

RESCUE: recovery of unattributed expression patterns in spatial transcriptomics

Received: 17 September 2025

Accepted: 26 March 2026

Published online: 10 April 2026



Young Joo Lee¹, Seokjin Yeo², Alex W. Schrader³, JuYeon Lee³, Ian M. Traniello⁴, Marisa Asadian³, Amy Cash Ahmed⁵, Gene E. Robinson^{5,6,7}, Hee-Sun Han^{2,3,5,6}✉ & Sihai Dave Zhao^{1,5}✉

Spatial transcriptomics (ST) enables gene expression profiling while preserving the spatial architecture of intact tissue. Analyzing ST data often proceeds by first extracting cell-level information, typically through cell segmentation or cell-type deconvolution. However, a critical oversight has been that a substantial portion of molecular expression is systematically lost or unannotated by these methods. This lost expression can arise from diverse and biologically important sources like fragile or underrepresented cell types, subcellular structures like neurites, and extracellular expression. These omissions can result in biased analyses and incorrect or incomplete biological interpretations. We describe a new computational method, RESCUE, that can recover the unattributed spatial expression patterns missed by existing ST analysis methods and enable robust inference even when reference is incomplete. We validate RESCUE using MERFISH data from the honey bee brain and apply it to multiple ST datasets to demonstrate how it can reveal novel insights into complex tissue biology.

Profiling cellular expression within its spatial context has become possible with the recent development of spatial transcriptomics (ST) technologies, which measure gene expression in intact tissue sections^{1,2}. Analyzing ST data typically proceeds by extracting cell-level information from the spatial expression data, as cells are natural units of analysis. With high-resolution platforms such as MERFISH³, 10x Visium HD⁴, or 10x Xenium⁵, this is usually done using cell segmentation^{6–8}. With low-resolution platforms like 10x Visium⁹ and Slide-seq^{10,11}, where each spatial location can contain multiple cells, it is common to deconvolve the observed expression into separate contributions from different cell types, based on reference single-cell/nuclei RNA-seq (sc/snRNA-seq) data^{12–16}.

However, a major problem with these existing ST analysis methods is that they can fail to capture important expression patterns. First, they may omit entire cells or cell types. Cell segmentation algorithms

for high-resolution data have difficulty with cells with complex morphologies, such as fibroblasts, or cells lacking nuclei, like erythrocytes^{17,18}. Deconvolution methods for low-resolution data cannot account for cell types that are absent from the reference sc/snRNA-seq data because they are too fragile to survive tissue dissociation^{19–21}. Second, existing ST approaches can overlook expression patterns unique to cellular substructures. Segmentation methods simply drop transcripts located in substructures that are difficult to segment, such as the axons and dendrites of neurons. These fragile structures are also typically destroyed during tissue dissociation and, therefore, are not represented in reference single-cell data used in deconvolution. Indeed, it has been estimated that over 40% of the total transcriptome of an adult mouse brain may be missed by sc/snRNA-seq²². Similarly, reference snRNA-seq data, by definition, ignores cytoplasmic expression. Finally, existing methods do not capture extracellular tissue

¹Department of Statistics, University of Illinois Urbana-Champaign, Urbana, IL, USA. ²Department of Bioengineering, University of Illinois Urbana-Champaign, Urbana, IL, USA. ³Department of Chemistry, University of Illinois Urbana-Champaign, Urbana, IL, USA. ⁴Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. ⁵Carl R. Woese Institute of Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA. ⁶Neuroscience Program, University of Illinois Urbana-Champaign, Urbana, IL, USA. ⁷Department of Entomology, University of Illinois Urbana-Champaign, Urbana, IL, USA.

✉ e-mail: hshan@illinois.edu; sdzhao@illinois.edu

expression patterns, which cannot be studied by segmentation or deconvolution.

These difficult-to-capture expression patterns can be critically important. First, they can bias analyses that do not take them into account. An example with dramatic implications is that deconvolution estimates of cell-type proportions can be invalid if the reference is missing any cell types²³, as we illustrate later in this work. Second, there may be important biological processes that are unique to the overlooked expression patterns. For example, important missed cell types include adipocytes⁵, which play crucial roles in metabolic disease and cancer²⁴, and neutrophils²⁵, which are core components of the immune system. Important missed substructural expression includes axonal or dendritic transcripts that are central to synaptic development and plasticity^{26–30} and diseases such as amyotrophic lateral sclerosis^{31,32}, as well as cytoplasmic transcripts that regulate key processes such as metabolism, signaling, and stress responses^{33–35} and are involved in Alzheimer's disease^{36,37}. Important missed extracellular expression patterns include extracellular RNAs that mediate intercellular communication in cancer^{38–40}.

We describe a new computational method, RESCUE, that can recover the spatial expression patterns missed by single-cell inference on ST data. Our approach is based on making a fundamental correction to the class of statistical models currently used for integrating ST data and reference data, such as sc/snRNA or cell segmentation results. Briefly, our method partitions observed gene expression into one component explainable by the reference data, which we refer to as “canonical” expression, and another component that is not, which we refer to as “unattributed” expression. Our purified canonical ST expression component can be used to correct existing cell-type deconvolution methods. Our recovered unattributed ST expression component gives us the new ability to specifically target biology that is

distinct from our reference data, as discussed above. The unattributed expression recovery step in RESCUE is analogous to negative selection in experimental biology^{41–44}, where we identify and remove canonical expression. We demonstrate that RESCUE accurately recovers unattributed expression, enables robust deconvolution of ST data despite incomplete reference data, and reveals biologically meaningful patterns overlooked by existing methods. While we focus here on ST data, we note that RESCUE can also be directly applied to non-spatial modalities, such as bulk RNA-seq data.

