Table S1. Summary of single-cell datasets used in this manuscript.

Dataset	Single-cell technique	Organism	Tissues	Samples	Number of Single-cells	Dimensionality for R_1 and R_2	Dimensionality of Input to NLDR	Identifier/Source
Samusik_01	Mass-cytometry	Mouse	Bone marrow	1	53173 (excluding ungated cells)	38	38	https://figshare.com/s/9c3a0136f12b97f1dadd (from ref[1])
PBMC3k	Single-cell RNAseq, 10x	Human	Periferal blood	1	2638	13714	50 PCs	https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc3k
Adipose_tissue	Single-cell RNAseq, 10x	mouse	Adipose tissue	Pooled sample	11420	17327	50 PCs	GEO: GSM3717977
Gastric_cancer	Single-cell RNAseq, 10x	Human	Gastric antral mucosae	5 (4 patients)	10638 (use sample IDs CAG1,EGC,IMS3,IMW1,NAG3 to reduce data size)	22763	50 PCs	GEO: GSE134520

PCs, principal components after PCA.

Table S2. Comparison of geodesic and Euclidean distances in high dimensional space for calculating R_1.

Dataset	Distance Type	Panoramap	t-SNE	UMAP
PBMC3k	Geodesic	0.8079	0.7118	0.7122
	Euclidean	0.6091	0.4659	0.4419
Samusik_01	Geodesic	0.9003	0.5854	0.7191
	Euclidean	0.91	0.5358	0.69
Adipose_tissue	Geodesic	0.7990	0.5249	0.7447
	Euclidean	0.7766	0.4601	0.6701
Gastric_cancer	Geodesic	0.6172	0.4449	0.5836
	Euclidean	0.5979	0.4191	0.5574

R_1 calculated using geodesic distance or Euclidean distance in high dimensional space are generally similar. Panoramap performs better than t-SNE and UMAP with both geodesic distances and Euclidean distances. Representatives from one experiment are shown in the table. Distance Type denotes the distance metric used in high dimensional space. Euclidean distances are used in all embedding spaces.

