

Article

STEAs: Histologically Validated and Reference-Independent Major Cell-Type Annotation for Spatial Transcriptomics Reveals Relevant Cellular Organization and Architecture of Tumor Microenvironment

Qian Li ^{1,2}, Qingyang Zhang ^{1,2}, Fanhong Zeng ^{1,2}, Irene Oi-Lin Ng ^{1,2,*}  and Daniel Wai-Hung Ho ^{1,2,*} 

¹ Department of Pathology, The University of Hong Kong, Hong Kong; liqian0901@connect.hku.hk (Q.L.); qyzh@hku.hk (Q.Z.); fanhong95@connect.hku.hk (F.Z.)

² State Key Laboratory of Liver Research, The University of Hong Kong, Hong Kong

* Correspondence: iolng@hku.hk (I.O.-L.N.); dwhho@hku.hk (D.W.-H.H.); Tel.: +852-2255-2664 (I.O.-L.N.)

Simple Summary

Spatial transcriptomics (ST) enables in situ gene expression profiling while preserving spatial context, but current sequencing-based platforms require spot-level cell-type annotation or may even experience technical ambiguity, due to mixed cell-type signals within spots. To address this, we developed STEA, a fast and efficient reference-independent annotation method for major cell types. STEA infers cell-type composition using predefined or adaptively expanded marker gene sets, avoiding reliance on scRNA-seq references while maintaining accuracy and interpretability. It first accepts curated markers or expands them via cosine similarity, then optionally refines gene sets using Jensen–Shannon divergence to improve specificity. Finally, enrichment analysis is performed to estimate cell-type proportions. Extensive benchmarking across synthetic and real datasets from multiple platforms demonstrates that STEA achieves competitive performance compared to state-of-the-art methods, while remaining computationally efficient for typical research settings.

Abstract

Background: Recent advances in spatial transcriptomic technologies enable in situ gene expression profiling while preserving spatial context. This capability is particularly important for studying the tumor microenvironment (TME), where diverse and admixed cell populations interact within highly organized spatial niches that influence tumor progression and therapeutic response. However, the limited resolution of early spatial transcriptomic platforms results in each spatial spot capturing transcripts from multiple cell types, making accurate spot deconvolution or annotation a critical yet challenging step in downstream data analysis. The level of complexity will be particularly prominent in heterogeneous samples like the tumor microenvironments where multiple cell types are highly admixed and reliable single-cell reference atlases may usually be unavailable. **Methods:** In this paper, we developed our method called STEA, which is a novel and accurate reference-independent enrichment-based annotation algorithm for major cell type. Unlike the existing approaches, STEA does not require single-cell RNA sequencing datasets as reference, offering both flexibility and computational efficiency in execution. **Results:** We performed comprehensive benchmarking using a variety of simulated datasets across different platforms and scenarios and demonstrated the superior accuracy of STEA. Apart from synthetic data, we also evaluated multiple real datasets to further exemplify its practical applicability on both oncology-related and oncology-unrelated data. More importantly, we could confidently



Academic Editors: Jason Roszik and Subburaj Ilangumaran

Received: 18 March 2026

Revised: 15 April 2026

Accepted: 28 April 2026

Published: 29 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

demonstrate the high concordance between prediction of STEA and histological classification by experienced pathologist. Conclusion: Our STEA algorithm provides a practical reference-independent framework to complement the cutting-edge spatial transcriptomics in genomics studies, facilitating accurate downstream high-dimensional spatial characterization of cellular and molecular landscapes, reconstruction of tissue architecture as well as cell–cell communication in malignant and non-malignant scenarios. Taken together, our comprehensive evaluation demonstrates the robustness and reliability of STEA, highlighting its potential as a valuable tool for studying complex tissue organization, particularly within heterogeneous TME.

Keywords: annotation; reference-independent; spatial transcriptomics; molecular pathology

1. Introduction

The tumor microenvironment (TME) is a highly complex ecosystem composed of malignant cells, immune cells, stromal cells, and vascular components that interact within spatially organized niches. These interactions play critical roles in tumor progression, immune regulation, and therapeutic response. Therefore, understanding the spatial architecture and cellular heterogeneity of the TME is essential for elucidating mechanisms of cancer development and treatment resistance. Spatial transcriptomics (ST) has rapidly emerged as a transformative technology for investigating tissue organization and cellular heterogeneity, offering a unique ability to preserve spatial context while capturing gene expression profiles in situ [1,2]. Its growing adoption has enabled researchers to explore diverse biological problems, from spatial context-preserved cellular and molecular landscapes and interactions of human tumors to developmental biology, with unprecedented molecular comprehensiveness and spatial complexity [3–5]. Despite these advances, the resolution of current sequencing-based ST platforms remains a major limitation [4,6,7].

Sequencing-based spatial transcriptomic approaches, including Visium v1/v2 [8], Slide-seq v1 [9], DBIT-seq [10], and GeoMX DSP [11], often aggregate signals from several cells within a single spatial spot. For instance, the 10X Genomics Visium v1 platform, one of the most widely used ST technologies, features capture spots with a diameter of 55 μm [12]. This is considerably larger than the average mammalian cell size [13,14], meaning that each spot typically contains multiple cells or even partial fragments of distinct cells. As a result, the measured gene expression profiles represent a mixture of transcripts from heterogeneous cell populations [15,16]. To disentangle this complexity and recover meaningful biological insights, computational deconvolution tools are required to infer cell type composition within each spot [17].

Existing deconvolution methods fall into two main categories [18]. The first category leverages single-cell RNA sequencing (scRNA-seq) datasets as the reference. Representative methods include probability-based method: RCTD [19], CARD [20] and Cell2location [21], non-negative factorization (NMF)-based method: SPOTlight [22] and SpatialDWLS [23], deep learning-based method: Tangram [24], GraphST [25] and SpatialDDLs [26]. While effective in certain contexts, these approaches face significant challenges: the possibility and cost of obtaining scRNA-seq reference, potential batch effects between scRNA-seq and ST datasets, and computational inefficiency in terms of long runtime. Unless having matched single-cell and spatial profilings of the same samples, cellular composition and molecular expression of ST may diverge substantially from the publicly available single-cell reference. In practical settings, matched or at least public single-cell references are frequently unavailable, incomplete, or biased. The second category does not rely on

external scRNA-seq references, exemplified by methods such as STdeconvolve [27] and spicemix [28]. Although these approaches are reference-free and computationally efficient, they often stop at the level of latent components, which do not correspond directly to known biological cell types. Consequently, additional steps are required to assign cell type identities, adding complexity and potential sources of error to the computational processes. Taken together, the aforementioned factors could potentially lead to significant limitations on their practical performance.

To address the aforementioned challenges, we developed STEA, a fast and efficient reference-independent enrichment-based annotation method. STEA infers cell type composition directly using predefined marker gene sets, either provided along with the method or adaptively generated during execution. This approach avoids the need for scRNA-seq references but still yields accurate and biologically interpretable results within a reasonable timeframe and using typical computational resources. Extensive benchmarking shows highly competitive performance of STEA as compared with state-of-the-art deconvolution and scoring methods across multiple ST datasets (synthetic and real) and sequencing platforms. Importantly, beyond computational benchmarking using synthetic data, we further evaluated the inferred cellular architectures through systematic comparison with independent histopathological assessment. The evaluation results confirmed that our reference-independent annotation algorithm remains highly consistent with established morphological or histological classifications, thereby ensuring biological relevance.

2. Materials and Methods

2.1. Workflow of STEA

We developed the STEA package to perform annotation analysis. The first step involves providing either curated gene sets directly or classical marker genes. In the latter case, STEA will expand the initial set of marker genes by identifying new genes with similar expression patterns based on cosine similarity. The second step will optionally apply Jensen–Shannon divergence to evaluate the discriminative efficiency of the marker genes obtained in the first step. Finally, in the third step, enrichment analysis will be conducted to generate the annotation results. In essence, our methods are based on ranking genes within each cell (or spot) and evaluating whether a predefined gene set is enriched toward the top of the ranked list. Notably, our STEA algorithm builds upon several important concepts including cosine similarity, Jensen–Shannon divergence, marker-based annotation and rank-based enrichment scoring. Our original repurposing for existing concepts and innovative integration lead to the emergent functionality for annotation of major cell types.

