Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/44/eabd0855/DC1

# Supplementary Materials for

## Learning for single-cell assignment

Bin Duan, Chenyu Zhu, Guohui Chuai, Chen Tang, Xiaohan Chen, Shaoqi Chen, Shaliu Fu, Gaoyang Li, Qi Liu*

*Corresponding author. Email: qiliu@tongji.edu.cn

Published 30 October 2020, *Sci. Adv.* **6**, eabd0855 (2020)
DOI: 10.1126/sciadv.abd0855

**The PDF file includes:**

Figs. S1 to S4
Legends for tables S1 to S10

**Other Supplementary Material for this manuscript includes the following:**

(available at advances.sciencemag.org/cgi/content/full/6/44/eabd0855/DC1)

Tables S1 to S10

# Supplementary Material

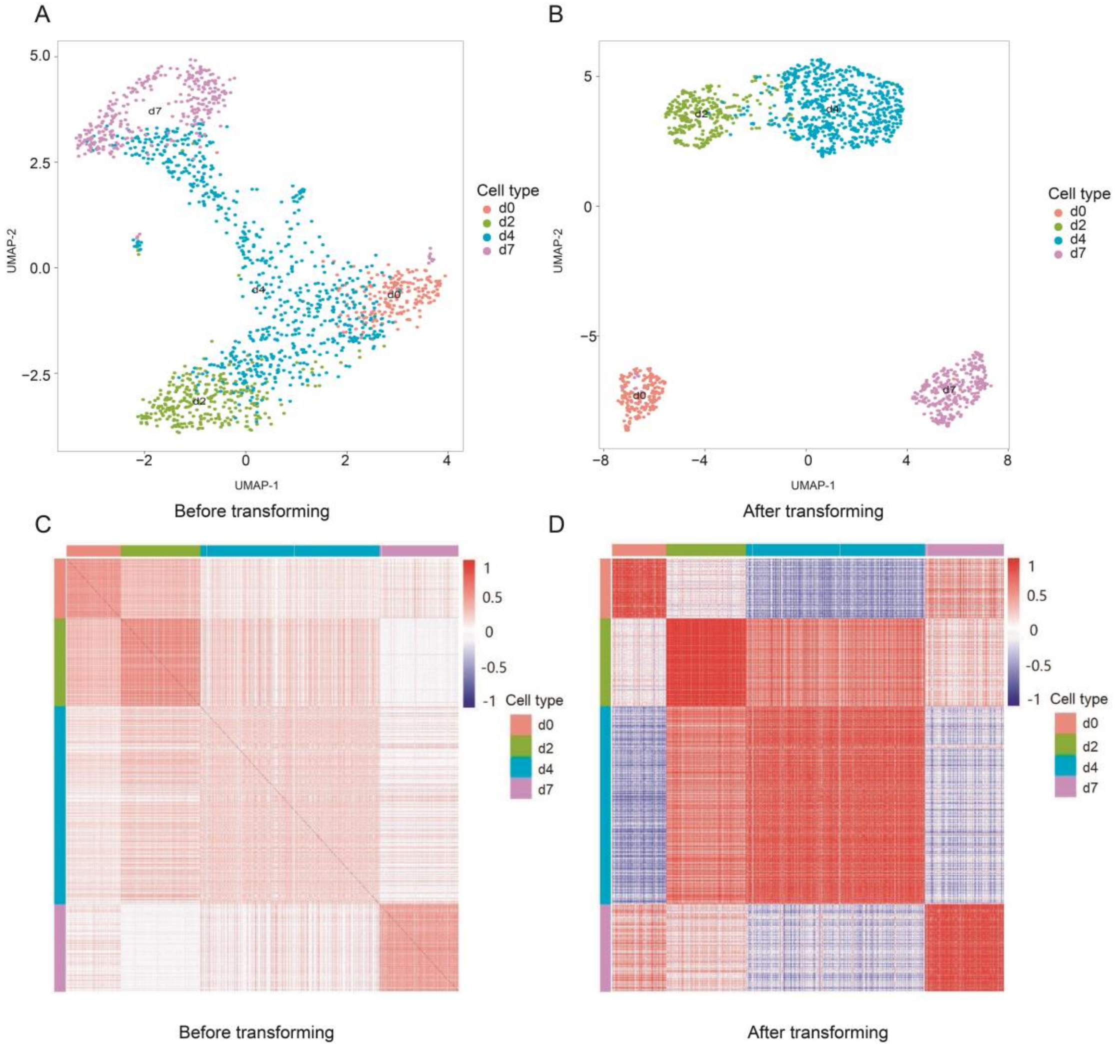
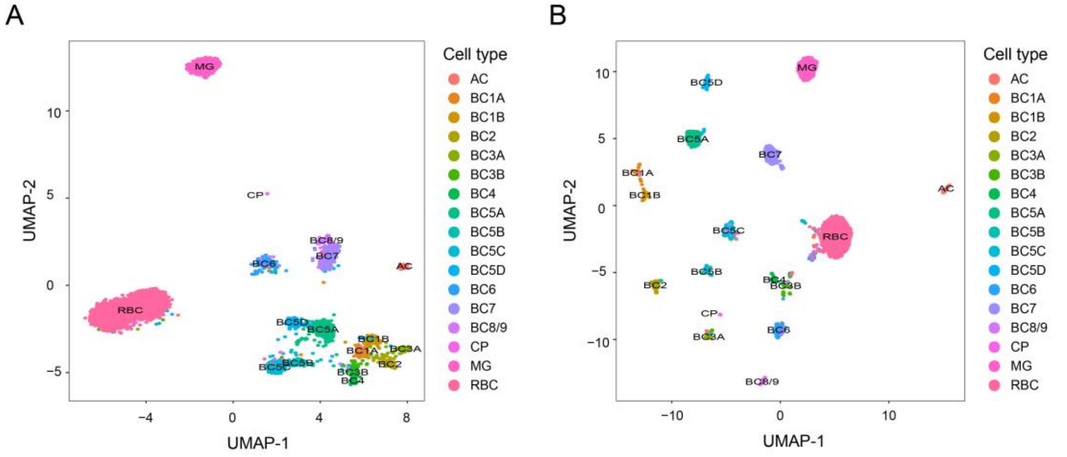
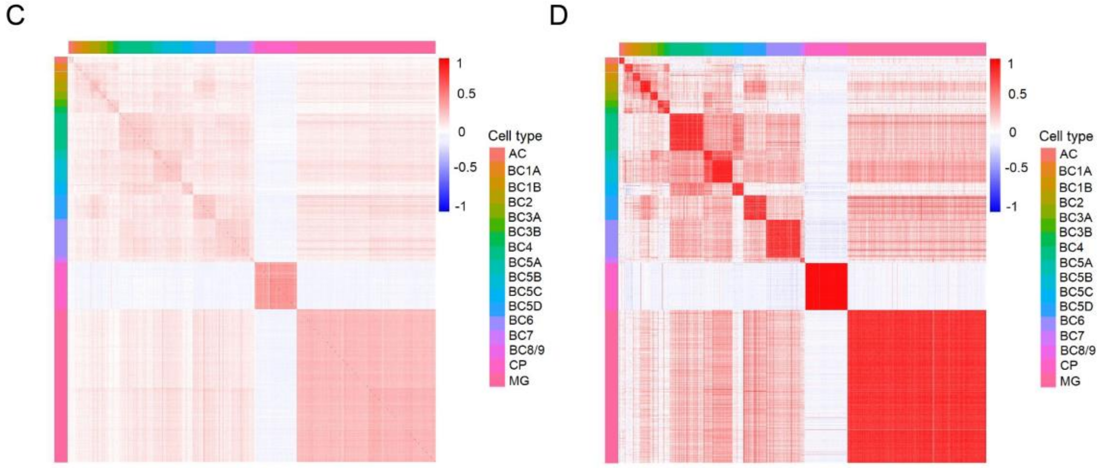


