

Cell-type deconvolution for bulk RNA-seq data using single-cell reference: a comparative analysis and recommendation guideline

Xintian Xu^{1,2,†}, Rui Li^{1,2,†}, Ouyang Mo^{1,2,†}, Kai Liu^{1,3}, Justin Li⁴, Pei Hao^{1,2,*}

¹Key Laboratory of Molecular Virology and Immunology, Shanghai Institute of Immunity and Infection, Chinese Academy of Sciences, 320 Yueyang Road, Xuhui District, Shanghai 200031, China

²University of Chinese Academy of Sciences, 1 Yanqihe East Road, Huairou District, Beijing 100039, China

³Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, 270 Dong'an Road, Xuhui District, Shanghai 200032, China

⁴Department of Mathematics, University of Connecticut, 352 Mansfield Road, Storrs, CT 06269, USA

*Corresponding author. Pei Hao, Key Laboratory of Molecular Virology and Immunology, Shanghai Institute of Immunity and Infection, Chinese Academy of Sciences, Shanghai 200031, China. E-mail: phao@siii.cas.cn

†Xintian Xu, Rui Li and Ouyang Mo contributed equally to this work.

Abstract

The accurate estimation of cell type proportions in tissues is crucial for various downstream analyses. With the increasing availability of single-cell sequencing data, numerous deconvolution methods that use single-cell RNA sequencing data as a reference have been developed. However, a unified understanding of how these deconvolution approaches perform in practical applications is still lacking. To address this, we systematically assessed the accuracy and robustness of nine deconvolution methods that use single-cell RNA sequencing data as a reference, evaluating them on real bulk data with cell proportions verified through flow cytometry, as well as simulated bulk data generated from five single-cell RNA sequencing datasets. Our study highlights the importance of several factors—including reference dataset construction strategies, dataset size, cell type subdivision, and cell type inconsistency—on the accuracy and robustness of deconvolution results. We also propose a set of recommended guidelines for software users in diverse scenarios.

Keywords: cell type deconvolution; immune infiltration; prediction accuracy; scRNA-seq reference; performance evaluation

Introduction

Immune cells play a crucial role in tumor development and inflammatory processes, and the cellular composition and proportions of immune infiltrates in tumors are directly associated with tumor progression and response to therapy [1]. Different immune cell populations are distributed in distinct locations within tumor tissues and exhibit diverse effects on antitumor immunity [2–5]. Notably, recent advances in immunotherapy have led to therapeutic advantages. However, its effectiveness is largely seen in a minority of patients with a “hot” tumor-immune microenvironment, defined by high levels of lymphocyte infiltration [6, 7]. Therefore, understanding the immune infiltration landscape in solid tumor patients, including the composition and proportions of different immune cell types, is of immense value for defining tumor subtypes, predicting disease progression, and estimating drug responses.

Single-cell sequencing, mass cytometry, and multiplexed spatial cellular phenotyping are commonly used experimental approaches for elucidating immune infiltration in tumor tissue samples [8]. However, these techniques bear the limitations of low throughput and/or high costs. Cell type deconvolution is a bioinformatic method that enables inference of the relative abundance of various cell types from the gene expression profiles

of tissues. The increasing availability of massive transcriptome data and data sharing has expanded the application scenarios for cell type deconvolution. With the aid of multiple deconvolution tools, scientists have been able to evaluate the composition and proportions of tissue immune cells in various clinical scenarios such as malignancies, autoimmune diseases, vaccination responses, infectious diseases, and inflammation [9–14].

The cell type deconvolution methods assume that the transcriptomic or methylation profiles of mixed tissues are linear combinations of expression or methylation levels from all constituent cell types, with mathematical approaches mainly including the least squares method (LS) and support vector regression (SVR) [15]. The former aims to minimize the squared differences between the fitted values and observed values without taking the error distribution into account, and the latter fits the data by finding an optimal hyperplane to maintain minimal errors [15, 16].

Early deconvolution methods inferred cell type proportions from bulk RNA-seq data, while, with the emergence of single-cell transcriptomic technology and the accumulation of data, more deconvolution methods using single-cell RNA sequencing (scRNA-seq) as a reference have been developed. A comparative study demonstrated that, in terms of performance, approaches based on single-cell references are comparable to the best-performing

Received: September 23, 2024. Revised: December 6, 2024. Accepted: January 20, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

For commercial re-use, please contact journals.permissions@oup.com

bulk methods [17]. From the user's perspective, deconvolution methods using scRNA-seq data as reference are more flexible and convenient compared to bulk deconvolution methods. On the one hand, the former allows users to break free from fixed cell types with a signature matrix provided by the tool itself, eliminating the need to supply additional tailored signature matrices for bulk RNA-seq data. On the other hand, the cell types involved in the prediction of proportions can be extended to nonimmune cells, such as neurons and tumor tissue cells. This flexibility is particularly useful in practical scenarios—e.g. when assessing immune infiltration in clinical samples where mixtures of immune and nonimmune cells are commonly observed.

Numerous benchmark studies have evaluated the performance of deconvolution methods from different perspectives. For instance, Hippen et al. [18] and Tran et al. [19] assessed the performance of deconvolution tools focusing on specific cancer types, while Avila Cobos et al. systematically evaluated the impact of several data preprocessing factors (such as data transformation and normalization) and cell type composition on the deconvolution results when applying deconvolution methods using scRNA-seq as a reference [17]. However, the performance of deconvolution tools using scRNA-seq data as a reference remains only partially understood, particularly in diverse application scenarios, such as situations where there is significant correlation between two or more cell types [16]. Consequently, selecting an appropriate deconvolution tool that aligns with a researcher's specific scientific inquiry poses a significant challenge.

Here, we assessed nine reference-based deconvolution algorithm tools using different mathematical approaches on both real and simulated bulk RNA-seq datasets, including four methods based on least squares, three methods based on support vector regression, and one algorithm based on elastic net and one based on Bayesian inference. Figure 1 provides an overview of the model evaluation process, factors affecting performance, and the metrics used for assessing each model. Among the least squares-based methods, Bisque estimates cell proportions from bulk RNA-seq data by using both a single-cell reference and Non-Negative Least Squares (NNLS) regression-transformed expression data [20]. Single-cell deconvolution of cell types (SCDC) focuses on integrating multiple scRNA-seq results and proposes a method called ENSEMBLE, which accurately deconvolutes bulk RNA-seq by utilizing multiple scRNA-seq reference datasets. It also uses Mutual Nearest Neighbor (MNN) to correct batch effects in scRNA-seq data [21]. Dampened Weighted Least Squares (DWLS) is a cell type-sensitive method for estimating cell scores, using a novel weighted least squares method to appropriately adjust the contribution of each gene, achieving better estimation of rare cell types [22]. MuSiC (Multi-Subject Single-Cell Deconvolution) transfers cell type-specific gene expression information from one dataset to another by weighting genes appropriately for intersample and intercell consistency [23]. Among the support vector regression-based methods, Bseq-SC adjusts individual gene expression samples by considering variations in cell type proportions, thereby removing differentially expressed genes caused by proportion differences, and performs cell type-specific differential gene expression analysis across groups, with the resulting basis matrix used for estimating cell proportions of each sample using CIBERSORT [24]. AutoGeneS uses multi-objective optimization to select discriminative genes, minimizing the average correlation coefficient while maximizing the Euclidean distance between cell types [25]. Cell Population Mapping (CPM) constructs its reference set from one or several related samples' scRNA-seq data and then uses this set to infer the cell composition in

additional batch samples [26]. Automated Deconvolution Augmentation of Profiles for Tissue Specific Cells (ADAPTS) implements a modular tool that provides a common interface for several popular deconvolution algorithms, using feature matrices to estimate cell type proportions present in heterogeneous samples [27]. In this study, we adopted the deconvolution algorithm Digital Cell Quantification (DCQ), which is based on elastic net in ADAPTS, and performed hierarchical deconvolution to improve the accuracy of cell type deconvolution [28]. In Bayesian inference-based methods, BayesPrism infers a joint posterior distribution of cell state proportion and gene expression from bulk RNA-seq data using Gibbs sampling, with explicit modeling of the differences in gene expression between single-cell reference and bulk data [29]. We explored the impact of reference dataset construction strategies, inconsistencies between the reference dataset and sample cell types, reference dataset size, and cell type subdivision on the deconvolution results. Root mean square error (RMSE) and Pearson correlation coefficient (PCC) served as the metrics for prediction accuracy. Guidelines were proposed based on the revealed deconvolution performance under diverse scenarios and execution time.

Materials and methods

Dataset selection and preprocessing

Five different datasets coming from different tissues were used to generate pseudo-bulk mixtures. We retained cells based on the following criteria: cells with the total number of RNA molecules detected >500 and $<10\,000$, the number of different RNA molecules detected >250 , and the logarithm of the number of gene counts per million RNA molecules >0.8 were retained. After screening out cells with $>20\%$ mitochondrial gene content, only genes with at least 1% of all cells having a Unique Molecular Identifier (UMI) or read count >1 were kept.