Figure S1. Panoramap better reveals cellular subtypes

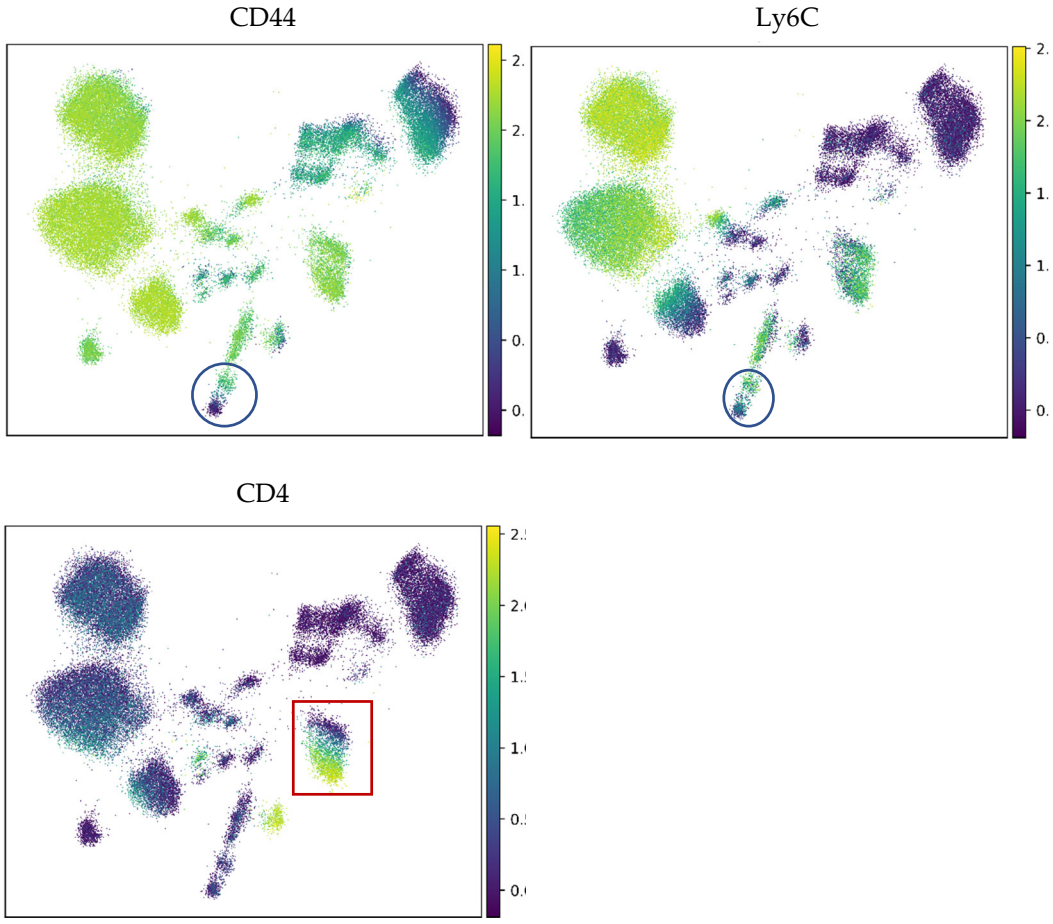


Figure S1. Panoramap embeddings of Samusik01 dataset show that CD8 T cells (within the blue circles) can be divided into two subgroups by CD44 and Ly6C. And pDCs (within the red rectangle) can be divided into two subgroups by CD4.

Figure S2. Panoramap preserves global structure best compared with t-SNE and UMAP in terms of R_1.

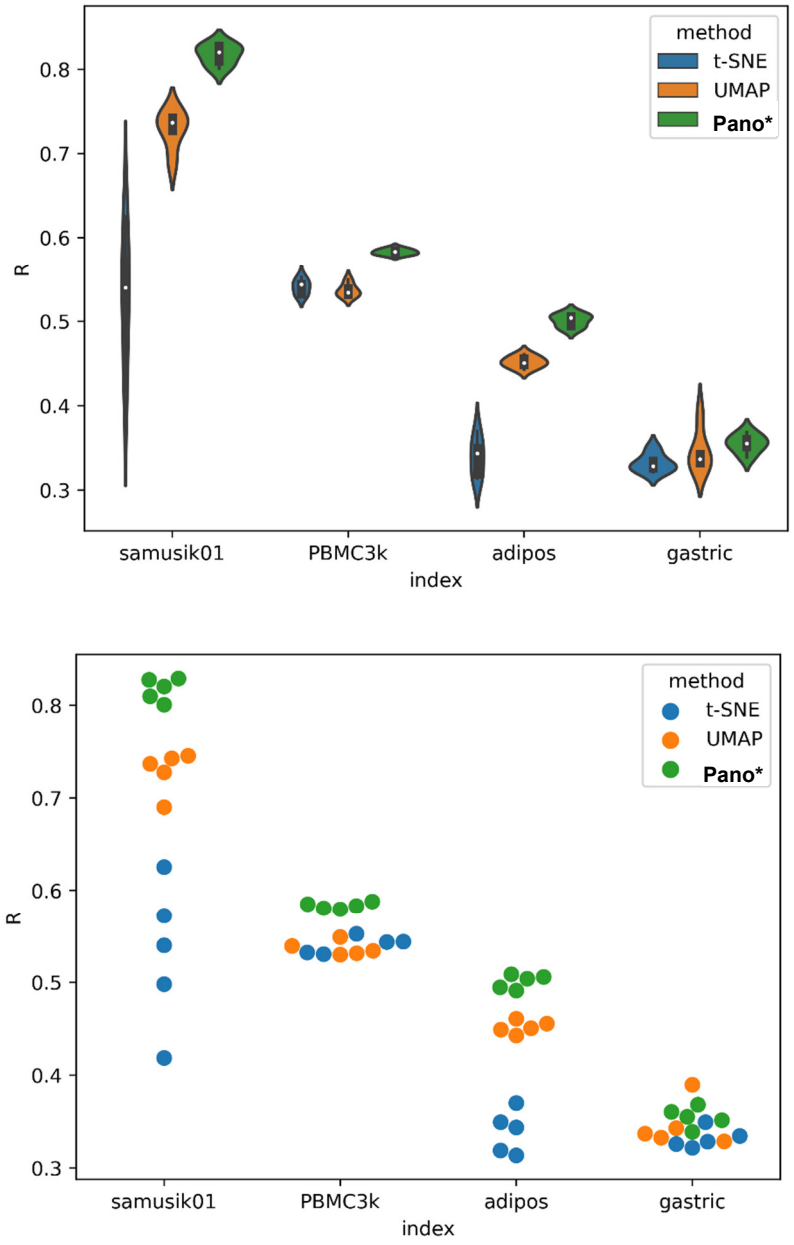
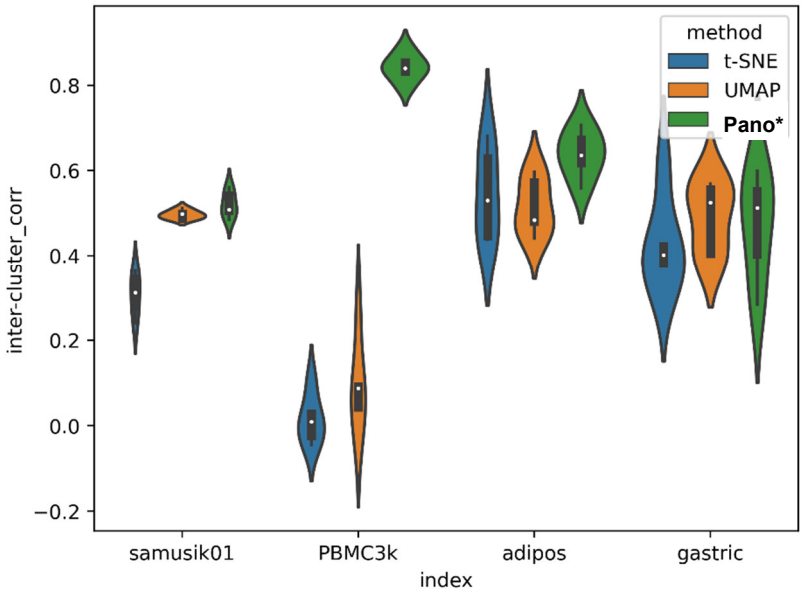