**Fig. S1 Comparison of the clustering results before and after transforming by scLearn for Klein (Mouse embryo stem cells) dataset. (A)** The visualization of clustering results by UMAP before DCA-based transforming. **(B)** The visualization of clustering results by UMAP after DCA-based transforming. **(C)** Similarity heatmap calculated with the Pearson correlation coefficient before transforming. **(D)** Similarity heatmap calculated with the Pearson correlation coefficient after transforming with scLearn.



Before transforming

After transforming



Before transforming

After transforming

**Fig. S2 Comparison of the clustering results before and after transforming by scLearn for Shekhar (Mouse retina) dataset. (A)** The visualization of clustering results by UMAP before DCA-based transforming. **(B)** The visualization of clustering results by UMAP after DCA-based transforming. **(C)** Similarity heatmap calculated with the Pearson correlation coefficient before transforming. **(D)** Similarity heatmap calculated with the Pearson correlation coefficient after transforming with scLearn.

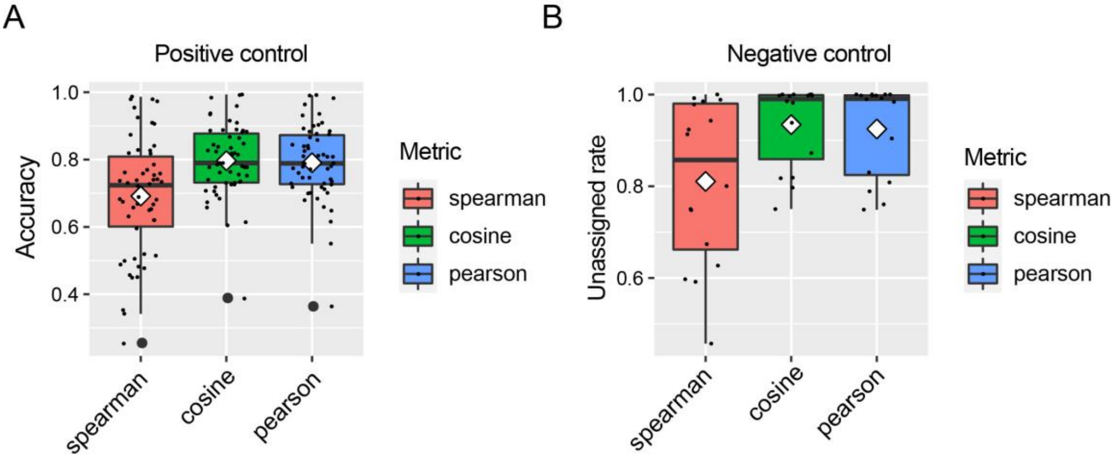


**Fig. S3 The comparison of different measures used after DCA. (A)** The comparison in positive control scenario for spearman, cosine and Pearson. The white diamond line represents the mean value. **(B)** The comparison in negative control scenario for spearman, cosine and Pearson. The white diamond line represents the mean value.

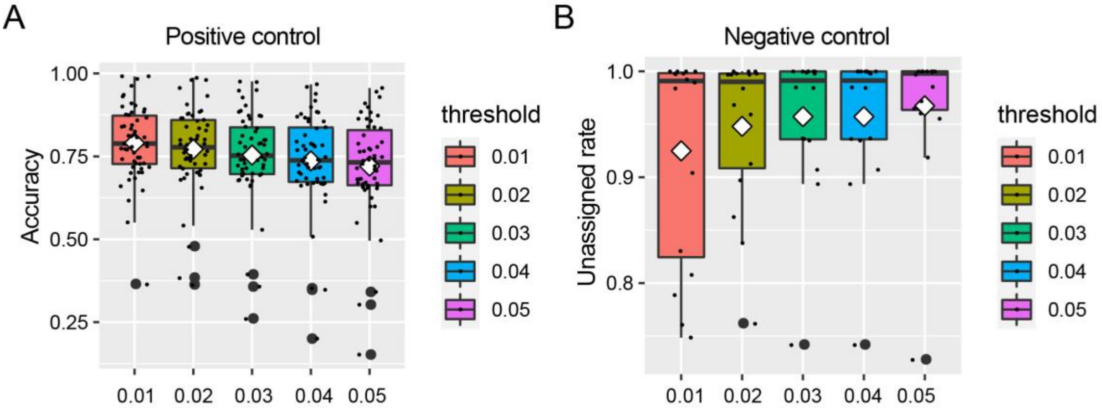


**Fig. S4 The comparison of different thresholds to determine unassigned cells**. **(A)** The comparison in positive control scenario. The white diamond line represents the mean value. **(B)** The comparison in negative control scenario. The white diamond line represents the mean value.

## Legends for Tables S1 to S10:

**Table S1** The test datasets for different evaluation scenarios in single cell type assignment

**Table S2** The intra-cluster compactness of 30 datasets before and after transforming.

**Table S3** The inter-cluster complexity of 30 datasets before and after transforming.

**Table S4** Benchmark results of positive control scenario. The accuracy is calculated as the proportion of correctly predicted cells among all query cells.

**Table S5** Benchmark results of negative control scenario. The unassigned rate is calculated as the proportion of predicted "unassigned" cells among all query cells.

**Table S6** A comprehensive benchmark result for both positive control scenario and negative control scenario.

**Table S7** The performance of confusion matrix for all the tested methods in real application scenario.

**Table S8** The accuracy and specificity for all the tested methods in real application scenario. The accuracy is calculated as the proportion of correctly predicted cells among common cell types. The specificity is calculated as the proportion of correctly predicted "unassigned" cells among all predicted "unassigned" cells.

**Table S9** The comparison of MDDM-based scLearn and DCA-based scLearn on three multi-label datasets.

**Table S10** The information of all pre-trained models for all tested datasets.