After quality control, single-cell data were normalized to find highly variable genes using the *FindVariableFeatures* function. Linear dimensionality reduction was performed using the *RunPCA* function, and the first 15 principal components were taken for subsequent analysis according to the *ElbowPlot*. Cell clustering was performed using the *FindNeighbors* and *FindClusters* functions. All these functions are part of the Seurat (v4.1.1) [30]. We made 2D t-distributed Stochastic Neighbor Embedding (t-SNE) plots for each dataset, annotating different types of cell subpopulations based on the original results of data source.

Pancreatic ductal adenocarcinoma (PDAC) and colorectal cancer (CRC) datasets were used to explore the effects of cell subtype subdivision. Ductal cells of the PDAC dataset were classified into four cell subtypes based on the published research [31], namely, Ductal-terminal_ductal_like, Ductal-CRISP3_high/centroacinar_like, Ductal-MHC_Class_II, and Ductal-APOL1_high/hypoxic. CD4 $^+$ T cells and CD8 $^+$ T cells from the CRC dataset were divided into 11 and 7 cell subtypes, respectively according to the published research [32].

Generation of pseudo-bulk mixtures

We established simulated bulk RNA-seq samples by sampling cells from the single-cell dataset and summing their total gene expression. Balanced sampling and unbalanced sampling were used for the simulation of the bulk dataset. Balanced sampling refers to the use of sampling with replacement so that the relative proportions of different cell types in the simulated bulk data are consistent with the single-cell reference, with sampling ratios of 25%, 50%, and 100% of the reference dataset, respectively. Unbalanced sampling refers to sampling with replacement at a fixed frequency of 100 and 500 for each cell type.

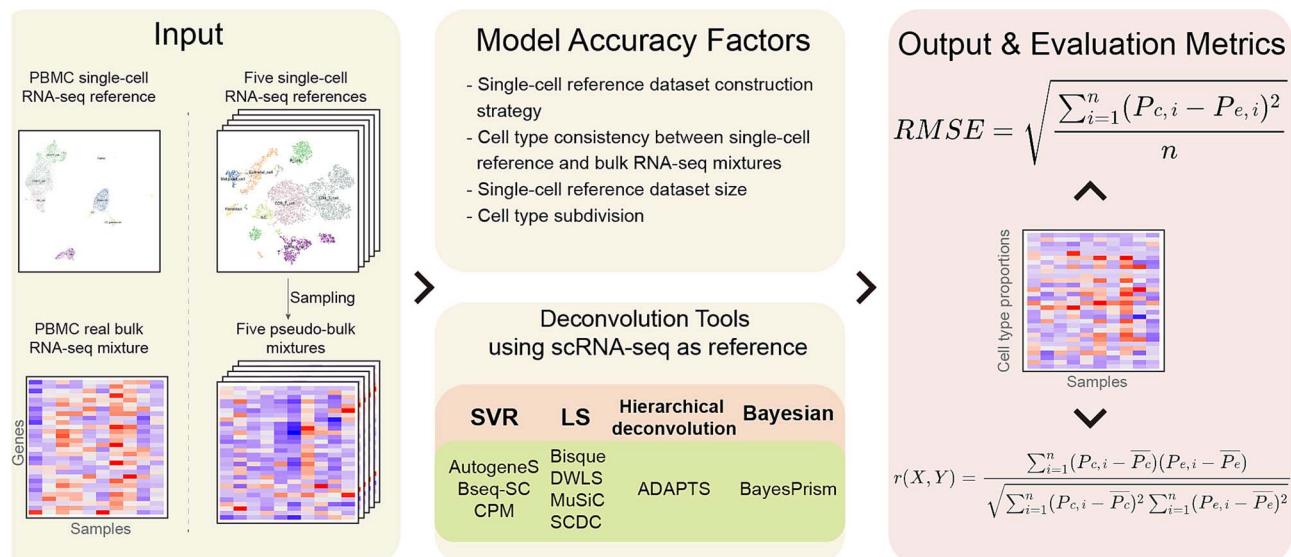


Figure 1. Graphical summary of deconvolution methods used and evaluation determinants. We conducted a comparative analysis using both real and simulated data for eight deconvolution tools using single-cell reference based on different deconvolution algorithms. The real data consisted of a PBMC single-cell reference dataset and PBMC real bulk RNA-seq data, with cell proportions in the bulk RNA-seq samples measured by flow cytometry. The simulated data comprised five single-cell datasets from different platforms and tissue sources, along with pseudo-bulk mixtures generated from these datasets. We assessed five impact factors that could affect deconvolution performance in practical scenarios, comparing the sample cell proportions output with real or simulated data using RMSE and Pearson correlation. Based on the influence of these impact factors on deconvolution performance and the performance of different tools in various practical scenarios, we proposed a set of recommendation guidelines. SVR: support vector regression; LS: least squares; P_c : computed proportion; P_e : expected proportion.

The construction of the simulated dataset was completed in R(v4.1.3) [33] using the strata function in the sampling package and the “rsqr” parameter to perform sampling with replacement. After sampling, all sampled cell gene expression values were summed as the expression value of the simulated bulk sample.

Evaluation of deconvolution methods with real RNA-seq dataset

We downloaded the RNA-seq expression matrix of peripheral blood mononuclear cell (PBMC) samples from GSE107011 [34] for which cell type proportions were measured by flow cytometry. Furthermore, we used the scRNA-seq dataset of PBMCs from GSE132044 [35] as a reference. Cell types were manually annotated according to the following markers: B cell (MS4A1), monocyte (CD14), LD granulocyte (low-density granulocyte) (FCGR3A), dendritic cell (FCER1A), platelet (PPBP), CD4⁺T cell (IL7R), CD8⁺T cell (CD8A), and NK cell (GNLY, NKG7).

Selection of deconvolution methods

Nine deconvolution methods that use single cell RNA-seq datasets as their respective references were chosen to be evaluated, including three SVR (AutoGeneS [25], Bseq-SC [24], CPM [26]), four least squares (Bisque [20], DWLS [22], MuSiC [23], SCDC [21]) and one penalized regression (ADAPTS [27]) method, and one Bayesian (BayesPrism [29]). The source codes of these methods were obtained from GitHub from the following commits: sdanzige/ADAPTS, amitfrish/scBio, xuranw/MuSiC, meichendong/SCDC, cran/BisqueRNA, shenorrLab/TRDF/bseqsc, dtsoucas/DWLS, theislabs/AutoGeneS, Danko-Lab/BayesPrism.

Software containing multiple analysis modes were selected for deconvolution based on a single cell reference mode. We ran ADAPTS with Digital Cell Quantification (DCQ) [28] method and selected the hierarchical augment model (hier.augment) for the prediction of cell proportions. The ct.sub parameter of SCDC was set to all cell types in the dataset. We ran CPM with absolute prediction after normalizing the expression matrix. Bisque was

run in markerless gene mode, with the specific parameter set to markers = NULL. We ran AutoGeneS with nusvr mode. Marker genes required for Bseq-SC were screened based on differentially expressed genes in single cell datasets. For each dataset, the top 30 genes with differential expression folds were selected.

Measurements of deconvolution performance

The elapsed time was measured with the “proc.time” function for R-based methods: ADAPTS, CPM, MuSiC, SCDC, Bisque, Bseq-SC, and DWLS. The elapsed time of AutoGeneS was measured with the “time.time” function.

We calculated Pearson correlation values and RMSE values for the predicted and known cell composition proportions for all methods in different samples from six datasets. A higher Pearson correlation value and a smaller RMSE value correspond to better deconvolution performance.

Statistical analysis

The visualization and data analysis were done in R4.1.3 using the R language packages: ggplot2 (v1.4.4) [36], corrplot (v0.92), ComplexHeatmap (v2.10.0) [37], circlize (v0.4.15) [38], RColorBrewer (v1.1.3), BuenColors (v0.5.6), reshape2 (v1.4.4) [39], and Seurat (v4.1.1) [30].

Results

Deconvolution performance on the real bulk RNA-seq dataset

We initially selected a set of 12 human PBMC bulk RNA-seq samples, for which the proportions of various cell types were determined using flow cytometry. As a reference dataset to assess the performance of the deconvolution methods, we used published 10x (v2) scRNA-seq data [35]. RMSE and PCC served as the metrics for prediction accuracy.

We used nine cell deconvolution methods to decipher the immune infiltration level of bulk RNA-seq samples, as summarized in Table 1. Among all deconvolution methods, BayesPrism

Table 1. Overview of single-cell reference-based cell type deconvolution tools used in the research.

