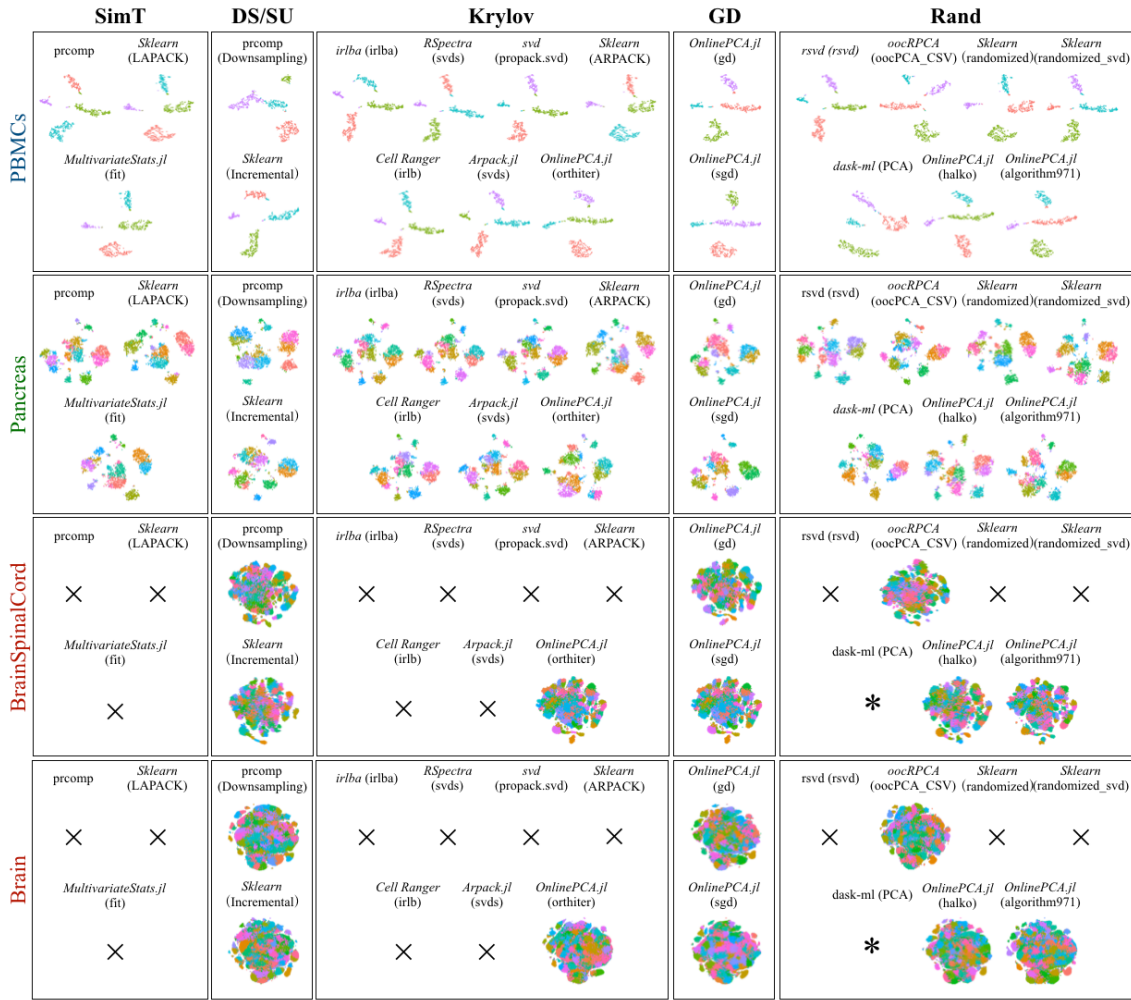


Results of clustering methods for all PCA implementations

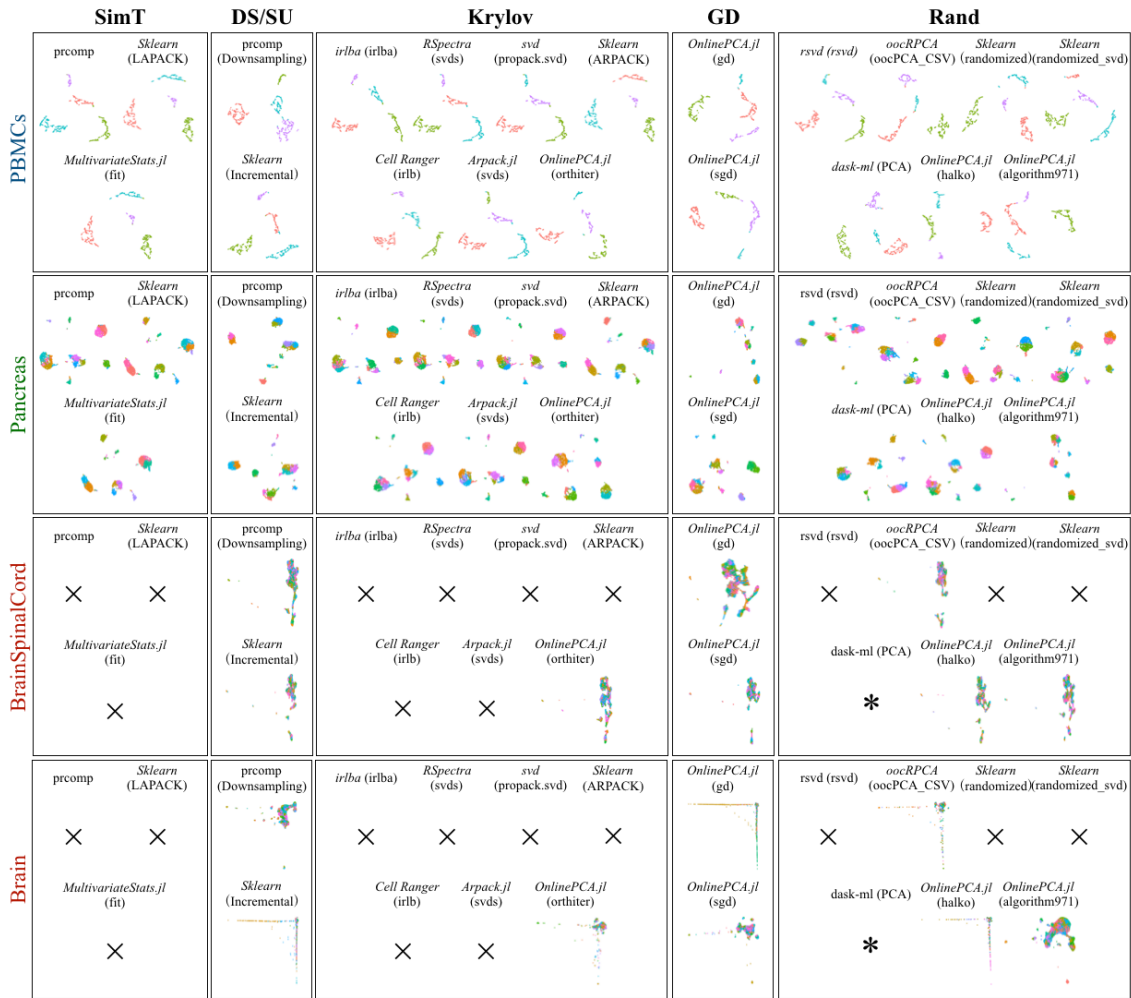
Here, we present the results of clustering performed for all the PCA implementations mentioned in the manuscript: *k*-means, Gaussian mixture model (GMM), HDBScan, and Louvain clustering. When performing HDBScan, principal components (PCs) were first converted to five dimensional UMAP coordinates, and then HDBScan was performed (**Figure S6-1 to 7**). When performing Louvain clustering, the Jaccard index was used as the similarity measure with the *k* nearest neighborhood cells. For the PBMCs, Pancreas, BrainSpinalCord, and Brain datasets, *k* values of 100, 75, 10, and 5 were used, respectively.

