Briefings in Bioinformatics, 2026, 27, bbag524
https://doi.org/10.1093/bib/bbag524
Published: 26 September 2026
Problem Solving Protocol

OXFORD

DeOPUS: cellular deconvolution via optimized power-transformed unmixing with shrinkage

Ha Nguyen[1], Khoi Nguyen[2], Phi Bya[2], Tarik Alafif[3], Tho T. Quan[4], Tin Nguyen[2,5,*]

[1]Department of Computer Science, Wayne State University, 5057 Woodward Avenue, Detroit, MI 48202, United States
[2]Department of Industrial and Systems Engineering, Wayne State University, 4815 Fourth Street, Detroit, MI 48201, United States
[3]Department of Computer Science, Jamoum University College, Umm Al-Qura University, Al-Abdiyyah District, Makkah 21955, Kingdom of Saudi Arabia
[4]Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology (HCMUT), Vietnam National University Ho Chi Minh City, 268 Ly Thuong Kiet Street, Dien Hong Ward, Ho Chi Minh City, Vietnam
[5]Department of Oncology, Karmanos Cancer Institute, Wayne State University School of Medicine, 4100 John R Street, Detroit, MI 48201, United States

*Corresponding author. Tin Nguyen, Department of Industrial and Systems Engineering, Wayne State University, 4815 Fourth Street, Detroit, MI 48201, United States. E-mail: tin@wayne.edu

Abstract

Single-cell RNA sequencing provides high-resolution insights into cellular heterogeneity but its widespread application is often constrained by high costs and technical complexity. Cellular deconvolution serves as a cost-effective alternative by computationally estimating cell-type proportions from bulk RNA-seq data. However, the extreme dynamic range and inherent heteroscedasticity of transcriptomic data pose significant challenges for accurate estimation. Here, we present deconvolution via optimized power-transformed unmixing with shrinkage (DeOPUS), a reference-based deconvolution method that introduces a hierarchical shrinkage transformation (HST) to robustly estimate cellular compositions. DeOPUS integrates multi-level adaptive priors, variance-stabilizing power transformations, and rank-based quantile normalization to mitigate the influence of outliers and high-variance technical noise. We systematically benchmark DeOPUS against eight state-of-the-art methods across 122 tissues and 12 organ systems. DeOPUS consistently outperforms all eight competitors by having mean Pearson ($r = 0.82$) and Spearman ($\rho = 0.77$) correlations. DeOPUS significantly surpasses the second-best method ($r = 0.75$, $\rho = 0.68$) while maintaining the lowest mean squared error (MSE = 0.007). Notably, DeOPUS ranks first in the vast majority of tissues and organ systems using all three metrics, demonstrating unrivaled robustness to increasing cellular complexity. More in-depth validation on 18 bulk datasets with experimentally determined cell-type proportions further confirms DeOPUS's strong performance. DeOPUS is the sole method to achieve positive correlations across all datasets, and achieves the highest accuracy for dominant cell-type identification. DeOPUS is available as an open-source R package at https://github.com/tinnlab/DeOPUS.

Keywords **single-cell analysis, bulk RNA-seq analysis, cell-type deconvolution, immune infiltration, tumor microenvironment**

Introduction

In a traditional bulk RNA sequencing (RNA-seq) experiment, a tissue sample containing hundreds of thousands to millions of cells is homogenized and sequenced. This yields a single expression profile that represents only the average expression across all constituent cells. The average expression is limited when interrogating rare yet functionally dominant populations, such as cancer stem cells or specialized neuronal subtypes, whose critical transcriptional signatures are often masked by the expression profiles of more abundant cells [1–6]. Therefore, accurate quantification of cell-type composition is essential for elucidating disease mechanisms, characterizing the tumor microenvironment, and identifying cell-type-resolved biomarkers [7–11].

Single-cell RNA sequencing (scRNA-seq) provides this cell-level resolution directly, but its widespread adoption in large-cohort studies remains limited by sequencing cost [12–14], dissociation-induced transcriptional artifacts [5, 15], and the practical impossibility of retrospectively profiling archived clinical specimens [16]. The computational inference of cell-type proportions from bulk expression data, termed *cellular deconvolution*, offers a cost-effective alternative that enables researchers to extract cell-type-level insights from vast repositories of existing bulk RNA-seq data, including The Cancer Genome Atlas (TCGA) [17], the Gene Expression Omnibus (GEO) [18, 19], and ArrayExpress [20, 21]. Because these repositories represent billions of dollars in experimental investment, cellular deconvolution can unlock cell-type-level knowledge without repeating costly single-cell assays [16].

Received: March 18, 2026. **Revised:** June 30, 2026. **Accepted:** August 30, 2026

The importance of deconvolution has driven the development of many methods over the past 16 years [16]. These approaches generally model the bulk expression profile of a sample as a weighted linear combination of the expression profiles of the constituent cell types. In this context, the weights are the cell-type proportions to be estimated, subject to non-negativity and a sum-to-one constraint. Methods differ primarily in how they estimate these proportions. They can be broadly classified into three categories: reference-free, semi-reference-free, and reference-based approaches. *Reference-free* methods [22–27] perform unsupervised inference of both the signature matrix and the cell-type proportions simultaneously from the bulk data alone. *Semi-reference-free* methods [28–30] require only a list of marker genes for each cell type. Overall, reference-free and semi-reference-free methods typically rely on non-negative matrix factorization or scoring-based strategies to jointly recover both the signature and the proportions, making them more flexible in data-scarce settings. However, they are generally less accurate than their reference-based counterparts [16], especially for well-annotated cell types and tissues.

Reference-based methods use single-cell data or a precomputed signature matrix to guide proportion inference. The most common approaches involve constrained least-squares (CLS) techniques [31–36], such as DeconPeaker [31] and FARDEEP [34], which minimize the discrepancy between observed and reconstructed bulk profiles via quadratic programming. Weighted CLS (W-CLS) variants [37–41], including MuSiC [37] and DWLS [38], extend this paradigm by down-weighting genes with high cross-subject variance in the reference. Support vector regression (SVR) frameworks [42–46], notably CIBERSORTx [47] and AutoGeneS [44], replace the least-squares objective with a margin-based loss that implicitly regularizes the solution. Recent regression approaches further refine this paradigm by optimizing informative gene selection (GEDIT [48]) or explicitly correcting cross-platform biases (ProM [49]), protocol-specific artifacts introduced by single-nucleus references (DeTREM [50]), transcriptome discrepancies across different cohorts (Harp [51]), and systemic differences in total cellular RNA yield (ReDeconv [52]).

Meanwhile, maximum-likelihood methods, such as RNA-Sieve [53] and DeMixT [54], model expression under parametric distributions. Bayesian approaches, including BayICE [55] and BayesPrism [56], extend this by incorporating prior distributions over cell-type proportions. Recent probabilistic advancements further enhance model reliability and cross-sample robustness. Specifically, BLEND [57] constructs sample-specific references, whereas UBD [58] introduces uncertainty estimation for predicted proportions. Concurrently, deep-learning models, such as DigitalDLSorter [59], DAISM-DNN [60], and Scaden [61], train end-to-end multilayer perceptrons on simulated bulk samples, bypassing signature matrices. To handle bulk-to-reference discrepancies, TAPE [62] utilizes an adaptive autoencoder, while DECODE [63], SCROAM [64], and OmicsTweezer [65] employ domain adaptation to bridge the simulation-to-real gap. Finally, generative architectures, including sc-CMGAN [66], BLUE [67], and DiffFormer [68], are increasingly leveraged for data augmentation and profile reconstruction.

Despite this methodological diversity, heteroscedasticity in RNA-seq data is the unifying barrier that has been addressed only partially. CLS and W-CLS methods, such as MuSiC [37] and DWLS [38], operate in the raw or minimally transformed expression space, where the loss function is dominated by high-abundance housekeeping genes. Gene weighting partially mitigates this issue, but because these weights are estimated strictly from reference data, they often fail to reflect the variance structure of the target bulk samples. SVR methods, such as CIBERSORT [42], implicitly regularize through a margin-based loss, but this regularization is applied uniformly across all genes and does not adapt to gene-specific variance patterns. Bayesian methods, such as BayesPrism [56], can in principle capture heteroscedasticity through their likelihood specification, but they are bounded by parametric forms that do not account for the irregular, non-parametric technical noise and platform shifts frequently encountered in real-world bulk profiles. Finally, deep learning methods, such as Scaden [61] or TAPE [62], learn nonlinear mappings that may implicitly capture some variance structure. However, their reliance on synthetic training data with broadly sampled proportions introduces a severe distributional mismatch that can be more problematic than the heteroscedasticity itself.

Here we introduce a novel method called Deconvolution via Optimized Power-transformed Unmixing with Shrinkage (DeOPUS). The method redefines deconvolution as a variance-moderated signal processing task [69]. DeOPUS overcomes the limitations of existing approaches by introducing a hierarchical shrinkage transformation (HST). By integrating local and global shrinkage priors within an adaptive power-transformation pipeline, DeOPUS selectively attenuates high-variance noise and outlier influence while preserving biologically informative marker signals. This systematic stabilization of the dynamic range ensures that the optimization objective reflects true cellular composition rather than transcript abundance extremes. Through rigorous benchmarking across diverse tissues and organ systems, we demonstrate that DeOPUS consistently outperforms state-of-the-art methods in both precision and stability.

Method

Figure 1 illustrates the workflow of DeOPUS, a reference-based framework that operates without a training or retraining phase. Users only need to provide two inputs: a bulk RNA-seq expression matrix (**B**) and a reference signature matrix (**C**), which is typically derived from high-resolution single-cell or single-nucleus RNA-seq atlases. While DeOPUS ideally utilizes raw count matrices, leveraging an internal depth normalization stage, it remains compatible with normalized data (e.g. TPM or CPM). This is due to the rank-based quantile normalization, which operates on relative rank order rather than absolute expression magnitudes. The foundational innovation of DeOPUS is the HST, which maps expression data into a robust, variance-stabilized feature space. HST represents an adaptive transformation that stabilizes variance and attenuates uninformative high-magnitude signals while increasing the separability of underlying cell types (Supplementary Figs S4 and S5).

Leveraging this transformation, we redefine deconvolution as a constrained optimization problem that minimizes the L^2 divergence between the HST-transformed bulk profile and the HST-transformed mixture model. DeOPUS solves this objective independently for each sample using the L-BFGS-B algorithm [70], enforcing non-negativity constraints to guarantee the biological plausibility of the results. In the final stage, the algorithm projects the raw estimates onto the unit simplex via sum-to-one normalization, yielding cell-type proportions. In the following sections, we provide the mathematical formulation of these components.