Tools	Deconvolution Method	Program	Published year	Input reference	Cell-type proportions	Expression matrix	Marker gene correction	Tool Features		
								Accuracy	Efficiency	Automation
Bseq-SC	SVR	R	2017	scRNA-seq	Y	Y	Y	7	5	Y
ADAPTS	Hierarchical deconvolution	R	2019	scRNA-seq/Signature matrix	Y	Y	Y	1	9	Y
CPM	SVR	R	2016	scRNA-seq	Y	Y	Y	2	8	Y
MuSiC	WNNLS	R	2016	Multi-scRNA-seq	Y	Y	Y	8	6	Y
DWLS	DWLS	R	2019	scRNA-seq	Y	Y	Y	5	7	Y
SCDC	WNNLS	R	2020	ENSEMBLE scRNA-seq	Y	Y	Y	9	4	Y
Bisque	Nu-SVR	R	2020	scRNA-seq/bulk-sorted RNA-seq	Y	Y	Y	6	2	Y
AutoGenes	Nu-SVR	R	2021	scRNA-seq	Y	Y	Y	3	2	Y
BayesPrism	Bayesian	R	2022	scRNA-seq	Y	Y	Y	1	9	Y

and DWLS performed best with the lowest RMSE and highest Pearson correlation values, as depicted in Fig. 2A and B. The DWLS prediction results showed Pearson correlation values above 0.6 for monocytes, CD8+ T cells, and CD4+ T cells, while BayesPrism achieved a remarkable correlation coefficient of 0.91 in B-cell prediction (see Fig. 2B). In contrast, MuSiC and SCDC performed the worst, with RMSE values close to 0.2 and Pearson correlation values below 0.4 (as shown in Fig. 2A and B).

To evaluate the performance of deconvolution tools in closely related cell types with similar gene expression profiles, we conducted pairwise comparisons of correlation coefficients between two groups of cell types and visualized the results using heatmaps arranged by Pearson correlation. As shown in Figs. 2C and D, DWLS demonstrated superior performance in distinguishing highly correlated cell types.

Impact of reference dataset construction strategies on deconvolution results

Based on the performance of the PBMC dataset, we found that the small sizes of real bulk datasets limit the range of scenarios available for evaluating deconvolution methods. Moreover, deconvolution results alone do not fully capture the performance of these methods and can be significantly affected by experimental errors. To address these limitations, we constructed simulated bulk datasets using single-cell data from various platforms to assess the performance of deconvolution methods across a broader range of scenarios, as summarized in Table 2.

We investigated the performance of different deconvolution methods on the SKIN dataset under four dataset construction strategies. These strategies can be grouped into two categories: balanced sampling and unbalanced sampling. As shown in Fig. 3A and B, all deconvolution methods, except Bisque, did not exhibit significant improvements between balanced and unbalanced sampling approaches, despite achieving slightly lower RMSE and better PCC values in balanced sampling. As illustrated in Figs. 3C and D, varying sampling ratios within both balanced and unbalanced groups did not significantly affect the deconvolution results. For both balanced and unbalanced sampling, the influence of sampling ratio and frequency on the prediction results was not significant. In terms of deconvolution performance, Bseq-Sc, DWLS, MuSiC, and SCDC achieved higher Pearson correlation values, while CPM had the lowest RMSE values (see Fig. 3A and B). To unify and simplify the benchmarking process, and considering the significant differences in the number of cells in different single-cell references, we will subsequently adopt the balanced strategy to construct simulated bulk datasets.

Impact of cell type inconsistency on deconvolution results

The inconsistency of cell types in the single-cell reference and the bulk sample is an important factor for the performance of the deconvolution methods. Since there have been studies of bulk samples with more cell types than the single-cell reference [17], we investigated the effect of cell type differences on the performance of the deconvolution methods when the bulk sample has fewer or the same cell types as the single-cell reference. The performance of all deconvolution methods across five bulk RNA-seq datasets revealed a consistent trend: as the number of cell types in the bulk sample becomes more consistent with the single-cell reference, the RMSE values of the predicted proportions relative to the true proportions gradually decrease, while the PCC values gradually increase. Deconvolution tools performed best when the cell types were constant between the bulk sample and

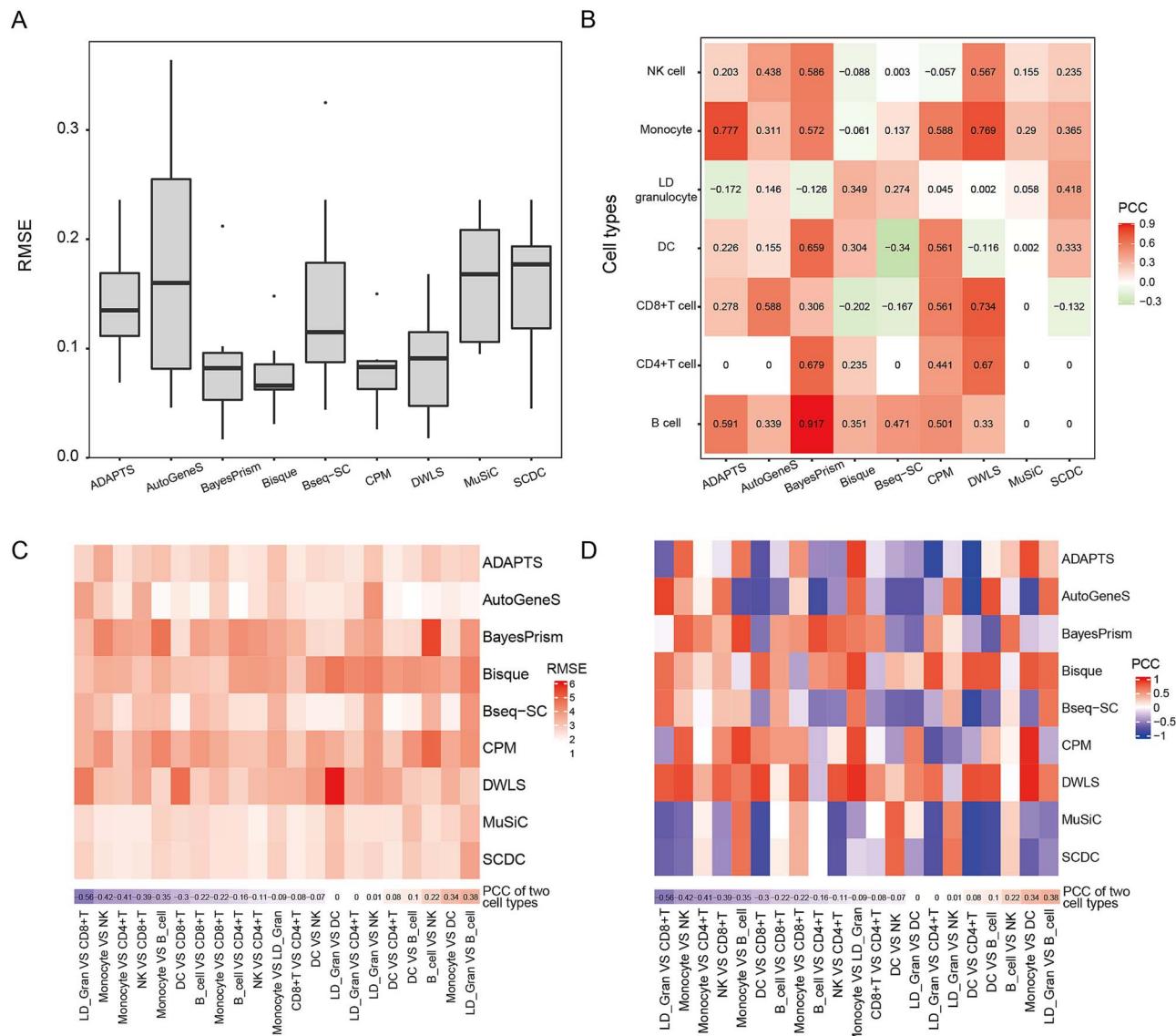


Figure 2. Performance of deconvolution methods on the real bulk dataset. (A) RMSE values of predicted and known proportions for nine deconvolution methods on all cell types; (B) PCC values of predicted cell proportions by deconvolution methods and ground truth determined using FACS on each cell type. Each cell type was predicted in 12 bulk samples; “0” values indicate that no correlation could be computed because all predictions were zero. (C) RMSE assessment of nine deconvolution algorithms across 21 distinct cell-type pairs. (D) PCC between predicted and actual cell proportions across 21 distinct cell-type pairs. The annotation bar beneath heatmaps shows true correlation coefficients between cell-type pairs, with numerical annotations arranged in ascending order to reflect the direction and strength of the correlation.

Table 2. Five single-cell RNA-seq datasets for constructing simulated bulk RNA-seq data.

Accession ID	Abbreviation	Tissue/Organ	Platform	Samples	Cell types	Cells	Genes	Median counts/Cell
1	E-MTAB-8142	SKIN	10× Chromium	3	22	79,583	12 704	6527
2	GSE146771	CRC	Smart-seq2	10	8(24)	10,456	15 160	19 952
3	GSE111672	PDAC	inDrop	2	15(18)	3659	19 736	2541
4	alddinger20	CERE	SPLiT-seq	5	21	4462	29 523	747.5
5	GSE158291	PTC	BD Rhapsody	4	7	3552	14 305	2091

Numbers in parentheses indicate the number of cell types after cell population subdivision.

the single-cell reference. However, the predictions of the Bisque method on the CRC, PDAC, and PTC datasets showed that, even when the cell types in the bulk RNA-seq datasets were consistent with the single-cell reference, the RMSE values either increased or did not significantly decrease (see Fig. 4A, Fig. S1A), indicating that this method may not be as reliable as others in such scenarios.

