DECODE: Deep learning-based common deconvolution framework for various omics data

Tianyi Zhao[1,2,#], Renjie Liu[2,3,1], Yuzhi Sun[3,1], Liyuan Zhang[3], Ruibang Luo[4], Liang Cheng[5,*], Yadong Wang[3,*]

[1] School of Medicine and Health, Harbin Institute of Technology, Harbin, China
[2] Harbin Institute of Technology Zhengzhou Research Institute, Zhengzhou, China
[#] These authors contributed equally: Tianyi Zhao, Renjie Liu, Yuzhi Sun.
[3] Faculty of Computing, Harbin Institute of Technology, Harbin, China
[4] Department of Computer Science, The University of Hong Kong, Hong Kong, China
[5] Harbin Medical University, Harbin, China
[*] These authors jointly supervised this work: Liang Cheng (liangcheng@hrbmu.edu.cn), Yadong Wang (ydwang@hit.edu.cn)

Supplementary

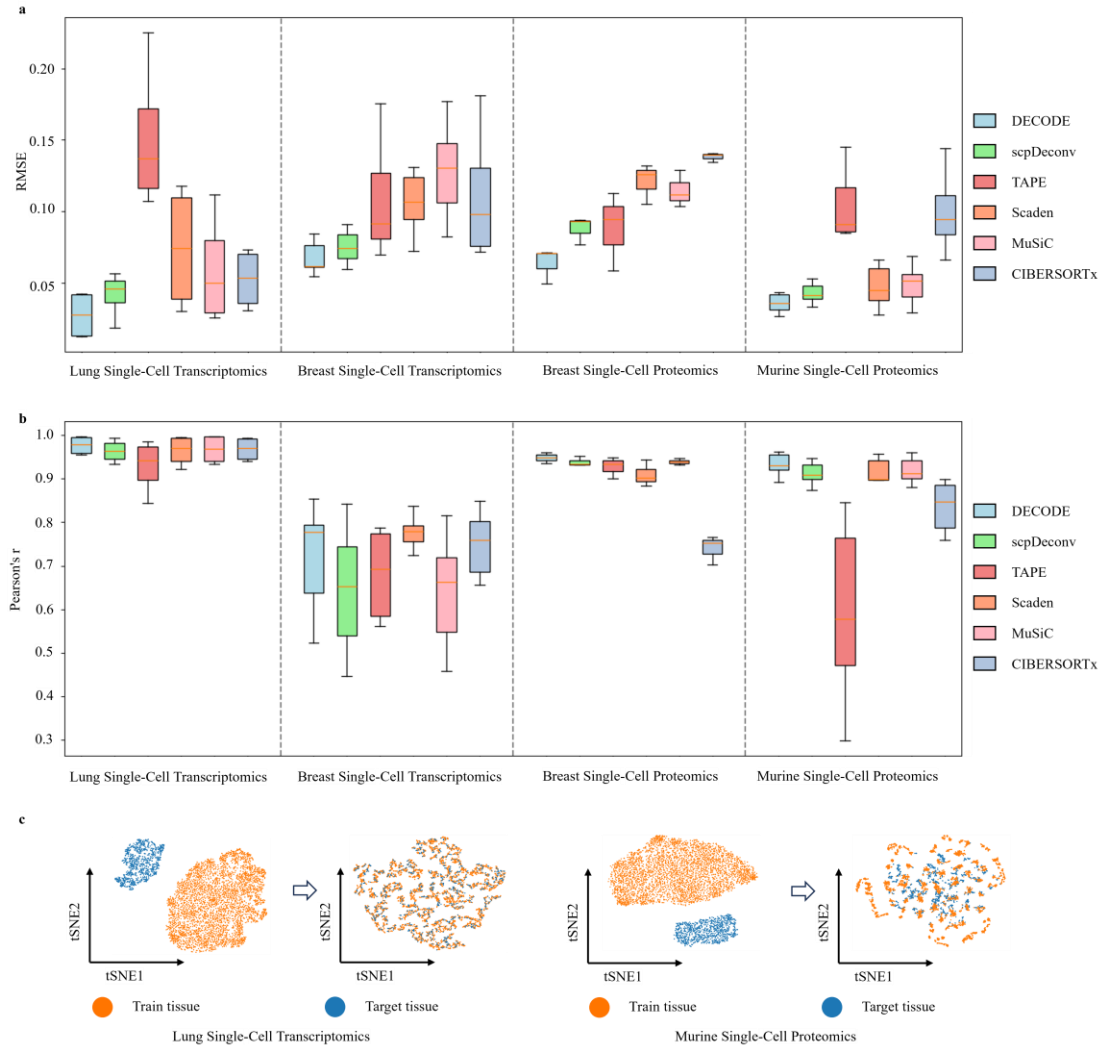

Fig. S1| Overview of DECODE's performance testing results for transcriptomics and proteomics data. (a-b) RMSE (a) and Pearson's r (b) results for DECODE, scpDeconv, TAPE, Scaden, MuSiC, and CIBERSORTx across the Lung Single-Cell Transcriptomics, Breast Single-Cell Transcriptomics, Breast Single-Cell Proteomics, and Murine Single-Cell Proteomics datasets. The RMSE values range from 0 and above, with smaller values indicating better performance. Pearson's r values range between -1 and 1, with values closer to 1 indicating better performance. (c) t-SNE plots of features processed with or without encoder blocks in Stage 2 (the orange and blue dots represent the training and testing datasets, respectively). The left and right pairs of plots present the results using the Lung Single-Cell Transcriptomics dataset and Murine Single-Cell Proteomics dataset, respectively. In each pair, the left and right sides present the t-SNE plot of the original omics features and the t-SNE plot after computation with the Stage 2 encoder.