Results

Overview of RESCUE

RESCUE is predicated on a new model that we propose for integrating ST and reference single-cell data, and can be applied to diverse ST platforms, including MERFISH and Visium, across different organisms. Most existing models^{12–16} posit that ST expression at each location is composed of a linear combination of expression patterns present in the reference, where the coefficients of this linear combination are interpreted as the proportions of each reference cell type in that location. However, as described in the introduction, there are numerous sources of ST expression that are left out of this model; we will refer to these as unattributed expression patterns, because they are not attributed to canonical expression patterns reflected in the reference.

To address this issue, we propose a corrected model (Fig. 1) that allows the ST profile to have both canonical and unattributed expression components. While it appears minor, this change has important implications. Under this model, results from the statistics literatures^{45–47} imply that under certain conditions it is possible to simultaneously recover unattributed expression and estimate the linear combination of reference patterns that constitute canonical

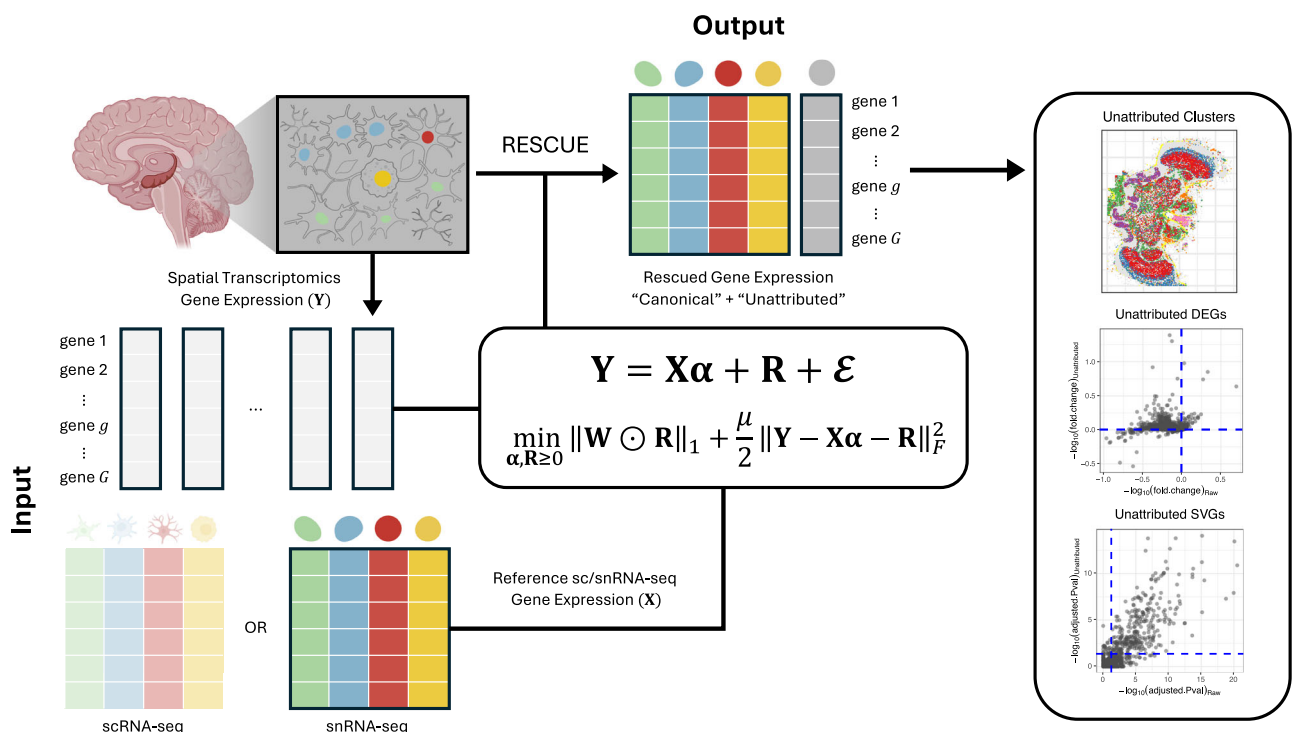


Fig. 1 | Schematic overview of RESCUE. RESCUE is a computational negative selection method designed to recover the spatial expression patterns missed by standard single-cell inference on spatial transcriptomics (ST) data. It takes as input ST data to be decomposed, along with a reference matrix constructed from single-cell/nuclei RNA-seq (sc/snRNA-seq) data from the same tissue, and performs robust sparse recovery via a regularized multivariate regression framework. RESCUE

partitions observed gene expression into a reference-explainable “canonical” component and a residual “unattributed” component. As a result, the purified canonical expression enables more accurate deconvolution, while the recovered unattributed expression enables downstream analyses such as clustering, differentially expressed gene (DEG), and spatially variable gene (SVG) analyses. Created with BioRender.com.

