Contents

Supplementary

Methods

Construction of Simulated Benchmarking Data

Because paired ground-truth data linking tissue-level expression profiles to cell-type-specific relative contributions to functional programs are generally unavailable in real datasets, we primarily evaluated model performance using simulation-based benchmarks. The simulations were designed to generate pseudo-bulk and pseudo-spot samples from single-cell data, each paired with controllable ground-truth labels of cell-type-specific functional program contributions.

For pseudo-bulk generation, we randomly mixed single cells according to sampled cell-type proportions. Specifically, for each pseudo-bulk sample, a cell-type proportion vector was drawn from a uniform random distribution, and a total of n cells were sampled from each cell type according to this vector, with n typically set to 100. The expression profiles of the sampled cells were summed to obtain the mixed expression input. The corresponding ground truth was calculated by summing single-cell functional program activity scores across the sampled cells, aggregating them by cell type, and normalizing each cell type's cumulative score by the total functional program activity.

For pseudo-spot generation, we used the spatial coordinates of cells in spatial transcriptomics data. Local cells were aggregated using fixed-size sliding windows across the tissue space to mimic spatial spots composed of multiple cells. The window size was chosen such that each pseudo-spot contained an average of approximately 6–7 cells, comparable to the cellular composition of 10x Visium spots. The expression profile of each pseudo-spot was obtained by summing the expression counts of all cells within the window. Its ground-truth functional contribution label was computed analogously by aggregating single-cell functional program activity scores by cell type and normalizing them to relative contributions.

Selection of KEGG Pathways as Functional Programs

We used KEGG pathways as representative functional programs because they provide well-curated gene sets with explicit biological annotations and are widely used for interpreting coordinated molecular activities in transcriptomic studies. In this study, a functional program was operationally defined as a predefined gene set representing a coherent biological process or signaling activity. KEGG pathways therefore offer a standardized and interpretable framework for quantifying pathway-level activity in single cells and for evaluating cell-type-specific contributions to functional signals in mixed expression profiles. For each dataset, we first identified cell-type-associated differentially expressed genes and performed enrichment analysis to select the top ten KEGG pathways, ensuring that the selected programs were both biologically interpretable and relevant to the cellular composition of each dataset. Importantly, the proposed framework is not restricted to KEGG pathways and can be applied to other predefined gene-set collections, such as GO biological processes, Reactome pathways, or custom functional signatures.

Binary classification of vaccination states

To evaluate whether FuncDECODE-inferred cell type–biological process contribution profiles contain phenotype-discriminative information, we performed three binary classification analyses based on vaccination time points: T1 versus T2, T3 versus T4, and T1 versus T3. For each task, samples from the two corresponding time points were selected and randomly split into training and test sets at a ratio of 7:3 with stratification by class label.

FuncDECODE-derived cell type–pathway contribution profiles were used as the primary input features. CIBERSORTx-inferred cell-type abundance profiles were included as a conventional cell-composition-based reference to provide a baseline for evaluating whether FuncDECODE captured additional functional information beyond cell-type abundance. For each feature set, an L1-regularized logistic regression model was trained. Features were standardized using z-score normalization, and the regularization strength was selected by five-fold cross-validation within the training set using AUC as the optimization metric. The discriminative ability of each feature set was evaluated on the held-out test set using receiver operating characteristic curves and the area under the ROC curve.

Risk stratification and cross-cohort validation using FuncDECODE-derived cell type–functional program features

Risk stratification using FuncDECODE-derived pathway-cell-type features. We used the TCGA-BRCA FuncDECODE pathway-cell-type features that were nominally associated with overall survival in the adjusted Cox analysis ($p < 0.05$) as input features for risk stratification. For TCGA samples, the selected FuncDECODE features were standardized and clustered into two groups using Ward-linkage agglomerative clustering. The two clusters were assigned as low- or high-risk groups according to their overall survival direction, with the cluster showing a higher event rate and shorter survival time labelled as high risk. Kaplan-Meier survival curves were then generated for overall survival and progression-free interval, and group differences were evaluated using the log-rank test. For comparison, the same two-cluster survival analysis was also applied to 21-gene panel expression features and CIBERSORTx-derived cell-type fractions. To assess transferability of the FuncDECODE risk structure, an XGBoost classifier was trained using TCGA FuncDECODE features and TCGA-derived risk labels, and then applied to METABRIC FuncDECODE features to predict high- and low-risk groups. METABRIC overall survival and progression-free interval were subsequently evaluated by Kaplan-Meier analysis and log-rank testing.

Cross-cohort comparison of FuncDECODE risk-associated pathway-cell-type changes. To interpret the FuncDECODE-defined risk groups, we quantified differential FuncDECODE contributions between high- and low-risk samples in both TCGA and METABRIC. For each shared pathway-cell-type feature, log2 fold change was computed as $\log_2((\text{mean High risk} + \varepsilon) / (\text{mean Low risk} + \varepsilon))$, and differences between risk groups were tested using a two-sided Mann-Whitney U test. TCGA and METABRIC log2 fold changes were paired by pathway-cell-type feature to assess cross-cohort directional concordance. Split bubble plots were used to visualize the paired effects, with the left and right halves representing TCGA and METABRIC, respectively; color indicated

log2 fold change and bubble size indicated statistical significance.

Spatial domain refinement using FuncDECODE-derived information

To evaluate whether the cell-type-specific biological process contribution profiles inferred by FuncDECODE could provide additional biologically meaningful information for spatial transcriptomics analysis, we designed a feature-guided spatial domain refinement strategy. The initial spatial domains were obtained using STAligner, and the FuncDECODE-derived cell type–pathway contribution matrix was used as an additional functional feature representation for each spatial spot. The refinement procedure was restricted to spatially adjacent domain boundaries and internal outlier spots, thereby avoiding global reassignment of spatial domains.

First, spatial neighborhood relationships were constructed based on the spatial coordinates of spots. For each spot, its k nearest spatial neighbors were identified using a KD-tree-based nearest-neighbor search. Based on the initial spatial domain labels, we counted the number of cross-domain neighboring edges between each pair of spatial domains. Domain pairs with sufficient contact frequency were regarded as adjacent domain pairs and were selected for boundary refinement.

For each adjacent domain pair, we identified boundary spots as those assigned to one domain but having a substantial fraction of neighboring spots from the other domain. To determine which FuncDECODE-derived feature best separated the two adjacent domains, each cell type–pathway contribution feature was evaluated by its ability to distinguish spots from the two domains using the area under the receiver operating characteristic curve. Because the discriminative direction was not fixed, the effective AUC was defined as the larger value between AUC and $1 - \mathrm{AUC}$. Only features with sufficient valid values and adequate discriminative ability were retained. The feature with the highest effective AUC was selected as the representative discriminative feature for that domain pair.

Next, a one-dimensional Gaussian model was fitted separately for the selected feature in the two adjacent domains. For each boundary spot, posterior probabilities of belonging to the two domains were estimated using the fitted Gaussian distributions and the empirical proportions of the two domains as priors. A spot was reassigned to the neighboring domain only when its posterior probability clearly favored the alternative domain by a predefined probability margin. This conservative criterion ensured that only boundary spots with strong functional evidence supporting reassignment were modified. The above procedure was iteratively applied across all adjacent domain pairs for a limited number of rounds until no further label changes occurred.

In addition to pairwise boundary refinement, we further performed a local smoothing step to correct isolated internal outlier spots. For each spatial domain, stable FuncDECODE-derived features were selected based on their low within-domain coefficient of variation and sufficient valid value coverage. For each spot, we examined the labels of its spatial neighbors. If the majority of neighboring spots belonged to a different domain and the current label was poorly supported by the local neighborhood, the spot was further evaluated using the stable features of the neighboring majority domain. Specifically, the spot was compared with the distribution of stable features in that domain using standardized z-scores. If a sufficient fraction of tested features fell within the expected range of the neighboring majority domain, the spot was reassigned to that domain. This procedure was also performed iteratively with a limited number of rounds.

Overall, this refinement strategy integrates local spatial continuity with FuncDECODE-derived

cell-type-specific biological process contribution features. Rather than replacing the original spatial domain detection method, it uses the functional contribution profiles inferred by FuncDECODE to refine uncertain boundary regions and isolated spatial outliers. Therefore, improvement in spatial domain consistency after refinement provides evidence that FuncDECODE captures biologically meaningful information that can be used as a general feature representation for downstream spatial transcriptomics analyses.