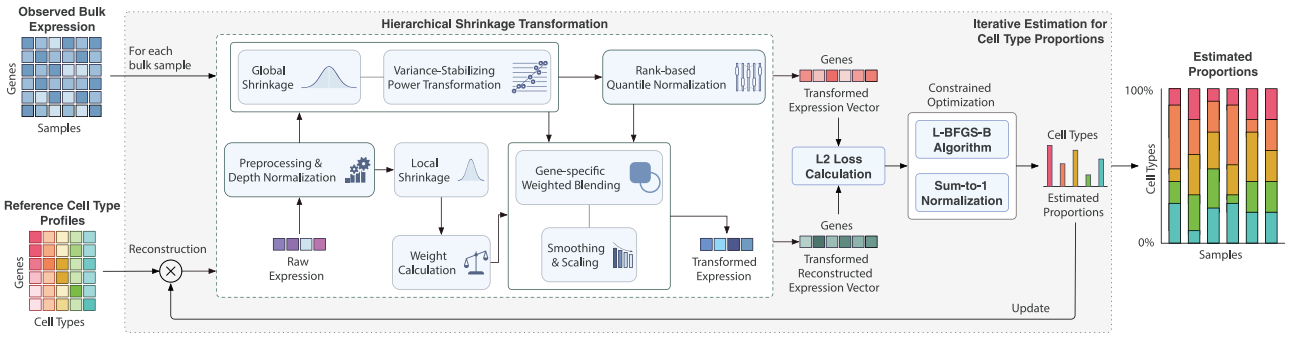

Figure 1 Overview of the DeOPUS deconvolution pipeline. **Input data:** A bulk RNA-seq expression matrix (genes by samples) and a reference cell-type expression matrix (genes by cell types) derived from scRNA-seq. **HST:** Expression values undergo preprocessing and depth normalization, variance-stabilizing power transformation using global shrinkage, rank-based quantile normalization, and gene-specific weighted blending, smoothing, and $\log_2$ scaling to stabilize variance and mitigate outlier effects. **Optimization:** The L^2 loss between the transformed bulk profile and the transformed reconstructed profile is minimized independently for each sample using the L-BFGS-B algorithm with non-negativity constraints. **Output:** Estimated cell-type proportions (sum to one) for each bulk sample.

Problem formulation

Let us denote $\mathbf{B} \in \mathbb{R}^{G\times N}$ (G genes by N samples) as the bulk expression matrix and $\mathbf{C} \in \mathbb{R}^{G\times K}$ (G genes by K cell types) as the reference signature matrix. The standard assumption is that the bulk expression $\mathbf{b}^i$ of sample i is a linear combination of the expression of its constituent cell types:

$$\mathbf{b}^i = p_1^i \mathbf{c}_1 + \ldots + p_K^i \mathbf{c}_K = \mathbf{C}\mathbf{p}^i \tag{1}$$

where $\mathbf{p}^i$ is the vector representing cell-type proportions for sample i, subject to $p_j^i \geq 0$ for all $j \in [1..K]$ and $\sum_{j=1}^K p_j^i = 1$. We note that although biological systems exhibit nonlinear complexities, the linear-sum model effectively avoids overfitting and remains the standard for top-performing deconvolution methods [16, 71]. This linear-sum assumption is further validated by its ability to account for observed patterns in co-expression analyses [72].

Rather than solving this equation directly in the original expression space, DeOPUS transforms the data to a new space as follows:

$$\mathcal{T}\left(\mathbf{b}^i\right) = \mathcal{T}\left(\mathbf{C}\mathbf{p}^i\right), \tag{2}$$

where $\mathcal{T} : \mathbb{R}^G \to \mathbb{R}^G$ is the HST, a deterministic gene-wise mapping that reduces heteroscedasticity, limits the influence of outliers, and emphasizes moderate-abundance signals that are most informative for distinguishing cell types. Because $\mathcal{T}$ is applied identically to both the bulk vector and the reconstructed mixture vector, the optimization preserves scale consistency between both sides of Equation (2).

Hierarchical shrinkage transformation

The HST executes four sequential steps: (i) preprocessing and depth normalization, (ii) variance-stabilizing power transformation using global shrinkage, (iii) rank-based quantile normalization, and (iv) blending, smoothing, and scaling. In the first step, the algorithm clips any negative expression values to zero and normalizes them using a mean-based scale factor. Second, the HST stabilizes variance by transforming the data with a modified power transformation, where a global shrinkage factor governs the data-adaptive exponent. Third, the process maps the power-transformed values to a standard normal distribution via their empirical ranks. Finally, the HST blends the power-transformed signal with the rank-based normalization—utilizing gene-specific weighting and local shrinkage—before applying a soft-thresholding smoothing operation. A final $\log_2$ transformation, informed by the original data range, scales the output to remain biologically interpretable and consistent with standard RNA-seq analysis. Supplementary Fig. S7 presents the ablation study results, demonstrating that each component of the HST pipeline is essential for maintaining deconvolution accuracy.

We apply the same transformation to both the bulk expression and the reconstructed profile ($\mathbf{C}\mathbf{p}$) so that both sides of Equation (2) are on the same scale. The four steps are described below.

Stage 1: Preprocessing and depth normalization

Given an expression vector $\mathbf{q} \in \mathbb{R}^G$, we first truncate negative values to zero: $q_g \leftarrow \max(q_g, 0)$ for all $g \in [1..G]$. To account for differences in sequencing depth, we compute a global scaling factor s based on the mean expression across all genes:

$$s = \max\left(1,\ \frac{100}{G}\sum_{g=1}^{G} q_g\right) \tag{3}$$

In this formula, we multiply the average expression by 100 so that the normalized values $\frac{q_g}{s}$ fall between 0 and 1 for most genes, preventing the subsequent power transformation from producing extreme magnitudes. If all expression values are zero, the scale factor is set to $s = 1$ to avoid division by zero. All subsequent computations use the normalized expression $\tilde{\mathbf{q}} \leftarrow \frac{\mathbf{q}}{s}$. Note that highly expressed genes could inflate s, but this potential vulnerability will be neutralized by the subsequent stages. Specifically, the power transformation (Stage 2) applies nonlinear compression to highly expressed genes using an exponentially decaying global shrinkage factor, while the rank-based quantile normalization (Stage 3) eliminates sensitivity to expression magnitudes entirely by mapping values to rank-based Gaussian scores. Consequently, the downstream optimization remains robust against dominant features.

Stage 2: Variance-stabilizing power transformation

The HST differentially weights genes based on their expression levels to account for varying signal-to-noise ratios. Low-abundance transcripts exhibit susceptibility to stochastic noise and technical dropouts; assigning these features uniform weight would propagate instability through the optimization process. Conversely, genes with ubiquitously

high expression across all cell types frequently lack discriminatory utility.

To address these heterogeneous reliability constraints, the HST incorporates a *global shrinkage factor* $\lambda_g^{(\text{global})}$ that modulates the strength of the power transformation. This factor decays exponentially from a value near 1 toward a floor of ϕ when the normalized expression increases:

$$\lambda_g^{(\text{global})} = (1-\phi)\, e^{-\delta \cdot \tilde{q}_g} + \phi \tag{4}$$

where $\phi = 0.8$ is the global shrinkage floor and $\delta = 0.3$ controls the rate of exponential decay. The floor ϕ guarantees that even the most highly expressed genes retain substantial regularization in the power transformation, preventing extreme values from dominating the transform.

Using the global shrinkage factor, we convert the normalized expression vector $\tilde{\mathbf{q}}$ into a variance-stabilized vector $\mathbf{z}$ as follows:

$$z_g = \frac{\left(\tilde{q}_g + \kappa\right)^{\mu\,\omega\,\lambda_g^{(\text{global})}} - \kappa^{\mu}}{\mu\,\lambda_g^{(\text{global})}} \tag{5}$$

where κ is a small positive offset to avoid singular behavior at zero and control the transform near low expression levels. The parameter μ is the base shape parameter that appears in both the exponent and in the denominator to regulate the overall dynamic range of the output. The factor ω modulates the exponent, slightly increasing its magnitude to enhance compression of high-expression values. Subtracting κ^{μ} centers the transformation near zero for unexpressed genes, while dividing by $\mu\,\lambda_g^{(\text{global})}$ stabilizes the scale of z_g across genes with different shrinkage levels. Because $\lambda_g^{(\text{global})}$ varies with expression, the effective exponent $\mu\,\omega\,\lambda_g^{(\text{global})}$ is smaller for highly expressed genes (where $\lambda_g^{(\text{global})} \approx \phi$) than for lowly expressed ones (where $\lambda_g^{(\text{global})} \approx 1$). This produces stronger compression in the high-expression regime, where heteroscedasticity is typically greatest. In the current implementation, we set $\kappa = 0.1$, $\mu = 2.0$, and $\omega = 1.5$, which provide stable behavior across datasets in our empirical evaluations.

Stage 3: Rank-based quantile normalization

Although the power transformation reduces heteroscedasticity, some outliers and distributional differences between bulk samples and reference profiles may persist. To further improve robustness, the HST constructs a complementary rank-based representation. Specifically, each power-transformed value z_g is converted into a Gaussian quantile score based on its rank among all G genes:

$$r_g = \frac{\text{rank}(z_g)}{G} \tag{6}$$

$$\tilde{z}_g = \Phi^{-1}\big(\text{clip}\big(r_g,\ 0.001,\ 0.999\big)\big) \tag{7}$$

where Φ^{-1} is the standard normal quantile function (the probit function) and the clipping to [0.001, 0.999] prevents infinite values at the tails of the distribution. The resulting $\tilde{z}_g$ is a *quantile-normalized* signal. Consequently, all expression vectors are mapped to approximate Gaussian scores, making the representation invariant to monotone nonlinear distortions in expression—a critical robustness property when comparing bulk and reference profiles processed by different protocols.

Stage 4: Blending, smoothing, and scaling

In the final stage, the power-transformed signal $\mathbf{z}$ and the quantile-normalized signal $\tilde{\mathbf{z}}$ are combined using a gene-specific weighted average. The weight depends on both expression level and local shrinkage. For highly expressed genes, z_g is considered reliable and receives greater weight. For lowly expressed genes, the globally shrunk rank-based signal $\tilde{z}_g$ provides a more stable alternative and therefore contributes more strongly.

We introduce the *local shrinkage factor*, $\lambda_g^{(\text{local})}$ which is a per-gene weight in $(0, 1)$ and decreases monotonically as expression increases, causing low-abundance genes to be pulled more strongly toward zero in the blending step:

$$\lambda_g^{(\text{local})} = \frac{\tau}{\tau + \tilde{q}_g} \tag{8}$$

where $\tau = 0.05$ is the local shrinkage parameter. When $\tilde{q}_g$ is small, $\lambda_g^{(\text{local})}$ is close to 1, meaning the gene is fully shrunk. When $\tilde{q}_g$ is large, $\lambda_g^{(\text{local})}$ approaches zero, indicating that high-confidence genes are not shrunk at all.

The per-gene weight w_g is defined as:

$$w_g = \min\!\left(1,\ \frac{q_g}{10 + \alpha\, q_g}\right) \cdot \left(1 - \lambda_g^{(\text{local})}\right), \tag{9}$$

where $\alpha = 0.01$ is a regularization constant. The first factor, $q_g/(10 + \alpha\, q_g)$, is a soft threshold that tends toward a ceiling of $1/\alpha = 100$ (capped at 1), ensuring that genes with negligible counts contribute minimal power-transformed signal. The second factor, $1 - \lambda_g^{(\text{local})}$, further down-weights genes with high local shrinkage (i.e. low-abundance genes), creating a consistent double penalization of unreliable signals.