2.2. Identification of Spatially Co-Localized Genes Using Cosine Similarity

To quantify the similarity between the spatial distribution of individual genes and predefined spatial patterns or gene signatures, we will compute the cosine similarity between a gene expression vector $g_i = (x_{1i}, x_{2i}, \dots, x_{Ni})$ across N spatial spots and a reference pattern $\lambda_k = (\lambda_{1k}, \lambda_{2k}, \dots, \lambda_{Nk})$. The similarity is defined as:

$$\cos(g_i, \lambda_k) = \frac{\sum_{j=1}^N x_{ji} \lambda_{jk}}{\sqrt{\sum_{j=1}^N x_{ji}^2} \sqrt{\sum_{j=1}^N \lambda_{jk}^2}}.$$

A high cosine similarity indicates that the gene and the reference pattern share a similar spatial distribution across the tissue. For each reference pattern, genes will be ranked according to their cosine similarity scores, allowing the identification of genes with spatial localization patterns resembling the reference.

2.3. Gene Pattern Similarity Assessment Based on Spatial Jensen–Shannon Divergence (JSD)

To evaluate the efficiency of candidate genes as spatial markers, we quantified the similarity between their spatial expression patterns and those of classical marker genes. A candidate gene was considered efficient if its spatial expression distribution closely matched that of a reference marker gene, indicating its ability to represent spatially localized cell types or functional regions.

For each gene P_g , we extracted its expression values across all spatial spots together with the corresponding 2D spatial coordinates (x_i, y_i) . The spatial domain was discretized into an $n_{\text{bins}} \times n_{\text{bins}}$ regular grid ($n_{\text{bins}} = 20$ by default). A 2D histogram was constructed by accumulating gene expression values within each grid cell. To avoid zero probabilities, a small pseudocount (10^{-10}) was added to all grid cells. The histogram was then normalized to obtain a 2D spatial probability distribution $P_g(x, y)$, summing to 1 over the grid. For each candidate gene g and its corresponding reference marker gene m , we computed the Jensen–Shannon divergence (JSD) between their spatial distributions:

$$\text{JSD}(P_g // P_m) = \frac{1}{2} \text{KL}(P_g // M) + \frac{1}{2} \text{KL}(P_m // M),$$

where $M = \frac{1}{2}(P_g + P_m)$, and KL denotes the Kullback–Leibler divergence. Lower JSD values indicate higher similarity between the spatial expression patterns of the candidate and marker genes, and thus higher marker efficiency.

To assess the statistical significance of the observed JSD, we performed permutation tests for each candidate gene. Specifically, we generated null expression patterns by permuting the expression values of gene g across spatial spots for $N_{\text{perm}} = 1000$ iterations. Two permutation strategies were considered: (i) random mode, in which expression values were fully permuted across all spots, and (ii) spatial blocks mode, in which the tissue was partitioned into coarse spatial blocks and expression values were reassigned between blocks while preserving within-block structure ($\text{block_n_bins} = 10$ by default).

For each permuted pattern, the corresponding 2D spatial distribution and its JSD relative to the marker gene were recomputed, yielding a null distribution of JSD values. The empirical p -value was calculated as:

$$p\text{-value} = \frac{\#\{\text{null JSD} \leq \text{observed JSD}\} + 1}{N_{\text{perm}} + 1}$$

Finally, p -values were adjusted for multiple testing across genes using the Benjamini–Hochberg procedure.

2.4. Enrichment Method

The enrichment method belongs to the family of rank-based gene set enrichment approaches, including AUCCell [29], UCell [30], GSEA [31,32] and ssGSEA [33]. These methods share the common principle of evaluating whether genes from a predefined set are preferentially ranked toward the top of a cell-specific expression profile.

For each cell i and each gene set S , the algorithm proceeds as follows: Firstly, it will compute a rank vector R_i across N genes (higher values indicate higher relative expression or rank within that cell). Then, it will order genes by descending rank to obtain a permutation $\pi(1), \dots, \pi(N)$. It will also define weights $w_j = R_{i, \pi(j)}^\alpha$ to emphasize high-ranked genes via an exponent parameter α ($\alpha = 0.25$ by default). The parameter α controls the contribution of highly ranked genes to the enrichment score. Specifically, gene weights are defined as a power transformation of their ranks, such that higher values of α increase the relative contribution of top-ranked genes, while lower values reduce this effect and yield a more uniform weighting across the gene set.

Two cumulative distribution functions will be constructed along the ordered gene list:

1. a weighted positive CDF for genes in S :

$$P^+(j) = \frac{\sum_{k=1}^j w_k \cdot 1[\pi(k) \in S]}{\sum_{k=1}^N w_k \cdot 1[\pi(k) \in S]}$$

2. a uniform negative CDF for genes not in S :

$$P^-(j) = \frac{\sum_{k=1}^j 1[\pi(k) \notin S]}{N - |S|}$$

3. Define the enrichment score for cell i and gene set S as the mean pointwise difference between the two CDFs:

$$ES_i(S) = \frac{1}{N} \sum_{j=1}^N (P_i^+(j) - P_i^-(j))$$

This implementation differs from classical GSEA, which typically uses the maximum deviation of a running-sum statistic. Here, the score reflects the average area between positive and negative cumulative distributions. To ensure numerical stability, denominators were checked to avoid division by zero, and ES values were optionally normalized across all cells. This approach generates a gene set $\times$ cell-matrix of enrichment scores suitable for downstream analysis.

2.5. Deriving Expected Annotations of the Mouse Olfactory Bulb ST Dataset

Since the original study did not provide explicit cell-type annotations, we derived a robust surrogate using consensus clustering, based on the principle that cells belonging to the same biological type should consistently co-cluster across a range of clustering parameters. The procedure was as follows: We performed Seurat FindClusters clustering at multiple resolutions (0.1 to 1.0, step 0.1), yielding 10 independent clustering results. For each pair of cells, we calculated the proportion of runs in which they were assigned to the same cluster, resulting in a consensus matrix with values ranging from 0 to 1. Hierarchical clustering of this consensus matrix revealed a clear block-diagonal structure, in which each compact, high-consensus block represents a highly stable cell population that remains cohesive across resolutions. These well-defined blocks were interpreted as biologically meaningful cell types and were used as the ground truth for deconvolution and annotation analyses.

2.6. Benchmarking Methodology

All methods used in this study, namely RCTD [19] (v2.2.1), CARD [20] (v1.1), SPOT-light [22] (v1.7.2), Tangram [24] (v1.0.4), GraphST [25] (v1.1.1), SpatialDDLs [26] (v1.0.3), STdeconvolve [27] (V1.3.2), SpatialDWLS [23] (v1.1.2) and Cell2location [21] (v0.1.5) were executed according to the recommended steps outlined in their respective manuals, including default or developer-recommended parameter selection. Major cell types defined according to the cell-type deconvolution were used for accuracy comparison with the annotation of STEA, against the ground-truth major cell-type identities designated in the simulation or reported in their original publication.

2.7. Evaluation Metrics

To assess the performance of different deconvolution methods, we computed the accuracy which was defined as the proportion of correctly predicted cells/spots over the total number of cells/spots.

2.8. Code Availability

STEA is implemented as an open-source Python package, with source code available at <https://github.com/yiranzhimo/stea-project> (accessed on 27 April 2026). A summary of key parameter settings is provided in Supplementary Table S1.

3. Results

3.1. Overview of STEA

Cell types are functionally defined by the expression of a specific gene set, known as a cell type signature [34,35]. Elevated or specific expression of this signature under particular conditions distinguishes different cell types, tissues, or states [36]. Therefore, distinct cell types can be identified by variations in their marker gene expression.

In this study, we present STEA, a novel reference-free computational approach designed for the annotation of spatial transcriptomics (ST) dataset at multi-cellular pixel resolution (Figure 1). STEA leverages the enrichment-based methodology, requiring classical marker genesets as input. Alternatively, correlation analysis is employed to identify additional candidate marker genes exhibiting same patterns analogous to those classical markers. Subsequently, Jensen–Shannon divergence could be optionally employed to evaluate whether these candidate genes can effectively annotate a cell toward its intended identity, as similar to the classical markers. Only genes that show similar reliability and biological relevance as known canonical markers are selected to proceed. The final step is enrichment analysis, which involves ranking all genes in descending order of normalized expression values, computing an enrichment score from the positional distribution of gene set members within the ranked list, and evaluating this distribution against background genes using the cumulative distribution function (CDF). A high enrichment score indicates these markers are strongly and coordinately expressed in that spot, which implies that the proportion or abundance of that cell type in the spot is relatively high.