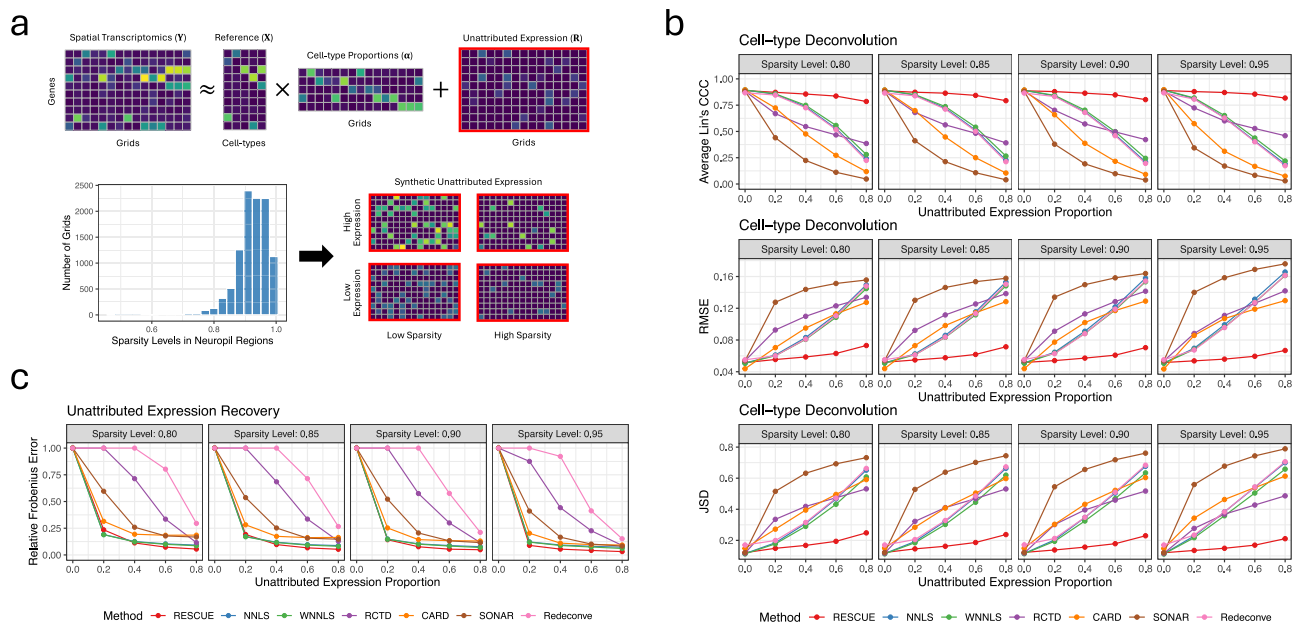


Fig. 2 | Simulation results. **a** Schematic of the simulation framework. Spatial transcriptomics (ST) data are generated as the sum of reference-driven canonical expression and synthetic unattributed expression. The distribution of sparsity levels observed in neuropil regions of the honey bee brain motivates the simulated sparsity regimes. Representative examples of synthetic unattributed expression under low- and high-sparsity/expression levels are shown. **b** Cell-type

deconvolution performance across methods, evaluated using average Lin's concordance correlation coefficient (CCC; top), root mean squared error (RMSE; middle), and Jensen-Shannon divergence (JSD; bottom) as a function of the sparsity/proportion levels of the unattributed expression. **c** Unattributed expression recovery performance measured by relative Frobenius error under the same simulation settings. Source data are provided as a Source Data file.

expression. The conditions are that only relatively few genes show non-zero, or sparse expression, in the unattributed component, which is reasonable given the nature of unattributed expression and the read depth/detection efficiency of ST technologies (Methods).

Leveraging this insight, we developed RESCUE, an algorithm that simultaneously deconvolves ST data while rescuing spatial unattributed expression. RESCUE solves a regularized robust multivariate regression problem and can be applied to both low- and high-resolution ST data. Unlike existing deconvolution methods^{12–16}, it is robust to incomplete references, such as snRNA-seq data which lack cytoplasmic expression profiles. We emphasize that the interpretation of the unattributed component depends on both the biological context of the tissue and the choice of the reference dataset (Methods). As a result, the unattributed component may comprise multiple biological sources, including expression from cell types missing in the reference, substructural expression, and extracellular RNA. RESCUE is designed to recover these signals jointly rather than to assign them to predefined categories. The rescued unattributed expression can then be further decomposed using standard unsupervised approaches, such as clustering and non-negative matrix factorization (NMF), to identify distinct expression patterns within the unattributed component. In addition, differentially expressed gene (DEG) and spatially variable gene (SVG) analyses can be applied to characterize the biological and spatial features of the recovered unattributed signals. Careful consideration of biological structures and tissue organization is therefore essential when interpreting the unattributed expression. We illustrate this in a broad range of examples below.