In contrast, ADAPTS and CPM generally outperformed other methods, demonstrating lower RMSE values when the bulk sample and the single-cell reference had inconsistent cell types, as shown in Fig. 4A and Fig. S1A. These results suggest that these two methods may be preferable when the bulk sample contains fewer cell types than the single-cell reference. AutoGeneS, bseqsc,

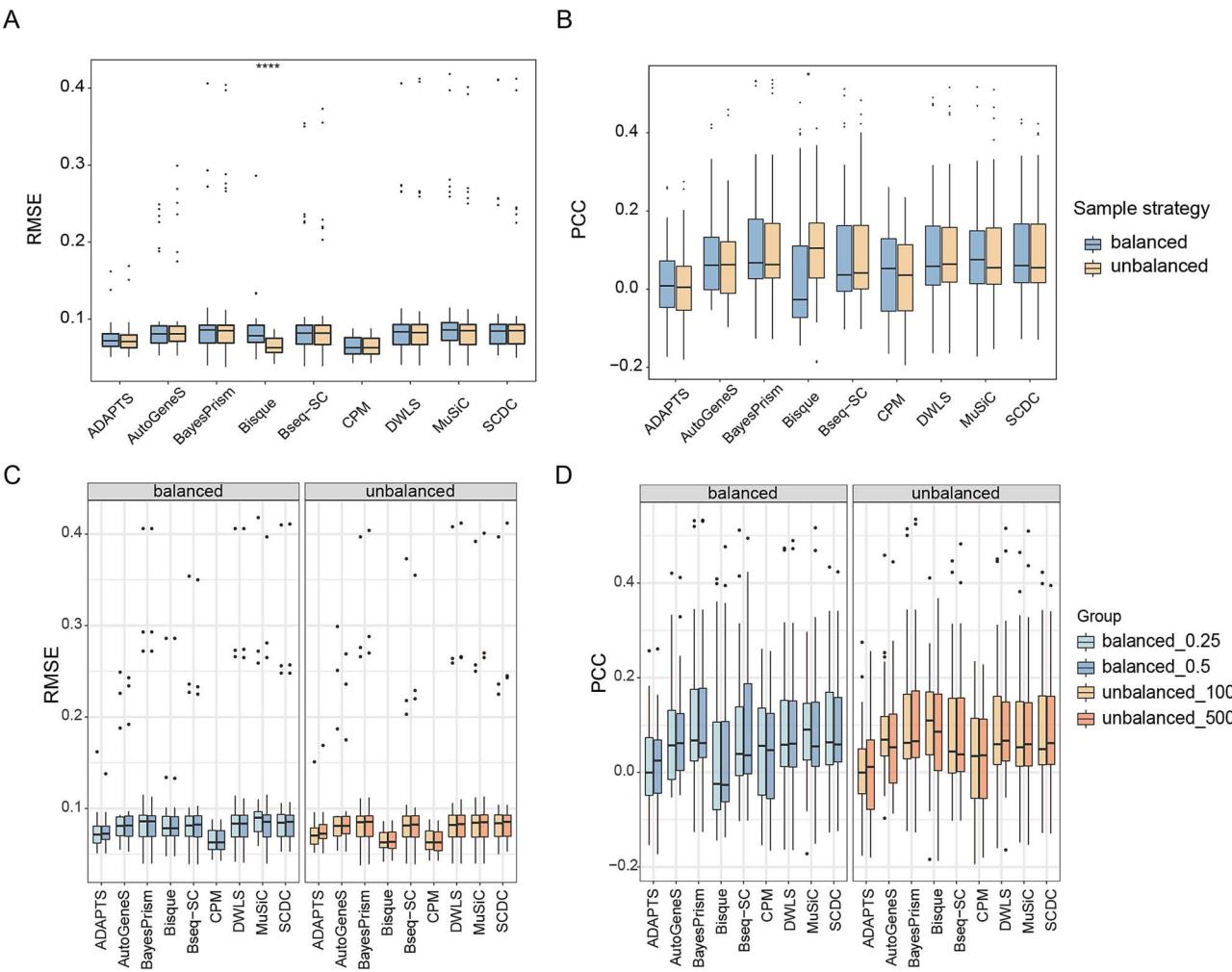


Figure 3. Comparison of prediction outcomes on the pseudo-bulk dataset constructed under four dataset construction strategies. Performance of different deconvolution methods on pseudo-bulk tissue mixtures constructed under four sampling strategies: 25% proportional balanced sampling (BS_0.25), 50% proportional balanced sampling (BS_0.5), 100 frequency unbalanced sampling (UBS_100), and 500 frequency unbalanced sampling (UBS_500). one hundred pseudo-bulk samples were constructed for each sampling method from the SKIN dataset. Each boxplot contains the evaluating metric of all cell types under the method. (A) RMSE values of predicted proportions and known proportions under balanced versus unbalanced sample strategy; (B) Pearson correlation (r) values of predicted proportions and known proportions under balanced versus unbalanced sample strategy. (C) Detailed RMSE comparison within balanced and unbalanced groups, stratified by sampling parameters. (D) PCC comparison within balanced and unbalanced groups. Each boxplot represents the distribution of evaluation metrics across all cell types for the respective method.

DWLS, MuSiC, and SCDC performed well in terms of Pearson correlation values across most datasets, with the exception of the CERE dataset (shown in Fig. 4B, Fig. S1B). All methods had low Pearson correlation values on the CERE dataset, probably because the CERE dataset has a lower median number of reads per cell.

We also assessed how the deconvolution tools performed on datasets with different sampling ratios. Under both sampling ratios, the RMSE and Pearson correlation values for all methods exhibited similar variation trends with the number of cell types. The RMSE and Pearson correlation values were also relatively close, with only small differences observed in specific conditions. This indicates that, in a balanced sample scenario, the sampling ratio has little impact on the accuracy of cell type proportion predictions.

Impact of dataset size on deconvolution results

Previous research has shown that deconvolution methods perform best when the cell types in the bulk and single-cell references are consistent [17]. To further investigate the impact of dataset size on deconvolution algorithms, we assumed that the

reference cell types for both bulk and single cells were consistent. We increased the dataset size by augmenting the number of cell types in both the single-cell and bulk simulation datasets, constructing pseudo-bulk datasets with balanced sampling rates of 25% and 100%, respectively.

The trend in RMSE across the nine deconvolution methods and five datasets was generally consistent: as the number of cell types increased, the RMSE decreased. Most methods achieved their lowest RMSE values when the number of cell types in the bulk sample was the largest, as shown in Figs 5A and S2A. We believe that increasing the number of cell types helps reduce the deviation between the predicted and known proportions to some extent. However, in some cases, the RMSE values followed an opposite trend. For instance, as shown in Fig. 5A, in Bisque, when the number of cell types was highest in the PTC, the RMSE value increased significantly.

The Pearson correlation heatmap of predicted and known cell type proportions reveals that the number of cell types and dataset size are not necessarily related to the performance of the deconvolution methods. As shown in Fig. 5B, AutoGenes,

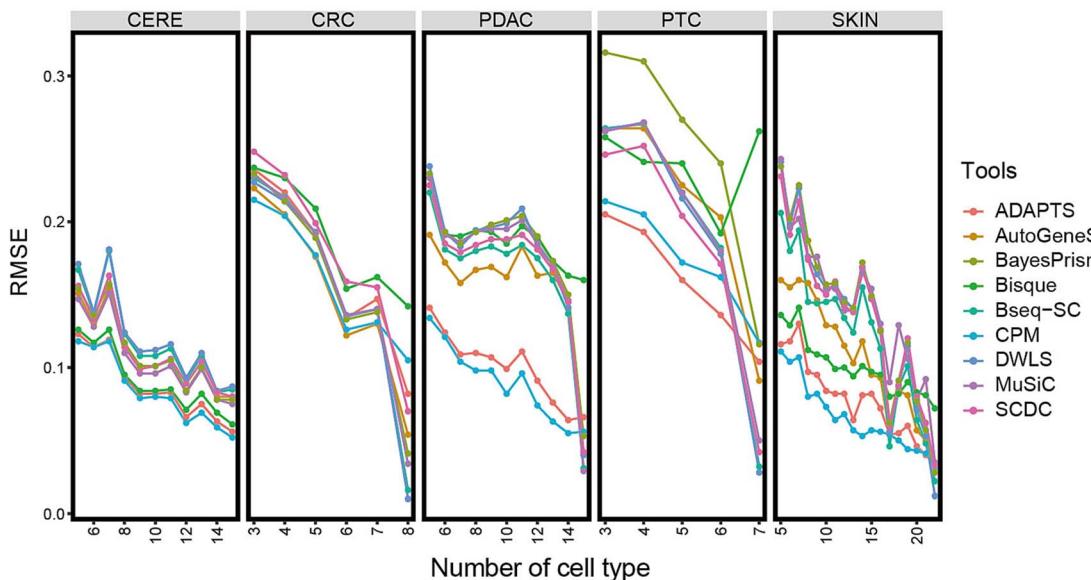
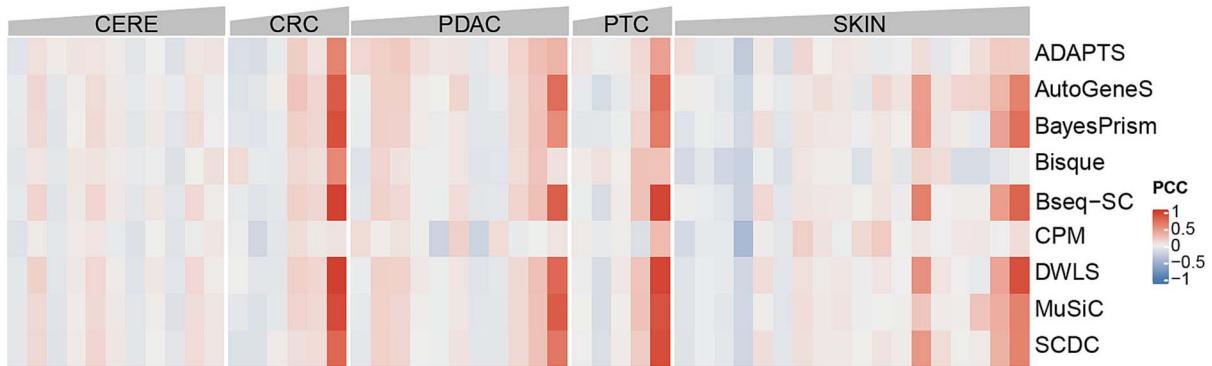
A**B**