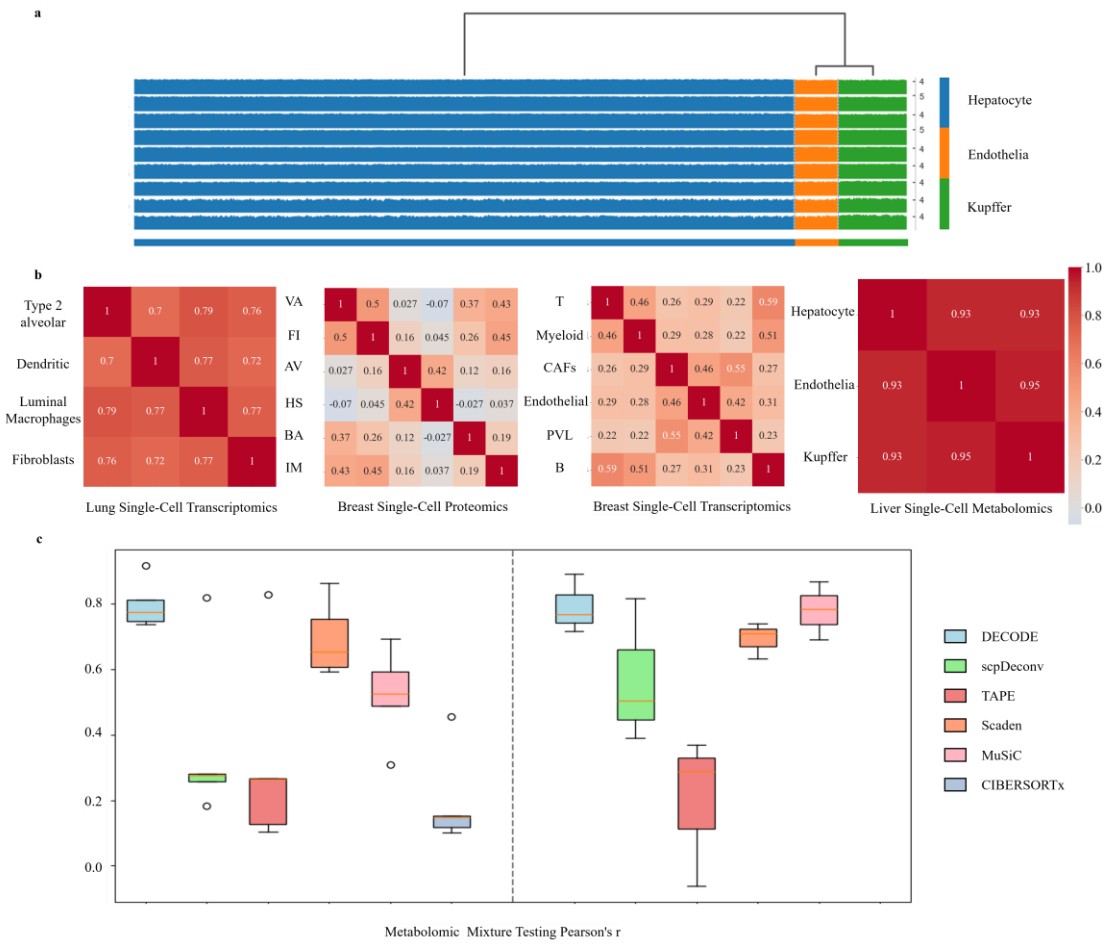

Fig. S2| Overview of DECODE's performance for metabolomics test data. (a) Abundance of differentially abundant metabolites among different cell types in the liver dataset. The vertical axis represents the abundance of metabolites, with each row corresponding to a metabolite. The colour scale on the right side represents differentially abundant metabolites of different cell types. (b) Heatmap of the cell type Kendall similarity of the Lung Transcriptomics, Breast Proteomics, Breast