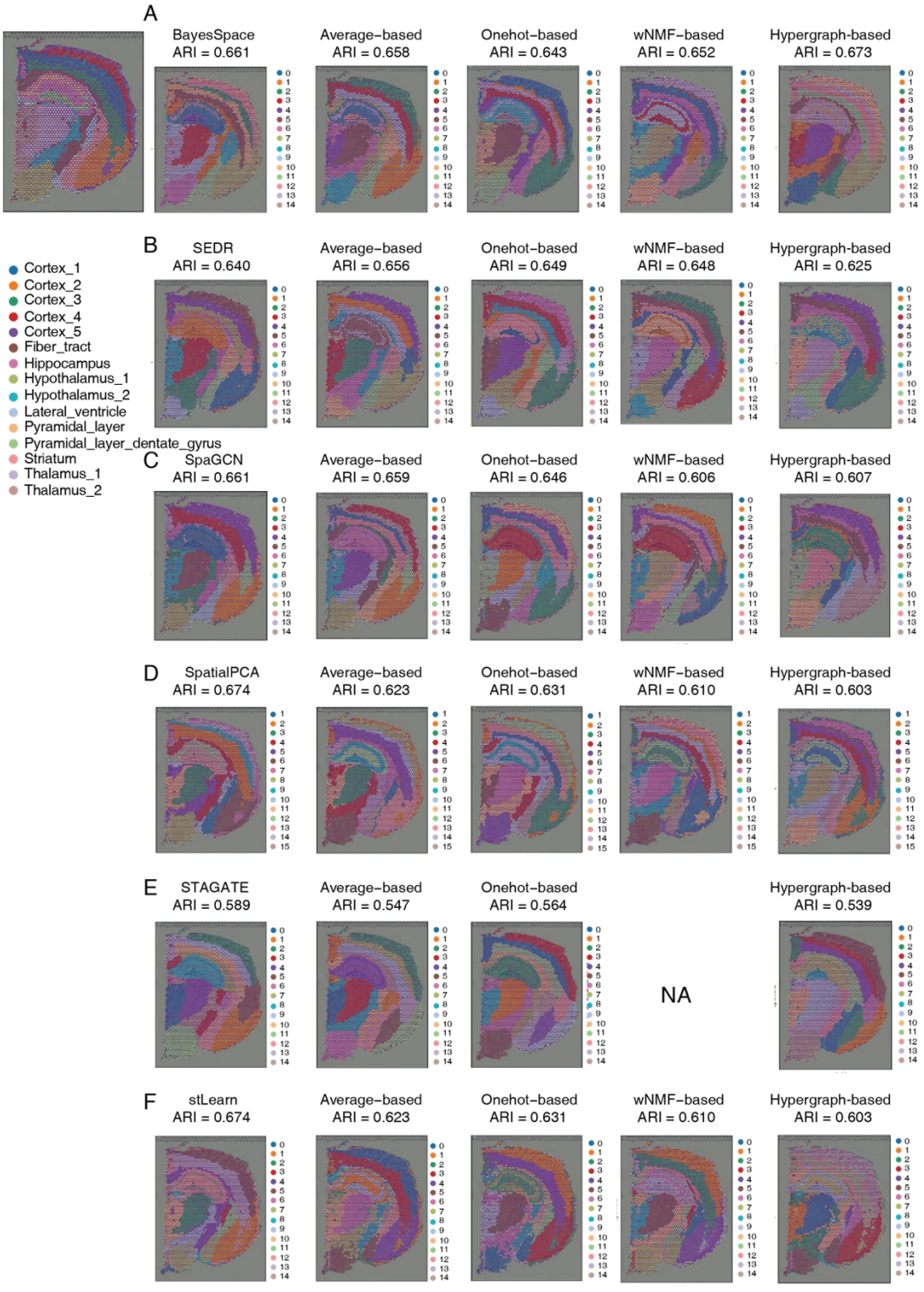

Figure S12. Clustering visualization of baseline algorithms and consensus strategies on Mouse Brain data. (*A*-*F*) showcase the application of BayesSpace (*A*), SEDR (*B*), SpaGCN (*C*), SpatialPCA (*D*), STAGATE (*E*), and stLearn (*F*) as the baseline algorithm for Mouse Brain data, presenting the consensus clustering results under 4 consensus strategies. Notably, the wNMF-based strategy cannot be applied to STAGATE clustering results due to the irreversibility of the matrix. The upper left represents the ground truth.