Simulations

To evaluate the performance of RESCUE, we conducted simulation studies on synthetic data designed to mimic real ST measurements with varying degrees of unattributed expression. We simulated ST data using real scRNA-seq data from the honey bee brain⁴⁸ by introducing synthetic unattributed expression of a small number of genes

(Methods). Unlike other tissues where cell bodies and non-cell-body structures are spatially intermingled, honey bee brains provide a unique benefit for simulating unattributed expression due to the existence of the neuropil region, where axons and dendrites are spatially segregated from cell bodies. Molecular expression in these regions is largely absent in single-cell sequencing data due to the structural fragility of neurites. During the tissue dissociation step, a necessary step in single-cell sequencing, neurites are mostly sheared off and lost. In contrast, ST measurements performed on intact tissue sections preserve these structures and therefore capture molecular expression within neuropils.

To guide the choice of sparsity levels used in our simulations, we examined empirical expression patterns in the neuropil region of our in-house MERFISH data. We found that neuropil regions exhibit highly sparse gene expression, with an average sparsity level of 0.92 across spatial locations (Fig. 2a). Here, sparsity refers to the structure of the unattributed component (i.e., that only a subset of genes is present at a given spatial location), rather than to low expression levels (Methods). Based on this empirical distribution, we simulated unattributed expression across a range of sparsity levels by varying the proportion of zero entries (ranging from 0.8 to 0.95). In addition, we varied the proportion of unattributed expression (ranging from 0 to 0.8) to cover different levels of the overall magnitude of unattributed expression. Importantly, these parameter ranges fall within those observed in real biological samples and are not artificially chosen to favor RESCUE.

We first compared the deconvolution performance of RESCUE with existing methods. Figure 2b summarizes the accuracy of cell-type proportion estimates across methods, measured using the average Lin's concordance correlation coefficient⁴⁹ (CCC), root mean squared error (RMSE), and Jensen-Shannon divergence (JSD) (Methods). As expected, the performance of standard deconvolution methods deteriorates sharply as the level of unattributed expression increases. In contrast, RESCUE maintains high accuracy across all conditions, demonstrating robustness to both high unattributed content and

reduced sparsity. These results indicate that failing to explicitly model unattributed expression can lead to systematic biases in cell-type proportion estimates.

We also evaluated the accuracy of unattributed expression recovery (Fig. 2c) using the relative Frobenius error between the true and estimated unattributed expression matrices (Methods). Overall, RESCUE achieves the lowest recovery error even under high unattributed expression and low sparsity levels, whereas other methods deviate substantially from the ground truth, especially when unattributed signals become strong. These results indicate that explicitly modeling unattributed expression enables accurate recovery of hidden transcriptional signals across a wide range of sparsity and signal-strength regimes.

To assess robustness and generalizability, we expanded the simulation study by varying noise levels via library size variability across spots and varying the effective grid resolution via changes in the average total UMI count per spot (Supplementary Fig. 1). As expected, overall performance tends to decrease with increasing library size variability and increase with higher average total UMIs. Across these regimes, RESCUE remains among the top-performing methods, demonstrating robustness across a range of noise levels and grid sizes.

RESCUE accurately recovers the unattributed transcriptome in real tissues

Using a custom MERFISH platform^{50,51}, we generated single molecule-resolution ST data from coronal sections of a honey bee brain, targeting 130 genes relevant to aggression (Supplementary Data 1)⁵². Transcripts in axons and dendrites, which are collectively termed as neurites, have been difficult to profile due to segmentation challenges. Cell segmentation methods, such as Cellpose⁶ and Baysor⁷, can only identify the boundaries of neuronal cell bodies, also called somata. Although individual axons and dendrites can be stained in 2D cell cultures⁵³, in native tissues they extend across multiple imaging planes and therefore cannot be reliably segmented as individual cellular units. As a result, neurite-associated expression has traditionally been ignored or incorrectly merged into nearby cytosolic expression.

We first experimentally validated that RESCUE can accurately recover gene expression in the neuropil region of the honey bee brain (Fig. 3a). This validation is possible because the physical separation of neurites and somata in the brain makes it simple for existing algorithms to separate somatic from unattributed expression. Cell soma were segmented using Cellpose⁶, revealing that only 73.5% of transcripts were localized within cells, with the remaining 26.5% comprising the unattributed expression. Using honey bee brain scRNA-seq data⁴⁸ as a reference, we then applied RESCUE and competing methods to our MERFISH data after dividing the brain into grids of size $20 \times 20 \mu\text{m}^2$ (Methods). For each grid, we estimated the proportion of canonical expression relative to total expression and obtained a ground-truth estimate by identifying transcripts localized within cell bodies via cell segmentation. RESCUE's estimates closely matched the ground truth (Fig. 3b) and were more accurate than those from other methods, as quantified by Lin's CCC and RMSE (Methods). Notably, RESCUE relies only on reference scRNA-seq data and does not use cell segmentation information.

To further clarify the source of discrepancies among methods, we evaluated canonical and unattributed expression separately in anatomically well-defined regions. Specifically, we examined grids corresponding to pure cell bodies and compared the average canonical expression of each gene against the ground truth derived from segmentation (Fig. 3c). RESCUE closely matched the true somatic expression levels, whereas competing methods systematically underestimated canonical expression. This behavior reflects the fact that existing deconvolution approaches are designed to estimate cell-type proportions rather than expression levels, and under the subtraction-based framework (Methods), expression that is not well explained by

the reference profiles is left as residual rather than attributed to the canonical component. Conversely, in grids containing no identifiable cell bodies, RESCUE accurately recovered unattributed expression in neuropil-only regions (Fig. 3d), while competing methods were significantly less accurate. Together, these results demonstrate that RESCUE effectively partitions the full transcriptome into canonical and unattributed components in real tissue, capturing valuable expression signals that would otherwise be lost using conventional approaches.