Transcriptomics and Liver Metabolomics datasets. (c) Pearson's r values for DECODE, scpDeconv, TAPE, Scaden, MuSiC, and CIBERSORTx using the bone marrow (left) and liver (right) datasets.

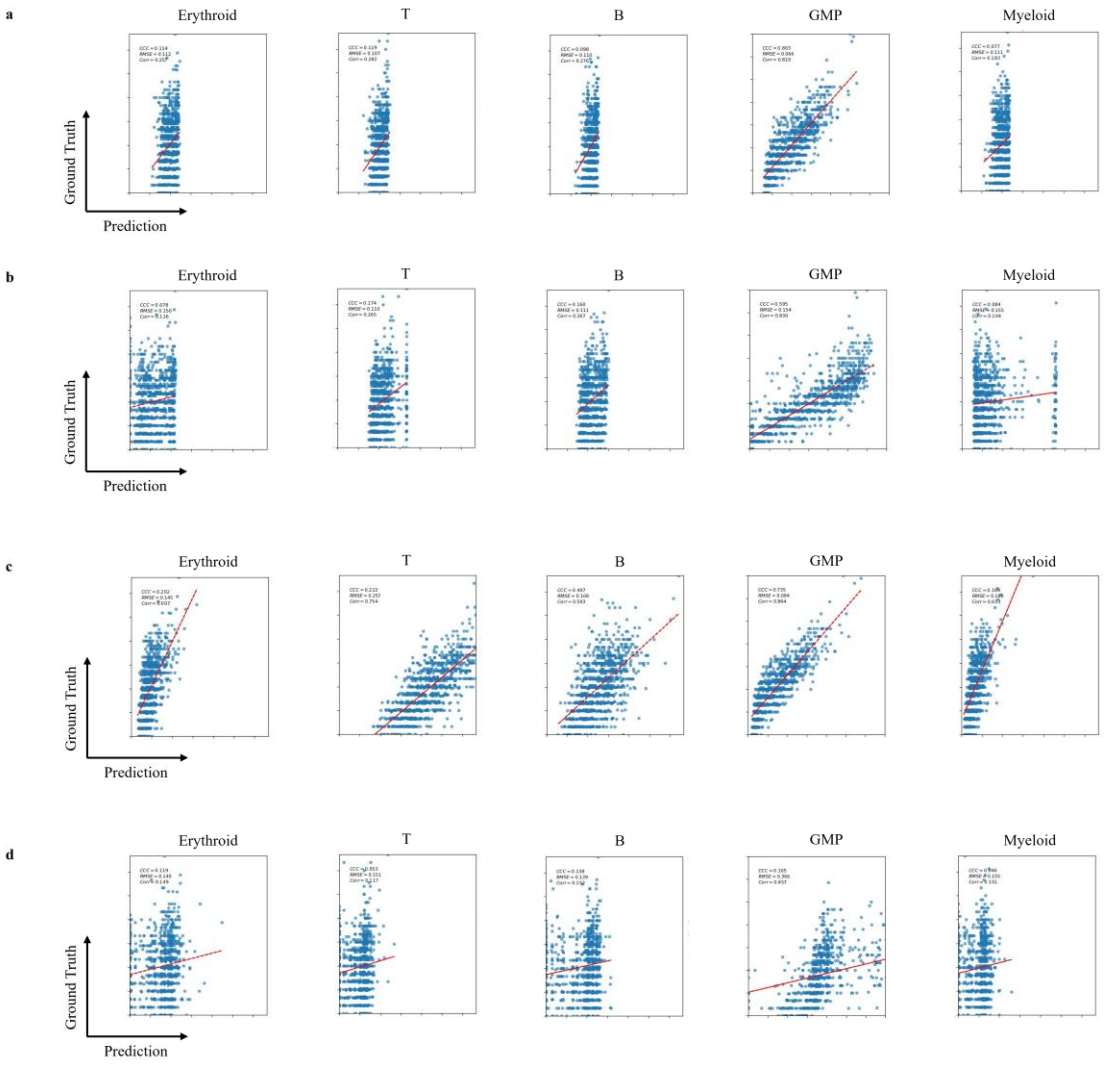

Fig. S3| Scatter plots of true and predicted values for different methods. (a-d) The scatter plots show the predicted proportions (y-axis) and the true proportions (x-axis) of different cell types in the bone marrow metabolomics dataset using scpDeconv (a), TAPE (b), Scaden (c), and CIBERSORTx (d).