Figure 4. Performance of deconvolution methods in pseudo-bulk datasets with inconsistent cell types compared with the single cell reference. The performance of different deconvolution methods on 1000 pseudo-bulk samples; the sum of all bulk samples in each dataset is 100. In heatmaps, the number of bulk sample cell types increases from left to right in each dataset, and the number of cell types in each pseudo-bulk sample is less than or equal to the single-cell RNA sequencing reference. (A) The RMSE values of the predicted proportions and the known proportions on datasets constructed with 25% proportional balance sampling. Each point contains metrics of all samples with that value for the number of cell types. (B) The PCC values of the predicted proportions and the known proportions on datasets constructed with 25% proportional balance sampling. Each square contains metrics of all samples with that value for the number of cell types.

Bseq-SC, BayesPrism, and DWLS performed well on large datasets for SKIN, CRC, and PDAC; Bisque and CPM performed best on small CERE and PTC datasets. The Pearson correlation values of CPM on SKIN, CRC, and CERE tended to drop as the number of cell types increased, whereas the Pearson correlation values of DWLS, MuSiC, and SCDC on PTC tended to decrease and subsequently increase (see Fig. 5B). Therefore, we believe that the predictive accuracy of deconvolution methods does not have a linear relationship with the size of the dataset.

In terms of the performance of the deconvolution methods under different dataset sizes, AutoGeneS, ADAPTS, and CPM outperformed the others, with ADAPTS performing particularly well on the smallest PTC dataset and CPM performing best on the two largest datasets, SKIN and CERE, as shown in Figs 5B and S2B. Conversely, MuSiC, Bseq-SC, and SCDC performed the worst across all datasets, with PTC (see Fig. 5B) and SKIN (see Fig. S2B) datasets showing a large number of negative Pearson correlation values. Increasing sampling ratios did not significantly improve prediction accuracy, with MuSiC,

Bseq-SC, and SCDC showing lower prediction accuracy on the 100% sampling ratio of the bulk dataset, as seen in Fig. S2B, possibly due to the exponentially larger dataset introducing more redundant information.

Impact of cell type subdivision on deconvolution results

Tissues typically comprise multiple cell types, and a single cell type can be functionally divided into several cell subtypes. Cells in each subtype tend to express genes in a more linear manner, with relatively less variation between them. Similar expression profiles can pose a considerable challenge to the performance of deconvolution methods, necessitating high sensitivity and resolution of the method for rare subtype cell characteristics. To assess the resolution of different deconvolution methods for cell subtypes, we classified ductal cells in the PDAC dataset into four cellular subtypes based on the original literature [31] and divided the CD4+ and CD8+ T cells in the CRC dataset into 11 and 7 subtypes [32]. The performance of different

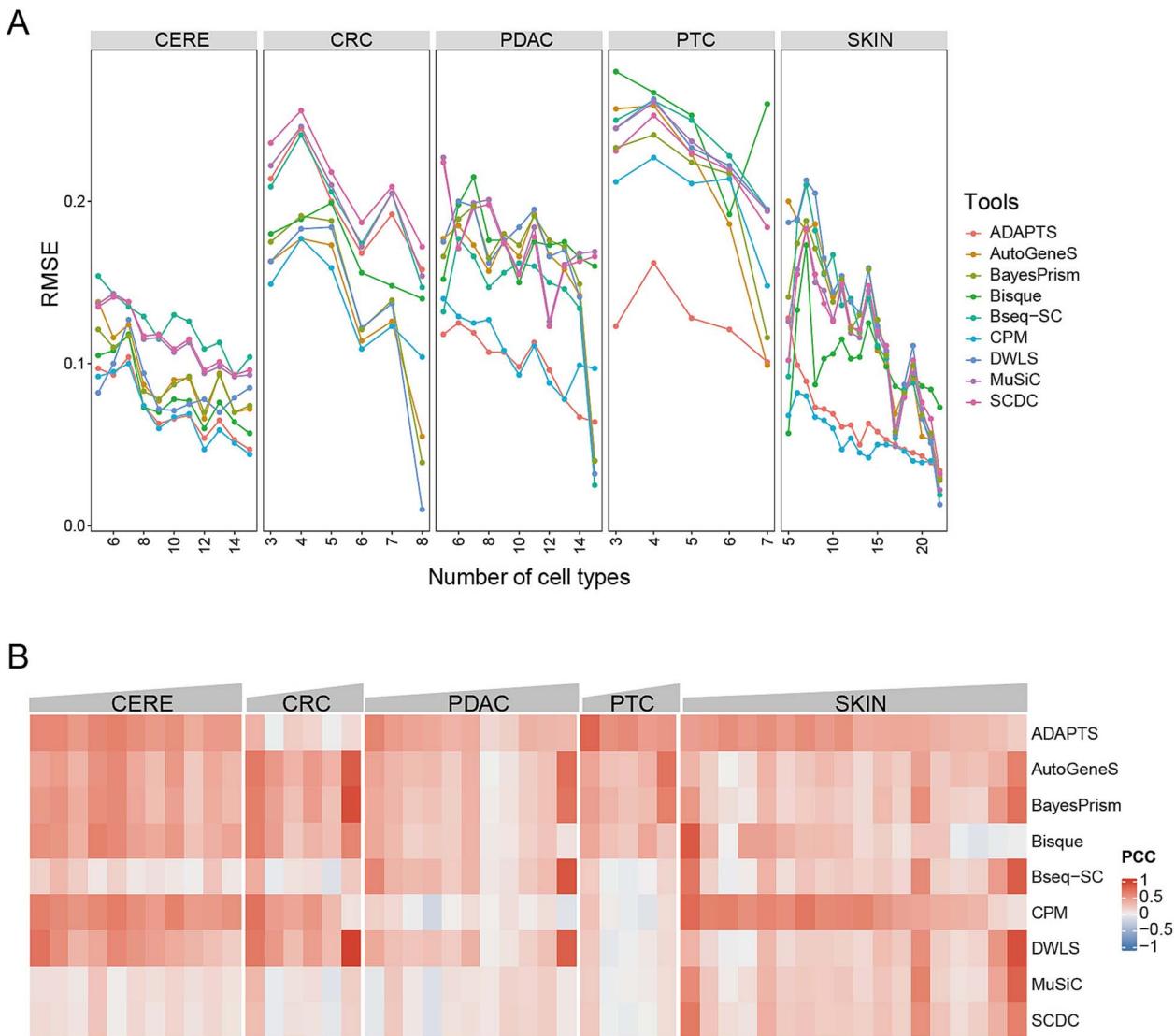


Figure 5. Performance of deconvolution methods in pseudo-bulk datasets with a consistent cell type of single cell reference. The performance of different deconvolution methods on 1000 pseudo-bulk samples, the sum of all bulk samples in each dataset is 100. In heatmaps, the number of bulk sample cell types increases from left to right in each dataset, whereas the number of cell types in each pseudo-bulk sample is equal to the single-cell RNA sequencing reference. (A) The RMSE values of the predicted proportions and the known proportions on datasets constructed with 25% proportional balance sampling. Each point contains metrics of all samples with that value for the number of cell types. (B) The Pearson correlation (r) values of the predicted proportions and the known proportions on datasets constructed with 25% proportional balance sampling. Each square contains all samples with that value for the number of cell types.

methods on two datasets before and after cell type subdivision was evaluated.