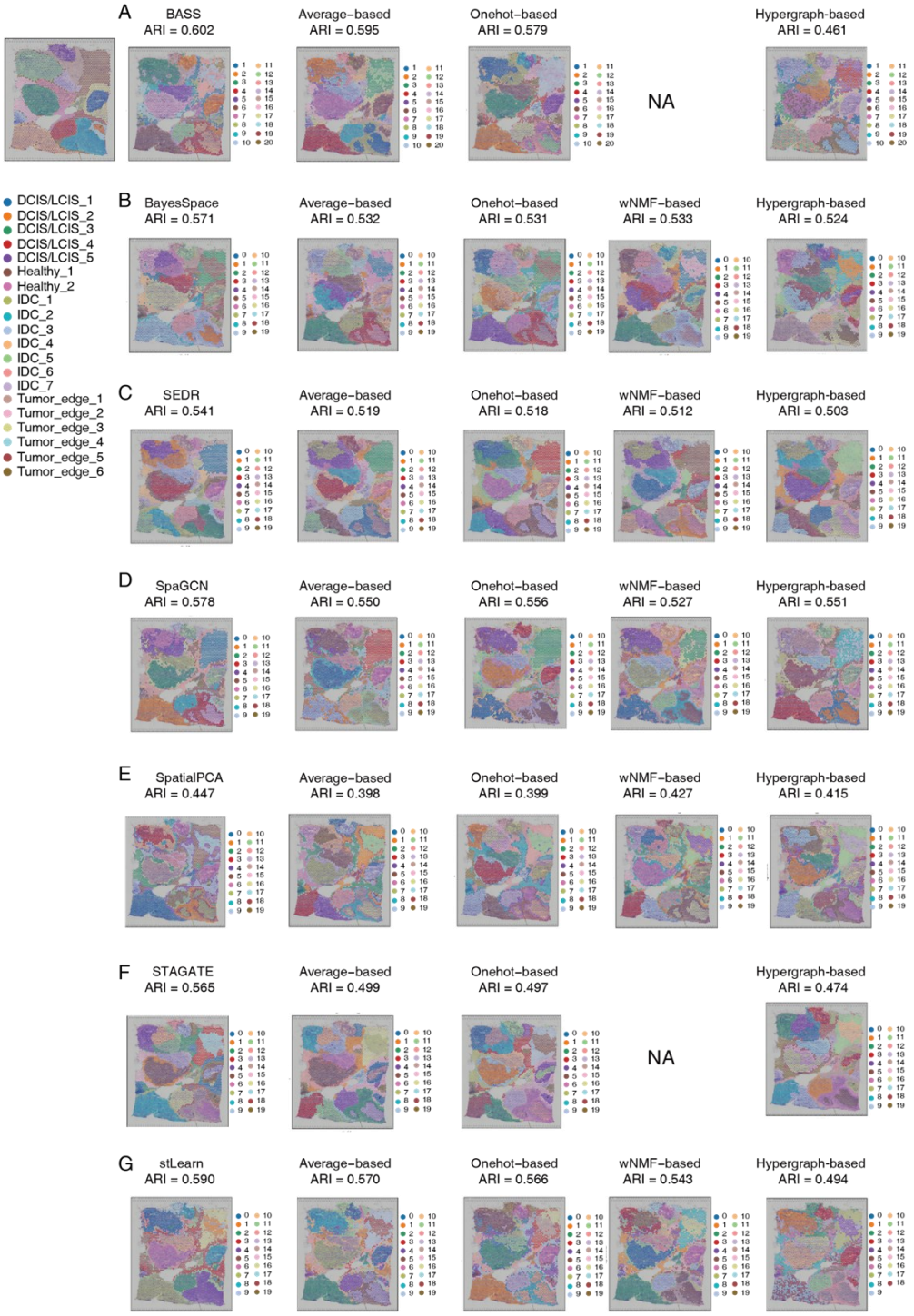

Figure S13. Clustering visualization of baseline algorithms and consensus strategies on Human Breast Cancer data. (*A*-*G*) showcase the application of BASS (*A*), BayesSpace (*B*), SEDR (*C*), SpaGCN (*D*), SpatialPCA (*E*), STAGATE (*F*), and stLearn (*G*) as the baseline algorithm for Human Breast Cancer data, presenting the consensus clustering results under 4 consensus strategies. Notably, the wNMF-based strategy cannot be applied to BASS and STAGATE clustering results due to the irreversibility of the matrix. The upper left represents the ground truth.