Tumor region annotation based on cancer-associated marker gene scores

In the prostate cancer spatial transcriptomics dataset, tumor regions were annotated based on detectable cancer-associated marker genes. We first filtered a predefined set of tumor-associated genes to retain 27 genes that were present in the dataset, including KLK2, KLK3, TRPM4, H2AFJ, KLK4, KRT18, TSPAN1, FXYD3, STEAP2, GOLM1, ABCC4, SPDEF, SORD, KRT8, TFF3, MLPH, DHRS7, PDLIM5, PRAC1, VIM, NUPR1, CLDN3, TPD52, NKX3-1, AGR2, SLC44A4 and RAMP1. We then calculated a tumor signature score for each spot using the scanpy.tl.score_genes function. This function computes the average expression of the target gene set and subtracts the average expression of a control gene set with comparable expression levels, thereby quantifying the enrichment of tumor-associated transcriptional programs in each spot. Next, we summarized the average tumor signature score for each spatial cluster defined by the mclust_refined_feature_smooth annotation. Clusters 5 and 2, which showed the highest average tumor signature scores, were annotated as tumor regions, whereas the remaining clusters were annotated as non-tumor regions. The resulting tumor_region annotation was used for subsequent region-stratified spatial analyses.

Spatial autocorrelation analysis using Moran’s I

To evaluate whether FuncDECODE-inferred pathway-cell-type contributions were spatially organized, Moran’s I was calculated for each pathway-cell-type feature within each slide and tissue region. Tumor and non-tumor regions were analyzed separately. For each slide-region pair, spatial coordinates were used to construct an undirected k-nearest-neighbor graph with $k = 6$, and slide-region pairs with fewer than 20 spots were excluded. For a given FuncDECODE feature, Moran’s I was computed across spots as:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where n is the number of spots, x_i and x_j are the FuncDECODE contribution values at spots i and j, $\bar{x}$ is the mean contribution value across spots in the slide-region, w_{ij} is the spatial adjacency weight between spots i and j, and $S_0 = \sum_i \sum_j w_{ij}$ is the total spatial weight. Positive Moran's I values indicate that neighboring spots tend to have similar contribution values, reflecting spatial clustering, whereas values close to zero indicate spatial randomness.

To assess whether the observed spatial autocorrelation exceeded random expectation, we generated a shuffled null distribution for each slide-region-feature combination by randomly permuting FuncDECODE contribution values across spots within the same slide and region while keeping the spatial graph fixed. This procedure was repeated 20 times per feature. Observed Moran's I values were then compared with shuffled Moran's I values across all slide-region-feature combinations and summarized separately for tumor and non-tumor regions.

Tumor–non-tumor boundary axis analysis

For each treatment-naive prostate cancer section, tumor and non-tumor spots were defined using the region annotation. To construct a tumor-to-non-tumor spatial axis, spatially connected tumor components were first identified within each section. Specifically, the median nearest-neighbor distance among spots in each section was calculated, and tumor spots located within 1.5 times this distance were considered spatially connected. Tumor components containing fewer than 20 spots were excluded to reduce the influence of small isolated tumor annotations. The tumor-non-tumor interface was then defined using retained tumor components. Spots were considered boundary-near if a tumor spot had at least one neighboring non-tumor spot, or if a non-tumor spot had at least one neighboring tumor spot. For every spot, the Euclidean distance to the nearest boundary-near spot was calculated. Distances were assigned negative values for tumor spots and positive values for non-tumor spots, yielding a signed distance to the tumor boundary. Based on the signed distance, spots were assigned to five ordered boundary bins: deep tumor, inner tumor, boundary, adjacent non-tumor, and distant non-tumor. The bin width was set to the median nearest-neighbor distance within each section. For each section, FuncDECODE cell type-functional program feature values were z-score normalized across spots. Section-level mean z-scores were then calculated for each feature within each boundary bin. To avoid dominance by sections with more spots, the final feature-bin value was obtained by averaging these section-level bin means across eligible sections. For each feature and each bin, spot-level significance was evaluated by comparing the section-normalized feature z-scores of spots in the current bin against spots in all other bins using a two-sided Mann-Whitney U test. These feature-bin mean z-scores and bin-wise significance values were used for downstream visualization.

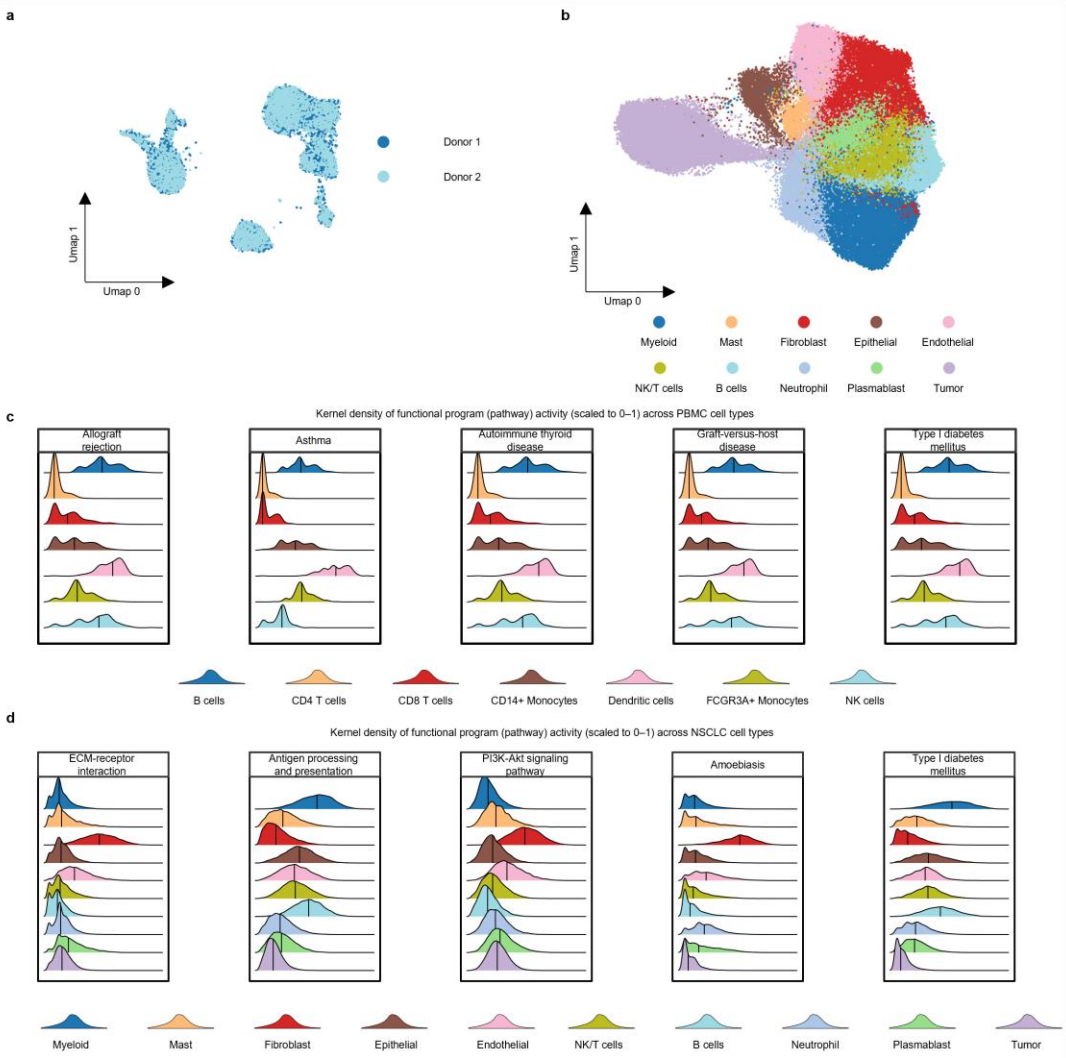
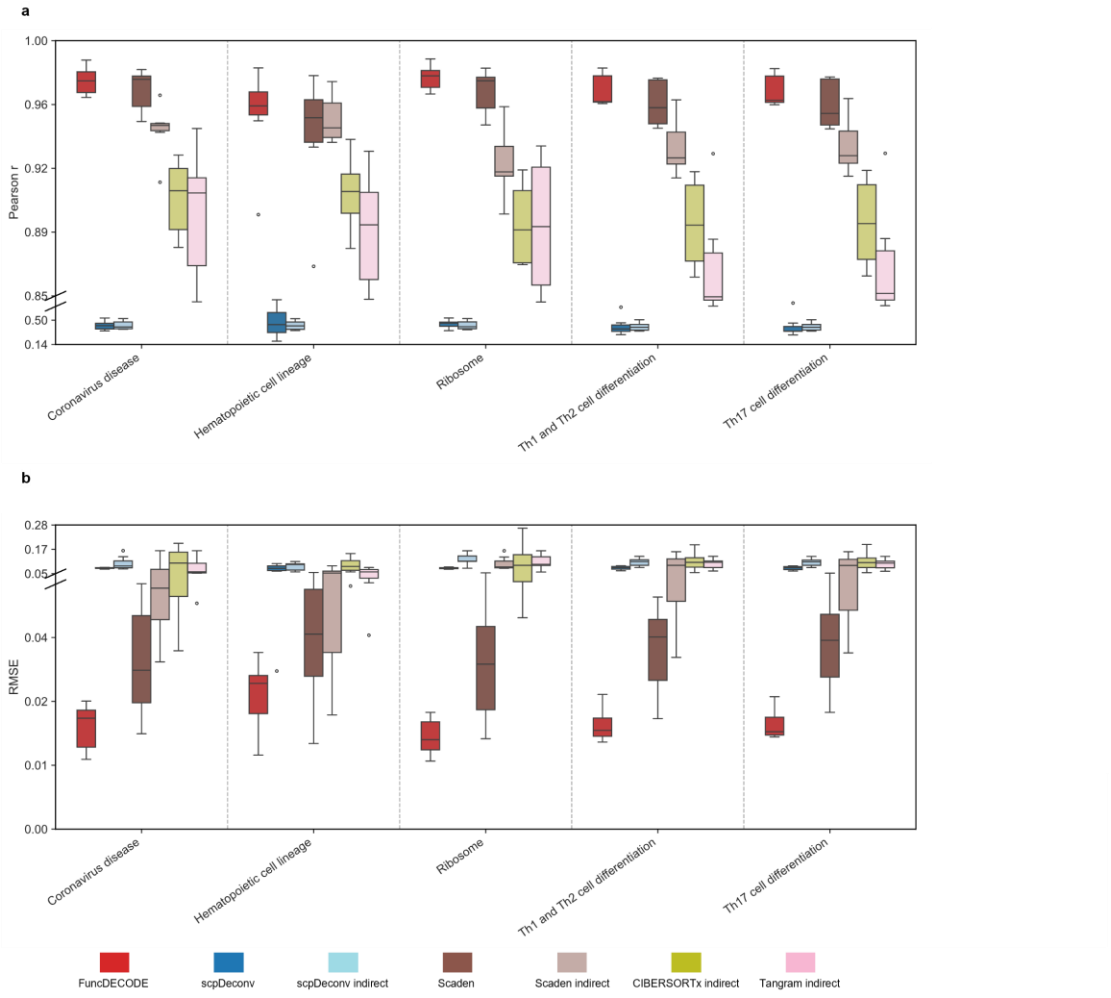