In the PDAC and CRC datasets, DWLS showed the most significant improvement in performance after cell type subdivision, suggesting that DWLS is the most suitable tool for deconvolution tasks after cell type subdivision. Additionally, ADAPTS and CPM also performed better on the subdivided datasets, exhibiting higher Pearson correlation values and smaller RMSE values. AutoGeneS, SCDC, and MuSiC did not show significant differences in performance between the two datasets. Regarding the performance of deconvolution methods, ADAPTS, AutoGeneS, and CPM consistently performed well both before and after cell type subdivision, as shown in Fig. 6A–D. Notably, as demonstrated in Fig. 6D, all deconvolution tools exhibited significantly lower RMSE values in the CRC dataset when the cell classification was refined from 8 to 24 cell types through cellular subdivision. This observation suggests that single-cell references with more granular cell

type classifications can enhance the performance of single-cell deconvolution tools. Overall, cell type subdivision can help deconvolution tools produce outputs that are closer to the true cell proportions. Deconvolution tools that performed well after cell type subdivision exhibited strong performance across various cell types within the subdivided dataset, as shown in Fig. S3A and B. The predicted proportions of ADAPTS, CPM, and DWLS for the four cell types were closer to the known proportions, indicating a higher predictive accuracy of these methods, as demonstrated in Fig. S3A.

Prediction accuracy of deconvolution methods in simulated bulk datasets

We evaluated the prediction accuracy of all deconvolution methods on five simulated datasets when the reference and bulk cell types were consistent. Of all the methods, ADAPTS and CPM performed best, achieving the highest Pearson correlation values

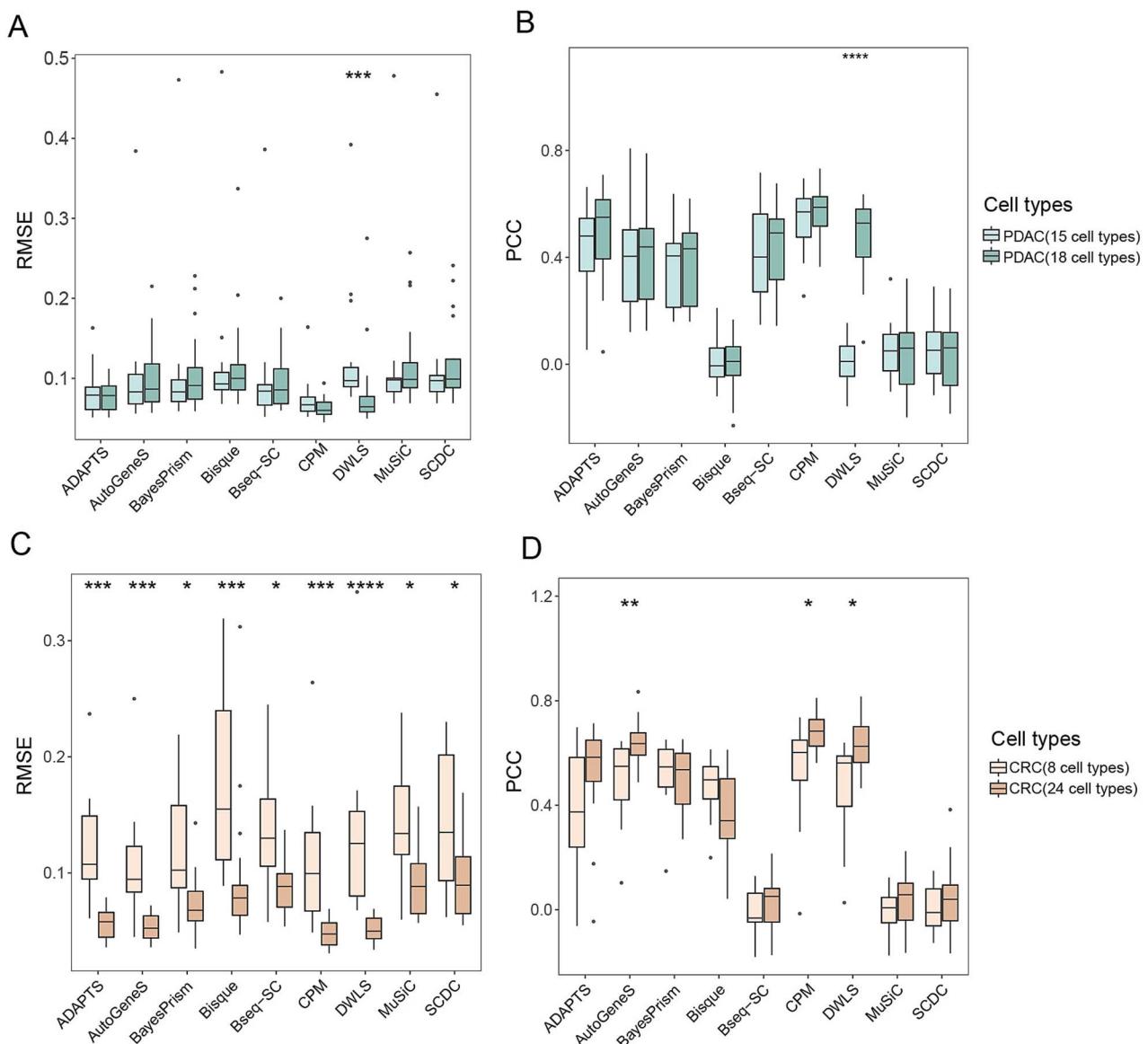


Figure 6. Impact of cell type subdivision on deconvolution results for the PDAC and CRC datasets. (A) Performance of nine methods before and after PDAC cell type subdivision, each boxplot represents the RMSE value of the predicted proportion and the known proportion of all cell types. (B) Performance of nine methods before and after PDAC cell type subdivision, each boxplot represents the Pearson correlation (r) value of predicted and known proportions of all cell types. (C) Performance of nine methods before and after CRC cell type subdivision, each boxplot represents the RMSE value of the predicted proportion and the known proportion of all cell types. (D) Performance of nine methods before and after CRC cell type subdivision, each boxplot represents the Pearson correlation (r) value of predicted and known proportions of all cell types.

and the smallest RMSE values, followed by AutoGeneS. MuSiC and SCDC showed the worst performance, as demonstrated in Fig. 7A and B. The performance of nine methods on both SKIN and CERE datasets was significantly better than on other datasets (see Fig. 7A and B). Under the two sampling ratios, in addition to Bisque, Bseq-SC and DWLS had obvious performance differences on some datasets, and the performance of the other methods was basically consistent.

Consideration of execution time

We measured the execution time of the deconvolution methods on the real and simulated datasets. As shown in Fig. 8A for real datasets and Fig. 8B for simulated datasets, the running time of ADAPTS, CPM, and DWLS on all datasets can be more than 10 times longer than other methods, among which ADAPTS was

the most time-consuming method. Bisque and Bseq-SC took the shortest time.

Discussion

This study collected single-cell datasets from six different tissue sources as a reference and systematically evaluated nine deconvolution methods for estimating cell type proportions in bulk RNA-seq datasets. Firstly, we assessed the performance of the deconvolution methods on real bulk samples using the PBMC dataset with matched flow cytometry results as ground truth. Next, we generated simulated bulk data from five single-cell datasets to investigate the impact of reference dataset construction strategies, inconsistency between the reference dataset and sample cell types, reference dataset size, and cell type subdivision on the performance of the deconvolution tools. Based on the

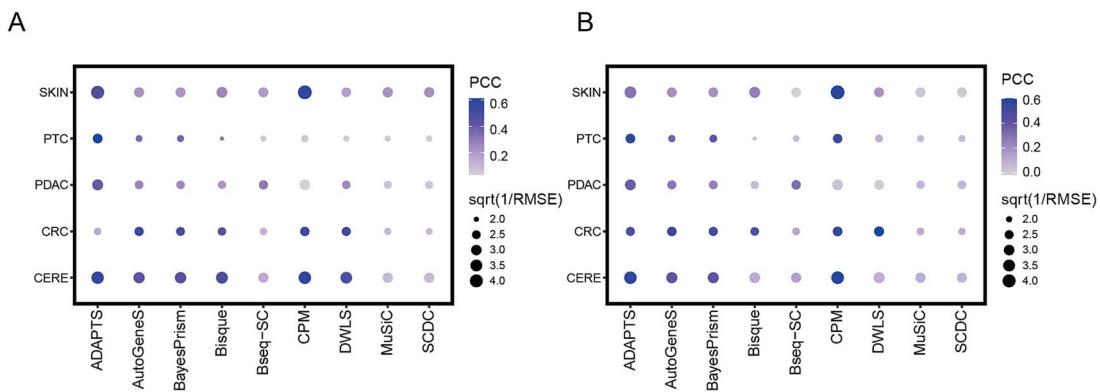


Figure 7. Accuracy comparison of deconvolution methods in simulated bulk datasets. Performance of different deconvolution methods on 1000 simulated bulk samples, each point contains 100 bulk samples. The color of the point represents the Pearson correlation (r), and the size of the point represents the reciprocal of the RMSE. (A) 25% balanced sampling group; (B) 100% balanced sampling group.