Importantly, transcripts outside segmented somata reflect genuine biological signals rather than technical artifacts: in our MERFISH sample (Fig. 3a), neuropil-associated transcripts form structured spatial patterns that are clearly segregated from DAPI-stained nuclei (Supplementary Fig. 2). Moreover, the MERFISH protocol includes gel embedding, expansion, and clearing steps that reduce autofluorescence and suppress nonspecific signal^{54,55}. These signals, which would have been ignored by standard cell segmentation approaches, highlight the importance of RESCUE in capturing overlooked transcriptomic information.

Finally, to investigate whether the unattributed expression recovered by RESCUE reflects structured biological organization rather than unstructured noise, we performed clustering analysis of the unattributed expression profiles across spatial locations. Clustering revealed anatomically distinct transcriptional niches within neuropil regions (Fig. 3e), indicating that the recovered unattributed component exhibits coherent spatial patterns. Specifically, distinct clusters localize to known neuropil structures, including the mushroom body calyces, antennal lobes, optic lobes, and regions proximal to the central complex. In contrast, spatial locations corresponding to somatic regions collapsed into a single cluster, consistent with the fact that canonical expression is explicitly removed, and only minimal unattributed signal remains in these regions. Examination of representative genes within the neuropil-associated clusters showed spatially structured expression patterns that are largely absent from the canonical component (Fig. 3f). These findings further support that the unattributed expression recovered by RESCUE captures biologically meaningful and spatially organized signals, revealing hidden tissue structure that is not accessible through conventional analyses.

RESCUE enables accurate deconvolution of ST data with incomplete reference

Here, we illustrate how RESCUE's estimated canonical expression component can be used to achieve accurate deconvolution of ST data when key cell types are missing from the reference, a challenge for all existing methods that has been noted in a recent study²³. We analyzed paired single-cell FFPE-seq (scFFPE-seq) and 10x Visium ST data from human breast cancer tissue⁵. In this ST dataset, adipocytes are manually annotated in approximately 10% of spatial spots (Fig. 4a). However, adipocytes are absent from the scFFPE-seq reference, likely due to their high lipid content interfering with fixation and processing. These cells are prone to floating, rupturing, or sticking to plastic surfaces during tissue dissociation, which hinders their recovery in single-cell sequencing protocols^{5,17}. As a result, reference-based methods that assume the observed expression is fully explained by the available reference cell types tend to misattribute adipocyte-derived transcripts to other cell types.

In adipocyte-rich regions, existing deconvolution methods fail to account for the missing adipocyte transcriptome and instead misattribute the signal to other reference cell types, such as stromal and endothelial cells (Fig. 4b). This systematic misassignment likely arises from transcriptional similarity, as shown by the strong correlation between adipocyte and stromal/endothelial profiles in pseudo-bulk measurements from Visium and scFFPE-seq (Fig. 4c). In contrast, RESCUE explicitly models unattributed expression and thus avoids forcing unassigned signal into existing reference cell types, which results in more accurate and biologically plausible estimates. As shown