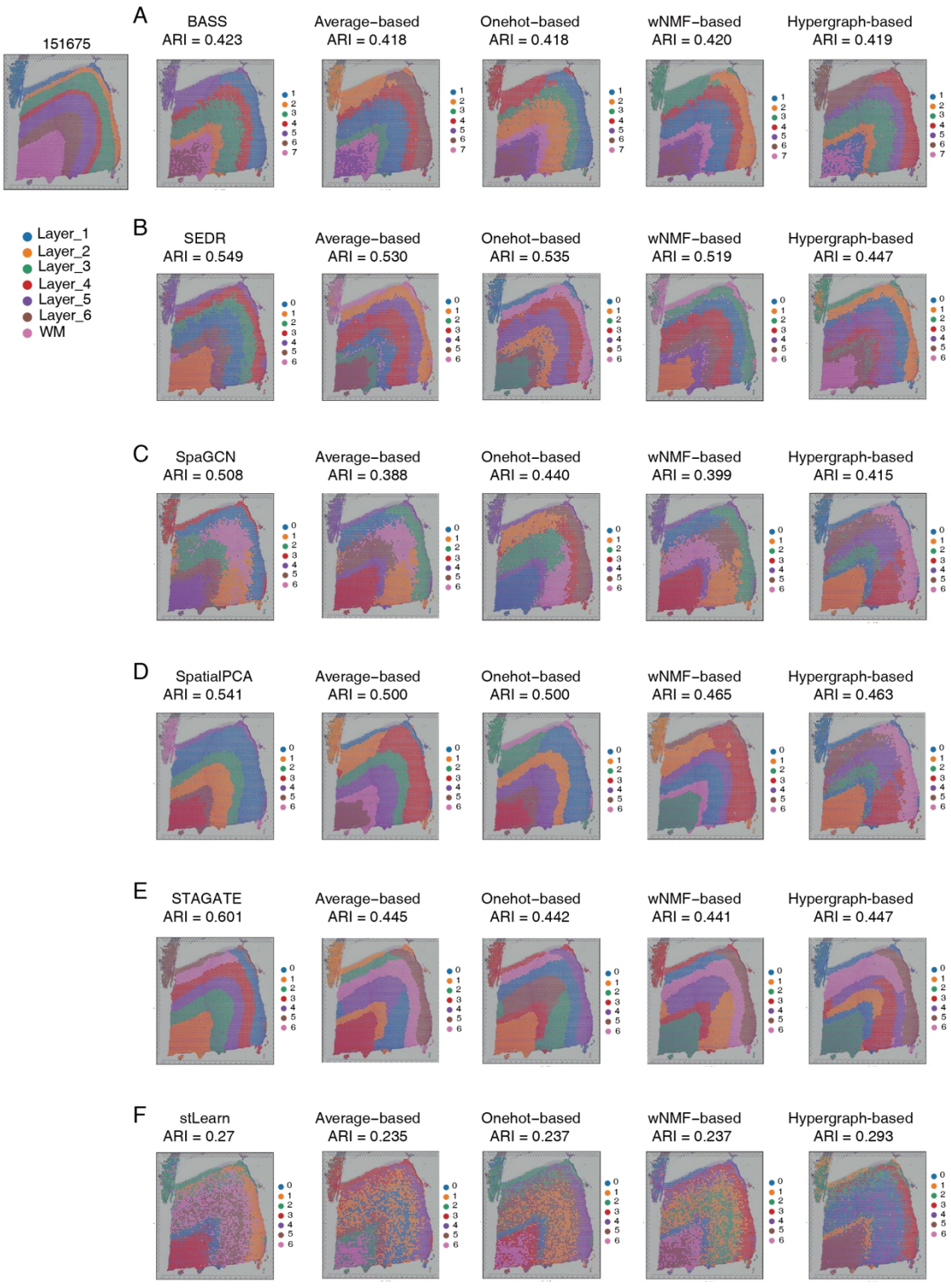

Figure S14. Clustering visualization of baseline algorithms and consensus strategies on 151675 sample data. (*A-F*) showcase the application of BASS (*A*), SEDR (*B*), SpaGCN (*C*), SpatialPCA (*D*), STAGATE (*E*), and stLearn (*F*) as the baseline algorithm for DLPFC sample 151675 data, presenting the consensus clustering results under 4 consensus strategies. The upper left represents the ground truth.