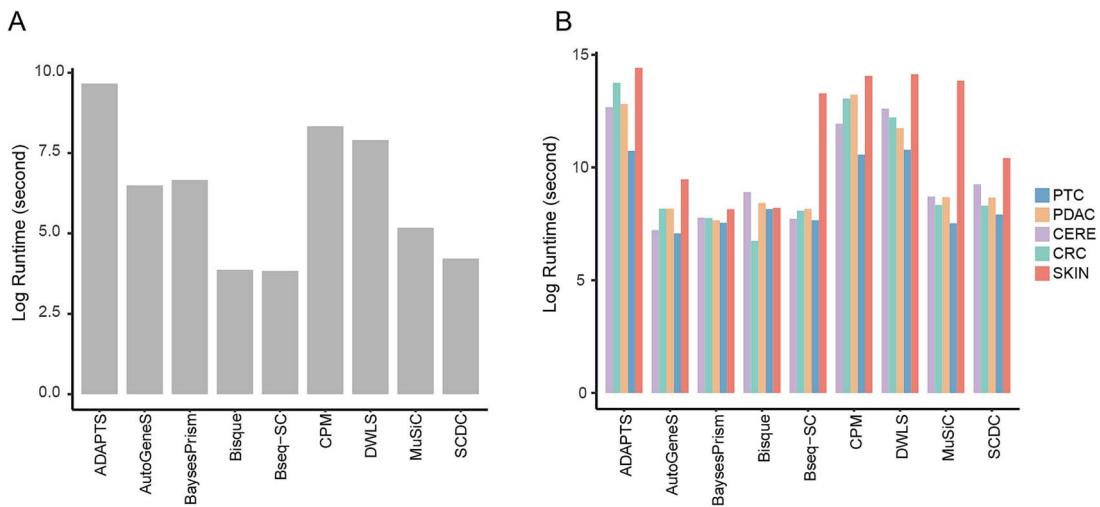


Figure 8. Comparison of execution time. (A) Time taken for deconvolution methods on real PBMC bulk dataset; (B) execution time of deconvolution methods on five simulated bulk datasets.

performance of the deconvolution tools on multiple datasets, we evaluated the performance of cell type proportion predictions and tracked the time taken by various methods. Finally, we used the recommended deconvolution tool to predict immune infiltration in large cohorts of two prevalent tumor types. The results demonstrated significant prognostic differences in sample grouping based on immune infiltration.

The performance of deconvolution methods varies on simulated datasets constructed using different sampling strategies. Balanced sampling, which ensures relative proportions of different cell types, leads to improved deconvolution tool performance. However, raising the sampling ratio has little to no impact on the accuracy of the predictions. We anticipate that expanding the dataset size will not improve the deconvolution methods' prediction accuracy as long as the expression characteristics of the same cell type in the single-cell reference data adequately capture the basic expression features of that cell type. On the contrary, it may introduce excessive noise and interfere with the performance of the deconvolution method. Inconsistencies between the single-cell reference dataset and the cell types in the bulk RNA-seq samples significantly affect the prediction accuracy of the deconvolution methods. Tests on five simulated datasets consistently showed that the deconvolution tools achieved the highest Pearson correlation and the smallest RMSE values between predicted and known proportions when the cell

types of the single-cell reference dataset and the bulk samples were consistent. This indicates that the additional cell types present in the single-cell reference, compared to the target samples, introduce extra information that may interfere with the deconvolution tool's classification of cell types. Furthermore, similarities in expression characteristics among some cell types can cause the predicted values to deviate significantly from the true values. Under the premise of ensuring consistency between the single-cell reference and the cell types in the bulk samples, the dataset size does not have a linear correlation with the predictive performance of the deconvolution tools. This relationship is determined by the algorithmic preferences of different tools and the representativeness of the dataset information. Cell type subdivision has a significant impact on some deconvolution tools, indicating that co-expression of genes affects the resolution of the deconvolution tool. After cell type subdivision, DWLS performs much better on the dataset, revealing its strong sensitivity to cell subtypes with linear gene expression.

Methods based on support vector regression outperform those based on ordinary least squares. Among the SVR-based methods, AutoGeneS and CPM exhibit good performance. AutoGeneS selects feature genes by minimizing correlations and maximizing distances between cell types. It uses Nu-SVR for deconvolution of given mixtures. By discarding a specified proportion of outliers

and performing regression on the remaining samples, AutoGeneS achieves robustness against technical and biological noise. Furthermore, limiting the number of feature genes to 400 enables great deconvolution efficiency while retaining accuracy [25]. CPM, on the other hand, employs a Markov model to infer the temporal dynamics of cell activation trajectories, allowing it to predict phenotypes for various cell states. It designates each cell state as a point in a multidimensional space defined by the cell state space and infers the abundance of each point in a given cell population based on scRNA-seq input. By incorporating cell state information, CPM achieves fine resolution in cell type identification [26]. DWLS outperforms the other three approaches based on ordinary least squares. This method introduces the concept of weights to adjust the contribution of each gene, building upon ordinary least squares. The introduction of a damping constant prevents infinite weights caused by low cell type proportions and/or low expression of marker genes [22]. In addition to the above two categories of methods, ADAPTS adopts a deconvolution algorithm called DCQ based on an elastic net. It employs a hierarchical strategy to avoid interference from cell types with similar expression profiles. The initial round of computation is used to estimate the major cell types, followed by the estimation of subpopulation abundances, enabling accurate estimation of cell type proportions in mixed samples [27].

For software users, the performance of immune infiltration predictions and execution time are the primary considerations. Based on the Pearson correlation and RMSE values of cell type predictions and known proportions, along with the execution time of nine deconvolution methods tested on real and simulated bulk datasets, we make the following recommendations: (i) use an expression matrix that includes all relevant cell types present in the mixture as a reference, ensuring consistency between the single-cell reference and bulk sample cell types. (ii) Avoid selecting excessively large single-cell datasets as references to prevent the introduction of excessive noise. (iii) Regardless of the running time, we recommend ADAPTS and CPM as they achieve the highest performance. If runtime is a concern, AutoGeneS is recommended. (iv) Certain tools perform best in specific scenarios: for datasets with cell types exhibiting high collinearity in gene expression, DWLS is recommended; for datasets where the cell types in the single-cell reference do not match those in the bulk samples, ADAPTS and CPM are recommended.

Through our evaluation, we have discovered that despite the development of numerous deconvolution methods along with specific optimizations for different problems, the performance of cell type deconvolution tools falls short of expectations when used to specific scenarios. There are notable deviations between the predictions of these tools and the true values, which are determined by factors such as the conformance of the reference dataset information, the variability of tissue gene expression, and the tool's resilience. The dependability of deconvolution outcomes is directly influenced by the selection and quality of benchmark datasets. In the future, it is necessary to establish more accurate and comprehensive benchmark datasets to enhance the reliability of deconvolution techniques. Furthermore, gene expression data can exhibit substantial differences across different cell types, necessitating optimization and improvement specific to different tissue or sample types. For deconvolution computations, traditional methods rely heavily on mathematical methodologies. With the accumulation of large-scale sequencing data, machine learning and deep learning techniques can be employed in the future to enhance the specificity and accuracy of deconvolution methods through analysis and modeling of extensive datasets.

Despite the fact that this study evaluated deconvolution methods on real bulk datasets and investigated the impact of different factors on deconvolution methods by constructing numerous simulated bulk datasets using five single-cell datasets, there are still some limitations in our evaluation. Three of the datasets used to construct simulated datasets in this study were derived from cancer tissues, which exhibit high heterogeneity, posing challenges for the determination of cell-type-specific genes and thus impacting deconvolution tools. Regarding the gradient settings for sampling proportions, we only implemented two gradients due to the large size of some datasets, which limited the comprehensiveness of our exploration of the impact of sampling proportions. Additionally, Bisque requires the manual provision of marker genes for cell annotation, which introduces human-induced bias into the prediction findings and renders it incomparable to other methods techniques.

In summary, we conducted a comprehensive evaluation of the performance of nine deconvolution algorithms through extensive testing on simulated and real bulk datasets in diverse practical application scenarios. This study revealed the impact of multiple factors on the prediction accuracy of deconvolution methods, assisting researchers in understanding the performance and limitations of various tools, thus enabling the selection of the most suitable tool for investigating specific research questions. Furthermore, the performance of different tools in diverse application scenarios highlighted the potential influence of algorithmic models on prediction performance, providing valuable insights for the development of new deconvolution tools.

Key Points

- We conducted a comprehensive evaluation of the performance of nine reference-based deconvolution algorithms through extensive testing on real and simulated bulk RNA-seq datasets and proposed a set of recommended guidelines for software users in diverse scenarios.
- Our study emphasizes the importance of aligning reference dataset construction strategies with the sample cell types to ensure accuracy and robustness in deconvolution while avoiding overly large datasets that may introduce noise.
- Automated Deconvolution Augmentation of Profiles for Tissue-Specific Cells (ADAPTS) and Cell Population Mapping (CPM) are recommended as the best-performing deconvolution tools, with Dampened Weighted Least Squares being suitable for scenarios with high collinearity in gene expression among cell types and ADAPTS and CPM being recommended when the reference dataset's cell types are inconsistent with the bulk sample.