The cluster assignment shown in **Figure S6-1 to 7** and the corresponding adjusted Rand index (ARI) values (**Figure S6-8**) show that the Louvain clustering results were stable across all datasets. Therefore, we finally selected Louvain clustering as the measure for benchmarking.



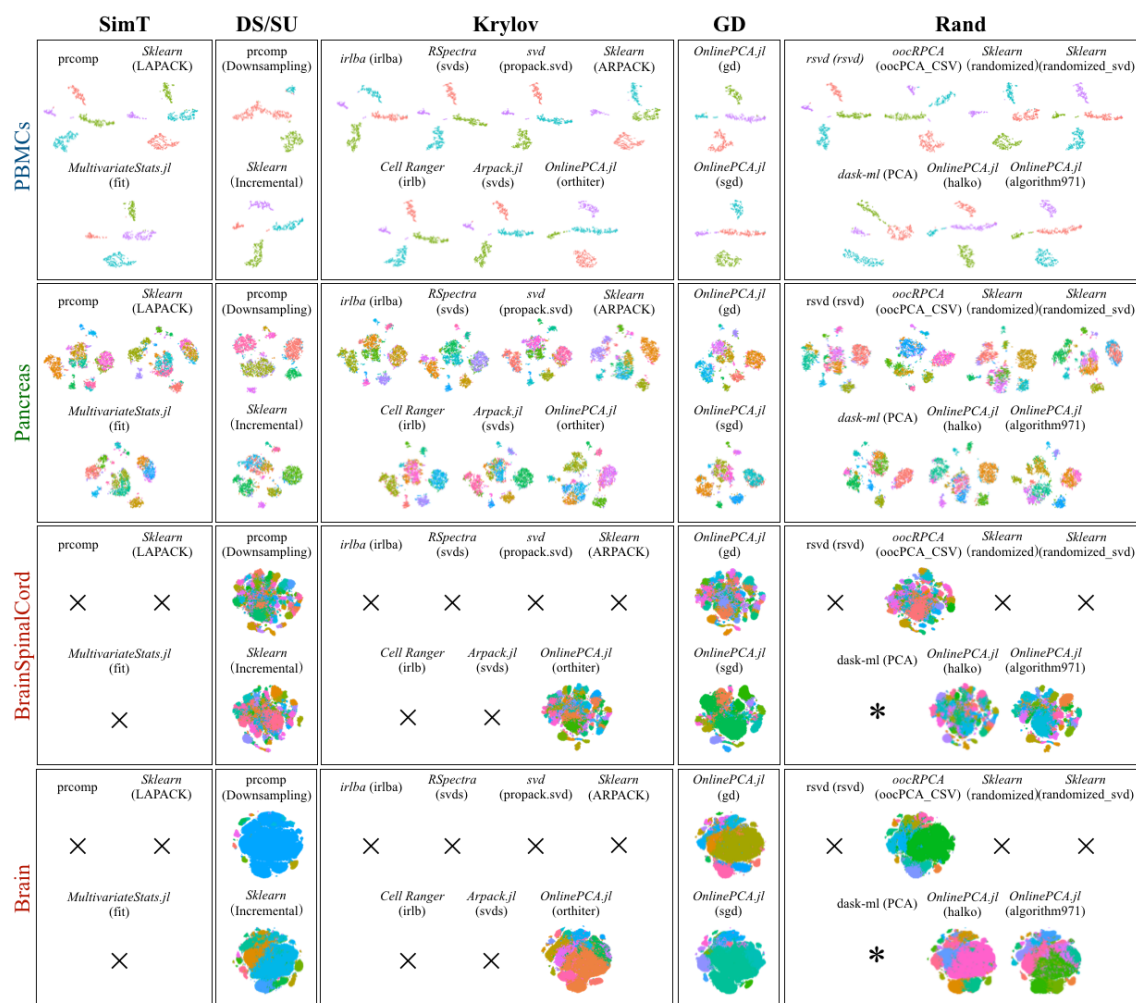
×: Out-of-memory
 *: Could not be finished within three days

Figure S6-1 | Results of k -means clustering The cluster label is visualized with two-dimensional t-SNE coordinates.