3.2. Benchmarking Using Simulated Sequencing-Based HCC ST Dataset

We utilized scRNA-seq data [37] to simulate our in-house HCC ST dataset to represent the admixed tumor microenvironment containing three major categories of cells: malignant cells, stromal cells (cancer-associated fibroblasts [CAFs], and tumor-associated endothelial cells [TECs]) and immune cells (B cells, T cells, and tumor-associated macrophages [TAMs]). In the simulated process, we generated pseudo-spots based on different underlying major cell types, which mainly stratified into malignant-predominant and non-malignant-predominant scenarios. The first scenario happened when the major constituting cell type was malignant cells (>50%), whereas immune cells and stromal cells were present in minority in the mixture of cells within the spot. Conversely, the second scenario was having non-malignant and malignant cells as majority and minority, respectively. Therefore, there were 6 types of pseudo-spots altogether, each with one major constituting cell type (malignant, CAF, TEC, B, T and TAM). Moreover, in reality, cell size and density also varied significantly across cell types and within the tissue architecture. For instance, B cells typically measure 8–10 μm in diameter, while typical tumor cells can reach over 20 μm . To mimic the physiological cell size heterogeneity in tissue and account for it in the generation of pseudo-spot data, we utilized different sizes of cell mixtures (a pool of either 3, 5 or 10 cells) to construct a pseudo-spot. More importantly, we added another layer of

variability by introducing variation in the proportion of major cell type, ranging from 60% to 100%. Taken together, variable numbers of malignant- and non-malignant-predominant pseudo-spots, synthesized based on different sizes of cell mixtures and proportions of major cell type were randomly grouped into a synthetic dataset of 10,000 pseudo-spots, with the whole process stochastically repeated for 100 times (i.e., 100 samples) to generate our overall in-house dataset for testing purpose (Figure 2a,b). Due to its extremely high computational cost and time-consuming model training process for this 100-repeat simulated dataset, Cell2location was not included in this part of benchmarking analysis.

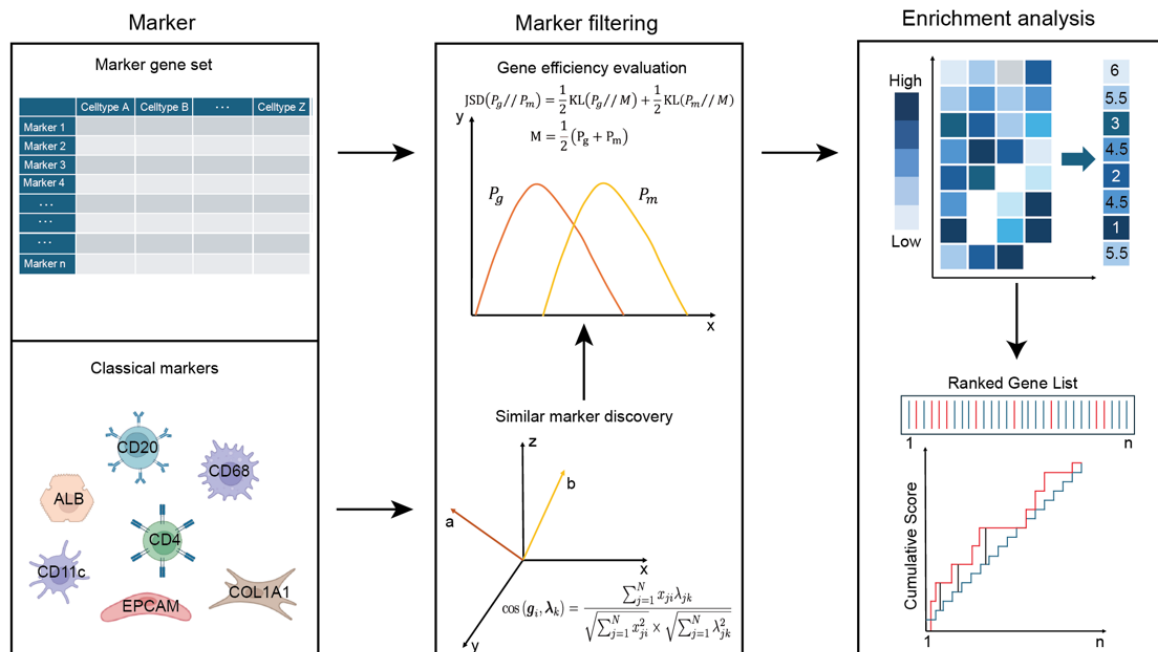


Figure 1. Schematic overview of STEA. STEA primarily identifies cell types using marker genes. If a marker gene list is unavailable, it can be supplemented by analyzing the distribution of canonical markers. After filtering and validation, single-spot enrichment analysis and annotation are then performed. In marker filtering step, the red curve represents the spatial distribution of the expression of gene g (P_g) and yellow curve represents the spatial distribution of the expression of gene m (P_m) in the gene efficiency evaluation process. The red color denotes expression vectors of gene a and gene b . In the enrichment step, dark blue color indicates high gene expression levels. In the ranked gene list, red lines mean genes within the target gene list while light blue lines indicate background genes.

For reference-free methods, STEA significantly outperformed STdeconvolve. Among reference-based methods, RCTD performed the best, followed by SpatialDWLS and SpatialDDLS. Intriguingly, in terms of both accuracy and stability of prediction, STEA achieved highly comparable performance as RCTD and substantially outperformed many of the remaining methods (Figure 2c,d). Given that STEA is a reference-independent method, its performance remarkably demonstrates its high flexibility but without compromising its robustness. It should be noted that STdeconvolve produces components (topics) rather than explicit cell-type identities for individual spots. Therefore, to enable a fair comparison with ground-truth annotations, we followed the recommended workflow and assigned biological cell-type identities to each component/topic by performing transcriptional correlation analysis between topic-specific expression profiles and the ground-truth cell-type signatures (Figure S1a,b).