The combined signal $t_g^{(\text{blend})}$ is then computed as:

$$t_g^{(\text{blend})} = w_g\, z_g + \left(1 - w_g\right)\, \lambda_g^{(\text{global})}\, \tilde{z}_g. \tag{10}$$

Genes with high w_g are represented primarily by z_g (the power-transformed signal), while genes with low w_g are represented primarily by $\lambda_g^{(\text{global})}\, \tilde{z}_g$ (the globally shrunk quantile signal).

A smoothing step is then applied to the combined output using a rational function that attenuates large values while leaving small values nearly unchanged:

$$t_g = \frac{t_g^{(\text{blend})}}{1 + \delta\, \left| t_g^{(\text{blend})} \right|}, \tag{11}$$

where $\delta = 0.3$. For small $|t_g^{(\text{blend})}| \ll 1/\delta$, the denominator is $\sim$1 and the signal passes through with minimal distortion; for large values, the denominator grows proportionally, bounding the output.

Finally, the smoothed signal is linearly rescaled to match a $\mathbf{log}_2$-scale dynamic range. This normalization ensures that the magnitude of the transformed features is consistent across different expression profiles:

$$t_g \leftarrow \frac{t_g}{\max(\mathbf{t})} \cdot \log_2(\max(\mathbf{q}) + 1). \tag{12}$$

This step normalizes the smoothed vector to unit maximum absolute value and rescales it so that the output range is comparable with a standard $\mathbf{log}_2$ transformation of the original counts. Any values that remain undefined after this step (e.g. due to all entries being zero) are set to zero.

Overall, HST is governed by seven hyperparameters: ϕ (global shrinkage floor), δ (decay rate), μ (base shape parameter), ω (exponent multiplier), κ (offset), τ (local shrinkage scale), and α (regularization constant). Supplementary Section 4 provides a comprehensive explanation of the parameters, their default values and practical ranges (Supplementary Table S1). We also provide a sensitivity analysis demonstrating that DeOPUS is highly robust to variations in these hyperparameters (Supplementary Fig. S2).

Optimization objective and loss function

Following the definition of the HST, deconvolution of each bulk sample $\mathbf{b}^{(i)}$ is formulated as a constrained optimization problem in the transformed space. The goal is to identify a proportion vector $\mathbf{p} \in \mathbb{R}^K$ that minimizes the L^2 distance between the transformed bulk profile and the transformed mixture model:

$$\hat{\mathbf{p}}^{(i)} = \arg\min_{\mathbf{p}^{(i)}} \sum_{g=1}^{G} \left[\mathcal{T}(\mathbf{b}^{(i)})_g - \mathcal{T}(\mathbf{C}\mathbf{p}^{(i)})_g \right]^2 \quad \text{subject to} \quad 0 \le p_k^{(i)} \le 100, \quad k = 1, \ldots, K. \tag{13}$$

The squared L^2 loss provides a smooth, differentiable surface well suited for gradient-based optimization. While the upper bound constraint $p_k \le 100$ is not restrictive in practice, given that proportions are renormalized to the unit simplex post-optimization, it effectively regularizes the search trajectory by preventing unbounded parameters. In this implementation, the objective function assigns uniform weight to all genes. When coupled with the HST, this approach provides implicit variance stabilization without the need for explicit gene-specific noise estimation.

We note that although the underlying mixture model $\mathbf{Cp}$ is linear, the overall objective is nonlinear in $\mathbf{p}$ because the transformation $\mathcal{T}$ is applied to the reconstructed profile and iteratively updates its internal shrinkage weights based on the current value of $\mathbf{p}$. Consequently, closed-form linear solutions are inapplicable, necessitating iterative gradient-based methods.

Optimization algorithm and per-sample parallelization

We minimize the objective function in Equation (13) using the L-BFGS-B algorithm [70], a limited-memory quasi-Newton method designed for box constraints. This choice aligns with the problem's requirements: the parameter dimension K is typically low (ranging from a few to several dozen cell types), gradients are efficiently computable via the chain rule, and the algorithm natively enforces the constraints $0 \le p_k \le 100$ without auxiliary penalty terms.

Optimization commences from a uniform proportion vector $\mathbf{p}^{(0)} = (1/K, \ldots, 1/K)^\top$, serving as a natural, uninformative starting point that places equal prior weight on all cell types. The algorithm allows a maximum of 100 L-BFGS-B iterations per sample, though convergence typically occurs much sooner. In instances of numerical failure (e.g. due to degenerate input profiles), the system initializes the proportion vector to zero for subsequent handling. To accommodate large datasets, DeOPUS treats each bulk sample as an independent optimization task, distributing these processes across parallel computational units to substantially accelerate the overall throughput.

Output normalization

While the L-BFGS-B optimization enforces non-negativity and the upper bound $p_k \le 100$, it does not strictly constrain the proportions to sum to one (i.e. to lie on the probability simplex). To ensure biological interpretability, DeOPUS normalizes the raw estimates $\hat{\mathbf{p}}^{(i)}$ by their total sum after the optimization concludes:

$$\hat{p}_k^{(i)} \leftarrow \frac{\hat{p}_k^{(i)}}{\sum_{j=1}^{K} \hat{p}_j^{(i)}}, \quad \forall k \in [1..K] \tag{14}$$

This step yields final cell-type proportions representing fractional contributions summing to 1 (or 100%). If the total sum is zero (a degenerate case signaling the absence of a detectable cell-type signature), the resulting undefined values (`NaN`) are replaced by zero. The final output of DeOPUS is a matrix $\hat{\mathbf{P}} \in [0,1]^{K \times N}$ where each column contains the normalized cell-type composition of a bulk sample, with entries summing to unity.

Results

Validation using DeconBenchmark and CELLxGENE

We assessed the performance of DeOPUS through the benchmarking framework named DeconBenchmark [16] applied to scRNA-seq data from the CELLxGENE repository [73]. In total, we processed scRNA-seq data spanning 122 human tissues, 467 donors, and 1241 cell types available from the repository. For each tissue, we implemented a donor-pairing strategy to ensure rigorous validation: single-cell data from one donor was held out as the test set for generating pseudo-bulk samples, while data from a different donor in a separate dataset served as the reference to construct the reference expression matrix. This cross-dataset pairing ensures that the reference and test data are independently generated, preventing information leakage.

To construct the reference expression matrix, we first subsampled the training data to a maximum of 1000 cells per cell type to ensure computational tractability. We identified marker genes for each cell type using limma-voom differential expression analysis [74] with TMM normalization [75]. To ensure the reliability of these markers, we retained only genes with log-fold-change (log2FC) larger than 0.5 and adjusted *P*-value smaller than 5%. Furthermore, we included cell types in the final reference if they contained at least 10 cells and possessed at least one significant marker gene. The final signature for each cell type was defined as the mean expression profile across all its constituent cells in the reference data.

For each tissue, we generated a bulk RNA-seq dataset consisting of 512 pseudo-bulk samples. To create each sample, we first introduced realistic compositional sparsity by randomly selecting 80%–100% of the available cell types. We derived the initial cell-type proportions from natural frequencies observed in the test data and transformed them using a square-root function to attenuate extreme ratios, thereby preventing any single cell type from dominating the mixture. We then applied a random perturbation by scaling these proportions with a

factor uniformly sampled between 0.9 and 1.1, followed by renormalization to ensure they summed to one. Finally, we sampled ∼5000 cells with replacement according to these adjusted proportions and averaged their individual expression profiles to create each synthetic bulk sample. This multifaceted approach preserves biologically plausible cell-type compositions while introducing the controlled variability necessary for robust benchmarking.

In total, we generated 1046 datasets from 122 tissues, 467 donors, and 1241 cell types, providing a diverse and representative evaluation across human biology. The total number of generated datasets for any given tissue is constrained by the number of valid donor pairs available in the repository. Because complex tissues with a high number of cell types are inherently rare in single-cell atlases, they yield far fewer donor pairs for data generation. We compared the performance of DeOPUS with eight established methods: MuSiC [37], FARDEEP [34], AutoGeneS [44], AdRoit [76], CIBERSORT [42], Scaden [61], TAPE [62], and DECODE [63]. The former six methods rank among the top performers in a recent benchmark [16] and broadly cover the primary methodological categories, whereas the latter two deep-learning approaches represent recently developed, state-of-the-art entries. We executed each method using its default parameters and author-recommended procedures. For methods requiring additional inputs, such as signature matrices or cell-type-specific markers, we provided the appropriate information derived from the reference scRNA-seq data. All evaluated methods operate within acceptable computational bounds (under 20 min and 5 GB of RAM per dataset). DeOPUS maintains a memory usage of <1 GB, while runtime peaks at 10 min for large datasets (Supplementary Section 9 and Tables S2–S5).

Following deconvolution, we quantitatively assessed the performance of each method by comparing estimated cell-type proportions against the true proportions across all tissues and samples. We utilized three complementary metrics: (i) Pearson correlation that measures the linear agreement between estimated and true cell-type proportions (the higher the better); (ii) Spearman correlation that captures rank-based concordance and outlier robustness (the higher the better); and (iii) MSE that quantifies the magnitude of estimation errors (the smaller the better). These metrics collectively capture both the direction and magnitude of deconvolution accuracy.

Accuracy across 122 tissues and 12 organ systems

Figure 2 shows the assessment results of the nine deconvolution methods across 1046 datasets (122 tissues and 12 organ systems). Overall, DeOPUS consistently outperforms the competing methods across all three metrics. Specifically, Fig. 2A illustrates the distributions of correlation coefficients and MSE values. DeOPUS achieves the highest median Pearson (0.91) and Spearman correlations (0.84), as well as the lowest median MSE (0.002). Notably, DeOPUS also exhibits the lowest performance variability, as demonstrated by the smallest interquartile ranges across all three metrics (Pearson: 0.78–0.96; Spearman: 0.70–0.93; MSE: 0.0006–0.005), further indicating its consistency. These findings are corroborated by Fig. 2B, which displays the average values of the three metrics. In this comparison, DeOPUS maintains the highest average Pearson (0.82) and Spearman correlations (0.77), along with the lowest average MSE (0.007). Finally, the advantage of DeOPUS remains consistent across individual tissues; it ranks first in 66 tissues for Pearson correlation, 78 tissues for Spearman, and 66 tissues for MSE (Fig. 2C). Supplementary Data S1–S3 provide the detailed results for the CELLxGENE data analysis.

Regarding the performance of other models, MuSiC emerges as the second-best method, with median Pearson and Spearman correlations of 0.87 and 0.77, and a median MSE of 0.003. MuSiC ranks first in 36, 46, and 22 tissues for each respective metric. However, its wider interquartile ranges suggest greater sensitivity to tissue variability (Pearson: 0.63–0.96; Spearman: 0.53–0.91; MSE: 0.0008–0.009). Meanwhile, FARDEEP, AutoGeneS, TAPE, and AdRoit achieve moderate performance. These methods respectively yield median Pearson correlations of 0.84, 0.80, 0.79, and 0.77; median Spearman correlations of 0.73, 0.71, 0.67, and 0.67; and median MSE values of 0.004, 0.003, 0.005, and 0.005. These methods rank first in relatively few tissues: FARDEEP in 8, 12, and 5 tissues; AutoGeneS in 10, 24, and 17 tissues; TAPE in 20, 31, and 6 tissues; and AdRoit in 17, 7, and 19 tissues across the respective metrics. CIBERSORT demonstrates lower and more variable performance (median Pearson: 0.75; Spearman: 0.63; MSE: 0.006), and ranks first in 17, 23, and 16 tissues. Scaden yields median Pearson, Spearman, and MSE of 0.67, 0.57, and 0.005, ranking first in 2, 2, and 6 tissues. Meanwhile, DECODE yields median Pearson, Spearman, and MSE values of 0.62, 0.53, and 0.004, respectively, while ranking first in 13, 24, and 14 tissues.