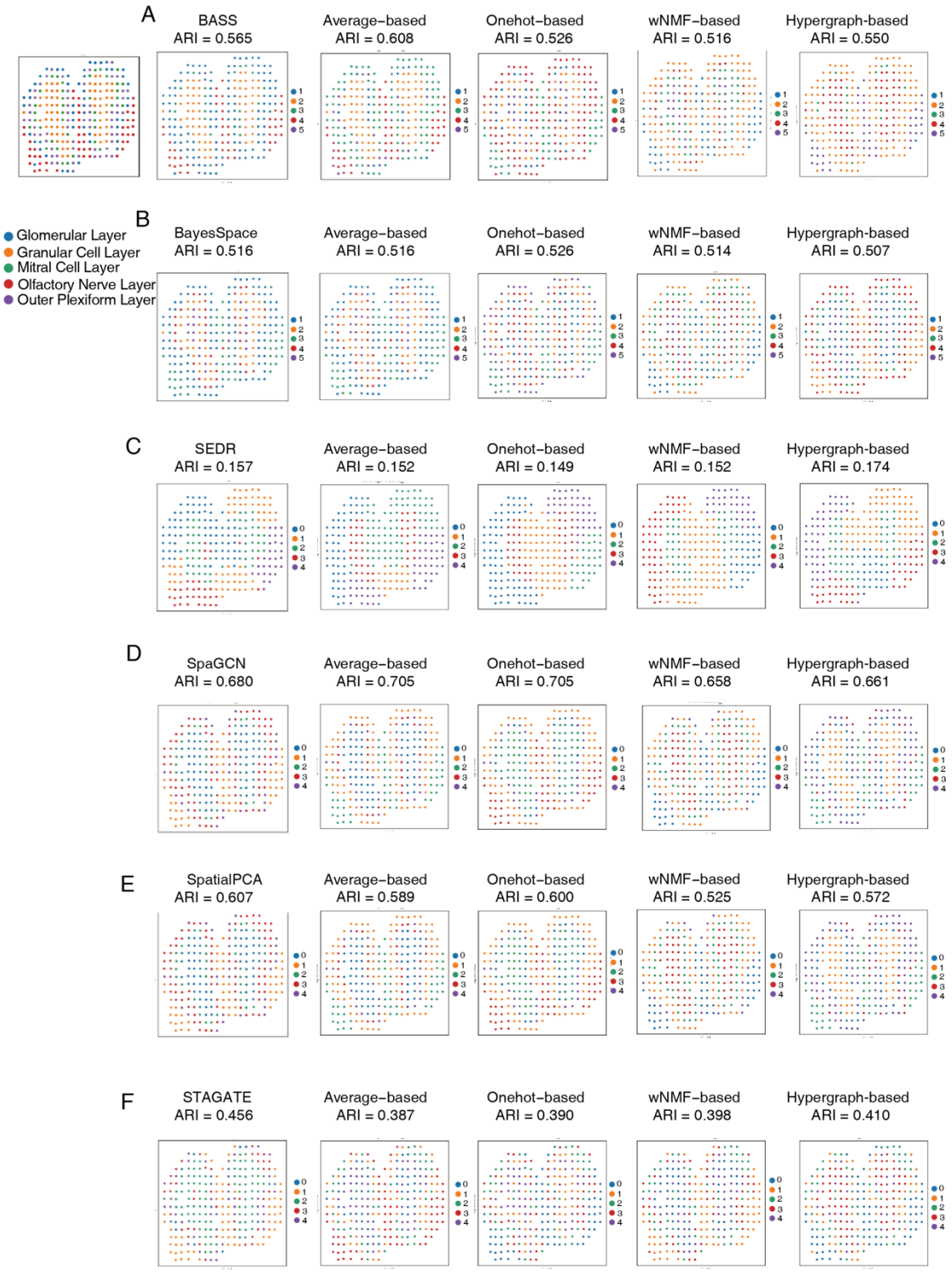

Figure S15. Clustering visualization of baseline algorithms and consensus strategies on Mouse Olfactory Bulb data. (*A*-*F*) showcase the application of BASS (*A*), BayesSpace (*B*), SEDR (*C*), SpaGCN (*D*), SpatialPCA (*E*), and STAGATE (*F*) as the baseline algorithm for Mouse Olfactory Bulb data, presenting the consensus clustering results under 4 consensus strategies. The upper left represents the ground truth.