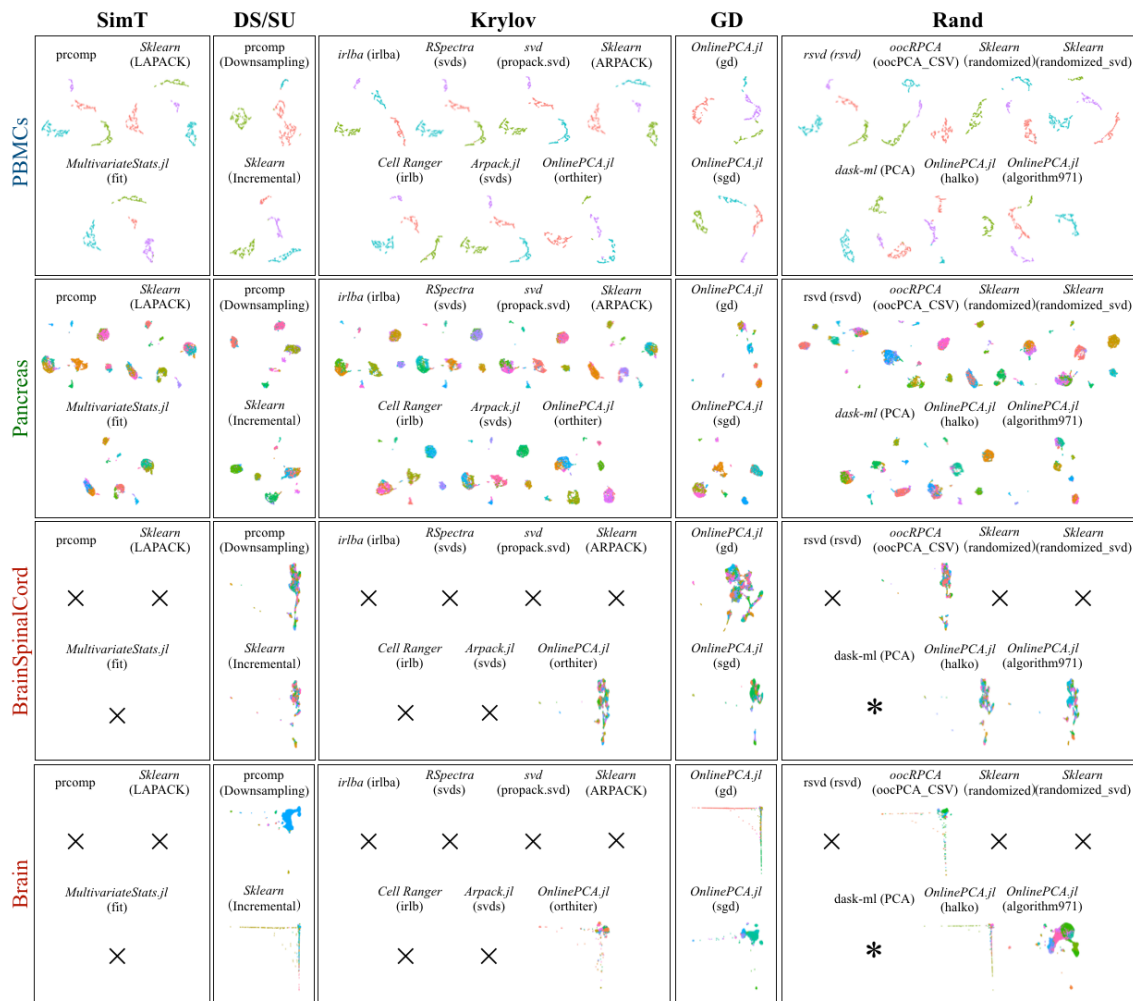
×: Out-of-memory
 *: Could not be finished within three days

Figure S6-2 | Results of k -means clustering The cluster label is visualized with two-dimensional UMAP coordinates.