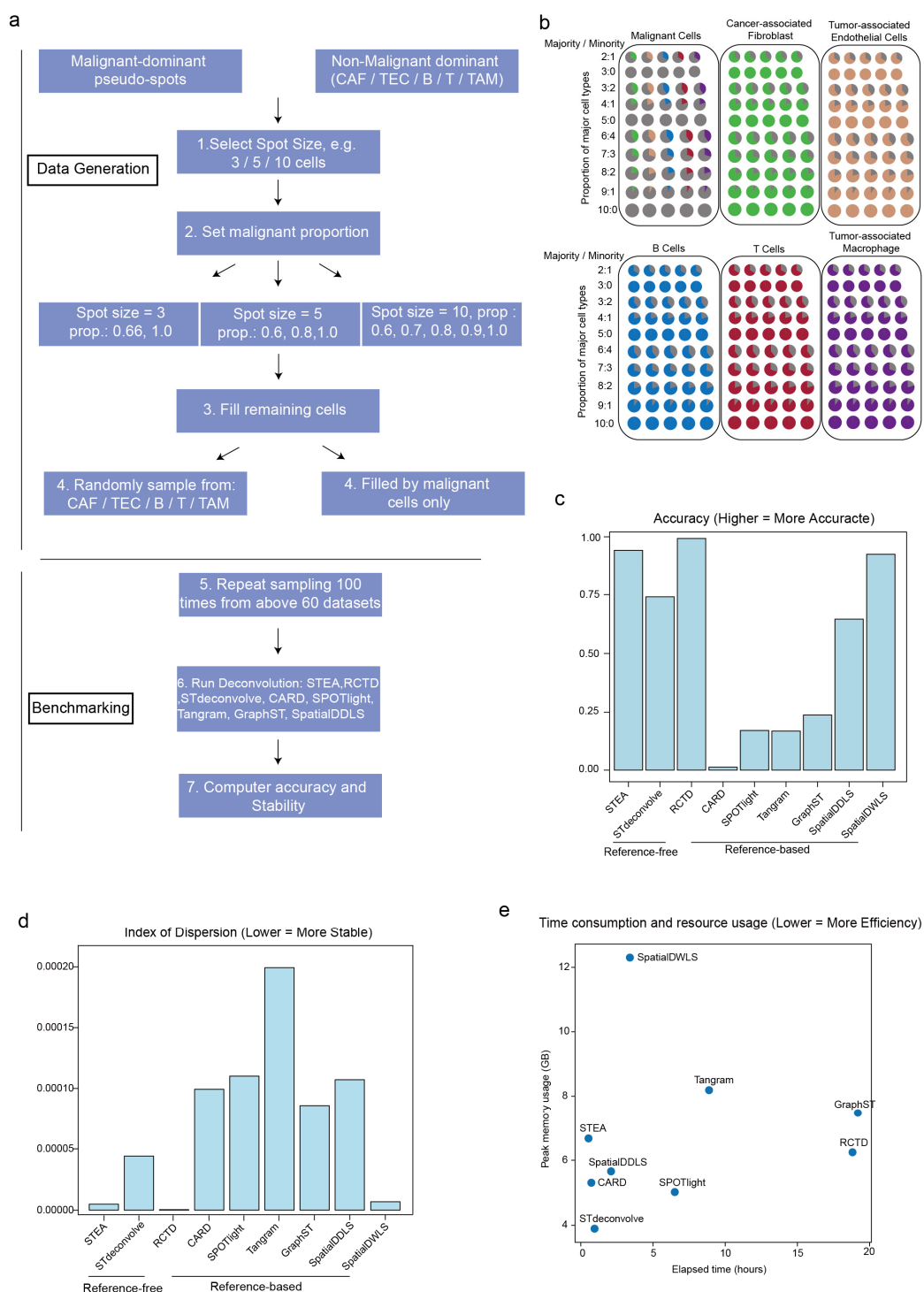


Figure 2. Benchmarking on simulated sequencing-based HCC ST dataset. **(a)** Outline of data generation process and benchmarking analysis. **(b)** Schematic diagram of malignant pseudo-spots, CAFs pseudo-spots, TECs pseudo-spots, B cells pseudo-spots, T cells pseudo-spots and TAM pseudo-spots. Gray represents malignant cells; green represents CAFs; brown represents TECs; blue represents B cells; red represents T cells; and purple represents TAM. **(c)** The box plot compares annotation accuracy among STEA, STdeconvolve, RCTD, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs and SpatialDWLS on simulated ST datasets. **(d)** The bar chart illustrates index of dispersion from STEA, STdeconvolve, RCTD, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs and SpatialDWLS on 100 cycles. **(e)** The scatter plot shows resources usage and time consumption among STEA, STdeconvolve, RCTD, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs and SpatialDWLS on simulated ST datasets.

Beyond predictive performance, we further assessed computational efficiency under a single-core CPU setting to ensure fair comparisons of running time and resource usage (Figure 2e). This setup minimizes variability introduced by hardware heterogeneity and highlights intrinsic algorithmic efficiency. Methods such as SpatialDWLS, Tangram, GraphST, and RCTD showed relatively poorer computational efficiency, characterized by longer runtimes and higher CPU or memory consumption. In contrast, STEA, together with CARD, SPOTlight, and SpatialDDLs, demonstrated moderate computational costs, requiring shorter runtimes and fewer computational resources. Although STdeconvolve was the most computationally efficient method, its inferior accuracy limited its overall applicability. Taken together, STEA achieves a favorable balance between predictive accuracy and computational efficiency.

3.3. Benchmarking Using Simulated Image-Based ST Datasets

We also benchmarked STEA and other deconvolution methods using a simulated image-based ST dataset [23]. Ground truth annotations show that excitatory neurons (eNeuron) constitute the main proportion of each spot, which mixed with minority cell types including inhibitory neurons (iNeuron), astrocytes, oligodendrocytes cells (Olig), microglia cells, and endothelial-mural cells (endo_mural) (Figure 3a,b). This dominance is further supported by the strong spatial expression of eNeuron marker genes across the tissue (Figure 3c).

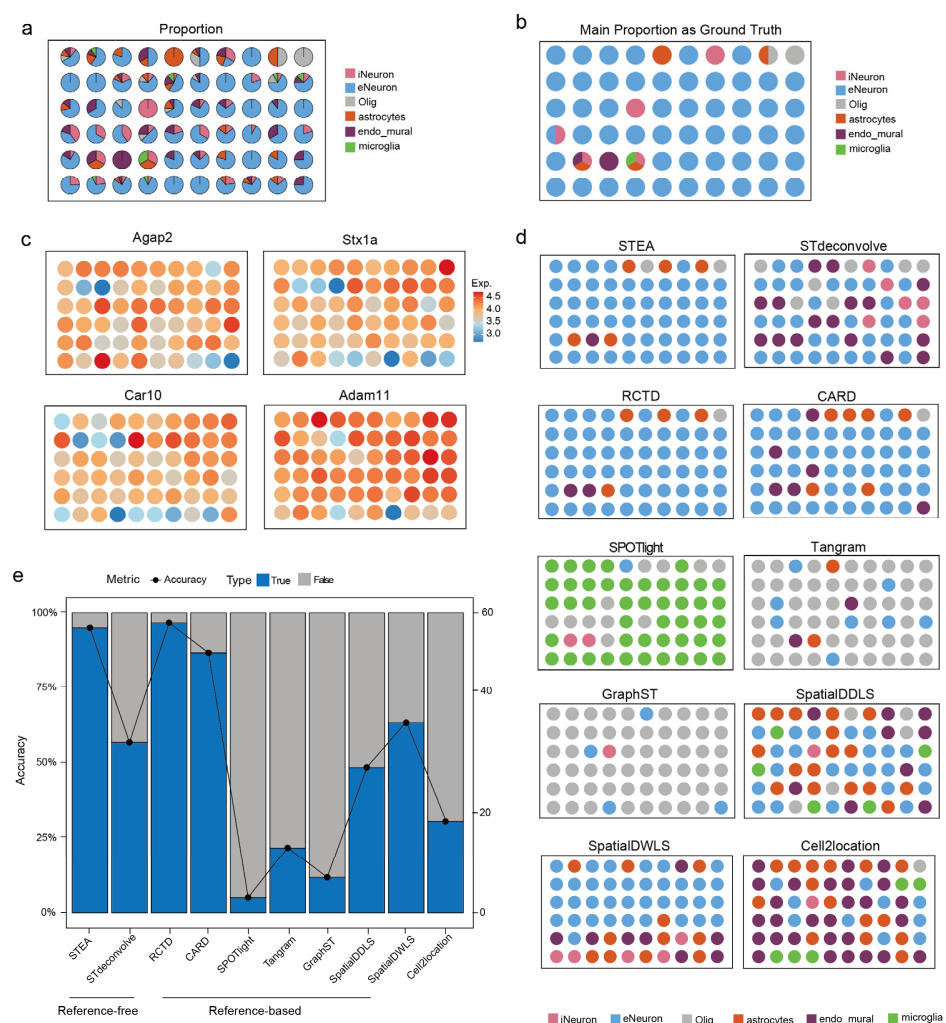


Figure 3. Benchmarking on simulated image-based ST datasets. (a) The cell type proportion of the simulated image-based ST dataset. eNeuron indicates excitatory neurons cells, iNeuron indicates inhibitory

neurons cells, Olig indicates oligodendrocytes, and endo mural indicates endothelial-mural cells. **(b)** The ground truth of the simulated image-based ST dataset. **(c)** The expression of eNeuron marker genes shows the distribution. **(d)** Results of major cell-type annotation of STEA, STdeconvolve, RCTD, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs, SpatialDWLS and Cell2location on simulated image-based ST datasets. **(e)** Accuracy of cell type annotation of different methods.

In the reference-free setting, STEA outperformed STdeconvolve (Figure 3d,e; Figure S2a,b). More importantly, STEA produced cell-type prediction results that were nearly identical to the ground truth, indicating its robustness. For reference-based methods, the deconvolution results revealed varying proportions of different cell types across them. This indicates that the deconvolution varied significantly, with some methods provided apparently better definition as compared to the others. For example, most spots were mistakenly predicted as microglia cells by SPOTlight, while oligodendrocytes cells were frequently and wrongly identified by Tangram and GraphST. SpatialDDLs, SpatialDWLS and Cell2location also produced inaccurate and heterogeneous results compared to the ground truth. Results of RCTD, and CARD appeared to be more reasonable and accurate compared to the others. In general, they correctly identified eNeurons which were predominantly found in the data, with the prediction on minority cell types slightly better by RCTD. Taken together, STEA's results were generally accurate and demonstrated high competitiveness as compared to both reference-free and reference-based methods in defining different major and minor cell types (Figure 3d,e).

3.4. Benchmarking Using Real ST Dataset on Intraductal Papillary Mucinous Neoplasm (IPMN)

Next, apart from aforementioned simulated datasets, we also accomplished benchmarking on another intricate IPMN ST dataset [38]. This dataset consists of two samples and they primarily comprise epithelial and fibroblast cells but there were also other immune and stromal cells (Figure 4a,c).