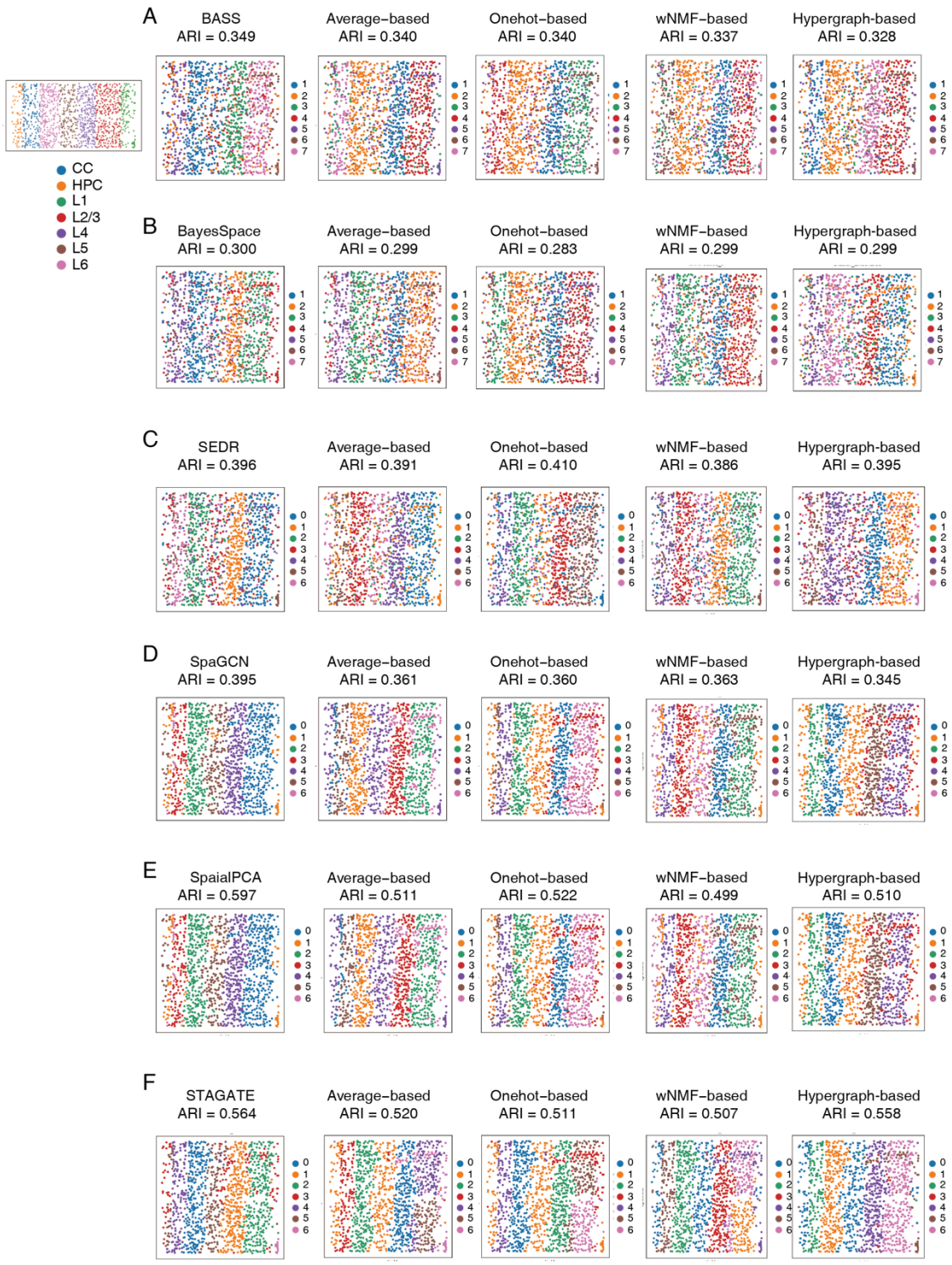

Figure S16. Clustering visualization of baseline algorithms and consensus strategies on Mouse Cortex data. (*A*-*F*) showcase the application of BASS (*A*), BayesSpace (*B*), SEDR (*C*), SpaGCN (*D*), SpatialPCA (*E*), and STAGATE (*F*) as the baseline algorithm for Mouse Cortex data, presenting the consensus clustering results under 4 consensus strategies. The upper left represents the ground truth.

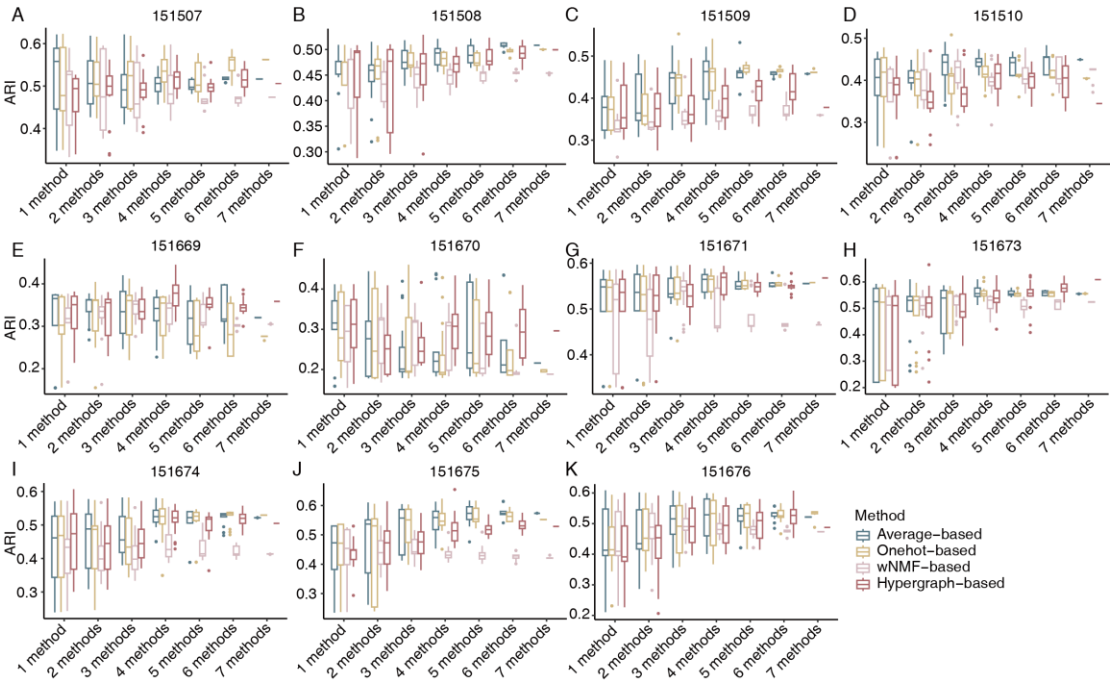

Figure S17. Integration performance of 4 consensus strategies with gradually increasing baseline algorithms on DLPFC data. (*A*-*K*) Accuracies of 4 consensus strategies under different numbers of baseline algorithm of DLPFC sample 151507 (*A*), 151508 (*B*), 151509 (*C*), 151510 (*D*), 151669 (*E*), 151670 (*F*), 151671 (*G*), 151673 (*H*), 151674 (*I*), 151675 (*J*), and 151676 (*K*). We randomly selected 1-6 baseline algorithms for DLPFC data, and performed consensus clustering for each scenario 20 times.