Fig. S1| Visualization of benchmark datasets, functional program distributions. (a) UMAP visualization of the PBMC dataset, with cells colored by donor information. (b) UMAP visualization of the NSCLC single-cell dataset used to generate the training dataset, colored by cell type. (c) Distribution of AUCell-derived scores for eight KEGG pathway gene sets across cell types in the PBMC dataset. (d) Distribution of AUCell-derived scores for eight KEGG pathway gene sets across cell types in the NSCLC dataset.



206

207 **Fig. S2| Additional information for benchmarking cell-type-specific relative contribution**
208 **inference in PBMC datasets.** Boxplots of Pearson's r (a) and RMSE (b) for five PBMC pathways
209 across different methods.

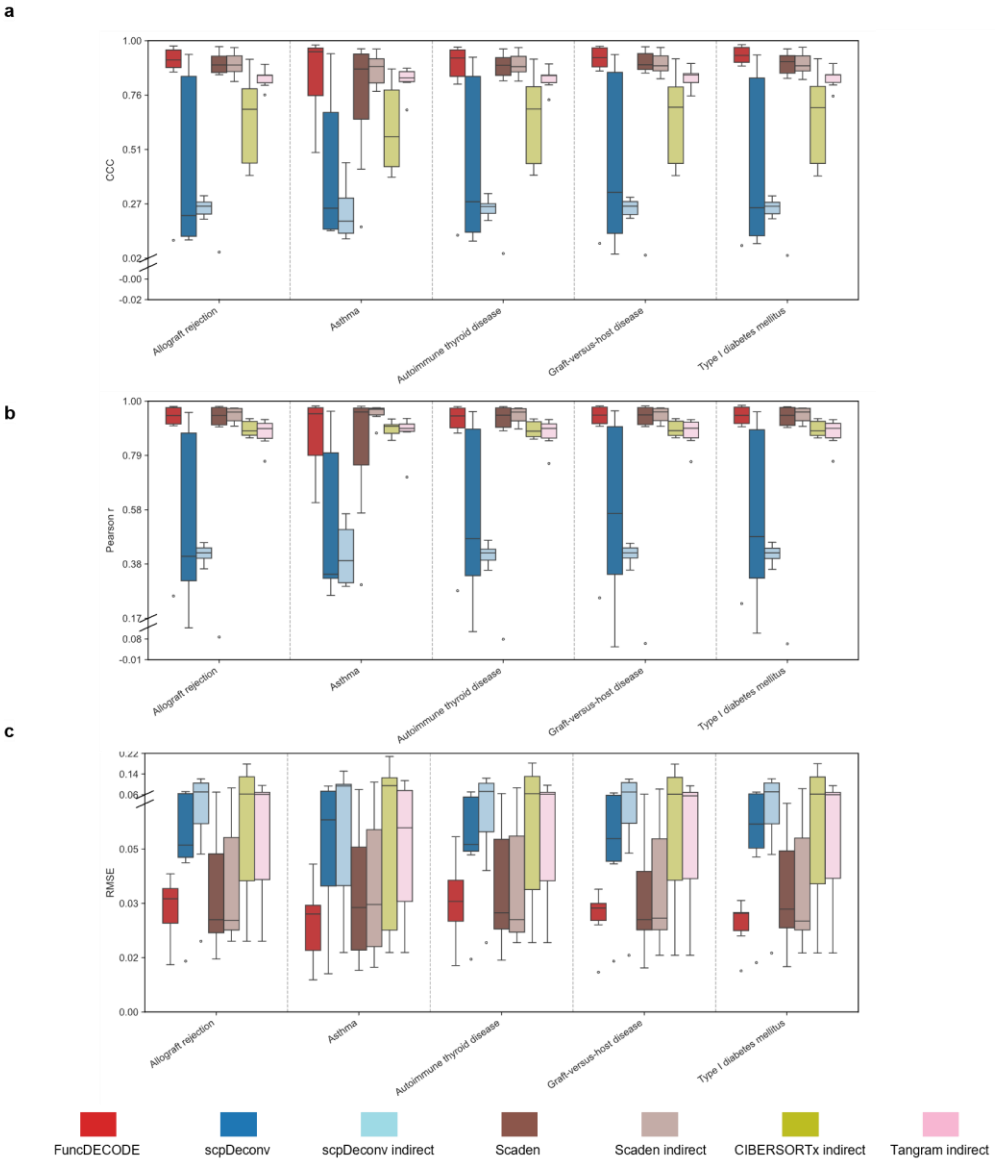


Fig. S3| Additional information for benchmarking cell-type-specific relative contribution inference in PBMC datasets. Boxplots of CCC (a), Pearson's r (b), and RMSE (c) for five PBMC pathways across different methods.

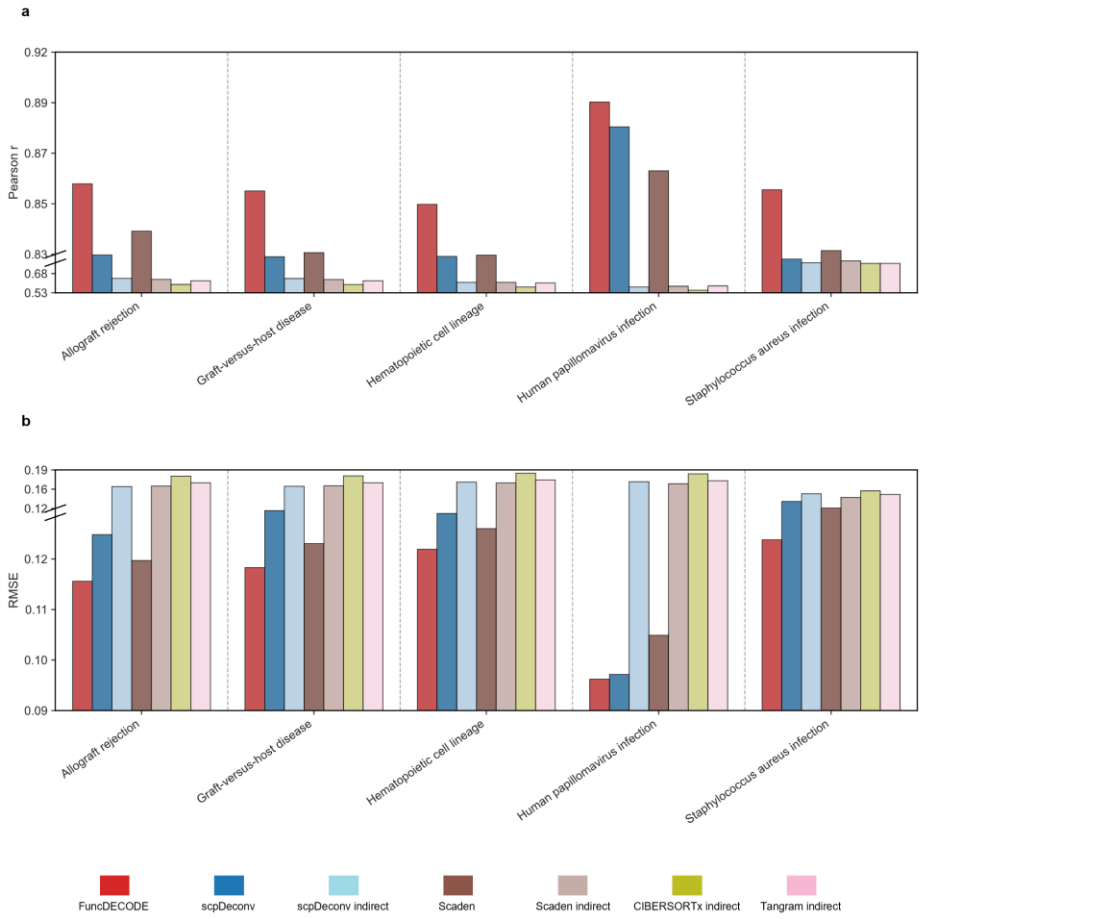


Fig. S4| Additional information for benchmarking cell-type-specific relative contribution inference in NSCLC datasets. Bar plots of Pearson's r (a) and RMSE (b) for five NSCLC pathways across different methods.