Deconvolution results of both samples generally exhibited high consistency, but performance of individual methods significantly varied (RCTD was used in the original study to analyze the IPMN ST dataset and therefore it was omitted in the comparison) (Figure 4a–d; Figure S3a–d). Notably, SPOTlight exhibited a strong bias towards monocytes and fibroblasts, resulting in spurious detections. Tangram and GraphST predominantly but inaccurately identified many epithelial cells, with very few fibroblasts. The results of SpatialDDLs were highly heterogeneous, scattering across the tissue without retaining endogenous pattern of enrichment of cell types within the tissue locality. Importantly, most of their predictions were incorrect, raising concerns about their accuracy. In sample 1, STdeconvolve assigned almost all fibroblasts to the myeloid lineage. In sample 2, most fibroblasts were predominantly classified as endothelial or epithelial cells (Figure 4a,c; Figure S3a,b). The results of CARD, SpatialDWLS, Cell2location and STEA were comparable and substantially resemble the ground truth, suggesting their robust cell type prediction. In consistent with the results using simulated datasets, STEA performed well using real ST data and its performance was satisfactory and demonstrated marked margin in outperforming many of the evaluated methods.

3.5. Benchmarking Using Real ST Dataset on Mouse Olfactory Bulb

We further evaluated the performance on a sequence-based ST dataset derived from the mouse olfactory bulb [8], which consists of four distinct layers: the granule cell layer (GCL), the mitral cell layer (MCL), the glomerular layer (GL) and the nerve layer (ONL) (Figure 5a, top panel). The expected cell types include granule cells (GCs), olfactory sensory neurons (OSNs), periglomerular cells (PGCs), and mitral/tufted cells (M-TCs) (Figure 5a, bottom panel). The distinctive expressions of representative marker genes support the

designated cell types and their spatial localization (Figure 5b). To mitigate the influence of variable cell type proportions, cell numbers were normalized across reference cell types.

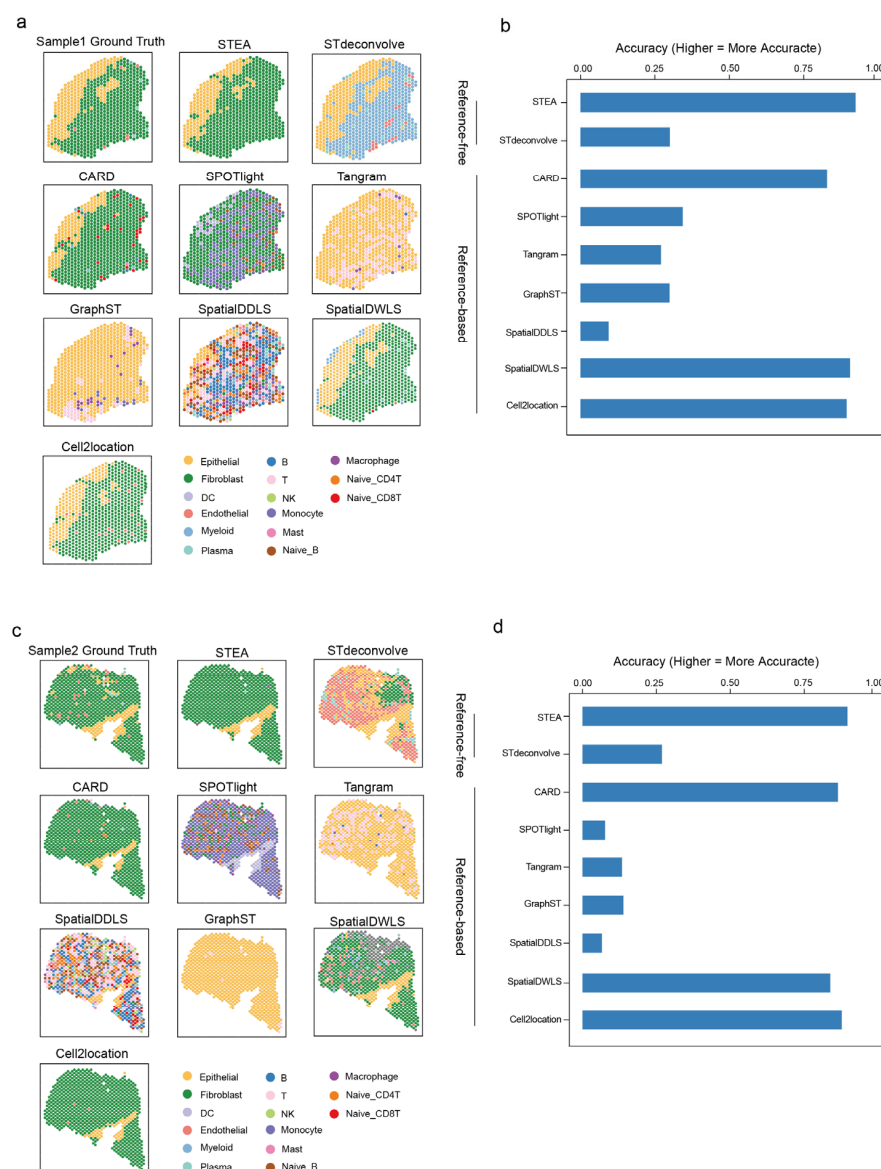


Figure 4. Benchmarking on IPMN ST datasets. (a) The Ground Truth of IPMN sample 1 ST dataset and deconvolution results of STEA, STdeconvolve, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs, SpatialDWLS and Cell2location. (b) The comparison of deconvolution accuracy among STEA, STdeconvolve, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs, SpatialDWLS and Cell2location on IPMN sample 1 ST dataset. (c) The Ground Truth of IPMN sample 2 ST dataset and deconvolution results of STEA, STdeconvolve, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs, SpatialDWLS and Cell2location. (d) The comparison of deconvolution accuracy among STEA, STdeconvolve, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs, SpatialDWLS and Cell2location on IPMN sample 2 ST dataset.

Benchmarking analyses of deconvolution methods revealed the concerns of several issues. For reference-free method, STdeconvolve misclassified PGCs as M-TCs. In the reference-based setting, SPOTlight tended to misclassify GCs as PGCs, while Tangram and GraphST assigned nearly all cells as GCs. SpatialDDLs identified OSNs ubiquitously throughout the tissue instead of confining within the expected ONL. SpatialDWLS misjudged OSNs as EPL-Ins, while Cell2location misidentified PGCs as M-TCs. RCTD and CARD performed better than the aforementioned methods, but they also suffered from

general misclassifications on various cell types, limiting their overall accuracy. While differentiating between M-TCs and PGCs was challenging as many methods failed this task in the benchmarking, STEA accomplished this with remarkable clarity and hierarchical precision over neighboring cell types. Moreover, STEA outperformed the other methods in identifying OSNs that clearly exhibited their expected enrichment in ONL. Taken together, its superior performance was further demonstrated by its accurate identification of cell types and clear delineation of tissue architecture (Figure 5c–e; Figure S4a,b).