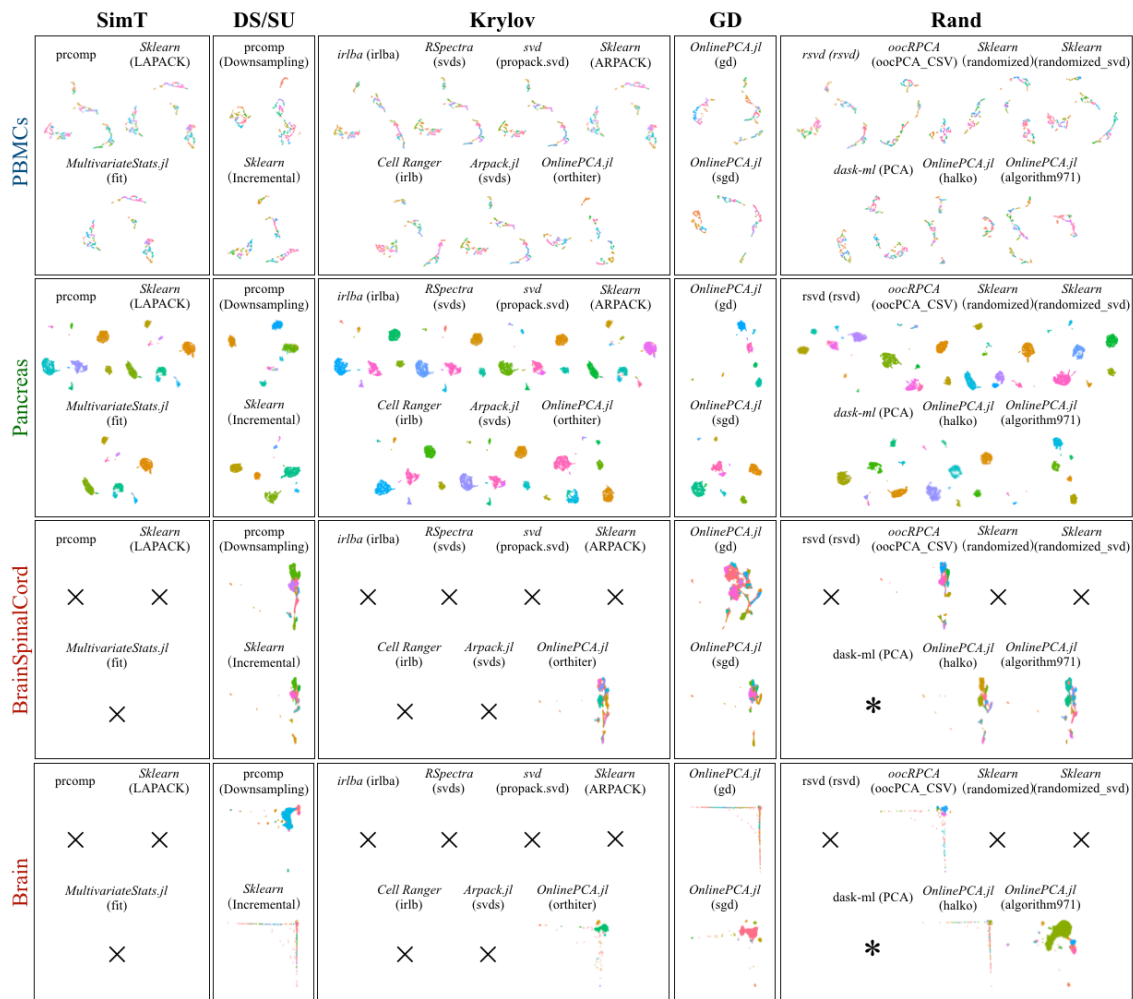
×: Out-of-memory
 *: Could not be finished within three days

Figure S6-3 | Results of GMM clustering The cluster label is visualized with two-dimensional t-SNE coordinates.