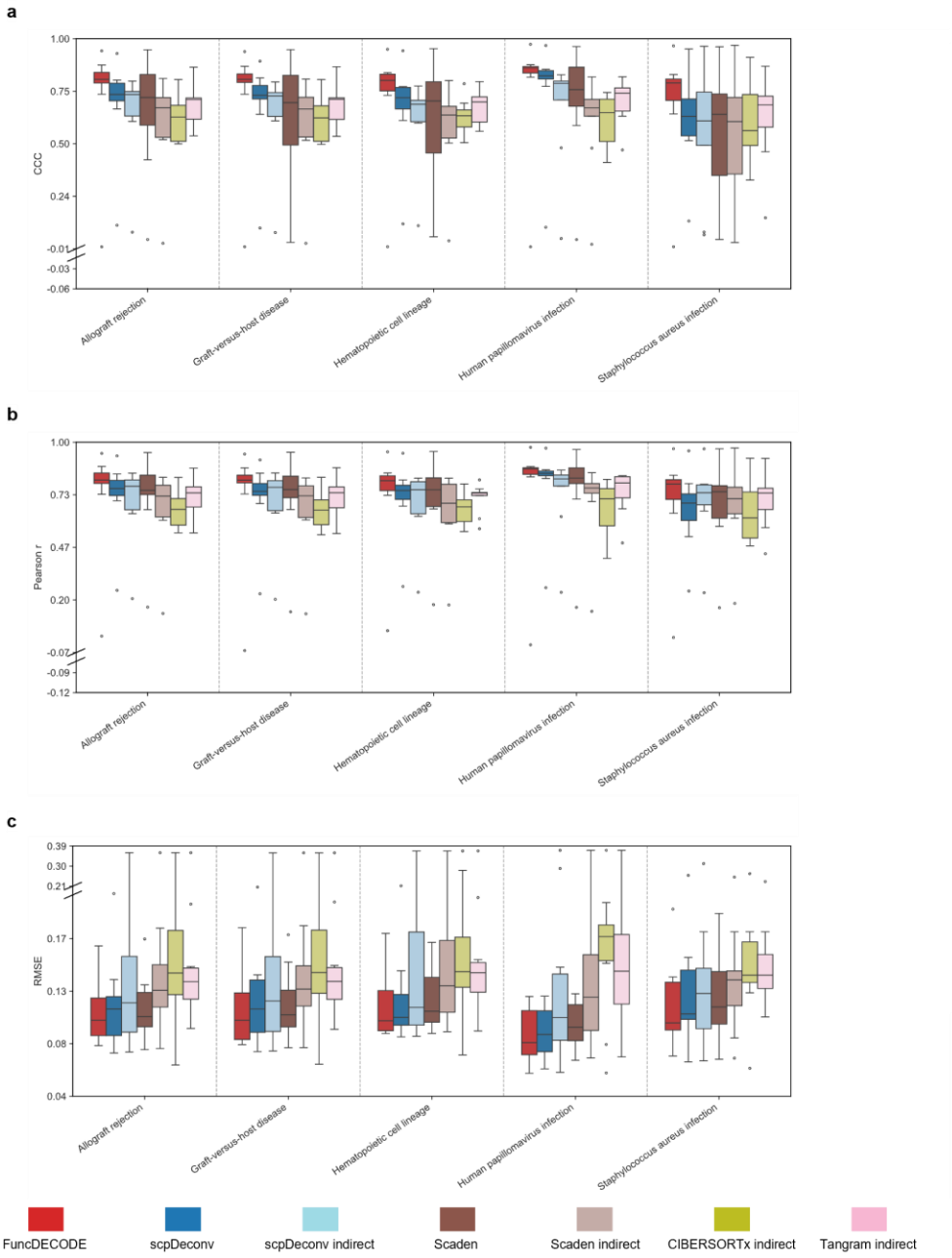


Fig. S5| Additional information for benchmarking cell-type-specific relative contribution inference in NSCLC datasets. Boxplots of CCC (a), Pearson's r (b), and RMSE (c) for five selected NSCLC pathways across different methods.

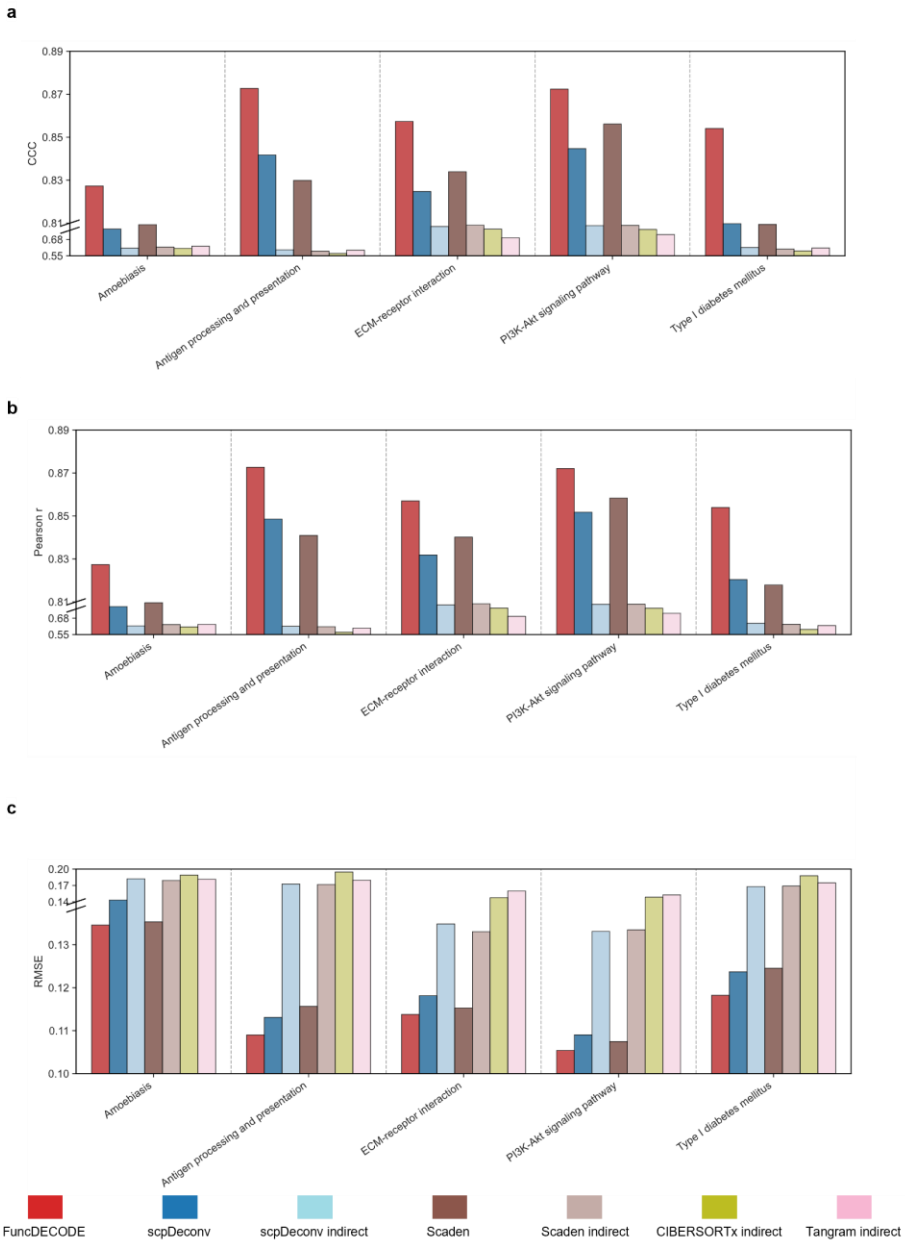


Fig. S6| Additional information for benchmarking cell-type-specific relative contribution inference in NSCLC datasets. Bar plots of CCC (a), Pearson's r (b), and RMSE (c) for the remaining five selected NSCLC pathways across different methods.