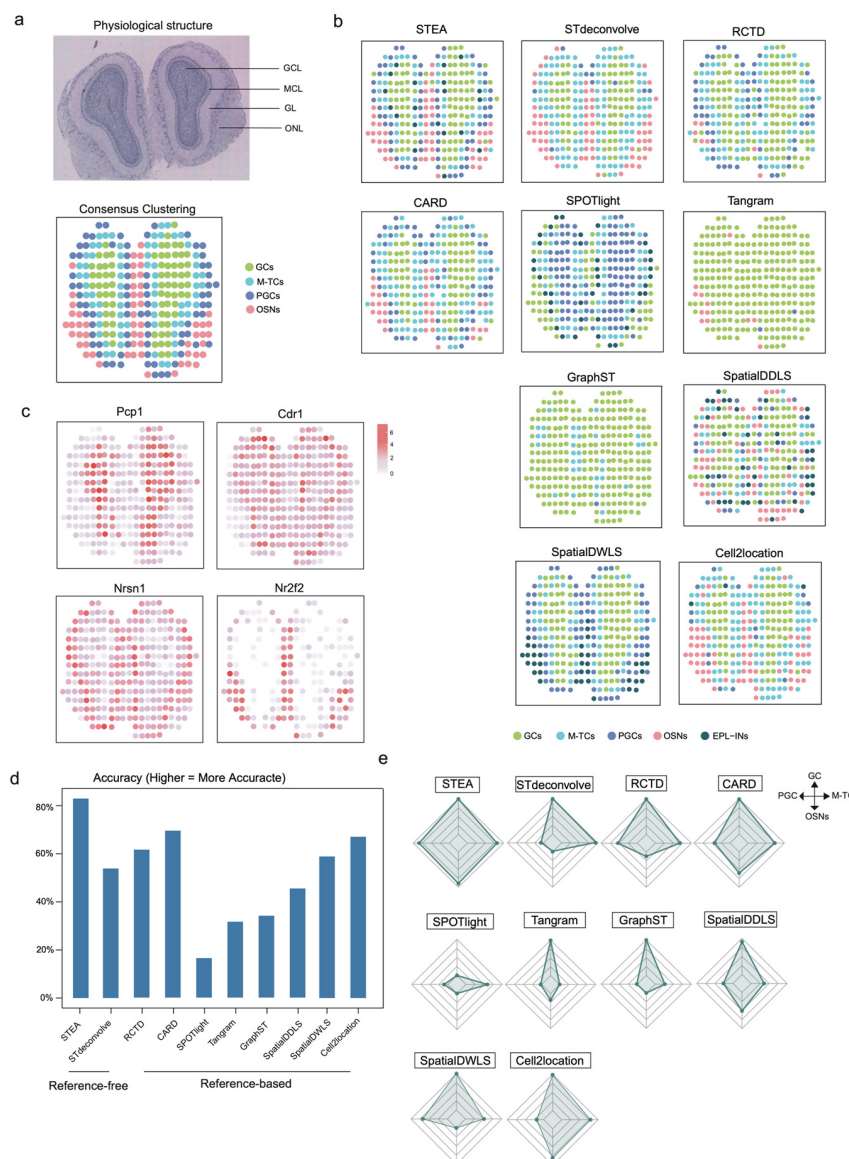


Figure 5. Benchmarking on mouse olfactory bulb ST datasets. (a) The physiological structure of mouse olfactory bulb and the ground truth of mouse olfactory bulb ST datasets. (b) The expression of marker genes shows layers distribution in mouse olfactory bulb. (c) Deconvolution results of STEA, STdeconvolve, RCTD, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs, SpatialDWLS and Cell2location on mouse olfactory bulb ST datasets. GCs indicates granule cells. OSNs indicates olfactory sensory neurons. PGCs indicates periglomerular cells. M-TCs indicates mitral/tufted cells. EPL-Ins indicates external plexiform layer interneurons. (d) The bar chart indicates deconvolution accuracy from STEA, STdeconvolve, RCTD, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs, SpatialDWLS and Cell2location. (e) The radar plot shows the deconvolution accuracy of different cell types from STEA, STdeconvolve, RCTD, CARD, SPOTlight, Tangram, GraphST, SpatialDDLs, SpatialDWLS and Cell2location.

3.6. Cell-Type Annotation by STEA Demonstrated High Concordance with Pathologist's Histological Classification Based on Hepatocellular Carcinoma (HCC) ST Dataset

To further validate the performance of STEA, we analyzed multiple HCC ST samples and compared the annotation results with histological regions annotated by pathologist [39,40]. In the hematoxylin and eosin (HE) images, pathologist manually classified tumor regions, immune regions, fibrous regions, immune and fibrous regions, and necrotic regions. Importantly, there were distinct cell types that characterized these regions: for instance, tumor regions were predominantly composed of tumor cells, while immune and fibrous regions contained mainly immune cells and fibroblasts.

Across the eight samples (Figure 6), STEA consistently detected a high abundance of tumor cells in tumor regions. In the tumor regions of samples 2, 7 and 8, STEA revealed subtle infiltration of immune cells. In the immune regions of samples 1 and 6, B cells, T cells, and TAMs were all detected as expected. Fibroblasts were also found to be distributed and enriched within the fibrous regions of samples 3 and 8. For all samples with necrotic regions, they were determined to have marked depletion of tumor cells while immune and stromal cells were also found to be enriched in samples 1 and 4. It seems prediction of STEA can faithfully capture the cell type characteristics in making precise annotation and shows strong concordance with histological classification by pathologist across the tested samples.

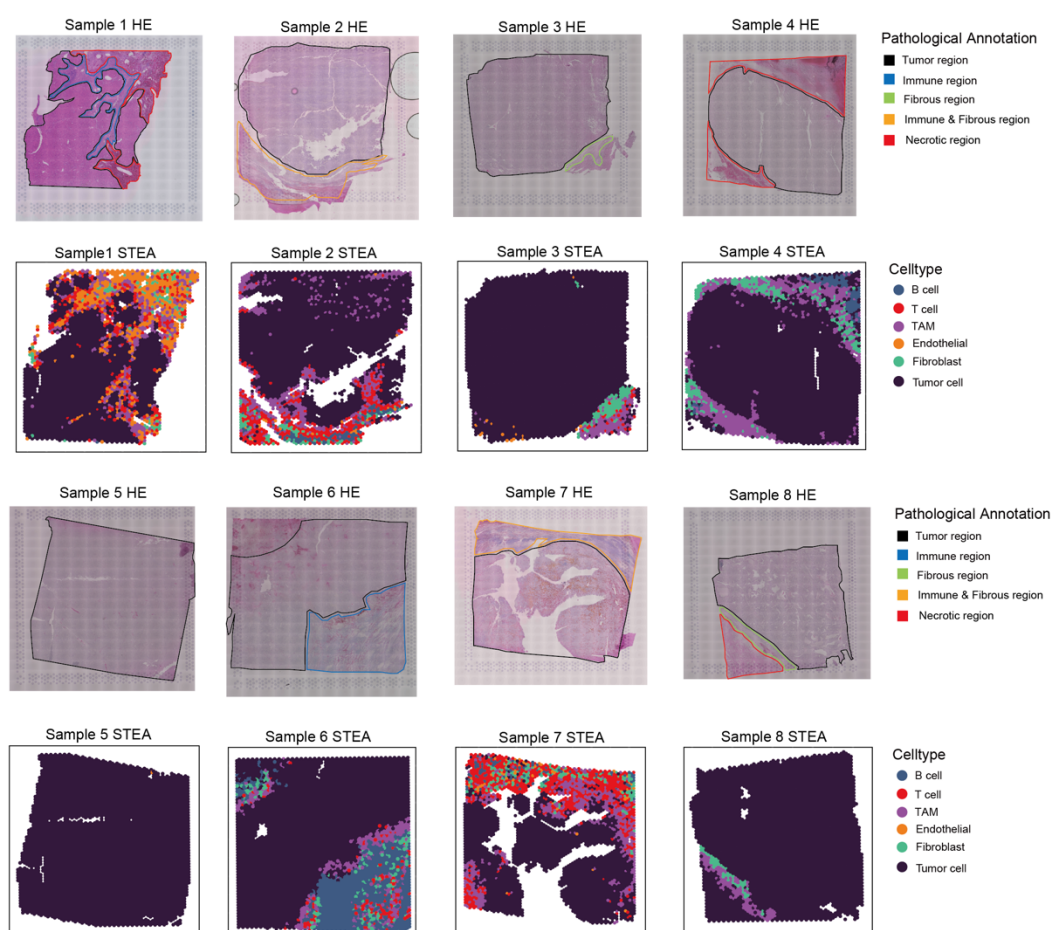


Figure 6. Pathological annotation and annotation results of STEA on eight HCC samples. Annotation results of STEA (major cell types) was compared against the histological annotation by pathologist (tumor region, immune region, fibrous region, immune and stromal region, and necrotic region).

4. Discussion

The emergence of ST technology has driven the development of a wide range of deconvolution tools, with different underlying principles and assumptions. Yet, their performance remains highly sensitive to data quality, choices of parameters and the selection of scRNA-seq reference that even minor variations can significantly impact the results. Here, we present our reference-independent enrichment-based cell-type annotation method called STEA. Based on various benchmarking results using a variety of synthetic and real ST datasets pinpointing different biological scenarios, STEA accurately identified the true or expected cell types and its performance were among the topmost methods, with consistent and marked margins in surpassing many of the counterparts. Unlike many other tools that commonly displayed cell type bias favoring the frequent and consistent misclassification of certain spurious identities, we did not notice major inclination or tendency of STEA in making unjust prediction. This could be demonstrated in the overall unbiased deconvolution using datasets of heterogeneous backgrounds. Importantly, classification by STEA also accords well with pathologist's annotation, which partly provides the essential basis in technically and professionally supporting its cell type prediction.