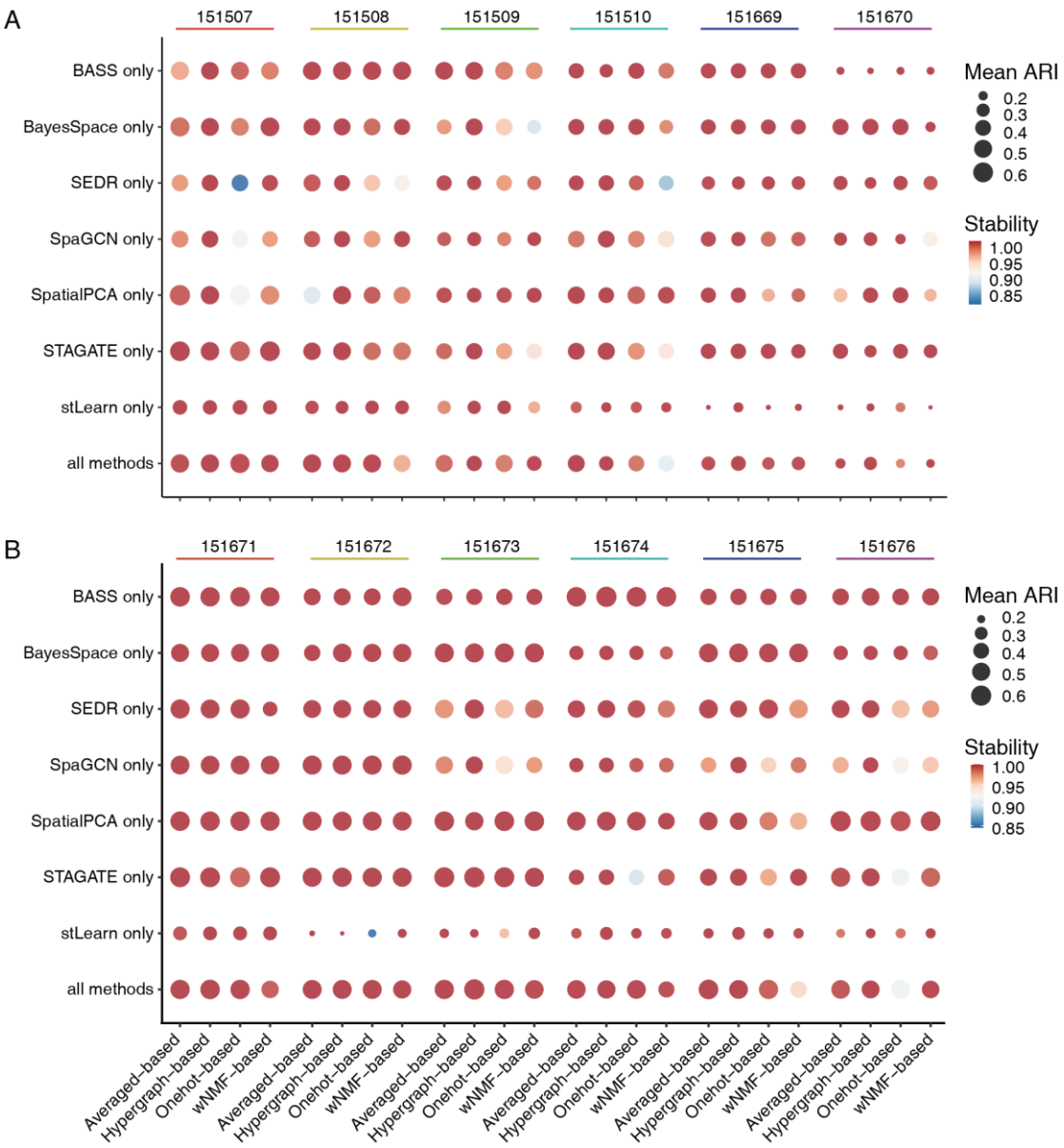

Figure S18. Accuracy and stability evaluation of consensus clustering on DLPFC. (*A*,*B*) Averaged ARI of 4 consensus strategies under 'single method' and 'all methods' situations of the first 6 samples (*A*) and the last 6 samples (*B*) of DLPFC data. The size of the dots represents the mean ARI of 10 repeat experiments, color of the spots indicates stability.