Acknowledgements

The authors highly appreciate all members of the Laboratory of Molecular Virology and Immunology for their contributions to the improvement of this manuscript.

Supplementary data

Supplementary data are available at *Briefings in Bioinformatics* online.

Conflict of interest: None declared.

Funding

This work was supported by grants from the National Science and Technology Major Project of China [2023ZD0502401], and from the National Natural Science Foundation of China [32270695].

Data availability

The five public datasets used for generating pseudo-bulk mixtures in this article can be found at their respective sources:

SKIN [40]: <https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-8142> (accessed on 10 June 2022).

CERE [41]: <https://www.covid19cellatlas.org/aldinger20> (accessed on 28 June 2022).

PDAC [31]: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11672> (accessed on 23 June 2022).

CRC [32]: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146771> (accessed on 27 June 2022).

PTC [42]: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15829> (accessed on 28 June 2022).

References

- Galon J, Costes A, Sanchez-Cabo F. et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006; **313**:1960–4. <https://doi.org/10.1126/science.1129139>.
- Fridman WH, Pages F, Sautes-Fridman C. et al. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* 2012; **12**:298–306. <https://doi.org/10.1038/nrc3245>.
- Tosolini M, Kirilovsky A, Mlecnik B. et al. Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, th2, treg, th17) in patients with colorectal cancer. *Cancer Res* 2011; **71**: 1263–71. <https://doi.org/10.1158/0008-5472.CAN-10-2907>.
- Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer* 2016; **16**:582–98. <https://doi.org/10.1038/nrc.2016.73>.
- Fridman WH, Zitvogel L, Sautes-Fridman C. et al. The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol* 2017; **14**:717–34. <https://doi.org/10.1038/nrclinonc.2017.101>.
- Galluzzi L, Chan TA, Kroemer G. et al. The hallmarks of successful anticancer immunotherapy. *Sci Transl Med* 2018; **10**:10. <https://doi.org/10.1126/scitranslmed.aat7807>.
- Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discov* 2019; **18**:197–218. <https://doi.org/10.1038/s41573-018-0007-y>.
- Finotello F, Rieder D, Hackl H. et al. Next-generation computational tools for interrogating cancer immunity. *Nat Rev Genet* 2019; **20**:724–46. <https://doi.org/10.1038/s41576-019-0166-7>.
- Newman AM, Alizadeh AA. High-throughput genomic profiling of tumor-infiltrating leukocytes. *Curr Opin Immunol* 2016; **41**: 77–84. <https://doi.org/10.1016/j.coim.2016.06.006>.
- Chen JC, Cerise JE, Jabbari A. et al. Master regulators of infiltrate recruitment in autoimmune disease identified through network-based molecular deconvolution. *Cell Syst* 2015; **1**: 326–37. <https://doi.org/10.1016/j.cels.2015.11.001>.
- Kurupati R, Kossenkov A, Haut L. et al. Race-related differences in antibody responses to the inactivated influenza vaccine are linked to distinct pre-vaccination gene expression profiles in blood. *Oncotarget* 2016; **7**:62898–911. <https://doi.org/10.18632/oncotarget.11704>.
- Muema DM, Mthembu M, Schiff AE. et al. Contrasting inflammatory signatures in peripheral blood and Bronchoalveolar cells reveal compartment-specific effects of HIV infection. *Front Immunol* 2020; **11**:864. <https://doi.org/10.3389/fimmu.2020.00864>.
- Manthey CL, Moore BA, Chen Y. et al. The CSF-1-receptor inhibitor, JNJ-40346527 (PRV-6527), reduced inflammatory macrophage recruitment to the intestinal mucosa and suppressed murine T cell mediated colitis. *PLoS One* 2019; **14**: e0223918. <https://doi.org/10.1371/journal.pone.0223918>.
- Sturm G, Finotello F, Pettprez F. et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* 2019; **35**:i436–45. <https://doi.org/10.1093/bioinformatics/btz363>.
- Chen Z, Wu A. Progress and challenge for computational quantification of tissue immune cells. *Brief Bioinform* 2021; **22**:bbaa358. <https://doi.org/10.1093/bib/bbaa358>.
- Avila Cobos F, Vandesompele J, Mestdagh P. et al. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* 2018; **34**:1969–79. <https://doi.org/10.1093/bioinformatics/bty019>.
- Avila Cobos F, Alquicira-Hernandez J, Powell JE. et al. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun* 2020; **11**:5650. <https://doi.org/10.1038/s41467-020-19015-1>.
- Hippen AA, Omran DK, Weber LM. et al. Performance of computational algorithms to deconvolve heterogeneous bulk ovarian tumor tissue depends on experimental factors. *Genome Biol* 2023; **24**:239. <https://doi.org/10.1186/s13059-023-03077-7>.
- Tran KA, Addala V, Johnston RL. et al. Performance of tumour microenvironment deconvolution methods in breast cancer using single-cell simulated bulk mixtures. *Nat Commun* 2023; **14**:5758. <https://doi.org/10.1038/s41467-023-41385-5>.
- Jew B, Alvarez M, Rahmani E. et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat Commun* 2020; **11**:1971. <https://doi.org/10.1038/s41467-020-15816-6>.
- Dong M, Thennavanan A, Urrutia E. et al. SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Brief Bioinform* 2021; **22**:416–27. <https://doi.org/10.1093/bib/bbz166>.
- Tsoucas D, Dong R, Chen H. et al. Accurate estimation of cell-type composition from gene expression data. *Nat Commun* 2019; **10**:2975. <https://doi.org/10.1038/s41467-019-10802-z>.
- Wang X, Park J, Susztak K. et al. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* 2019; **10**:380. <https://doi.org/10.1038/s41467-018-08023-x>.
- Baron M, Veres A, Wolock SL. et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst* 2016; **3**:346–360.e4. <https://doi.org/10.1016/j.cels.2016.08.011>.
- Aliee H, Theis FJ. AutoGeneS: automatic gene selection using multi-objective optimization for RNA-seq deconvolution. *Cell Syst* 2021; **12**:706–715.e4. <https://doi.org/10.1016/j.cels.2021.05.006>.
- Frishberg A, Pesheh-Yaloz N, Cohn O. et al. Cell composition analysis of bulk genomics using single-cell data. *Nat Methods* 2019; **16**:327–32. <https://doi.org/10.1038/s41592-019-0355-5>.
- Danziger SA, Gibbs DL, Shmulevich I. et al. ADAPTS: automated deconvolution augmentation of profiles for tissue specific cells. *PLoS One* 2019; **14**:e0224693. <https://doi.org/10.1371/journal.pone.0224693>.
- Altbaum Z, Steuerman Y, David E. et al. Digital cell quantification identifies global immune cell dynamics during influenza

- infection. *Mol Syst Biol* 2014;10:720. <https://doi.org/10.1002/msb.134947>.
29. Chu T, Wang Z, Pe'er D. et al. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Can* 2022;3:505–17. <https://doi.org/10.1038/s43018-022-00356-3>.
 30. Hao Y, Hao S, Andersen-Nissen E. et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
 31. Moncada R, Barkley D, Wagner F. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol* 2020;38:333–42. <https://doi.org/10.1038/s41587-019-0392-8>.
 32. Zhang L, Li Z, Skrzypczynska KM. et al. Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell* 2020;181:442–459.e29. <https://doi.org/10.1016/j.cell.2020.03.048>.
 33. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2022.
 34. Monaco G, Lee B, Xu W. et al. RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep* 2019;26:1627–1640.e7. <https://doi.org/10.1016/j.celrep.2019.01.041>.
 35. Ding J, Adiconis X, Simmons SK. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol* 2020;38:737–46. <https://doi.org/10.1038/s41587-020-0465-8>.
 36. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Gentleman R, Hornik K, Parmigiani G (eds.), New York: Springer-Verlag, 2016. <https://doi.org/10.1007/978-3-319-24277-4>.
 37. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32:2847–9. <https://doi.org/10.1093/bioinformatics/btw313>.
 38. Gu Z, Gu L, Eils R. et al. Circlize implements and enhances circular visualization in R. *Bioinformatics* 2014;30:2811–2. <https://doi.org/10.1093/bioinformatics/btu393>.
 39. Wickham H. Reshaping data with the reshape package. *J Stat Softw* 2007;21:1–20. <https://doi.org/10.18637/jss.v021.i12>.
 40. Reynolds G, Vegh P, Fletcher J. et al. Developmental cell programs are co-opted in inflammatory skin disease. *Science* 2021;371:371. <https://doi.org/10.1126/science.aba6500>.
 41. Aldinger KA, Thomson Z, Phelps IG. et al. Spatial and cell type transcriptional landscape of human cerebellar development. *Nat Neurosci* 2021;24:1163–75. <https://doi.org/10.1038/s41593-021-00872-y>.
 42. Peng M, Wei G, Zhang Y. et al. Single-cell transcriptomic landscape reveals the differences in cell differentiation and immune microenvironment of papillary thyroid carcinoma between genders. *Cell Biosci* 2021;11:39. <https://doi.org/10.1186/s13578-021-00549-w>.