Figure 3 presents the assessment results at the organ-system level. To achieve this, we group the 122 tissues into 12 organ systems and calculate the mean values of the three assessment metrics for each method (Fig. 3A). DeOPUS demonstrates robust performance across all 12 organ systems, ranking first in 10 systems for Pearson correlation, 11 systems for Spearman correlation, and 6 systems for MSE. Importantly, DeOPUS ranks first across all three metrics simultaneously in 6 systems: cardiovascular, integumentary, nervous, reproductive, respiratory, and skeletal. Specifically, DeOPUS achieves its highest Pearson correlation in the nervous and skeletal systems (0.87 for both), its highest Spearman correlation also in the nervous system (0.83), and its lowest MSE in the skeletal system (0.002).

The heatmap of Pearson correlations (Fig. 3B) provides a comprehensive view of the performance trend. DeOPUS consistently exhibits the darkest red cells across most organ systems, indicating the highest correlations. The performance advantage of DeOPUS over the second-best method, TAPE, is most pronounced in the respiratory system, where DeOPUS outperforms TAPE by 0.35 in Pearson correlation (0.82 versus 0.47). Similar patterns emerge in the nervous system, where DeOPUS (0.87) outperforms TAPE (0.73) by 0.14, and the lymphatic system, where the difference is 0.13 (0.71 versus 0.58). In the integumentary system, DeOPUS achieves a score of 0.83 compared with TAPE's 0.75, while DECODE shows notably weak performance (0.33).

The Spearman correlation heatmap (Fig. 3C) reveals similar patterns, with DeOPUS achieving the highest Spearman correlations in 11 out of 12 systems. MuSiC and TAPE have the second highest average Spearman correlation. The largest advantage is observed in the respiratory system (0.74 versus 0.57 for MuSiC, a difference of 0.17). The endocrine system is the only category where DeOPUS does not rank first. Instead, MuSiC achieves the highest Pearson (0.88) and Spearman (0.81) correlations, compared with DeOPUS (0.86 and 0.80, respectively). Notably, Scaden exhibits negative correlations in the cardiovascular system (Pearson: −0.05; Spearman: −0.03), as indicated by gray cells in the heatmaps, highlighting the limitations of deep learning approaches in systems with limited training data. Altogether, these results demonstrate that DeOPUS maintains consistent superiority across diverse organ systems, whereas competing methods show greater variability in system-specific performance.

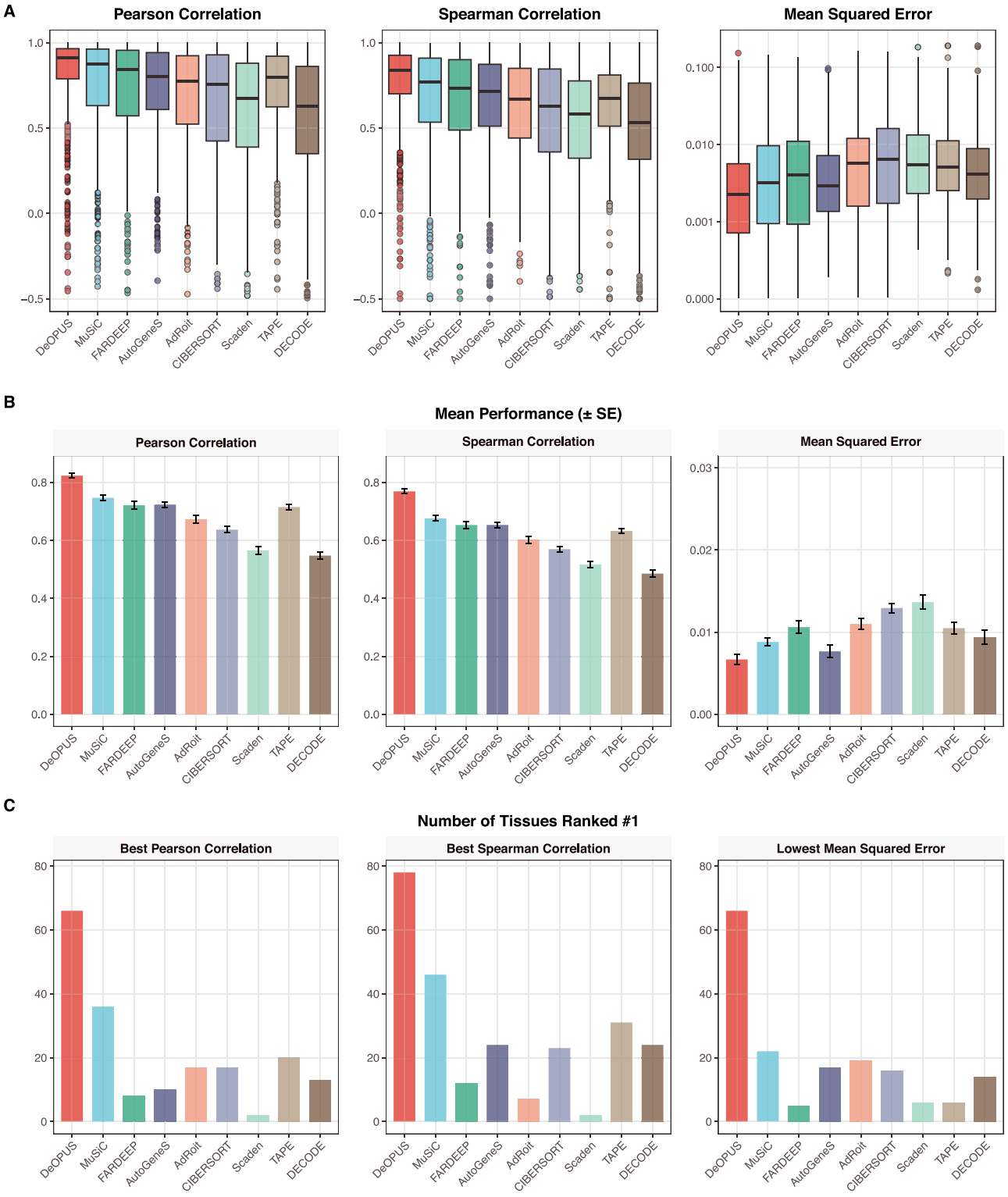

Figure 2 Assessment of DeOPUS, MuSiC, FARDEEP, AutoGeneS, AdRoit, CIBERSORT, Scaden, TAPE, and DECODE across 1046 datasets (122 tissues) using three evaluation metrics. (A) Distributions of Pearson correlation (left), Spearman correlation (middle), and mean squared error (MSE) (right) between estimated cell-type proportions and ground truth across 122 tissues. DeOPUS outperforms all other methods, achieving the highest median Pearson (0.91) and Spearman (0.84) correlations, along with the lowest median MSE (0.002). (B) Average Pearson, Spearman, and MSE values across all datasets. DeOPUS consistently maintains the highest average Pearson (0.82) and Spearman (0.77) correlations and the lowest average MSE (0.007). (C) The frequency with which each method ranks first across the 122 tissues (highest Pearson and Spearman correlations, or lowest MSE). DeOPUS ranks first in 66 tissues for Pearson correlation, 78 for Spearman correlation, and 66 for MSE. Across all assessment metrics, DeOPUS significantly outperforms the competing methods in the vast majority of tissues.

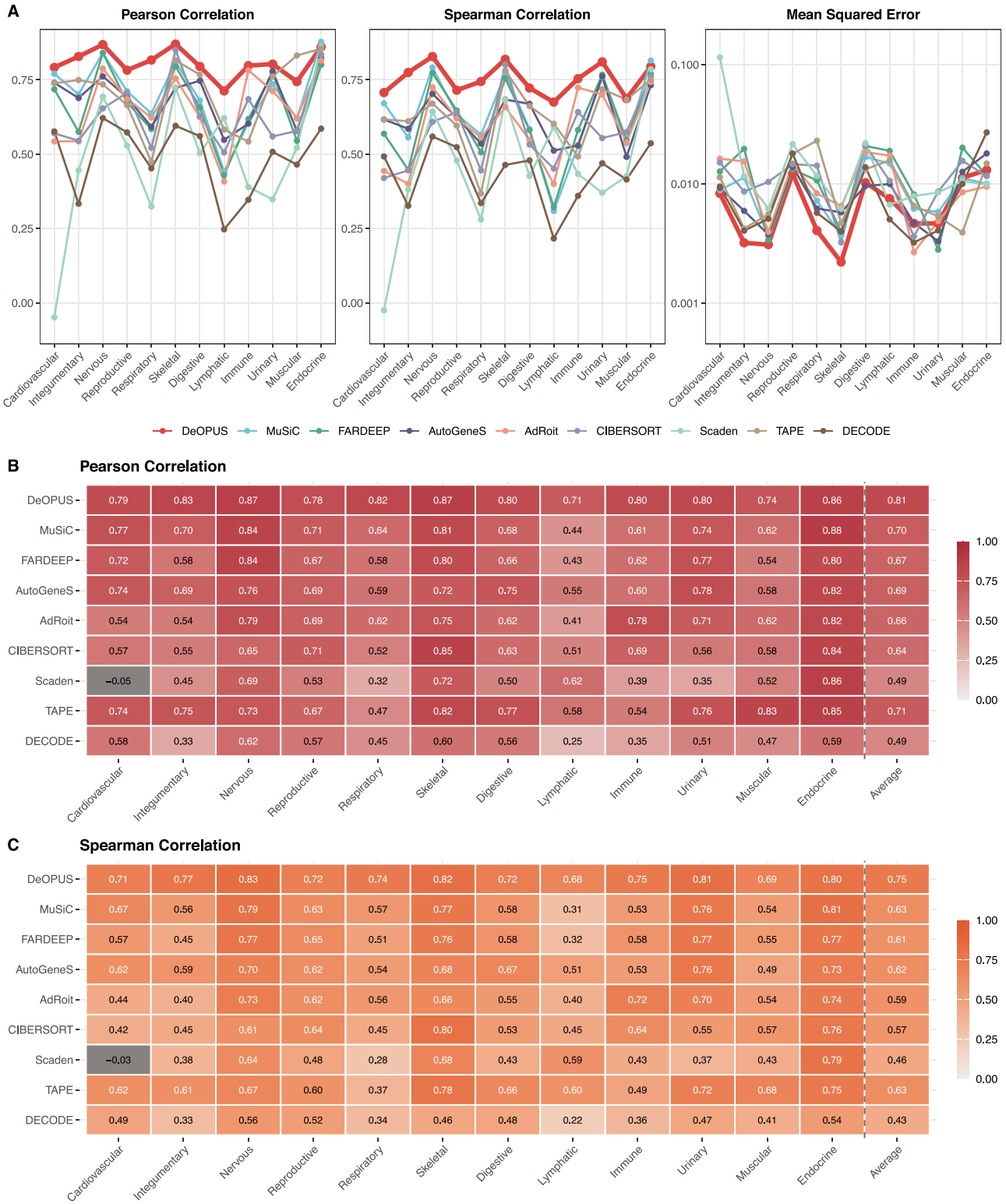

Figure 3 Assessment of the deconvolution methods across 12 organ systems using three evaluation metrics. (A) Average Pearson, Spearman correlation, and MSE. The horizontal axes represent the organ systems while the vertical axes denote the assessment metrics. DeOPUS demonstrates robust performance across all 12 organ systems. It ranks first in 10 systems for Pearson correlation, in 11 systems for Spearman correlation, and in 6 systems for MSE. MuSiC achieves the highest correlations in the endocrine system while TAPE achieves the highest Pearson correlation in the muscular system. (B) Pearson correlation heatmap comparing methods across organ systems. Higher values (darker red) indicate stronger correlation between estimated cell-type proportions and ground truth. DeOPUS consistently exhibits the highest correlations in all systems except the endocrine and muscular systems. (C) Spearman correlation heatmap. The results confirm the same trend as the Pearson correlations, with DeOPUS outperforming all other methods in 11 out of 12 organ systems.