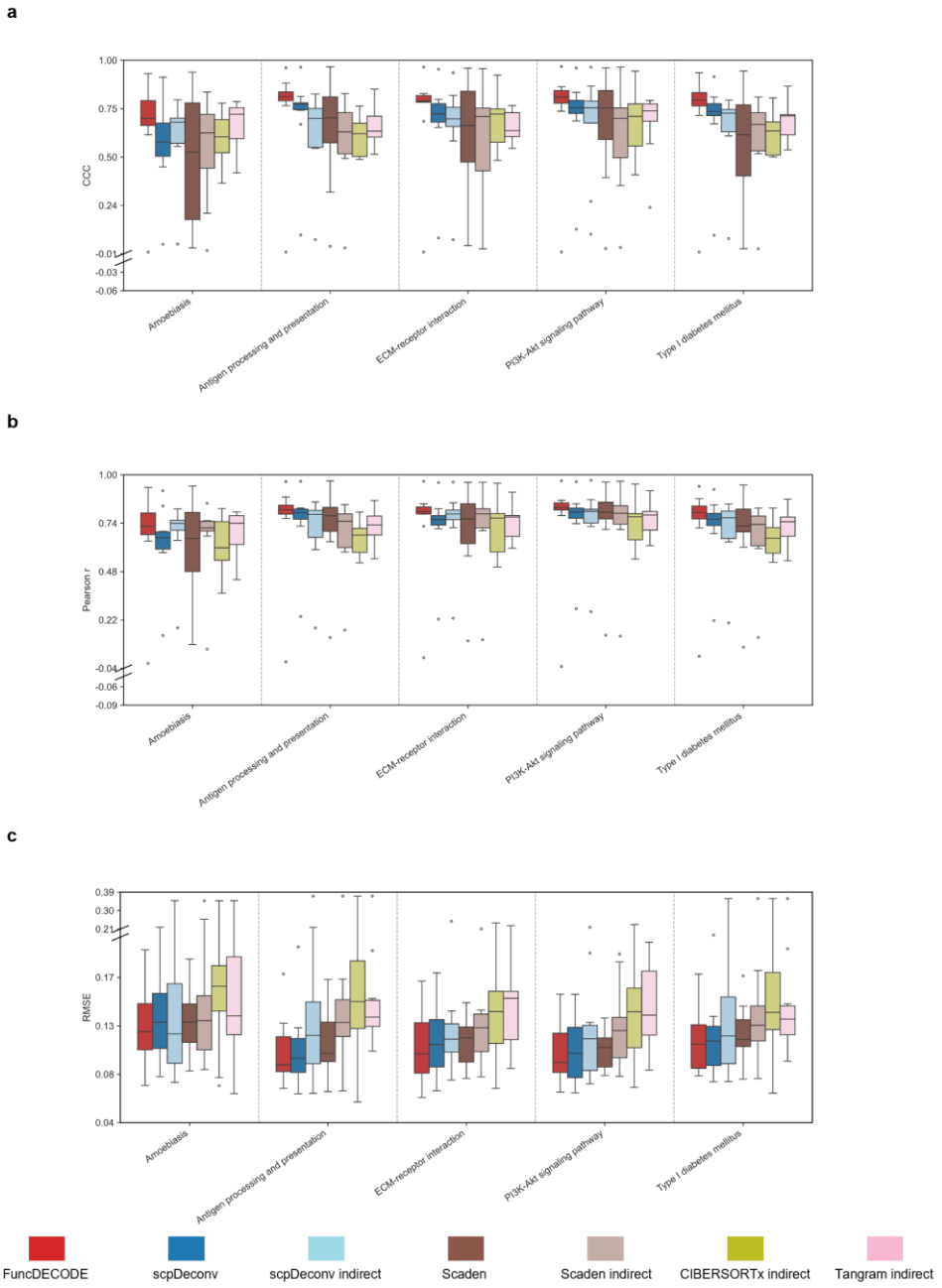


Fig. S7| Additional information for benchmarking cell-type-specific relative contribution inference in NSCLC datasets. Boxplots of CCC (a), Pearson's r (b), and RMSE (c) for the remaining five selected NSCLC pathways across different methods.

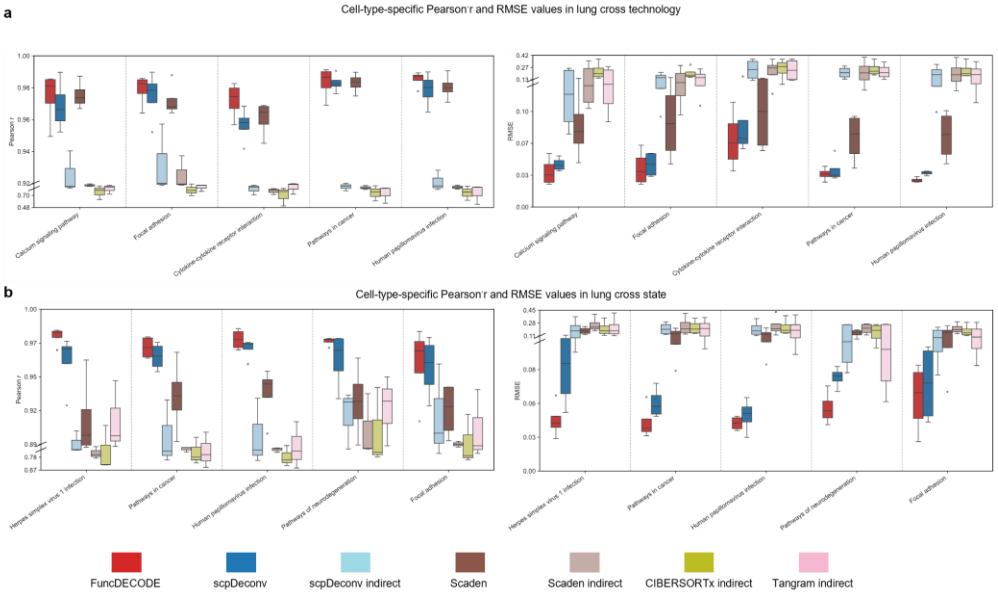


Fig. S8| Pearson's r and RMSE results for bulk-data cross-technology and cross-state evaluations. Pearson's r and RMSE results corresponding to the five KEGG pathways shown in Fig. 3 c, d for the bulk-data cross-technology (a) and cross-state evaluations (b).

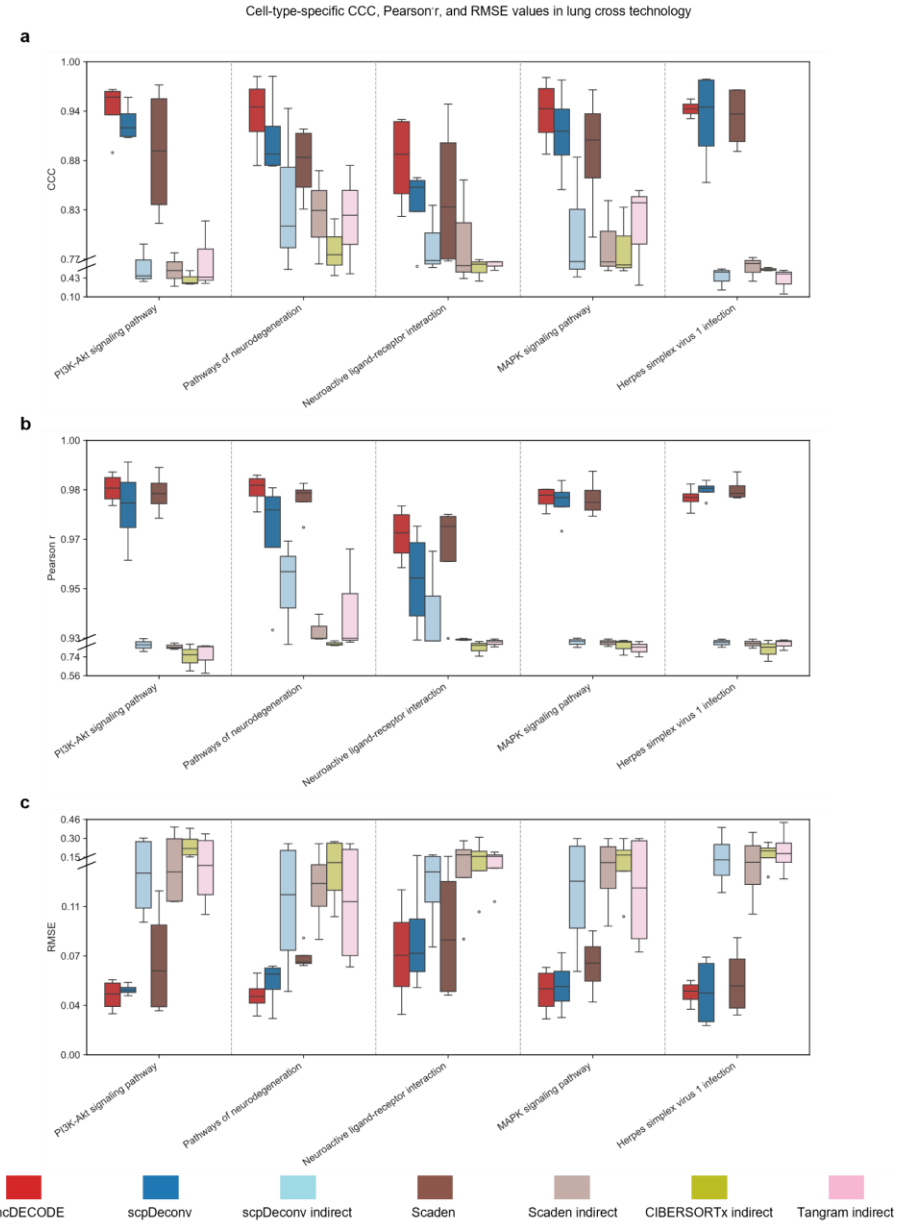


Fig. S9| Additional benchmark results for five KEGG pathways in the bulk-data cross-technology evaluation. CCC (a), Pearson's r (b), and RMSE (c) results for five additional KEGG pathways in the bulk-data cross-technology evaluation.