Despite the apparent strengths of STEA, it also has some limitations. First, its reliance on conventional marker genes or predefined gene sets can yield suboptimal outcomes when these markers lack sufficient signal intensity. While STEA could identify the overall patterns in the IPMN datasets, it hardly detected sparsely distributed minor populations. We believe the gene expression signal intensities for different cell types are highly divergent and therefore the strengths for their corresponding marker signatures are unequal, resulting in differential efficiencies in identifying the presence of different cell types (i.e., some minor cell types are more difficult to detect due to their weaker strength of marker signatures). This highlights a potential vulnerability to misclassification when signal intensity is low. The adaptive marker optimization process of STEA could partly enhance the efficiency of marker signatures and reduce the reliance on a few key markers (the presence of multiple correlated weak markers may somehow compensate for the absence of key markers). Moreover, current cell-type deconvolution methods often face challenges in estimating accurate cell-type proportions for individual spots. Although different methods report the extent or likelihood of cell type enrichment in spots, they are not fully equivalent to their exact biological proportions. While precise cell type proportion estimation can provide useful information for spatial transcriptomics analysis, many studies primarily utilize deconvolution tools to determine the major cell types of the spots for spot-level annotation. Therefore, the aim of our STEA algorithm is to make use of the extent of enrichment for different cell-type-related gene sets to determine the major cell type of the spots. Its designated purpose is spot annotation but not for accurate estimation of continuous cell type proportion. Importantly, the advent of subcellular-level ST technologies offers an unprecedented way to determine the spatially resolved gene expression profiles at single-cell resolution. Findings from those newer platforms reveal that individual spots may no longer be whole-cell mixtures but fractions of single cells or even discrete cellular compartments instead. This may ease the necessity for precise determination of cell type proportions, but general cell-type annotation should still be required, just the same as the case of single-cell sequencing. Next, due to incompetence of GPU or multi-core utilization for many of the tested tools, we intentionally conducted all the computational efficiency benchmarking using single CPU core to ensure a standardized and comparable evaluation across methods. The suboptimal utilization of GPU-competent tools, e.g., Cell2location, Tangram and GraphST, cannot fully reflect their true operational performance.

5. Conclusions

In summary, STEA provides a robust framework for major cell-type annotation, validated across diverse ST platforms. Through comprehensive benchmarking and histopathological comparison, STEA demonstrates reliable performance and maintains biological plausibility, preserving the spatial and functional integrity of complex tissues such as TME. Regarding its future development, we envision that further optimization of the current STEA functionalities, together with an enhanced cell segmentation process, is expected to enable more precise and holistic annotation of different transcriptomic datasets in cancer and other complex tissues.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers18091425/s1>, Figure S1: Deconvolution results of STdeconvolve on simulated sequencing-based HCC ST dataset. (a). The deconvolution results. (b). The heatmap illustrates the transcriptional correlation between ground truth cell types and deconvolved components; Figure S2: Deconvolution results of STdeconvolve on simulated imaged-based ST datasets. (a). The deconvolution results. (b). The heatmap illustrates the transcriptional correlation between ground truth cell types and deconvolved components; Figure S3: Deconvolution results of STdeconvolve on intraductal papillary mucinous neoplasm datasets. (a). The deconvolution results of sample 1. (b). The heatmap illustrates the transcriptional correlation between ground truth cell types and deconvolved components. (c). The deconvolution results of sample 2. (d). The heatmap illustrates the transcriptional correlation between ground truth cell types and deconvolved components; Figure S4: Deconvolution results of STdeconvolve on mouse olfactory bulb datasets. (a). The deconvolution results. (b). The heatmap illustrates the transcriptional correlation between ground truth cell types and deconvolved components; Table S1: The summary of key parameters setting; Table S2: Marker gene lists.

Author Contributions: Study concept and design: D.W.-H.H.; Acquisition of data: Q.L., F.Z. and Q.Z.; Analysis and interpretation of data: Q.L. and D.W.-H.H.; Acquisition and analysis of clinical samples: I.O.-L.N.; Drafting of the manuscript: Q.L. and D.W.-H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Hong Kong Research Grants Council Theme-based Research Scheme (T12-716/22-R), Innovation and Technology Commission grant to State Key Laboratory of Liver Research (ITC PD/17-9), Health and Medical Research Fund (10212956 and 07182546), RGC General Research Fund (17100021 and 17117019). I.O.L. Ng is Loke Yew Professor in Pathology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Datasets used in this study were derived from public sources. The scRNA-seq dataset for generating the synthetic HCC ST datasets are freely available at the Gene Expression Omnibus (GEO) under accession number GSE125449. The simulated image-based ST dataset can be found at the website (https://github.com/rdong08/spatialDWLS_dataset, accessed on 27 April 2026). The IPMN ST datasets are available from the Gene Expression Omnibus (GEO) under accession number GSE233254. The MOB ST datasets are available at the website (<http://www.spatialtranscriptomicsresearch.org>, accessed on 28 January 2025) and the scRNA-seq dataset can be found at the Gene Expression Omnibus (GEO) under accession number GSE121891. The HCC ST datasets are available at PMID:38344450 and PMID: 40205519. All marker gene sets are provided as Supplementary Table S2.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Marx, V. Method of the Year: Spatially Resolved Transcriptomics. *Nat. Methods* **2021**, *18*, 9–14. [[CrossRef](#)] [[PubMed](#)]
2. Moses, L.; Pachter, L. Museum of Spatial Transcriptomics. *Nat. Methods* **2022**, *19*, 534–546. [[CrossRef](#)]