Robustness to cellular complexity

A critical challenge in cellular deconvolution is maintaining accuracy as the number of constituent cell types increases. Greater cellular complexity renders the deconvolution problem increasingly difficult because algorithms must distinguish among increasingly similar cell populations. To evaluate the robustness of DeOPUS under these conditions, we examine the method's accuracy as tissue complexity increases, with the number of cell types per tissue ranging from 2 to 44 (Fig. 4).

We also evaluate the ability of each method to estimate the proportions of low-abundance cell populations by focusing on datasets that have rare cell types (proportion <5%). We stratify the results into four bins based on the true cell-type proportions in each sample: $\leq$ 1%, (1%, 5%], (5%, 10%], and > 10%. DeOPUS achieves the lowest MSE across all proportion categories (Supplementary Fig. S3). This advantage is notable in the rarest bin ($\leq$ 1%), where DeOPUS yields an MSE of 0.00059, compared with 0.00096 of FARDEEP. Similarly, in the (1%, 5%] bin, DeOPUS maintains the lowest MSE (0.00026), outperforming MuSiC (0.00049). DeOPUS remains the best method for highly abundant cell types (> 10% bin).

Figure 4A shows the Pearson correlations of the methods with respect to tissue complexity. In this figure, the horizontal axes represent the number of cell types per tissue, while the vertical axes represent the correlation. Individual tissues are represented by dots, with regression lines indicating the expected performance trends. While the Pearson correlations of most competing methods decline as tissues become more diverse, the regression line for DeOPUS remains flat and consistently above 0.8. Notably, DeOPUS achieves Pearson correlations exceeding 0.85 even in many tissues with >30 cell types—a threshold where other methods fail significantly. In contrast, the Pearson correlations of DECODE, Scaden, and CIBERSORT drop to 0.28, 0.34, and 0.51, respectively. Although MuSiC and FARDEEP exhibit relatively stable trends, their overall correlation levels remain consistently and notably lower than those of DeOPUS.

This robustness is further corroborated by Fig. 4B, which shows the same trend whereby DeOPUS maintains a consistent Spearman correlation at ∼0.8. While MuSiC and FARDEEP exhibit lower yet stable performance, the remaining methods, AutoGeneS, AdRoit, CIBERSORT, Scaden, TAPE, and DECODE, show clear trends of declining performance with respect to tissue complexity. The degradation observed in these methods suggests significant limitations in handling high-dimensional compositional spaces. DECODE and Scaden, e.g. exhibit the steepest declines, likely because these approaches struggle to generalize when training data is insufficient to capture the complex relationships among many cell types.

On both metrics, DeOPUS produced negative values in only 24 of 1046 datasets (2.3%), compared with 29 (2.8%) for TAPE, 42 (4.1%) for AutoGeneS, 50 (4.9%) for MuSiC, and between 5.4% and 11.6% for the remaining methods. These 24 datasets span 16 distinct tissues across eight organ systems. The primary driving factor is high transcriptomic similarity (collinearity) among the cell types in the utilized reference matrices. When reference profiles are highly collinear, the proportion of one cell type can easily be misassigned to a closely related neighbor (more details in Supplementary Section 2).

To evaluate the methods under compositional mismatch, we analyze datasets where the cell-type proportion correlation between the training reference donor and testing donor is below 0.7. Supplementary Fig. S1 shows the Pearson correlation, Spearman correlation, and MSE in this scenario. While all methods show reduced accuracy under this distribution shift, DeOPUS remains the top-performing model, achieving the highest median correlations, lowest median MSE, and tightest interquartile ranges among all nine methods. This demonstrates the robustness of DeOPUS against biological shifts.

The unique resilience of DeOPUS is rooted in its ability to leverage the additional information provided by diverse cell-type signatures. In standard regression frameworks, increasing the number of cell types often amplifies the influence of outlier gene expression, which creates a noisy optimization landscape. DeOPUS overcomes this through its HST, which utilizes local and global shrinkage priors to adaptively down-weight non-informative or noisy genes. By stabilizing the optimization process and preventing overfitting to spurious patterns, DeOPUS effectively converts increased complexity into greater discriminatory power. This robustness is particularly relevant for applications involving complex tissues such as peripheral blood mononuclear cells (PBMCs), the tumor microenvironment, and developing organs, where dozens of distinct cell types may coexist. Collectively, these results demonstrate that DeOPUS not only achieves superior overall accuracy but also maintains consistent performance across tissues of varying cellular complexity, making it a reliable choice for diverse biological applications.

Immune infiltration estimation using 18 bulk datasets

We evaluated the performance of DeOPUS using real bulk datasets with experimentally determined cell-type proportions. Accurate estimation of immune infiltration levels from bulk tumor expression profiles is essential for understanding the tumor microenvironment, predicting responses to immunotherapy, and identifying prognostic biomarkers. This validation provides critical evidence that DeOPUS maintains its superiority when applied to data that capture the full complexity of true biological samples, including platform-specific biases, batch effects, and biological variability not fully represented in simulated data.

We assembled a comprehensive collection of bulk datasets with experimentally measured cell-type proportions from multiple sources (Table 1). These include datasets from the Tumor Deconvolution DREAM Challenge [77] and NIH GEO. These encompass PBMC samples, whole blood samples, *in vitro* admixtures with known mixing proportions, and pseudo-bulk profiles aggregated from scRNA-seq data. Ground-truth cell-type proportions were determined through flow cytometry (FACS), known *in vitro* mixing ratios, or annotated single-cell profiles. The datasets spanned multiple experimental platforms, utilizing both microarray and sequencing technologies that provide a rigorous test for cross-platform generalizability.

To support these analyses, we constructed a unified single-cell reference from the Tabula Sapiens project, encompassing 17 immune and stromal cell types commonly found in the tumor microenvironment: macrophages, monocytes, fibroblasts, endothelial cells, neutrophils, B cells, naïve B cells, memory B cells, $CD4^+$ T cells, memory $CD4^+$ T cells, naïve $CD4^+$ T cells, regulatory T cells, $CD8^+$ T cells, memory $CD8^+$ T cells, naïve $CD8^+$ T cells, myeloid dendritic cells, and immature NK cells. For each cell type, we obtained marker genes from the Cell Annotation Platform and Cell Taxonomy databases, and computed reference expression profiles as the mean expression across all cells of each type.

We applied the deconvolution methods to infer immune infiltration levels using the curated single-cell reference and marker gene sets. Predicted cell-type proportions were compared against the

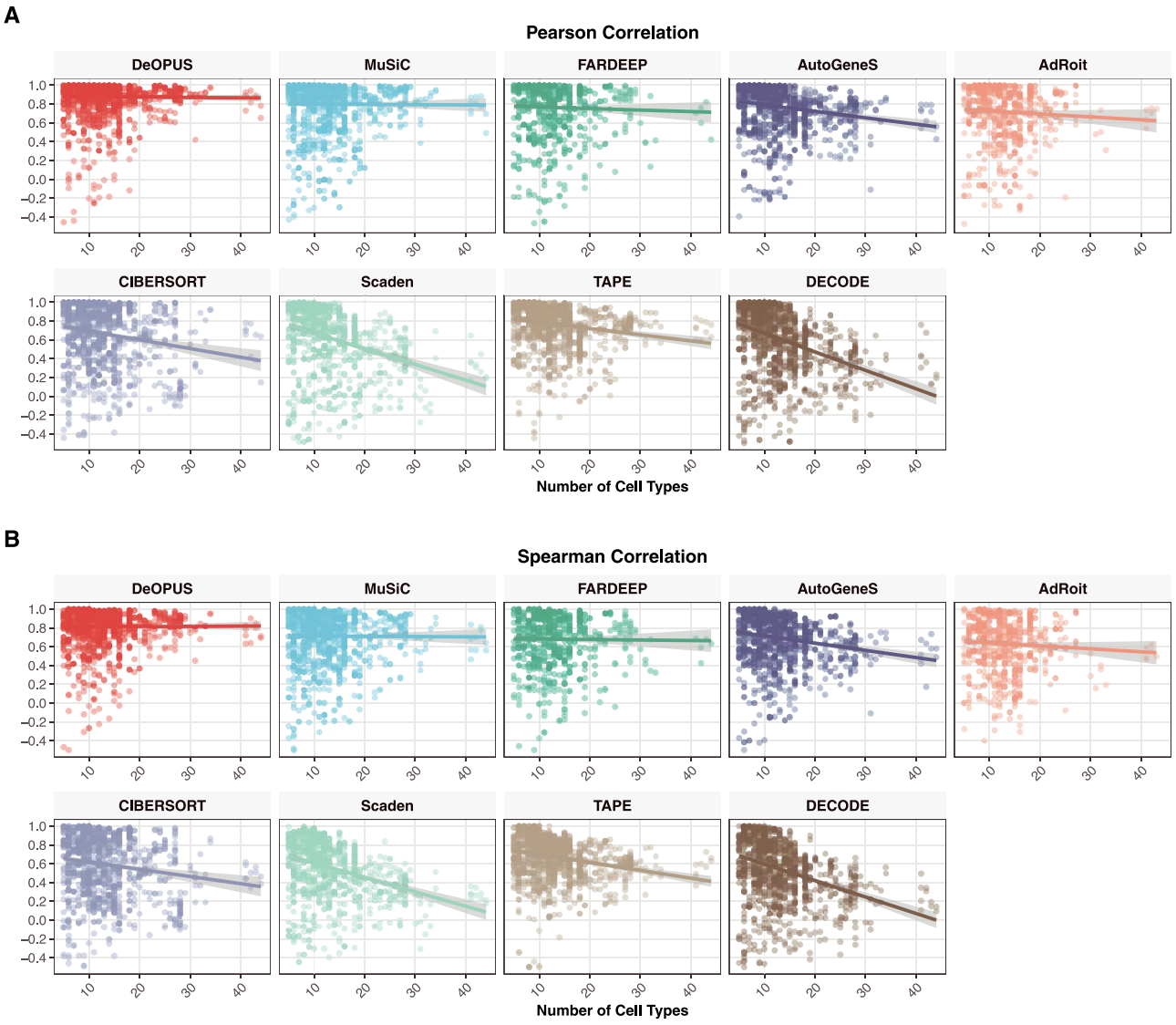

Figure 4 Robustness of deconvolution methods with respect to cellular complexity. (A) Pearson correlations of the nine methods across an increasing number of cell types. The horizontal axes represent the number of cell types per tissue, while the vertical axes represent the correlation. Dots represent individual tissues, and regression lines indicate expected performance trends with 95% confidence intervals. The regression line for DeOPUS remains flat, with the method maintaining Pearson correlations above 0.85 even in many tissues with >30 cell types. In contrast, the Pearson correlations of DECODE, Scaden, and CIBERSORT drop to ∼0.28, 0.34, and 0.51, respectively. Although MuSiC and FARDEEP exhibit relatively stable trends, their overall correlation levels remain consistently and notably lower than those of DeOPUS. (B) Spearman correlations of the competing methods across an increasing number of cell types. These results confirm the same trend, whereby DeOPUS maintains a consistent Spearman correlation at ∼0.8. While MuSiC and FARDEEP exhibit lower yet stable performance, the remaining methods, AutoGeneS, AdRoit, CIBERSORT, Scaden, TAPE, and DECODE, show clear declines in performance as tissue complexity increases.

experimentally validated proportions using both Pearson (r) and Spearman correlations (ρ). For datasets where ground-truth annotations were provided at a coarser granularity than our reference, we aggregated predicted proportions according to the predefined cell-type mapping before computing evaluation metrics. For example, we aggregated naïve and memory B cell predictions to match "B cells" ground-truth labels.