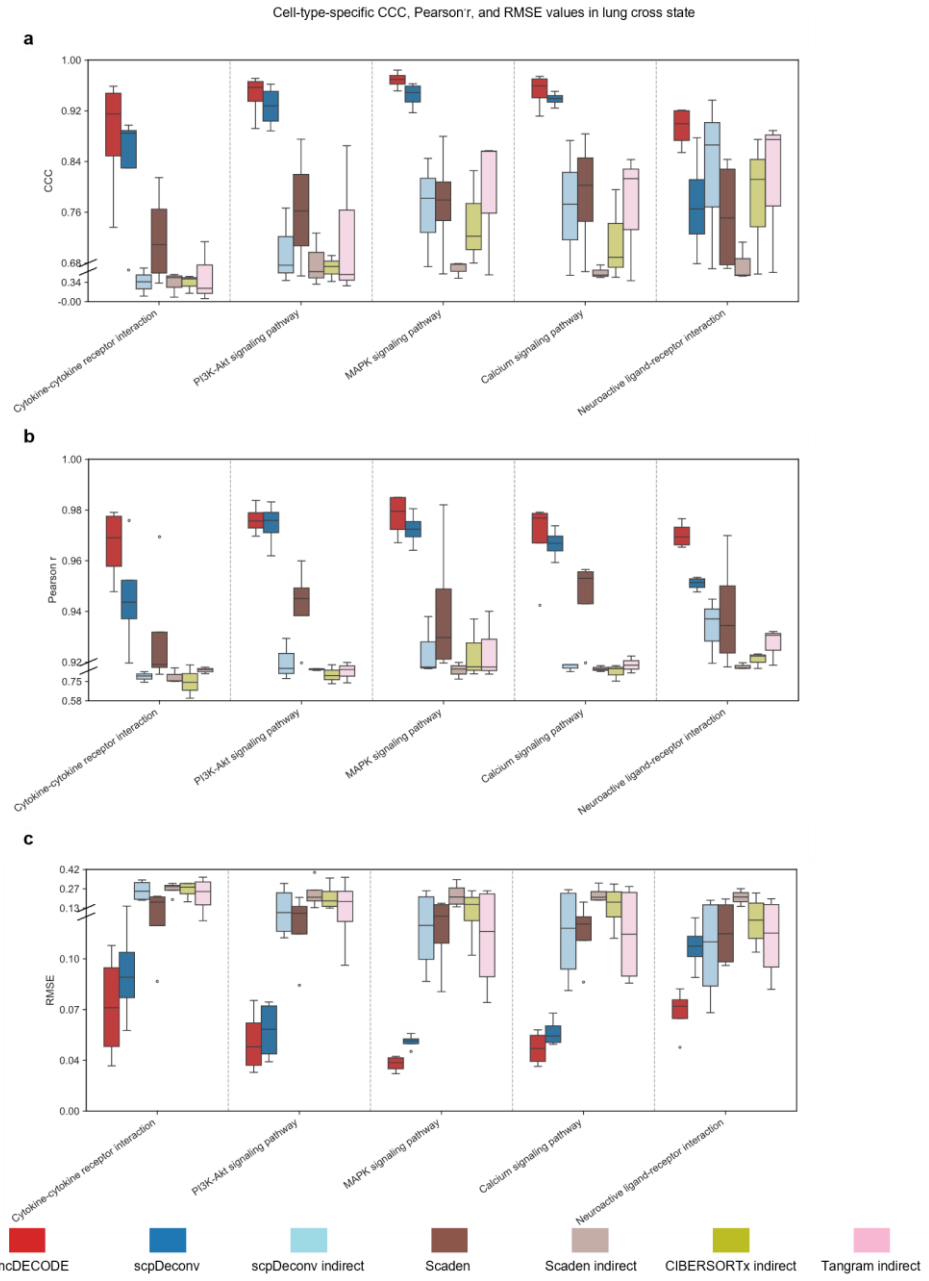


Fig. S10| Additional benchmark results for five KEGG pathways in the bulk-data cross-state evaluation. CCC (a), Pearson's r (b), and RMSE (c) results for five additional KEGG pathways in the bulk-data cross-state evaluation.

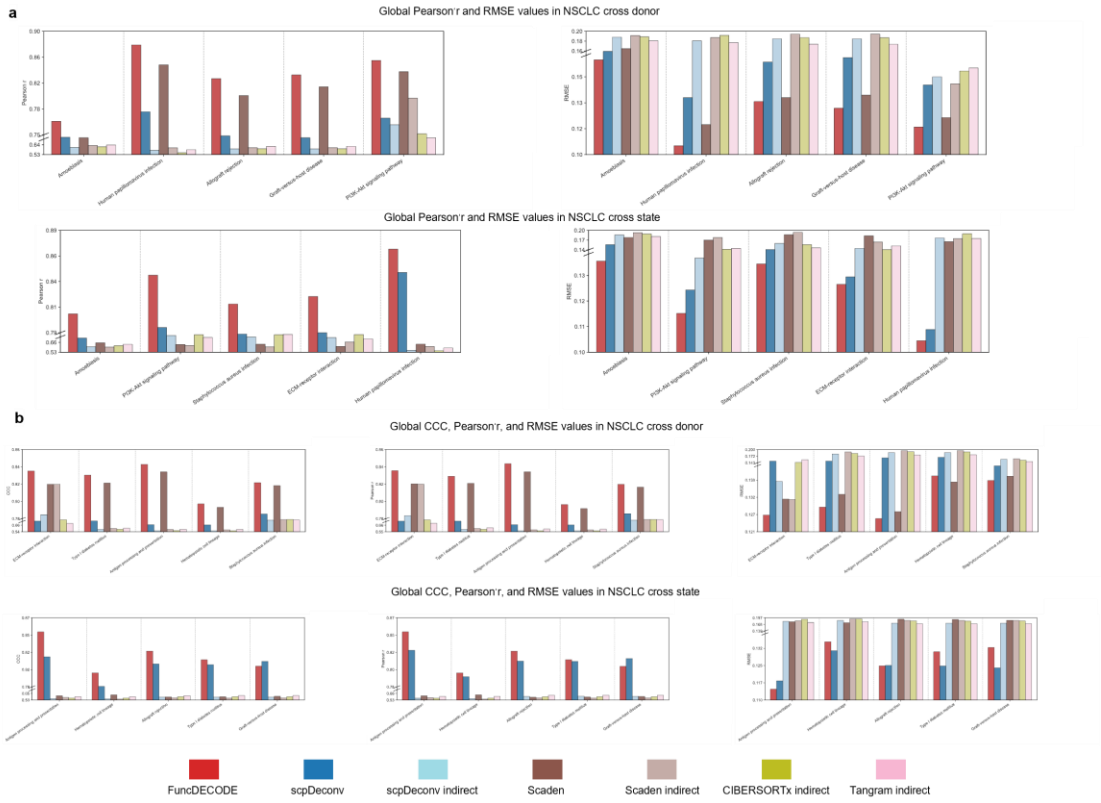
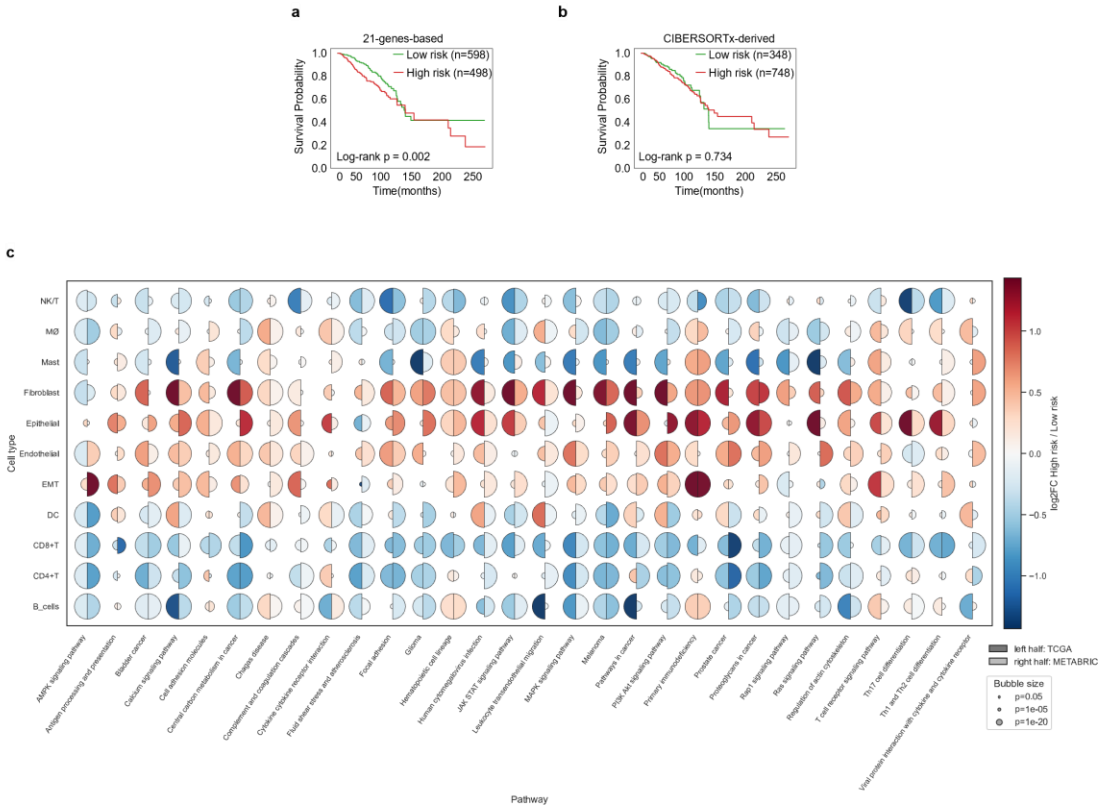