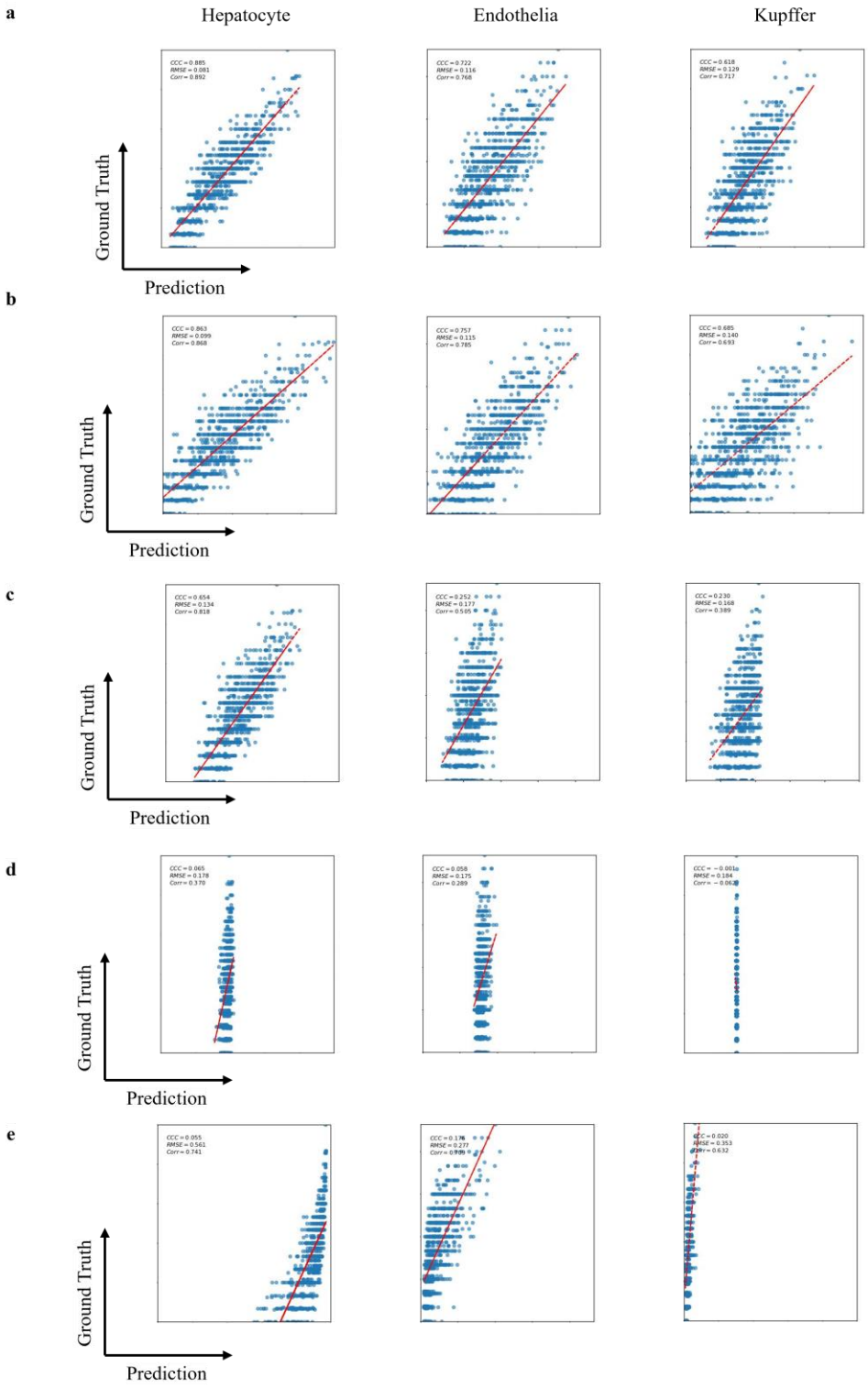

Fig. S4| Scatter plots of true and predicted values for different methods. (a-d) The scatter plots show the predicted proportions (y-axis) and the true proportions (x-axis) of different cell types in the liver metabolomics dataset using DECODE (a), MuSiC (b), scpDeconv (c), TAPE (d), and Scaden (e).