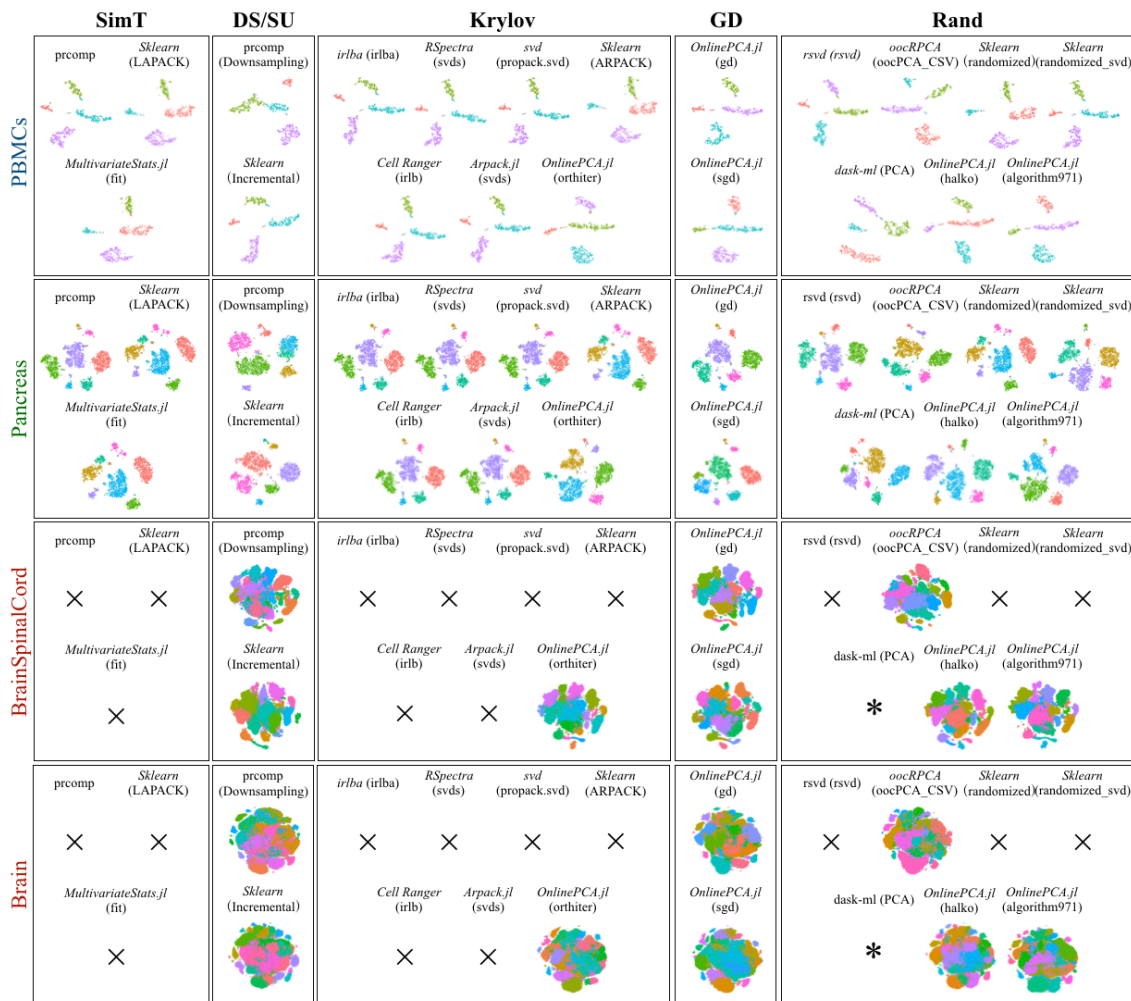
×: Out-of-memory
 *: Could not be finished within three days

Figure S6-4 | Results of GMM clustering The cluster label is visualized with two-dimensional UMAP coordinates.