Fig. S11| Additional benchmark results for spatial transcriptomic-data cross-donor and cross-state evaluations. Pearson's r and RMSE results (a) corresponding to the five KEGG pathways shown in Fig. 3 e, f, together with CCC, Pearson's r, and RMSE results (b) for five additional KEGG pathways in the spatial transcriptomic-data cross-donor (top) and cross-state (bottom) evaluations.



252

253 **Fig. S12| Survival stratification by alternative feature sources and cross-cohort concordance**
254 **of FuncDECODE-derived risk-associated features.** (a, b) Kaplan–Meier survival curves of
255 TCGA breast cancer patients stratified using 21-gene panel expression features (a) or
256 CIBERSORTx-derived cell-type proportions (b). P values were calculated using the log-rank test.
257 (c) Bubble heatmap showing log2 fold changes of FuncDECODE-derived cell type–pathway
258 features between high-risk and low-risk groups in TCGA and METABRIC. The left and right halves
259 of each bubble represent TCGA and METABRIC, respectively. Bubble color indicates the log2 fold
260 change between high-risk and low-risk groups, and bubble size indicates statistical significance.

261

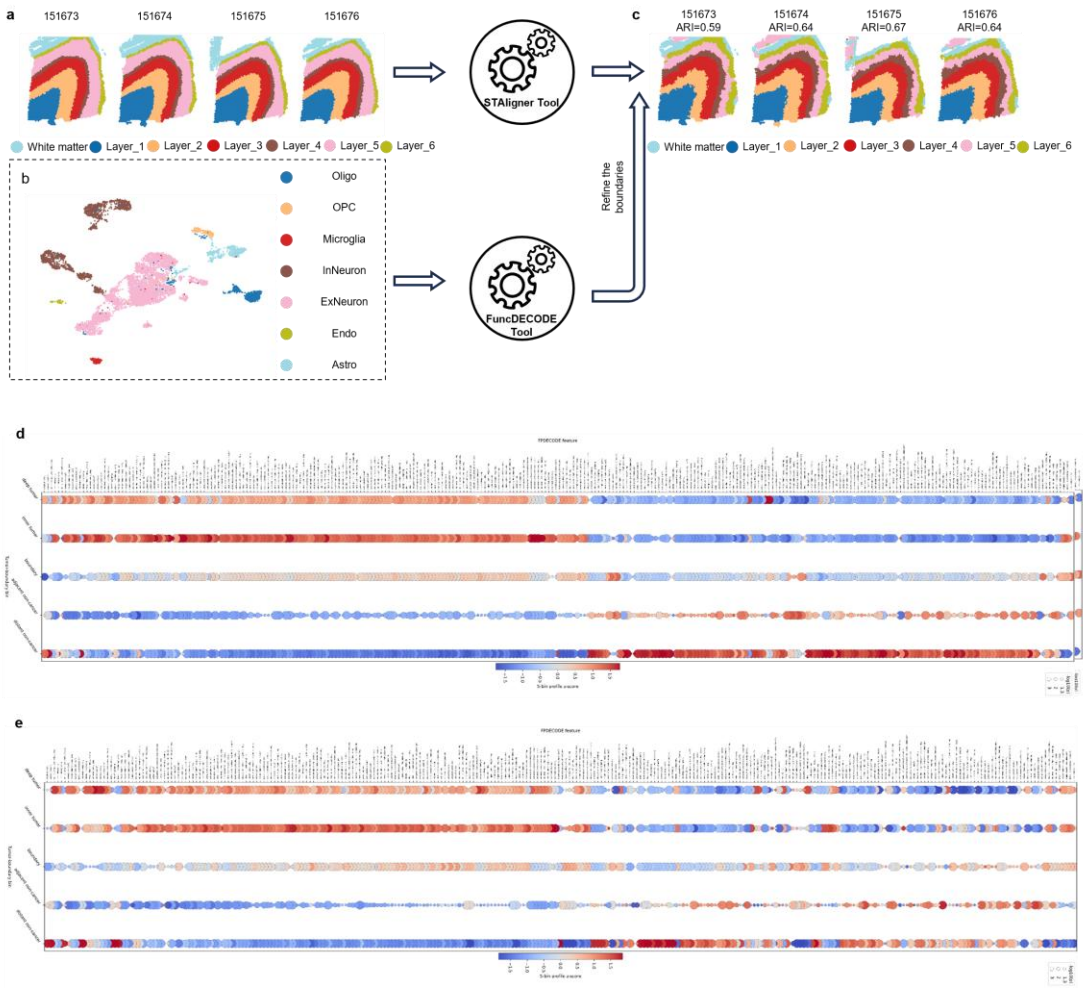


Fig. S13| FuncDECODE refines spatial tissue architecture and uncovers treatment-associated microenvironmental remodeling. (a) Spatial visualization of four human DLPFC sections colored according to the expert-annotated cortical layer labels provided with the original dataset. (b) UMAP visualization of the DLPFC reference single-nucleus dataset, colored by cell type annotations. (c) Spatial visualization of four DLPFC sections, colored by the final domain labels obtained after STAligner annotation and subsequent refinement with FuncDECODE-derived features. (d) Boundary-gradient bubble plot of FPDECODE cell type-functional program features in post-treatment tumors. Bubble color represents the mean section-normalized feature z-score within each boundary bin. Bubble size represents spot-level significance from a two-sided Mann–Whitney U test comparing spots in the current bin with spots in all other bins. (e) Boundary-gradient bubble plot of FPDECODE cell type-functional program features in CRPC tumors. Bubble color represents the mean section-normalized feature z-score within each boundary bin. Bubble size represents spot-level significance from a two-sided Mann–Whitney U test comparing spots in the current bin with spots in all other bins.