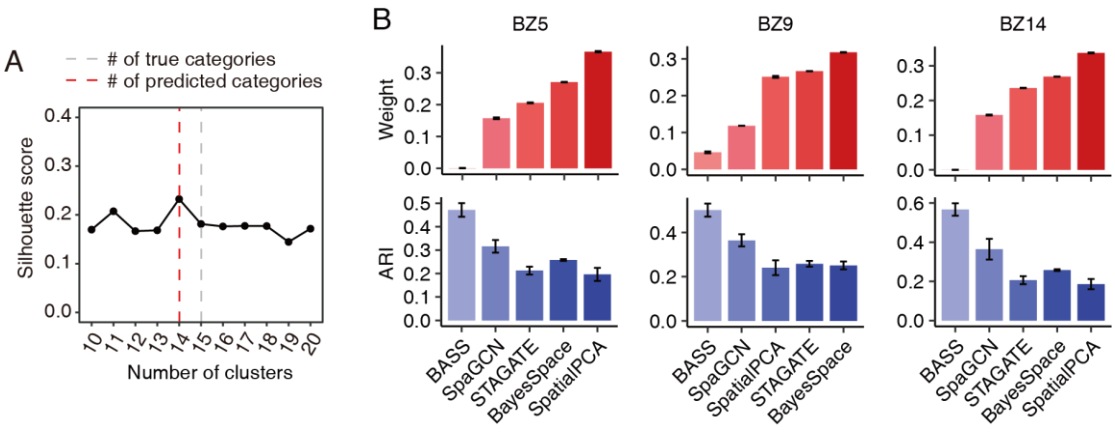

Figure S19. Determination of optimal clusters and weights in the wNMF-based consensus strategy. (A) The optimal number of clusters is determined based on various silhouette metrics, with the red dashed line indicating the predicted number of clusters and the gray dashed line representing the true number of classes. (B) The bar plots display the weights assigned to baseline algorithms in the wNMF-based consensus strategy for three samples (top) and the corresponding ARI of each baseline algorithm.

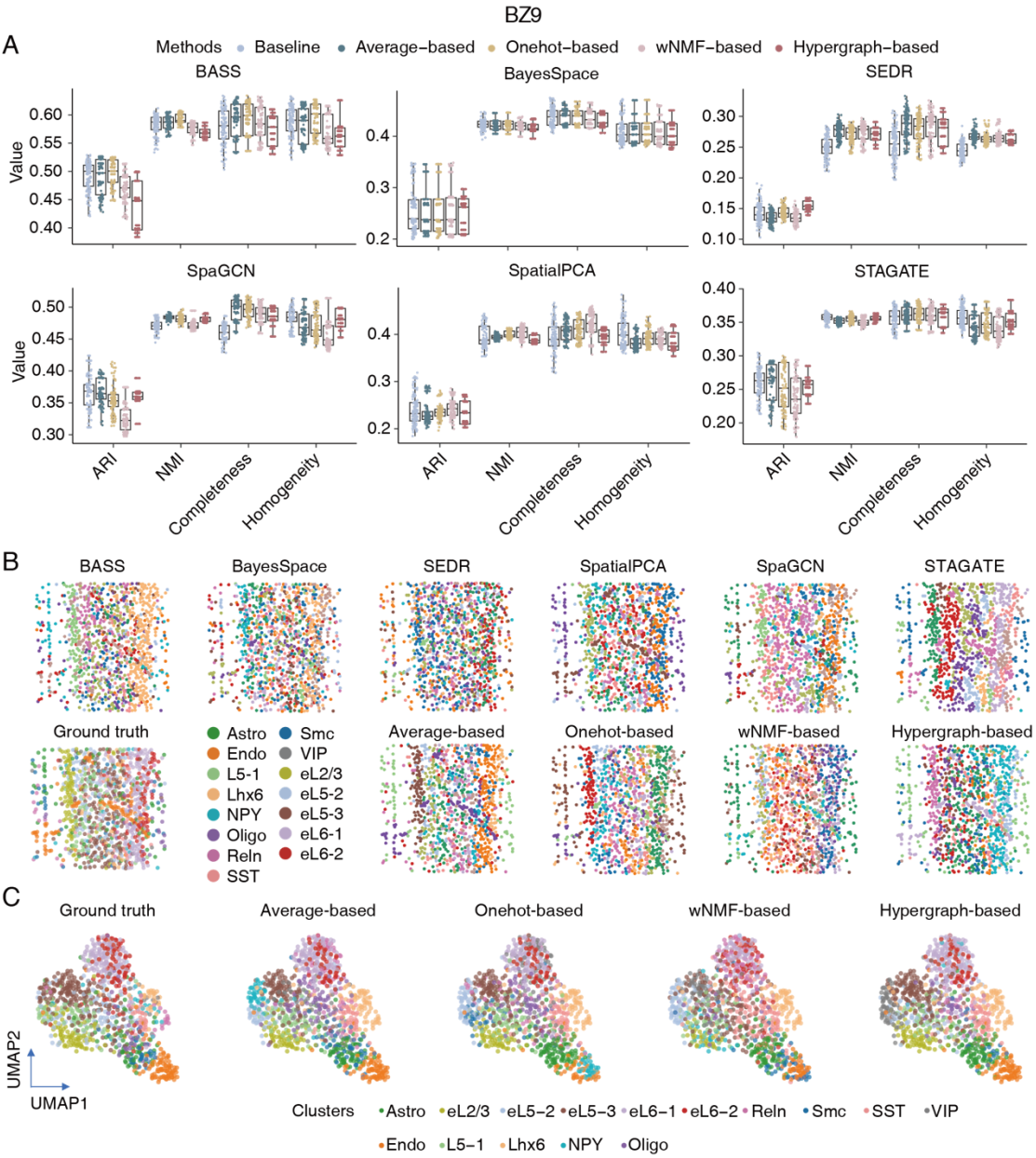

Figure S20. Performance of consensus clustering strategies on the Mouse Cortex data. (*A*) Consensus clustering accuracy of the BZ9 sample of the Mouse Cortex dataset based on a single baseline algorithm. Each baseline algorithm and consensus strategy were repeated 10 times. (*B*) Visualization of clustering results of the baseline algorithm (top) and 4 consensus strategies (bottom right) on the BZ9 sample of Mouse Cortex data. (*C*) UMAP visualization comparing ground truth cell type labels (leftmost) with clustering results from 4 consensus strategies.