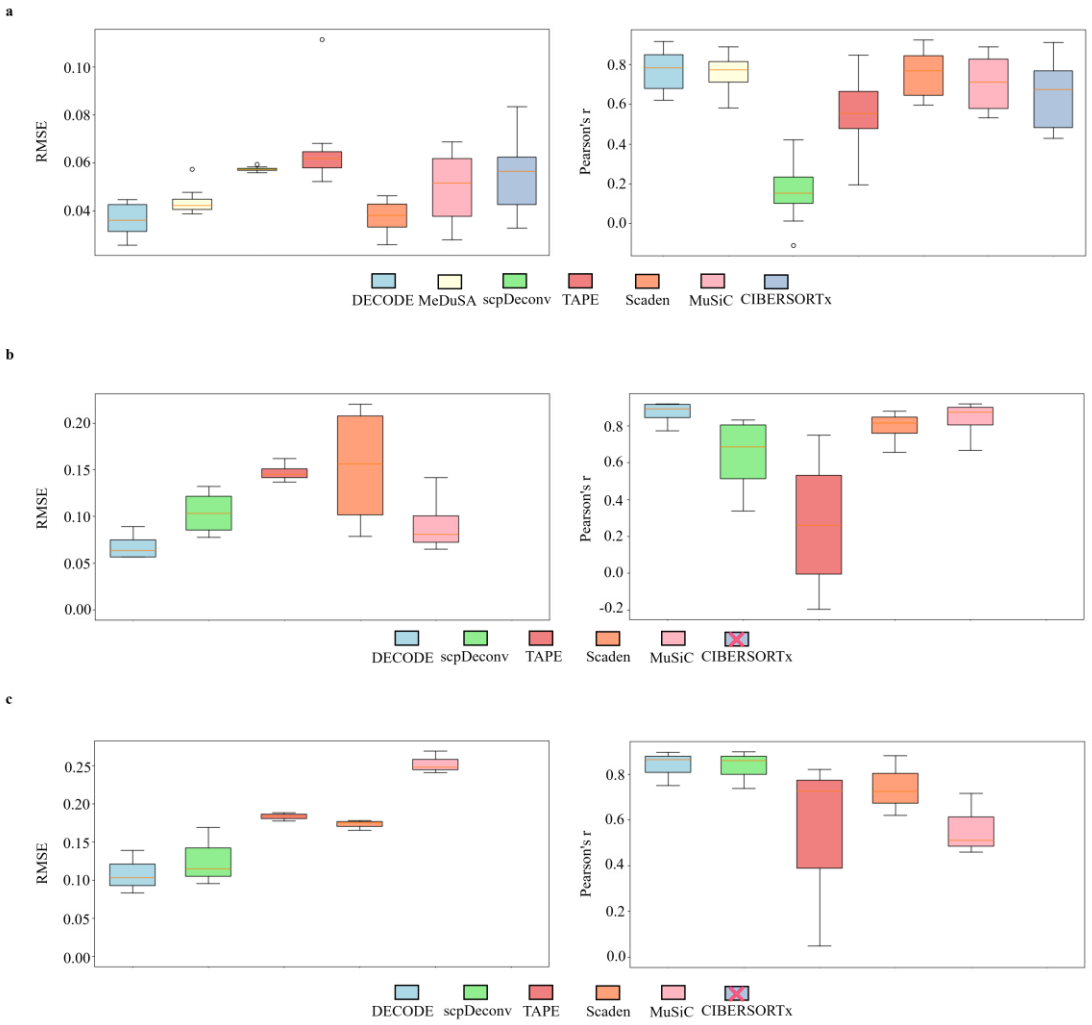

Fig. S5| Multiomics cell state deconvolution performance of different methods. (a) Comparison results of RMSE and Pearson's r for the monocyte pseudotime dataset (n = 10 cell states). (b) Comparison results of RMSE and Pearson's r for the drug treatment detection dataset (n = 4 cell states). (c) Comparison results of RMSE and Pearson's r for the cell cycle dataset (n = 3 cell states).