Figure 5 shows the evaluation results of the deconvolution methods across the 18 datasets. Figure 5A illustrates the average Pearson and Spearman correlations for each method. The vertical segments represent the standard error. DeOPUS outperformed all other methods on both metrics, demonstrating robust performance on real-world data. Across the 18 datasets, DeOPUS achieved the highest average Pearson ($r = 0.54$) and Spearman correlations ($\rho = 0.51$). The second best method is AutoGeneS, which yielded Pearson and Spearman correlations of 0.42 and 0.39, respectively, followed by TAPE with 0.40 and 0.32 average correlations. The remaining six methods, MuSiC, CIBERSORT, FARDEEP, Scaden, AdRoit, and DECODE, exhibited substantially lower average Pearson (0.14–0.32) and Spearman correlations (0.17–0.26). The large standard errors for the eight competing methods reflect significant variability across datasets, whereas DeOPUS maintained consistent performance throughout. Supplementary Data S4 provides the detailed results for the bulk data analysis.

Table 1 Real bulk expression datasets used for validation. Each dataset has known experimentally determined cell-type proportions via flow cytometry (FACS) or known in vitro mixing proportions. The second to fourth columns show the name, platform, and the number of samples while the last two columns detail the cell types and the description of each dataset

No	Dataset	Platform	Size	Cell types	Description
1	BALL_coarse	RNA-seq	17	B cells, CD4$^+$ T cells, CD8$^+$ T cells, NK cells, neutrophils, monocytic lineage, fibroblasts, endothelial cells	Coarse-grained *in vitro* admixtures generated by DREAM Challenge
2	BALL_fine	RNA-seq	17	Myeloid dendritic cells, endothelial cells, fibroblasts, macrophages, memory CD4$^+$ T cells, memory CD8$^+$ T cells, monocytes, naive B cells, naive CD4$^+$ T cells, naive CD8$^+$ T cells, neutrophils, NK cells, regulatory T cells, memory B cells	Fine-grained *in vitro* admixtures generated by DREAM Challenge
3	GSE11057 [78]	Microarray	4	Memory CD4$^+$ T cells, naive CD4$^+$ T cells	PBMC samples with sorted T cell subsets via flow cytometry (FACS)
4	GSE199324 [77]	RNA-seq	96	Naive B cells, memory CD4$^+$ T cells, naive CD4$^+$ T cells, regulatory T cells, memory CD8$^+$ T cells, NK cells, monocytes, myeloid dendritic cells, macrophages, fibroblasts, endothelial cells	PBMC samples with detailed immune cell characterization
5	GSE20300 [79]	Microarray	24	Neutrophils, monocytes	Whole blood from pediatric kidney transplant patients (stable versus acute rejection) with sorted cells via flow cytometry (FACS)
6	GSE64385 [29]	Microarray	10	HCT116 cancer cells, monocytes, neutrophils, NK cells, T cells	*In vitro* mixtures of five immune cell types (T, B, NK, monocytes, neutrophils) and HCT116 cancer cells at known proportions
7	GSE64655_coarse_all [80]	RNA-seq	14	B cells, NK cells, neutrophils, monocytic lineage	Coarse-grained blood and ovarian cancer ascites samples with immune cell profiling
8	GSE64655_coarse_challenge [80]	RNA-seq	14	B cells, NK cells, neutrophils, monocytic lineage	Coarse-grained blood and ovarian cancer ascites samples with immune cell profiling provided by DREAM Challenge
9	GSE64655_fine_all [80]	RNA-seq	14	NK cells, neutrophils, monocytes, myeloid dendritic cells	Fine-grained blood and ovarian cancer ascites samples with immune cell profiling
10	GSE64655_fine_challenge [80]	RNA-seq	14	NK cells, neutrophils, monocytes, myeloid dendritic cells	Fine-grained blood and ovarian cancer ascites samples with immune cell profiling provided by DREAM Challenge
11	GSE77343 [81]	Microarray	197	Neutrophils, lymphocytes, monocytes, eosinophils, basophils	Whole blood samples with cell counts via flow cytometry (FACS)
12	monaco	RNA-seq	12	NK cells, monocytes, B cells, CD4$^+$ T cells, CD8$^+$ T cells, neutrophils	PBMC samples generated by Monaco *et al*. immune study [82]
13	newman	Microarray	12	Neutrophils, monocytes, CD8$^+$ T cells, CD4$^+$ T cells, B cells, NK cells	Whole blood samples by Newman *et al*. deconvolution study [47]
14	SDY80	RNA-seq	46	Memory B cells, naive B cells, memory CD4$^+$ T cells, naive CD4$^+$ T cells, regulatory T cells, memory CD8$^+$ T cells, naive CD8$^+$ T cells, monocytes	PBMC samples generated by Tsang *et al*. immune study [83]
15	Wu_coarse_all_cells	RNA-seq	21	B cells, CD4$^+$ T cells, CD8$^+$ T cells, endothelial cells, fibroblasts, monocytic lineage, NK cells, neutrophils	Coarse-grained pseudo-bulk from Wu *et al*. BRCA single-cell data [84]; raw counts summed per cell type within each patient
16	Wu_coarse_challenge_cells	RNA-seq	21	B cells, CD4$^+$ T cells, CD8$^+$ T cells, endothelial cells, fibroblasts, monocytic lineage, NK cells	Coarse-grained pseudo-bulk from Wu *et al*. BRCA single-cell data [84]; QC-filtered to exclude patients with mis-clustering cells
17	Wu_fine_all_cells	RNA-seq	21	Endothelial cells, fibroblasts, macrophages, memory B cells, memory CD4$^+$ T cells, memory CD8$^+$ T cells, monocytes, myeloid dendritic cells, naive B cells, NK cells, regulatory T cells, naive CD4$^+$ T cells, naive CD8$^+$ T cells, neutrophils	Fine-grained pseudo-bulk from Wu *et al*. BRCA single-cell data [84]; raw counts summed per cell type within each patient
18	Wu_fine_challenge_cells	RNA-seq	21	Endothelial cells, fibroblasts, macrophages, memory B cells, memory CD4$^+$ T cells, memory CD8$^+$ T cells, monocytes, myeloid dendritic cells, naive B cells, NK cells, regulatory T cells, naive CD4$^+$ T cells, naive CD8$^+$ T cells, neutrophils	Fine-grained pseudo-bulk from Wu *et al*. BRCA single-cell data [84]; QC-filtered to exclude patients with mis-clustering cells