Figure S2. Violinplot and scatterplot show comparison of Global structure preservation R_1 in Samusik01, PBMC3k, adipose_tissue, gastric_cancer datasets using t-SNE, UMAP and Panoramap. Experiments were repeated for 5 times with 5 random seeds. Pano* denotes Panoramap.

Figure S3. Panoramap preserves global structure best compared with t-SNE and UMAP in terms of R_2.



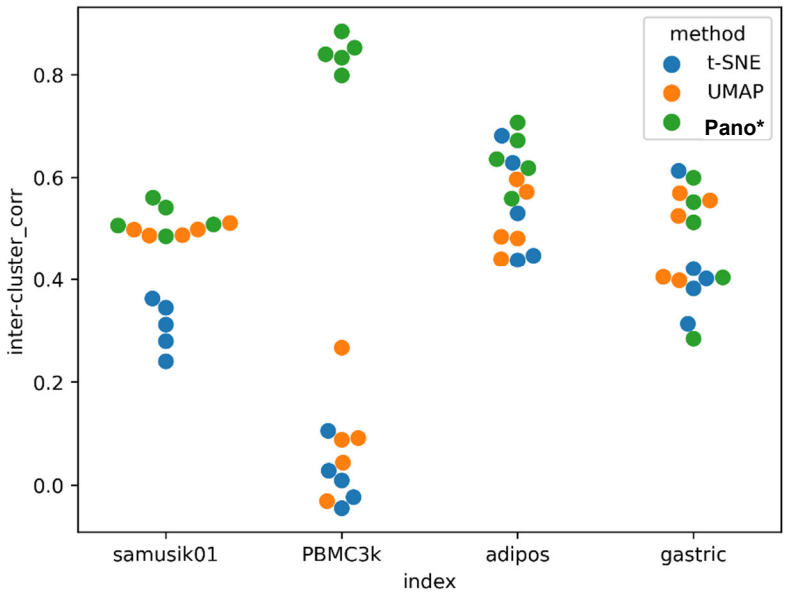


Figure S3. Violinplot and scatterplot show comparison of Inter-cluster distance preservation R2 in Samusik01, PBMC3k, adipose_tissue, gastric_cancer datasets using t-SNE, UMAP and Panoramap. Experiments were repeated for 5 times with 5 random seeds. Pano* denotes Panoramap.