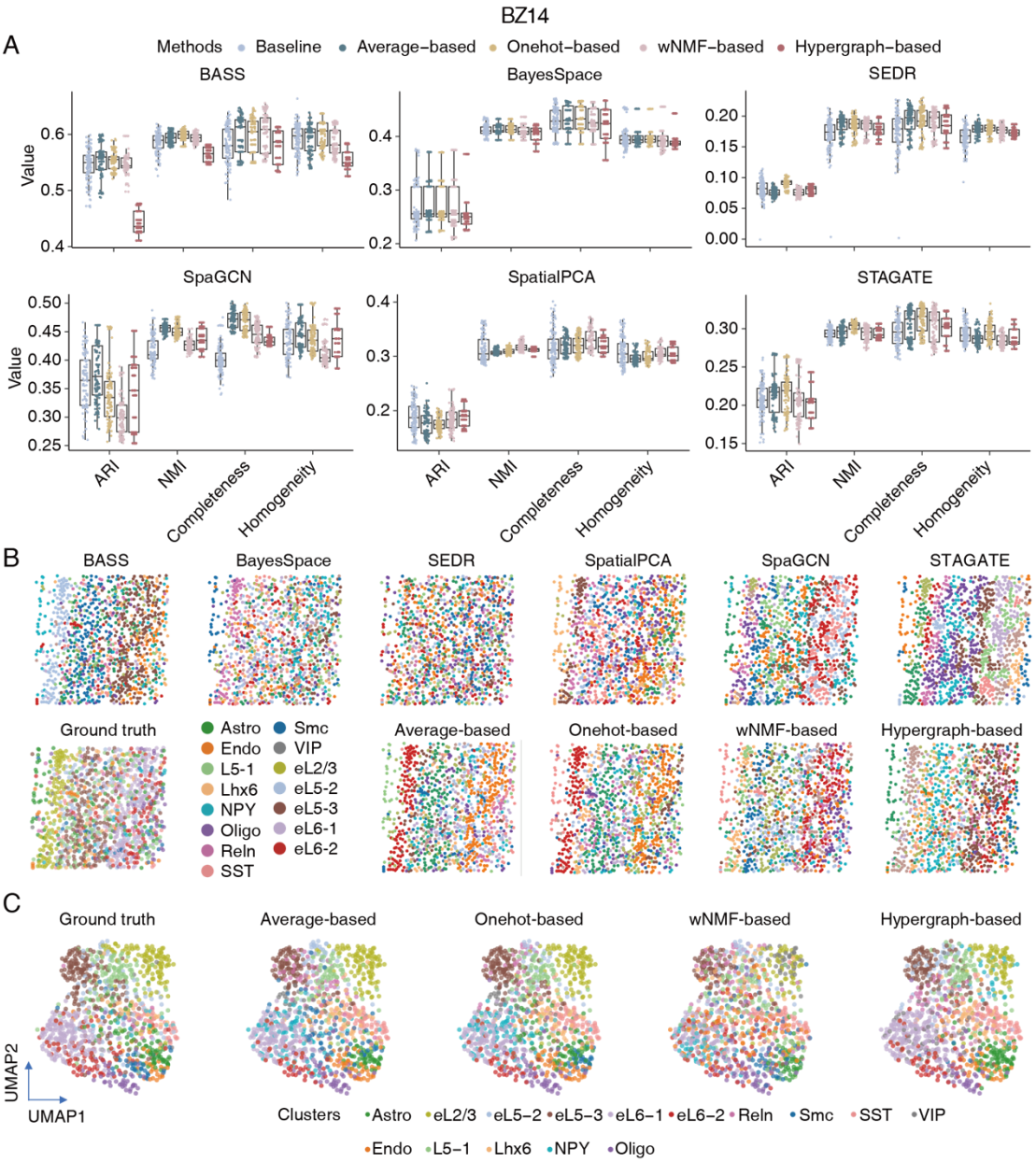

Figure S21. Performance of consensus clustering strategies on the Mouse Cortex data. (*A*) Consensus clustering accuracy of the BZ14 sample of the Mouse Cortex dataset based on a single baseline algorithm. Each baseline algorithm and consensus strategy were repeated 10 times. (*B*) Visualization of clustering results of the baseline algorithm (top) and 4 consensus strategies (bottom right) on the BZ14 sample of Mouse Cortex data. (*C*) UMAP visualization comparing ground truth cell type labels (leftmost) with clustering results from 4 consensus strategies.