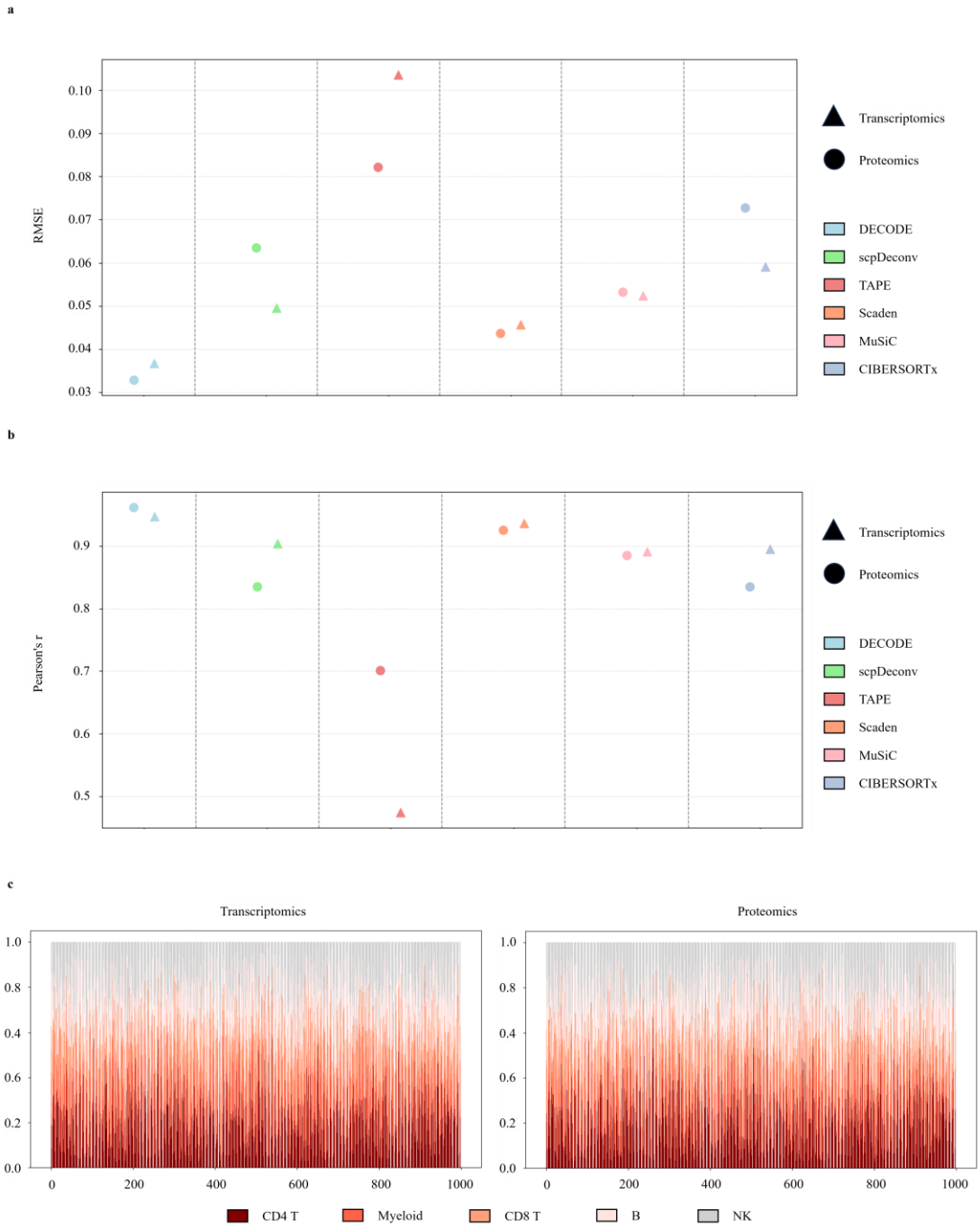

Fig. S6| Consistency assessment and integration results of multiomics deconvolution. (a-b) Scatter plot of the RMSE (a) and Pearson's r (b) values for DECODE, scpDeconv, TAPE, Scaden, MuSiC, and CIBERSORTx. The vertical axis represents the RMSE or Pearson's r value, while the horizontal axis represents the different methods. Triangles and circles represent the transcriptomics and proteomics results, respectively. (c) Predictive results of DECODE for 1,000 test samples (n = 5 cell types) using two molecular information sources. The vertical axis represents the cell abundance of each sample, whereas the horizontal axis represents the different samples. The left and right sides show the predictive results using transcriptomic and proteomic information, respectively. The more similar the two images on the left and right are, the greater DECODE's consistency.