Figure S4. Panoramap performs better than t-SNE and UMAP even when t-SNE and UMAP have optimized initialization

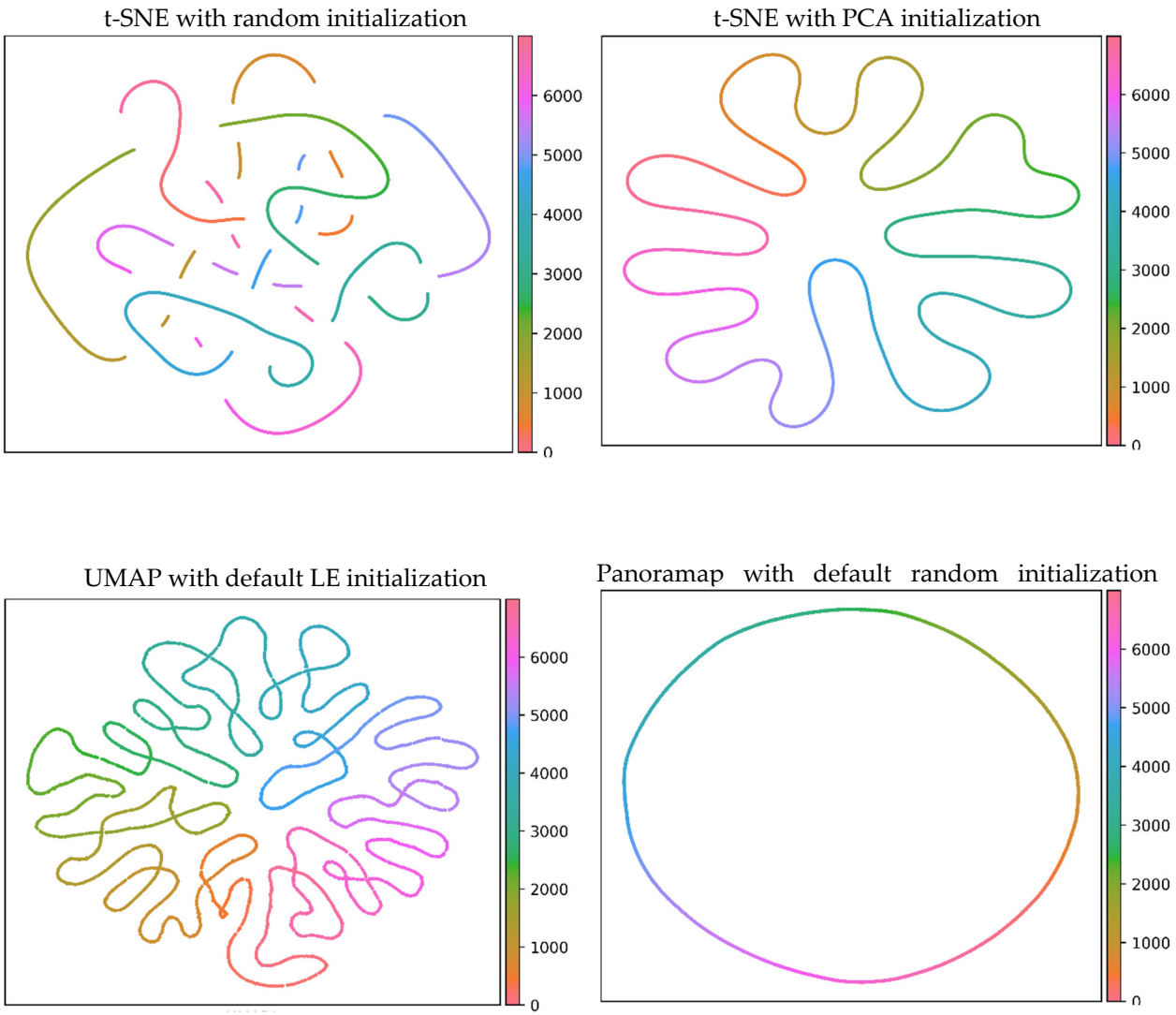


Figure S4. Kobak et al [2]proposed that initialization is critical for preserving global data structure in both t-SNE and UMAP. We tested the dataset (https://github.com/dkobak/tsne-umap-init/blob/master/tsne-umap-circle.ipynb) provided with the paper. The original data is a colorful circle. The t-SNE algorithm uses random initialization in default and the circle's topological structure is violated after t-SNE processing. T-SNE with PCA initialization and UMAP with the default Laplacian eigenmaps (LE) initialization preserve the topology, but give artificial structures like the curves. Panoramap uses random initialization and preserves both topological and global geometrical structures.