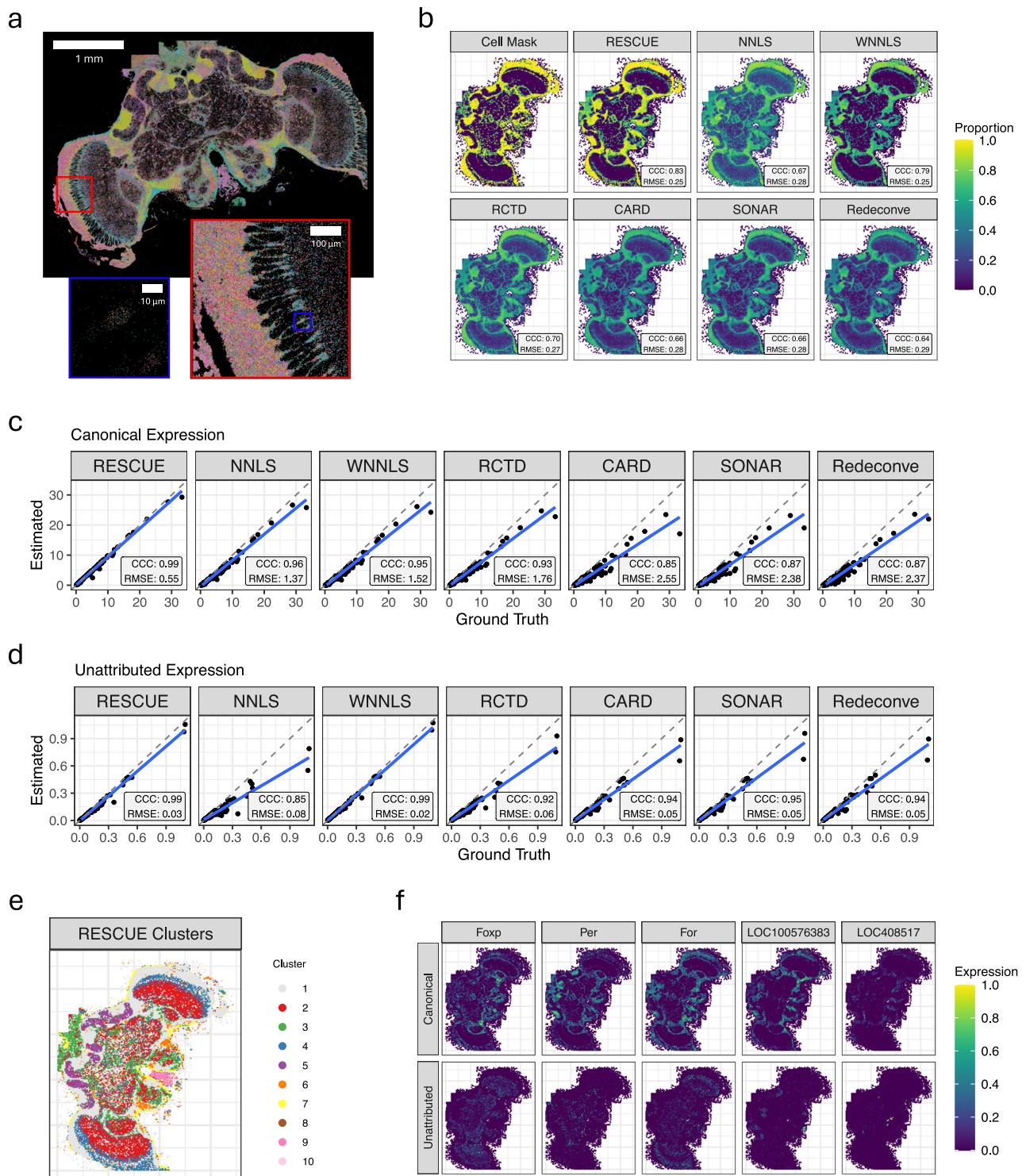


Fig. 3 | Application to the honey bee brain MERFISH data. **a** MERFISH profiling of a coronal honey bee brain section. Each dot represents an individual transcript, with genes shown in distinct colors. A substantial fraction of transcripts localize to neurites, rather than to cell bodies. **b** Comparison of segmentation-based ground-truth (Cell Mask) and estimated somatic proportions. Cell mask ratio is computed as the proportion of MERFISH spots in each grid lying within a cell. Lin's concordance correlation coefficient (CCC) and root mean squared error (RMSE) between ground-truth and estimated proportions are reported. **c** Comparison of estimated versus ground-truth canonical gene expression levels in grids corresponding to pure cell body regions. **d** Comparison of estimated versus ground-

truth unattributed gene expression levels in neuropil-only grids lacking segmented somata. For **c** and **d**, each point represents a gene averaged across grids; gray dashed lines indicate the identity line, and blue lines denote linear regression fits. Lin's CCC and RMSE between ground-truth and estimated average expression are reported. **e** Clustering of the unattributed expression profiles inferred by RESCUE reveals distinct transcriptional niches within neuropil regions, while somatic regions collapse into a single cluster (Cluster 1). **f** Spatial expression patterns of representative genes, shown separately for canonical and unattributed components. Expression values are min-max normalized separately for each panel. Source data are provided as a Source Data file.