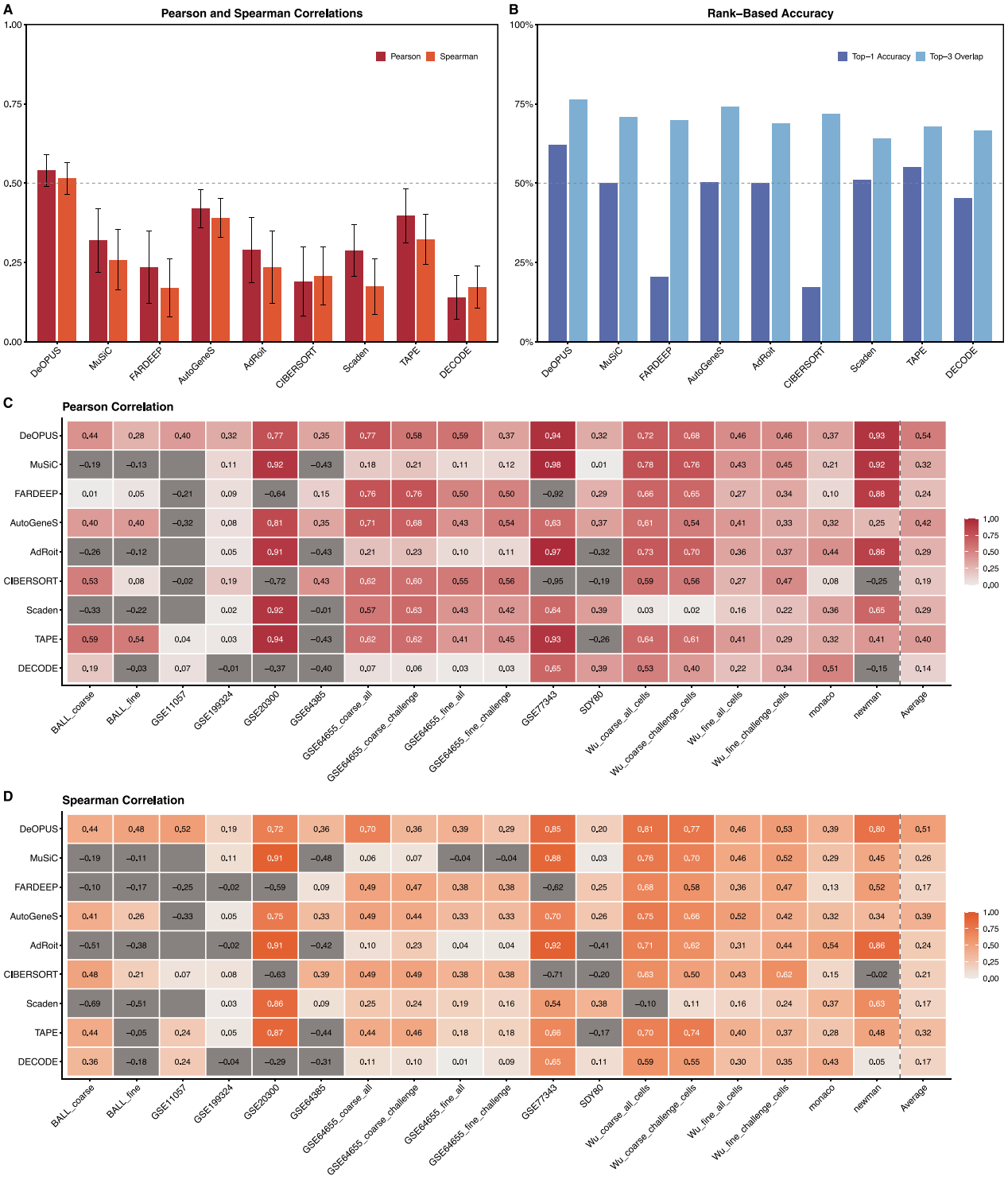

Figure 5 Performance assessment of nine deconvolution methods using 18 real datasets with known cell-type proportions. (A) Correlations between predicted and ground truth cell-type proportions, averaged across 18 real datasets. For each method, the first bar shows the average Pearson correlation while the second bar shows the Spearman correlation across 18 datasets. The vertical segments represent the standard error. DeOPUS outperforms all other methods in having the highest average Pearson and Spearman correlations (0.54 and 0.51, respectively). The second best method is AutoGeneS with Pearson and Spearman correlations of 0.42 and 0.39, respectively. (B) Rank-based accuracy metrics: Top-1 accuracy (correctly identifying the most abundant cell type) and Top-3 overlap (agreement between the three most abundant cell types in predictions versus ground truth). DeOPUS correctly identifies the most abundant cell type in 62% of the samples. The second best method using the Top-1 accuracy metric is TAPE with 55%. DeOPUS also has the highest Top-3 overlap (76%), but all methods perform well using this metric (64%–74%). (C) Pearson correlation heatmap for each dataset and each method. Higher values (darker red) indicate stronger linear agreement; gray indicates weak or negative correlations. (D) Spearman correlation heatmap for each dataset and method. Not only that DeOPUS has the highest average Pearson and Spearman correlations, but also is the only method that has positive correlations in each of the 18 datasets.

Beyond correlation metrics, we assessed the ability to correctly rank cell types by abundance (Fig. 5B). DeOPUS achieved the highest Top-1 accuracy, correctly identifying the most abundant cell type in 62% of the samples, while the Top-1 accuracy of other methods ranged from 17% (CIBERSORT) to 55% (TAPE). The Top-3 overlap metric, measuring agreement between the three most abundant cell types in predictions versus ground truth, reached ∼76% for DeOPUS. All other methods performed reasonably well on this metric (64%–74%), suggesting that most methods can identify highly abundant cell types.

Figure 5C shows the Pearson correlations for each method and dataset, which reveal substantial heterogeneity. Note that for the GSE11057 dataset, MuSiC, AdRoit, and Scaden predict a proportion of exactly zero for target cell types across all samples. Consequently, their correlation coefficients are mathematically undefined due to zero variance. DeOPUS achieved positive correlations across all datasets, with values ranging from 0.28 (BALL_fine) to 0.94 (GSE77343), and other datasets showing strong correlations above 0.7 (GSE20300: $r = 0.77$; GSE64655_coarse_all: $r = 0.77$; newman: $r = 0.93$; Wu_coarse_all_cells: $r = 0.72$). In contrast, other methods frequently exhibited negative correlations, indicating an inverse relationship with ground truth (Scaden: $r = -0.61$ and -0.45 on BALL_coarse and BALL_fine; CIBERSORT: $r = -0.65$ on GSE64655_fine_challenge; AdRoit: $r = -0.48$ on BALL_coarse; TAPE: $r = -0.43$ on GSE64385; DECODE: $r = -0.40$ on GSE64385). These negative values suggest systematic biases in cellular deconvolution in certain methods.

Spearman correlation results confirmed the above findings. Figure 5D shows that DeOPUS maintained positive correlations across all datasets, reaching high values of 0.85 (GSE77343), 0.81 (Wu_coarse_all_cells), and 0.80 (newman). This indicates that DeOPUS reliably identifies which cell types are more or less abundant, even when absolute proportion estimates deviate from ground truth. Other methods showed greater inconsistency, with several exhibiting strongly negative Spearman correlations (Scaden: $\rho = -0.69$ on BALL_coarse, $\rho = -0.51$ on BALL_fine; CIBERSORT: $\rho = -0.71$ on GSE64655_fine_challenge; AdRoit: $\rho = -0.51$ on BALL_coarse; TAPE: $\rho = -0.44$ on GSE64385; DECODE: $\rho = -0.31$ on GSE64385).

To assess biological interpretability, we examined how each method ranks cell-type markers. We ranked the genes by feature-importance scores (the HST weight for DeOPUS, and the cross-cell-type coefficient of variation for competing methods) and evaluated marker rank percentiles across the 18 datasets (Supplementary Section 7 and Fig. S6). DeOPUS assigned these markers the lowest median rank percentile among all methods, demonstrating that the genes it prioritizes during deconvolution tightly coincide with established cell-type markers. This alignment suggests that DeOPUS's gene-level weighting reflects a biologically relevant expression architecture, supporting the interpretability of its results.

Conclusion

We introduce DeOPUS, a reference-based deconvolution method that employs the HST to estimate cell-type proportions from bulk transcriptome data. To evaluate its performance, we compared DeOPUS against eight state-of-the-art methods, MuSiC, FARDEEP, AutoGeneS, AdRoit, CIBERSORT, Scaden, TAPE, and DECODE. Our analysis of 122 tissues from CELLxGENE demonstrates that DeOPUS outperforms existing methods, achieving the highest accuracy as measured by Pearson and Spearman correlations and MSE. Furthermore, the results show that DeOPUS greatly outperforms other methods in 10 out of 12 organ systems for Pearson correlation and 11 out of 12 organ systems for Spearman correlation, with the sole exception of the endocrine system, where MuSiC exhibits slightly higher values for both correlations. The analysis of the 18 bulk datasets further confirms these findings, showing that DeOPUS yields the highest average correlations and is the most consistent among the evaluated methods.

The robust performance of DeOPUS is attributed to three key features: (i) adaptive power transformation that stabilizes variance across the dynamic range of gene expression to prevent highly expressed genes from dominating the optimization; (ii) dual-level shrinkage priors that adaptively down-weight outlier genes while preserving biologically meaningful signals; and (iii) quantile normalization blending that provides robustness to distributional differences between reference and bulk data arising from batch effects or platform variations.

Despite its strong performance, DeOPUS shares limitations common to reference-based methods. Specifically, its accuracy depends on the reference matrix accurately representing cell types present in the bulk sample, and the linear mixing assumption may not fully capture nonlinear effects from cell–cell interactions. Substantial shifts in cell-type proportions demand careful reference curation [16, 85, 86], while unknown cell types and batch effects remain persistent hurdles [87, 88]. Direct validation on cancer data is also limited by a lack of matched, experimentally verified cell-type proportions; clinical correlations indicate biological relevance but do not guarantee individual cell-type precision [89]. Furthermore, because DeOPUS is optimized for transcriptomic data, extending it to chromatin accessibility requires algorithmic re-engineering. Future work will explore unknown content estimation, Bayesian uncertainty quantification, and adaptation to spatial transcriptomics, open chromatin, or DNA methylation.

Key Points

- DeOPUS introduces a hierarchical shrinkage transformation that addresses heteroscedasticity in cellular deconvolution.
- DeOPUS outperforms eight state-of-the-art methods across 122 tissues and 12 organ systems.
- DeOPUS maintains robust accuracy even in tissues with over 30 cell types, where most competing methods degrade substantially.
- Validation on 18 real bulk datasets with experimentally determined cell-type proportions confirms that DeOPUS is the sole method to achieve positive correlations across all datasets.
- DeOPUS is freely available as an open-source R package at https://github.com/tinnlab/DeOPUS.

Author contributions

H.N. developed the methodology, performed benchmarking experiments, and wrote the initial manuscript. K.N. helped with benchmark analysis. P.B. contributed to method development and helped with benchmark analysis. T.A. contributed to method development and reviewed the manuscript. T.Q. contributed to method development and reviewed the manuscript. T.N. wrote the manuscript and supervised the project. All authors reviewed the manuscript.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflicts of interest

No competing interest is declared.

Funding

This work is partially supported by the National Science Foundation (award numbers: 2614807 and 2203236), and the National Cancer Institute (award number: U01CA274573). Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the funding agencies.

Data and code availability

The single-cell data to generate the simulated benchmark datasets are available at the CELLxGENE repository https://cellxgene.cziscience.com/datasets. The 18 real datasets are available at Zenodo https://doi.org/10.5281/zenodo.19050845. The R code for DeOPUS and the benchmarking analysis is freely available on GitHub (https://github.com/tinnlab/DeOPUS).

References

1. Li T, Fan J, Wang B *et al.* TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* 2017;**77**:e108–10.
2. Rooney MS, Shukla SA, Wu CJ *et al.* Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 2015;**160**:48–61.
3. Gentles AJ, Newman AM, Liu CL *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 2015;**21**:938–45.
4. Mahmoud SMA, Lee AHS, Paish EC *et al.* Tumour-infiltrating macrophages and clinical outcome in breast cancer. *J Clin Pathol* 2012;**65**:159–63.
5. Denisenko E, Guo BB, Jones M *et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol* 2020;**21**:30.
6. Nguyen H, Tran D, Tran B *et al.* A comprehensive survey of regulatory network inference methods using single-cell RNA sequencing data. *Brief Bioinform* 2021;**22**:1–15.
7. Roerink SF, Sasaki N, Lee-Six H *et al.* Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* 2018;**556**:457–62.
8. Patel AP, Tirosh I, Trombetta JJ *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 2014;**344**:1396–401.
9. Puram SV, Tirosh I, Parikh AS *et al.* Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 2017;**171**:1611–24.
10. Li B, Severson E, Pignon J-C *et al.* Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol* 2016;**17**:174.
11. Maghsoudi Z, Nguyen H, Tavakkoli A *et al.* A comprehensive survey of the approaches for pathway analysis using multi-omics data integration. *Brief Bioinform* 2022;**23**:bbac435.
12. Yazar S, Alquicira-Hernandez J, Wing K *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* 2022;**376**:eabf3041.
13. Perez RK, Grace Gordon M, Subramaniam M *et al.* Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science* 2022;**376**:eabf1970.
14. Mathys H, Peng Z, Boix CA *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* 2023;**186**:4365–85.
15. Tran D, Nguyen H, Tran B *et al.* Fast and precise single-cell data analysis using hierarchical autoencoder. *Nat Commun* 2021;**12**:1029.
16. Nguyen H, Nguyen H, Tran D *et al.* Fourteen years of cellular deconvolution: methodology, applications, technical evaluation and outstanding challenges. *Nucleic Acids Res* 2024;**52**:4761–83.
17. Grossman RL, Heath AP, Ferretti V *et al.* Toward a shared vision for cancer genomic data. *N Engl J Med* 2016;**375**:1109–12.
18. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;**30**:207–10.
19. Barrett T, Wilhite SE, Ledoux P *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013;**41**:D991–5.
20. Brazma A, Parkinson H, Sarkans U *et al.* ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 2003;**31**:68–71.
21. Rustici G, Kolesnikov N, Brandizi M *et al.* ArrayExpress update–trends in database growth and links to data analysis tools. *Nucleic Acids Res* 2013;**41**:D987–90.
22. Zaitsev K, Bambouskova M, Swain A *et al.* Complete deconvolution of cellular mixtures based on linearity of transcriptional signatures. *Nat Commun* 2019;**10**:2209.
23. Chen L, Chiung-Ting W, Wang N *et al.* debCAM: a bioconductor R package for fully unsupervised deconvolution of complex tissues. *Bioinformatics* 2020;**36**:3927–9.
24. Li Z, Hao W. TOAST: improving reference-free cell composition estimation by cross-cell type differential analysis. *Genome Biol* 2019;**20**:190.
25. Li Z, Guo Z, Cheng Y *et al.* Robust partial reference-free cell composition estimation from tissue expression. *Bioinformatics* 2020;**36**:3431–8.
26. Xie F, Zhou M, Yanxun X. BayCount: a Bayesian decomposition method for inferring tumor heterogeneity using RNA-Seq counts. *Ann Appl Stat* 2018;**12**:1605–27.
27. Rahmani E, Zaitlen N, Baran Y *et al.* Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat Methods* 2016;**13**:443–5.
28. Dimitrakopoulou K, Wik E, Akslen LA *et al.* Deblender: a semi-/unsupervised multi-operational computational method for complete deconvolution of expression data from heterogeneous samples. *BMC Bioinformatics* 2018;**19**:408.
29. Becht E, Giraldo NA, Lacroix L *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* 2016;**17**:218.
30. Zhong Y, Wan Y-W, Pang K *et al.* Digital sorting of complex tissues for cell type-specific gene expression profiles. *BMC Bioinformatics* 2013;**14**:89.
31. Li H, Sharma A, Luo K *et al.* DeconPeaker, a deconvolution model to identify cell types based on chromatin accessibility in ATAC-Seq data of mixture samples. *Front Genet* 2020;**11**:392.
32. 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.
33. Arneson D, Yang X, Wang K. MethylResolver–a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents. *Commun Biol* 2020;**3**:422.