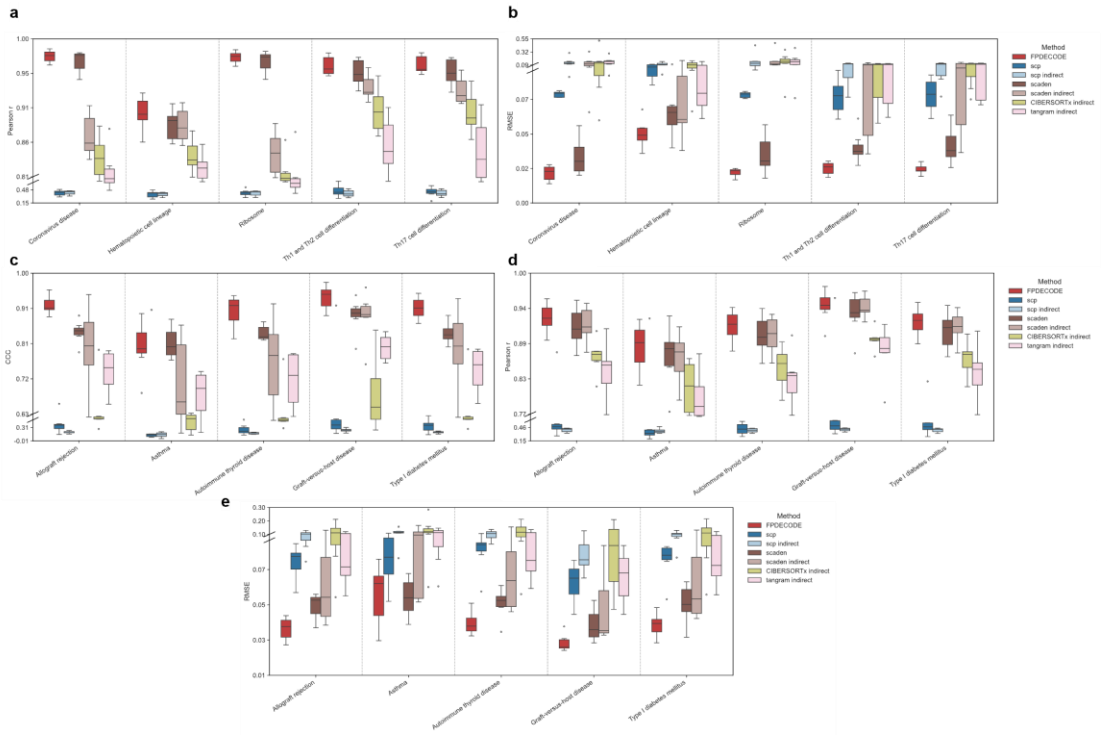


Fig. S14| Extended benchmarking of FuncDECODE using GSEApy-derived KEGG pathway annotations in the PBMC cross-donor experiment. (a–b) Pearson's r (a) and RMSE (b) boxplots showing the performance of FuncDECODE and competing methods for five KEGG pathways in the PBMC cross-donor evaluation using GSEApy-derived single-cell gene-set activity scores. (c–e) CCC (c), Pearson's r (d), and RMSE (e) boxplots showing the corresponding performance of FuncDECODE and competing methods for five additional KEGG pathways under the same experimental setting.

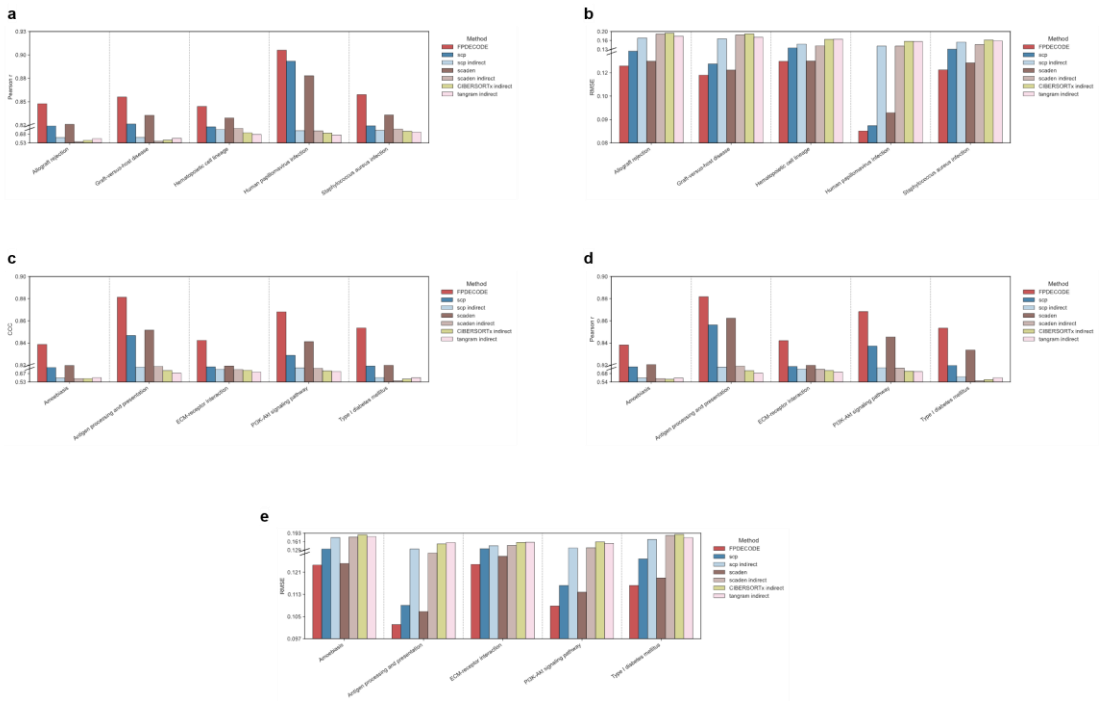


Fig. S15| Extended benchmarking of FuncDECODE using GSEApy-derived KEGG pathway annotations in the NSCLC spatial transcriptomics cross-section experiment. (a–b) Pearson's r (a) and RMSE (b) bar plots showing the performance of FuncDECODE and competing methods for five KEGG pathways in the NSCLC cross-section evaluation using GSEApy-derived single-cell gene-set activity scores. (c–e) CCC (c), Pearson's r (d), and RMSE (e) bar plots showing the corresponding performance of FuncDECODE and competing methods for five additional KEGG pathways under the same experimental setting.

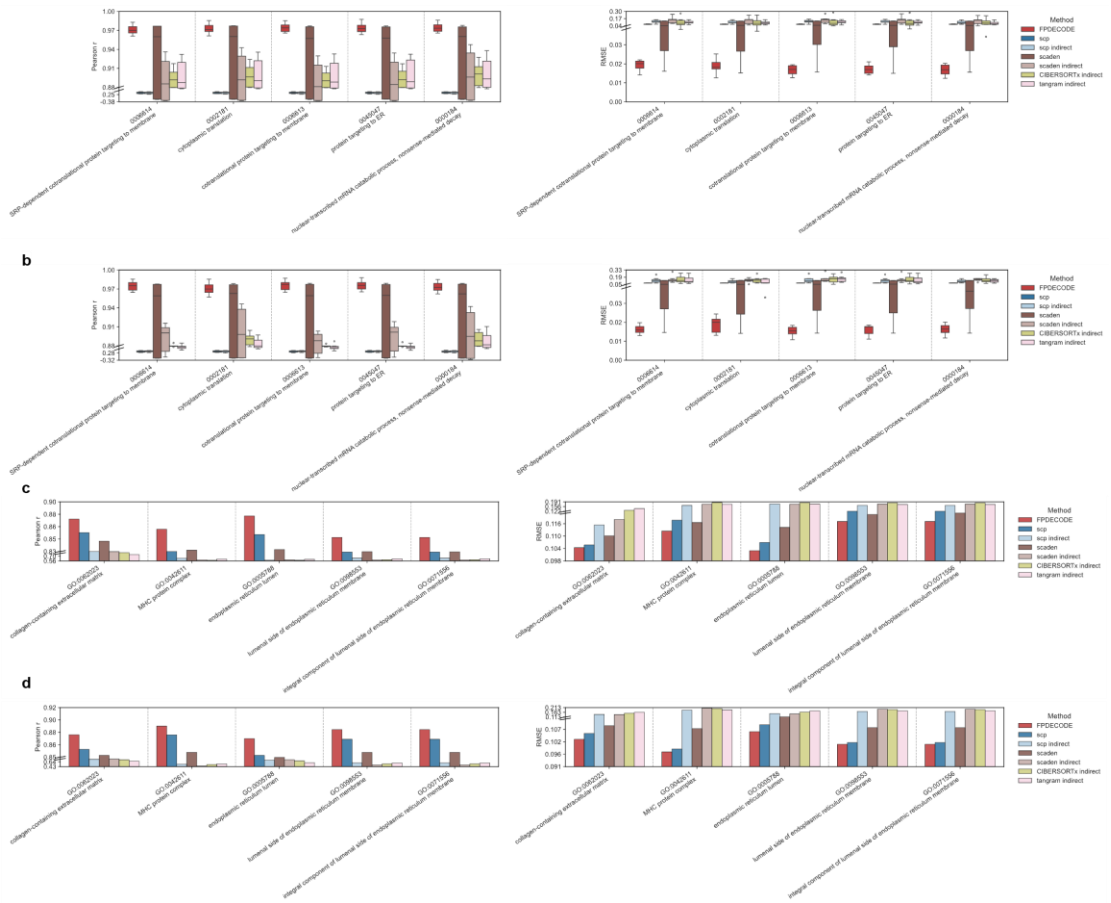


Fig. S16| Extended performance evaluation of FuncDECODE using GO-based functional program proxies. (a–b) Pearson's r and RMSE boxplots showing the performance of FuncDECODE and competing methods in the PBMC cross-donor experiment when functional programs were defined by the five selected Gene Ontology biological process (GO BP) terms. Single-cell functional activity scores were computed using AUCell (a) and GSEApy (b), respectively. (c–d) Pearson's r and RMSE bar plots showing the corresponding performance in the NSCLC spatial transcriptomics cross-section experiment when functional programs were defined by the five selected Gene Ontology cellular component (GO CC) terms. Single-cell functional activity scores were computed using AUCell (c) and GSEApy (d), respectively.