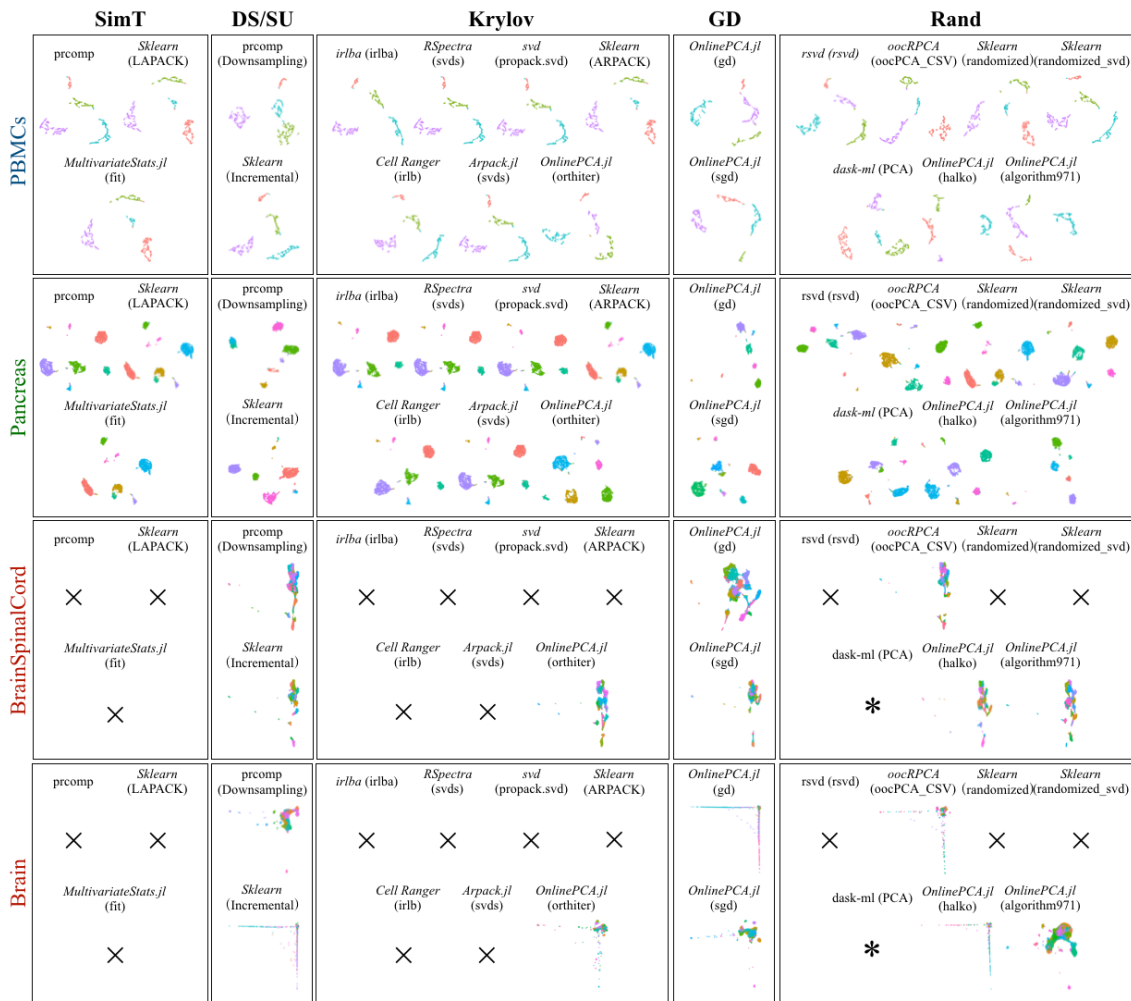
×: Out-of-memory
 *: Could not be finished within three days

Figure S6-5 | Results of HDBScan clustering The cluster label is visualized with two-dimensional UMAP coordinates.



×: Out-of-memory
 *: Could not be finished within three days

Figure S6-6 | Results of Louvain clustering The cluster label is visualized with two-dimensional t-SNE coordinates.



×: Out-of-memory
 *: Could not be finished within three days

Figure S6-7 | Results of Louvain clustering The cluster label is visualized with two-dimensional UMAP coordinates.