34. Hao Y, Yan M, Heath BR *et al.* Fast and robust deconvolution of tumor infiltrating lymphocyte from expression profiles using least trimmed squares. *PLoS Comput Biol* 2019;**15**:e1006976.
35. Finotello F, Mayer C, Plattner C *et al.* Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med* 2019;**11**:34.
36. Gong T, Szustakowski JD. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics* 2013;**29**:1083–5.
37. 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.
38. Tsoucas D, Dong R, Chen H *et al.* Accurate estimation of cell-type composition from gene expression data. *Nat Commun* 2019;**10**:2975.
39. Dong R, Yuan G-C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol* 2021;**22**:145.
40. Li H, Sharma A, Ming W *et al.* A deconvolution method and its application in analyzing the cellular fractions in acute myeloid leukemia samples. *BMC Genomics* 2020;**21**:652.
41. Racle J, de Jonge, Baumgaertner P *et al.* Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* 2017;**6**:e26476.
42. Newman AM, Liu CL, Green MR *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;**12**:453–7.
43. Zhang W, Hanwen X, Qiao R *et al.* ARIC: accurate and robust inference of cell type proportions from bulk gene expression or DNA methylation data. *Brief Bioinform* 2022;**23**:bbab362.
44. Aliee H, Theis FJ. AutoGeneS: automatic gene selection using multi-objective optimization for RNA-seq deconvolution. *Cell Syst* 2021;**12**:706–15.
45. Frishberg A, Peshes-Yaloz N, Cohn O *et al.* Cell composition analysis of bulk genomics using single-cell data. *Nat Methods* 2019;**16**:327–32.
46. Fernández EA, Mahmoud YD, Veigas F *et al.* Unveiling the immune infiltrate modulation in cancer and response to immunotherapy by MIXTURE-aAn enhanced deconvolution method. *Brief Bioinform* 2021;**22**:bbaa317.
47. Newman AM, Steen CB, Liu CL *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;**37**:773–82.
48. Nadel BB, Lopez D, Montoya DJ *et al.* The gene expression deconvolution interactive tool (GEDIT): accurate cell type quantification from gene expression data. *GigaScience* 2021;**10**:giab002.
49. Wang L, Izadmehr S, Sfakianos JP *et al.* Single-cell transcriptomic-informed deconvolution of bulk data identifies immune checkpoint blockade resistance in urothelial cancer. *iScience* 2024;**27**:109928.
50. O'Neill NK, Stein TD, Junming H *et al.* Bulk brain tissue cell-type deconvolution with bias correction for single-nuclei RNA sequencing data using DeTREM. *BMC Bioinformatics* 2023;**24**:349.
51. Nozari Z, Hüttl P, Simeth J *et al.* Harp: data harmonization for computational tissue deconvolution across diverse transcriptomics platforms. *Bioinformatics* 2025;**41**:btaf455.
52. Songjian L, Yang J, Yan L *et al.* Transcriptome size matters for single-cell RNA-seq normalization and bulk deconvolution. *Nat Commun* 2025;**16**:1246.
53. Erdmann-Pham DD, Fischer J, Hong J *et al.* Likelihood-based deconvolution of bulk gene expression data using single-cell references. *Genome Res* 2021;**31**:1794–806.
54. Wang Z, Cao S, Morris JS *et al.* Transcriptome deconvolution of heterogeneous tumor samples with immune infiltration. *IScience* 2018;**9**:451–60.
55. Tai A-S, Tseng GC, Hsieh W-P. BayICE: a Bayesian hierarchical model for semireference-based deconvolution of bulk transcriptomic data. *Ann Appl Stat* 2021;**15**:391–411.
56. 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 Cancer* 2022;**3**:505–17.
57. Huang P, Cai M, McKennan C *et al.* BLEND: probabilistic cellular deconvolution with individualized single-cell reference integration. *Genome Biol* 2025;**26**:328.
58. Cheng Y, Lin C, Li H *et al.* UBD: incorporating uncertainty in cell type proportion estimates from bulk samples to infer cell-type-specific profiles. *Brief Bioinform* 2026;**27**:bbaf711.
59. Torroja C, Sanchez-Cabo F. Digitaldlsorter: deep-learning on scRNA-Seq to deconvolute gene expression data. *Front Genet* 2019;**10**:978.
60. Lin Y, Haojun Li X, Xiao LZ *et al.* DAISM-DNNXMBD: highly accurate cell type proportion estimation with in silico data augmentation and deep neural networks. *Patterns* 2022;**3**:100440.
61. Menden K, Marouf M, Oller S *et al.* Deep learning-based cell composition analysis from tissue expression profiles. *Sci Adv* 2020;**6**:eaba2619.
62. Chen Y, Wang Y, Chen Y *et al.* Deep autoencoder for interpretable tissue-adaptive deconvolution and cell-type-specific gene analysis. *Nat Commun* 2022;**13**:6735.
63. Zhao T, Liu R, Sun Y *et al.* DECODE: deep learning-based common deconvolution framework for various omics data. *Nat Methods* 2026;**23**:596–608.
64. Guo X, Huang Z, Fen J *et al.* Highly accurate estimation of cell type abundance in bulk tissues based on single-cell reference and domain adaptive matching. *Adv Sci* 2024;**11**:2306329.
65. Yang X, Zhao F, Ren T *et al.* OmicsTweezer: a distribution-independent cell deconvolution model for multi-omics data. *Cell Genom* 2025;**5**:100950.
66. Nishikawa T, Lee M, Amau M. New generative methods for single-cell transcriptome data in bulk RNA sequence deconvolution. *Sci Rep* 2024;**14**:4156.
67. Zhu S, Wang Z, Bunting KD *et al.* Deconvolving cell-type-specific gene expression profiles from bulk RNA-seq samples. *PLoS Comput Biol* 2026;**22**:e1014101.
68. Liu Y, Sun J, Li H *et al.* A transformer-based deep diffusion model for bulk RNA-Seq deconvolution. *Biology* 2025;**14**:1150.
69. Bayat F, Libbrecht M. VSS: variance-stabilized signals for sequencing-based genomic signals. *Bioinformatics* 2021;**37**:4383–91.
70. Zhu C, Byrd RH, Peihuang L *et al.* Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization. *ACM Trans Math Softw* 1997;**23**:550–60.
71. Wolfram-Schauerte M, Vogel T, Tuoken H *et al.* Approaching the holistic transcriptome—convolution and deconvolution in transcriptomics. *Brief Bioinform* 2025;**26**:bbaf388.
72. Farahbod M, Pavlidis P. Untangling the effects of cellular composition on coexpression analysis. *Genome Res* 2020;**30**:849–59.
73. CZI Cell Science Program, Abdulla S, Aevermann B *et al.* CZ CELLxGENE discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Res* 2025;**53**:D886–900.
74. Ritchie ME, Phipson B, Di W *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47.
75. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40.

76. Yang T, Alessandri-Haber N, Fury W *et al.* AdRoit is an accurate and robust method to infer complex transcriptome composition. *Commun Biol* 2021;**4**:1218.
77. White BS, de Reyniès, Newman AM *et al.* Gentles, and tumor deconvolution DREAM challenge consortium. Community assessment of methods to deconvolve cellular composition from bulk gene expression. *Nat Commun* 2024;**15**:7362.
78. Abbas AR, Wolslegel K, Seshasayee D *et al.* Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One* 2009;**4**:e6098.
79. Shen-Orr SS, Tibshirani R, Khatri P *et al.* Cell type-specific gene expression differences in complex tissues. *Nat Methods* 2010;**7**:287–9.
80. Hoek KL, Samir P, Howard LM *et al.* A cell-based systems biology assessment of human blood to monitor immune responses after influenza vaccination. *PLoS One* 2015;**10**:e0118528.
81. Shannon CP, Balshaw R, Chen V *et al.* Enumerateblood–an R package to estimate the cellular composition of whole blood from Affymetrix gene ST gene expression profiles. *BMC Genomics* 2017;**18**:43.
82. Monaco G, Lee B, Weili X *et al.* RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep* 2019;**26**:1627–40.
83. Tsang JS, Schwartzberg PL, Kotliarov Y *et al.* Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* 2014;**157**:499–513.
84. Wu SZ, Al-Eryani G, Roden DL *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* 2021;**53**:1334–47.
85. Cobos FA, Alquicira-Hernandez J, Powell JE *et al.* Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun* 2020;**11**:1–14.
86. Li M, Yuqing S, Tang Y *et al.* Evaluating deconvolution methods using real bulk RNA-expression data for robust prognostic insights across cancer types. *Genome Biol* 2026;**27**:38.
87. Maden SK, Kwon SH, Huuki-Myers L *et al.* Challenges and opportunities to computationally deconvolve heterogeneous tissue with varying cell sizes using single-cell RNA-sequencing datasets. *Genome Biol* 2023;**24**:288.
88. Ivich A, Davidson NR, Grieshober L *et al.* Missing cell types in single-cell references impact deconvolution of bulk data but are detectable. *Genome Biol* 2025;**26**:86.
89. Luo S, Zhu M, Lin L *et al.* DECA: harnessing interpretable transformer model for cellular deconvolution of chromatin accessibility profile. *Brief Bioinform* 2025;**26**:bbaf069.

Briefings in Bioinformatics, 2026, **27**, bbag524
https://doi.org/10.1093/bib/bbag524
Problem Solving Protocol