3. Bassiouni, R.; Gibbs, L.D.; Craig, D.W.; Carpten, J.D.; McEachron, T.A. Applicability of Spatial Transcriptional Profiling to Cancer Research. *Mol. Cell* **2021**, *81*, 1631–1639. [\[CrossRef\]](#)
4. Williams, C.G.; Lee, H.J.; Asatsuma, T.; Vento-Tormo, R.; Haque, A. An Introduction to Spatial Transcriptomics for Biomedical Research. *Genome Med.* **2022**, *14*, 68. [\[CrossRef\]](#)
5. Lewis, S.M.; Asselin-Labat, M.-L.; Nguyen, Q.; Berthelet, J.; Tan, X.; Wimmer, V.C.; Merino, D.; Rogers, K.L.; Naik, S.H. Spatial Omics and Multiplexed Imaging to Explore Cancer Biology. *Nat. Methods* **2021**, *18*, 997–1012. [\[CrossRef\]](#)
6. Rao, A.; Barkley, D.; França, G.S.; Yanai, I. Exploring Tissue Architecture Using Spatial Transcriptomics. *Nature* **2021**, *596*, 210–220. [\[CrossRef\]](#)
7. Tian, L.; Chen, F.; Macosko, E.Z. The Expanding Vistas of Spatial Transcriptomics. *Nat. Biotechnol.* **2022**, *41*, 773–782. [\[CrossRef\]](#)
8. Ståhl, P.L.; Salmén, F.; Vickovic, S.; Lundmark, A.; Navarro, J.F.; Magnusson, J.; Giacomello, S.; Asp, M.; Westholm, J.O.; Huss, M.; et al. Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics. *Science* **2016**, *353*, 78–82. [\[CrossRef\]](#)
9. Rodriques, S.G.; Stickels, R.R.; Goeva, A.; Martin, C.A.; Murray, E.; Vanderburg, C.R.; Welch, J.; Chen, L.M.; Chen, F.; Macosko, E.Z. Slide-Seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution. *Science* **2019**, *363*, 1463–1467. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Liu, Y.; Yang, M.; Deng, Y.; Su, G.; Enniful, A.; Guo, C.C.; Tebaldi, T.; Zhang, D.; Kim, D.; Bai, Z.; et al. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **2020**, *183*, 1665–1681.e18. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Merritt, C.R.; Ong, G.T.; Church, S.E.; Barker, K.; Danaher, P.; Geiss, G.; Hoang, M.; Jung, J.; Liang, Y.; McKay-Fleisch, J.; et al. Multiplex Digital Spatial Profiling of Proteins and RNA in Fixed Tissue. *Nat. Biotechnol.* **2020**, *38*, 586–599. [\[CrossRef\]](#)
12. Lim, H.J.; Wang, Y.; Buzdin, A.; Li, X. A Practical Guide for Choosing an Optimal Spatial Transcriptomics Technology from Seven Major Commercially Available Options. *BMC Genom.* **2025**, *26*, 47. [\[CrossRef\]](#)
13. Hatton, I.A.; Galbraith, E.D.; Merleau, N.S.C.; Miettinen, T.P.; Smith, B.M.; Shander, J.A. The Human Cell Count and Size Distribution. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2303077120. [\[CrossRef\]](#)
14. Ginzberg, M.B.; Kafri, R.; Kirschner, M. On Being the Right (Cell) Size. *Science* **2015**, *348*, 1245075. [\[CrossRef\]](#)
15. Burgess, D.J. Spatial Transcriptomics Coming of Age. *Nat. Rev. Genet.* **2019**, *20*, 317. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Asp, M.; Bergenstråhle, J.; Lundeberg, J. Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration. *BioEssays* **2020**, *42*, 1900221. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Gaspard-Boulinc, L.C.; Gortana, L.; Walter, T.; Barillot, E.; Cavalli, F.M.G. Cell-Type Deconvolution Methods for Spatial Transcriptomics. *Nat. Rev. Genet.* **2025**, *26*, 828–846. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Li, H.; Zhou, J.; Li, Z.; Chen, S.; Liao, X.; Zhang, B.; Zhang, R.; Wang, Y.; Sun, S.; Gao, X. A Comprehensive Benchmarking with Practical Guidelines for Cellular Deconvolution of Spatial Transcriptomics. *Nat. Commun.* **2023**, *14*, 1548. [\[CrossRef\]](#)
19. Cable, D.M.; Murray, E.; Zou, L.S.; Goeva, A.; Macosko, E.Z.; Chen, F.; Irizarry, R.A. Robust Decomposition of Cell Type Mixtures in Spatial Transcriptomics. *Nat. Biotechnol.* **2021**, *40*, 517–526. [\[CrossRef\]](#)
20. Ma, Y.; Zhou, X. Spatially Informed Cell-Type Deconvolution for Spatial Transcriptomics. *Nat. Biotechnol.* **2022**, *40*, 1349–1359. [\[CrossRef\]](#)
21. Kleshchevnikov, V.; Shmatko, A.; Dann, E.; Aivazidis, A.; King, H.W.; Li, T.; Elmentaite, R.; Lomakin, A.; Kedlian, V.; Gayoso, A.; et al. Cell2location Maps Fine-Grained Cell Types in Spatial Transcriptomics. *Nat. Biotechnol.* **2022**, *40*, 661–671. [\[CrossRef\]](#)
22. Elosua-Bayes, M.; Nieto, P.; Mereu, E.; Gut, I.; Heyn, H. SPOTlight: Seeded NMF Regression to Deconvolute Spatial Transcriptomics Spots with Single-Cell Transcriptomes. *Nucleic Acids Res.* **2021**, *49*, e50. [\[CrossRef\]](#)
23. Dong, R.; Yuan, G.-C. SpatialDWLS: Accurate Deconvolution of Spatial Transcriptomic Data. *Genome Biol.* **2021**, *22*, 145. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Biancalani, T.; Scalia, G.; Buffoni, L.; Avasthi, R.; Lu, Z.; Sanger, A.; Tokcan, N.; Vanderburg, C.R.; Segerstolpe, Å.; Zhang, M.; et al. Deep Learning and Alignment of Spatially Resolved Single-Cell Transcriptomes with Tangram. *Nat. Methods* **2021**, *18*, 1352–1362. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Long, Y.; Ang, K.S.; Li, M.; Chong, K.L.K.; Sethi, R.; Zhong, C.; Xu, H.; Ong, Z.; Sachaphibulkij, K.; Chen, A.; et al. Spatially Informed Clustering, Integration, and Deconvolution of Spatial Transcriptomics with GraphST. *Nat. Commun.* **2023**, *14*, 1155. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Mañanes, D.; Rivero-García, I.; Relación, C.; Torres, M.; Sancho, D.; Jimenez-Carretero, D.; Torroja, C.; Sánchez-Cabo, F. SpatialDDLs: An R Package to Deconvolute Spatial Transcriptomics Data Using Neural Networks. *Bioinformatics* **2024**, *40*, btae072. [\[CrossRef\]](#)
27. Miller, B.F.; Huang, F.; Atta, L.; Sahoo, A.; Fan, J. Reference-Free Cell Type Deconvolution of Multi-Cellular Pixel-Resolution Spatially Resolved Transcriptomics Data. *Nat. Commun.* **2022**, *13*, 2339. [\[CrossRef\]](#)
28. Chidester, B.; Zhou, T.; Alam, S.; Ma, J. SpiceMix Enables Integrative Single-Cell Spatial Modeling of Cell Identity. *Nat. Genet.* **2023**, *55*, 78–88. [\[CrossRef\]](#)
29. Aibar, S.; Gonzalez-Blas, C.B.; Moerman, T.; Huynh-Thu, V.A.; Imrichova, H.; Hulselmans, G.; Rambow, F.; Marine, J.C.; Geurts, P.; Aerts, J.; et al. SCENIC: Single-Cell Regulatory Network Inference and Clustering. *Nat. Methods* **2017**, *14*, 1083–1086. [\[CrossRef\]](#)

30. Andreatta, M.; Carmona, S.J. UCell: Robust and Scalable Single-Cell Gene Signature Scoring. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3796–3798. [[CrossRef](#)]
31. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)] [[PubMed](#)]
32. Mootha, V.K.; Lindgren, C.M.; Eriksson, K.-F.; Subramanian, A.; Sihag, S.; Lehar, J.; Puigserver, P.; Carlsson, E.; Ridderstråle, M.; Laurila, E.; et al. PGC-1 α -Responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes. *Nat. Genet.* **2003**, *34*, 267–273. [[CrossRef](#)]
33. Barbie, D.A.; Tamayo, P.; Boehm, J.S.; Kim, S.Y.; Moody, S.E.; Dunn, I.F.; Schinzel, A.C.; Sandy, P.; Meylan, E.; Scholl, C.; et al. Systematic RNA Interference Reveals That Oncogenic KRAS-Driven Cancers Require TBK1. *Nature* **2009**, *462*, 108–112. [[CrossRef](#)] [[PubMed](#)]
34. Fleck, J.S.; Camp, J.G.; Treutlein, B. What Is a Cell Type? *Science* **2023**, *381*, 733–734. [[CrossRef](#)]
35. Zeng, H. What Is a Cell Type and How to Define It? *Cell* **2022**, *185*, 2739–2755. [[CrossRef](#)] [[PubMed](#)]
36. Kelley, K.W.; Nakao-Inoue, H.; Molofsky, A.V.; Oldham, M.C. Variation among Intact Tissue Samples Reveals the Core Transcriptional Features of Human CNS Cell Classes. *Nat. Neurosci.* **2018**, *21*, 1171–1184. [[CrossRef](#)]
37. Ma, L.; Hernandez, M.O.; Zhao, Y.; Mehta, M.; Tran, B.; Kelly, M.; Rae, Z.; Hernandez, J.M.; Davis, J.L.; Martin, S.P.; et al. Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. *Cancer Cell* **2019**, *36*, 418–430.e6. [[CrossRef](#)]
38. Sans, M.; Makino, Y.; Min, J.; Rajapakshe, K.I.; Yip-Schneider, M.; Schmidt, C.M.; Hurd, M.W.; Burks, J.K.; Gomez, J.A.; Thege, F.I.; et al. Spatial Transcriptomics of Intraductal Papillary Mucinous Neoplasms of the Pancreas Identifies NKX6-2 as a Driver of Gastric Differentiation and Indolent Biological Potential. *Cancer Discov.* **2023**, *13*, 1844–1861. [[CrossRef](#)]
39. Zeng, F.; Zhang, Q.; Tsui, Y.-M.; Ma, H.; Tian, L.; Husain, A.; Lu, J.; Lee, J.M.-F.; Zhang, V.X.; Li, P.-M.; et al. Multimodal Sequencing of Neoadjuvant Nivolumab Treatment in Hepatocellular Carcinoma Reveals Cellular and Molecular Immune Landscape for Drug Response. *Mol. Cancer* **2025**, *24*, 110. [[CrossRef](#)]
40. Cheung, T.-T.; Wai-Hung Ho, D.; Lyu, S.X.; Zhang, Q.; Tsui, Y.-M.; Ching-Yun Yu, T.; Man-Fong Sze, K.; Man-Fong Lee, J.; Lau, V.W.-H.; Yin-Lun Chu, E.; et al. Multimodal Integrative Genomics and Pathology Analyses in Neoadjuvant Nivolumab Treatment for Intermediate and Locally Advanced Hepatocellular Carcinoma. *Liver Cancer* **2024**, *13*, 70–88. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.