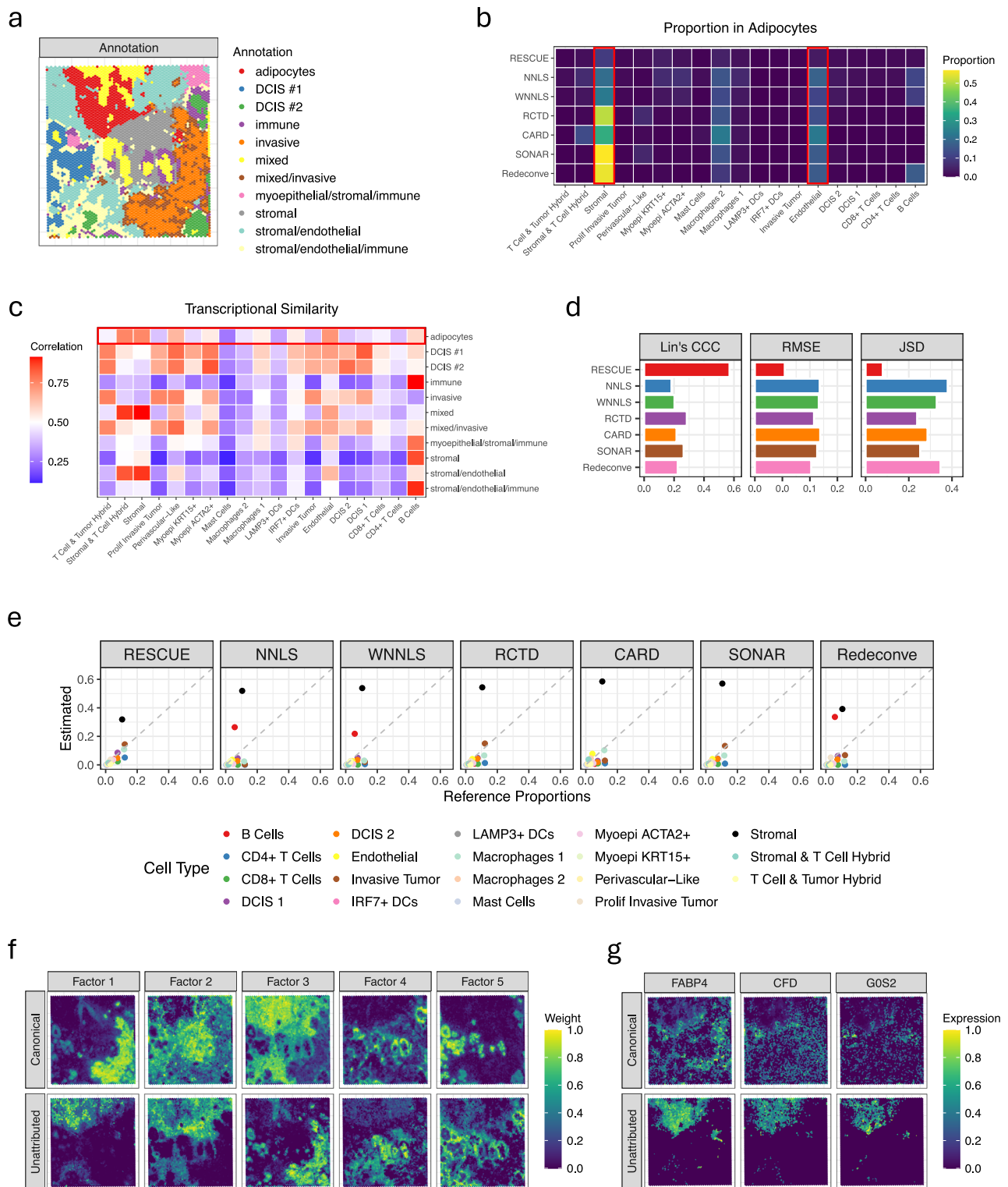


Fig. 4 | Application to the human breast cancer Visium data. a Manual annotation of Visium data from human breast cancer tissue. **b** Estimated cell-type proportions in adipocyte regions across methods. Red box highlights that existing deconvolution methods misattribute adipocyte-derived signal to other reference cell types due to the absence of adipocytes in the reference, whereas RESCUE mitigates such misassignment by modeling unattributed component. **c** Heatmap of correlations between pseudo-bulk Visium and scFFPE-seq reference cell types. Red box highlights high correlations between adipocytes and other cell types including stromal and endothelial cells. **d** Comparison of cell-type deconvolution accuracy across methods using Lin's concordance correlation coefficient (CCC), root mean

squared error (RMSE), and Jensen-Shannon divergence (JSD). **e** Comparison of estimated versus reference cell-type proportions across methods. Each point represents a cell type, with gray dashed lines indicating the identity line. **f** Non-negative matrix factorization (NMF) of the canonical and unattributed expression component inferred by RESCUE. Weight values are min-max normalized separately for each panel. **g** Spatial expression patterns of adipocyte marker genes, shown separately for canonical and unattributed components. Expression values are min-max normalized separately for each panel. Source data are provided as a Source Data file.

in Fig. 4d–e, RESCUE's estimated proportions closely align with the scFFPE-seq composition, while other methods exhibit significant distortions. Together, these results demonstrate that RESCUE improves deconvolution accuracy in settings with incomplete references by preventing the systematic misassignment of unattributed transcriptomes to incorrect cell types.

To further interpret the biological content of the unattributed expression recovered by RESCUE, we applied NMF to the unattributed component. This analysis revealed multiple latent factors capturing structured expression patterns that are not explained by the canonical component (Fig. 4f). Notably, the NMF Factor 1 (Unattributed) captures the spatial distribution of adipocyte-rich regions, effectively recovering a missing cell type absent from the reference. Genes with the highest loadings in this factor include well-established adipocyte markers such as *FABP4*, *CFD*, and *GOS2*^{5,56} (Fig. 4g). In addition, the NMF Factor 2 (Unattributed) is enriched for genes such as *DCN*, *COL14A1*, and *CIR*, which have been implicated in extracellular matrix (ECM) domains^{57–59}. Together, these findings indicate that the unattributed component captured by RESCUE is biologically meaningful, thereby supporting that explicitly modeling it leads to improved deconvolution in the presence of incomplete references.

Finally, to assess whether these deconvolution improvements generalize beyond the human breast cancer example, we benchmarked RESCUE using additional datasets, including human dorso-lateral prefrontal cortex (DLPFC)⁶⁰ and cichlid fish telencephalon⁶¹ Visium samples. In both datasets, reference profiles were derived from snRNA-seq, which inherently underrepresent cytoplasmic and process-localized RNA, leading standard deconvolution methods to misattribute unexplained ST signals. Across these datasets, RESCUE consistently outperformed competing methods (Supplementary Figs. 3–4), indicating that the benefits of explicitly modeling unattributed expression extend across tissues, species, and reference modalities.