Figure S5. Panoramap has the shortest test run time.

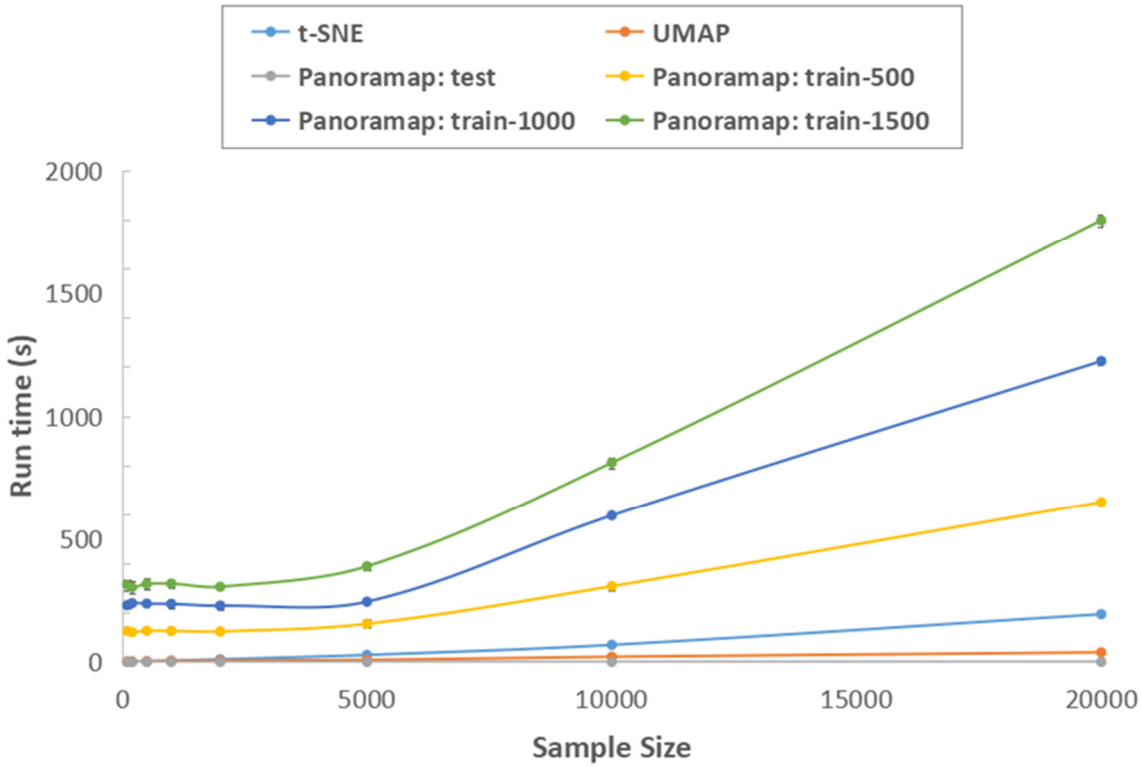


Figure S5. Panoramap has the shortest test run time compared with t-SNE and UMAP. The test run time is less than 0.1 second for all the experiments tested. The training time for Panoramap is dependent on the number of epochs and samples, e.g. about 600 seconds for 500 epochs and 20,000 samples. The time consumed for training Panoramap is comparable to other neural network algorithm and is about 10 times that of UMAP[3]. The times for t-SNE and UMAP are obtained using single core CPU, while times for Panoramap are calculated on one single V100 GPU. The run time experiments were done for 3 times with different random seeds.

1. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IW, Ng LG, Ginhoux F, Newell EW: **Dimensionality reduction for visualizing single-cell data using UMAP.** *Nature biotechnology* 2019, **37:**38-44.
2. Kobak D, Linderman GC: **Initialization is critical for preserving global data structure in both t-SNE and UMAP.** *Nature Biotechnology* 2021, **39:**156-157.
3. Tim Sainburg LM, Timothy Q Gentner: **Parametric UMAP: learning embeddings with deep neural networks for representation and semi-supervised learning.** *arXiv:200912981 [csLG]* 2020.