RESCUE uncovers layer-specific unattributed expression in the human DLPFC

In many neuroscience experiments, cell-level information is often characterized using snRNA-seq^{62,63}, which can leave a substantial fraction of biologically relevant expression unattributed. Maynard et al.⁶⁰ profiled the DLPFC of neurotypical adult donors using paired snRNA-seq and 10x Visium ST. We applied RESCUE to eight samples in which all six cortical layers and the white matter (WM) were annotated, to recover unattributed expression patterns not represented in the snRNA-seq reference (Fig. 5a). As shown in Fig. 5b, RESCUE reveals that Layer 1 and the WM contain substantially higher levels of unattributed expression compared to other cortical layers. This is consistent with the fact that Layer 1 has sparser cell densities and contains neuropil-rich regions⁵⁷, and that WM contains a substantial amount of axons and myelin⁶⁴, whose transcriptomic content is mostly cytoplasmic and thus is absent from the reference snRNA-seq data.

RESCUE enables us to study biology specific to the unattributed transcriptome of Layer 1. We first identified DEGs by comparing Layer 1 to all other layers using limma⁶⁵, based on either the raw observed ST expression or the unattributed expression recovered by RESCUE. These two analyses gave substantially different results (Fig. 5c). For example, the top five genes most enriched in Layer 1 based on unattributed expression had either fold-changes in the opposite direction (*MT-ND1*, *MT-ND2*, and *MT-CO1*) or negligible fold-changes (*MT3* and *CST3*) in the differential expression analysis using raw data. We also identified genes such as *CLU*, *GFAP*, and *MALAT1* that exhibit clear spatial enrichment within the unattributed component of Layer 1 (Fig. 5d). These genes are associated with local translation and neurite outgrowth, and their enrichment outside nucleus is consistent with known roles of astrocytic processes and non-somatic cellular structures in Layer 1 of the cortex^{66,67}.

To explore the underlying biological processes, we performed gene set enrichment analysis separately for DEGs derived from the unattributed expression and from the raw ST data (Fig. 5e). Notably, gene ontology (GO) terms enriched among the unattributed DEGs include mitochondrial processes: oxidative phosphorylation, energy derivation by oxidation of organic compounds, and generation of precursor metabolites and energy. In contrast, no mitochondria-related pathways reached statistical significance in the analysis based on observed ST expression; only the broad term, response to stress, passed the FDR < 0.05 threshold. These results align with the elevated metabolic demands of a neurite-enriched Layer 1, which is densely populated with synapses and reliant on local energy production to support axonal and dendritic activity^{68,69}. Collectively, these findings demonstrate that RESCUE can uncover previously obscured yet functionally important gene expression programs in Layer 1 of the human DLPFC.

RESCUE reveals spatially structured unattributed expression in the cichlid fish telencephalon

Johnson et al.⁶¹ collected paired snRNA-seq and Visium ST data from the dorsal telencephalon of individual cichlid fishes exhibiting distinct social behaviors. The dorsal telencephalon is characterized by prominent radial glial (RG) processes that originate in the ventricular zone and extend long projections toward superficial pallial regions^{70,71}. These RG projections span nearly the entire depth of the pallium and may play a role in shaping spatial gene expression profiles across distinct layers and compartments, but they do not survive tissue dissociation and thus their transcriptomes are not captured by snRNA-seq.

We applied RESCUE to this dataset to recover unattributed transcriptomic signals using the snRNA-seq data as a reference. To assess spatial heterogeneity, we performed SVG analysis using SPARK-X⁷² on both the observed ST expression and the unattributed expression. As shown in Fig. 6a, a substantial number of SVGs are uniquely identified in the unattributed transcriptome. In Fig. 6b, we show the expression maps of those SVGs that were undetectable in the observed ST expression yet exhibited clear spatial structure once nuclear signals were removed. Among these, *app*, *cntn1*, *dner*, and *eno1* mark distinct transcriptional programs that may operate along RG projections. These genes highlight transcriptional signals that are likely to reflect localized molecular activities along RG scaffolds: *app* encodes a transmembrane protein that promotes synapse formation and plasticity^{73,74}, *cntn1* (Contactin-1) is a GPI-anchored adhesion molecule essential for axon guidance and neurite outgrowth⁷⁵, *dner* mediates neuron-glia interactions that are critical for astrocyte differentiation⁷⁶, and *eno1* (α -enolase) promotes neuritogenesis and contributes to structural remodeling in the developing nervous system⁷⁷. Indeed, Johnson et al.⁶¹ found that bower-building behavior in cichlid fishes may be related to neuronal rebalancing driven by a subpopulation of RG cells and their processes, highlighting the importance of studying transcriptional processes in RG projections, which RESCUE enables.

To further characterize the structure of the unattributed transcriptome, we applied NMF to the unattributed expression component (Fig. 6c). Several unattributed factors are enriched for genes associated with synaptic and process-localized functions, including *slc1a2*, *calm2*, and mitochondrial ATP synthase (Fig. 6d), consistent with transcriptional activity in extended RG scaffolds and neuronal processes. In addition, one factor shows strong enrichment for ribosomal protein genes, suggesting localized protein synthesis within these structures³³. Notably, these structured transcriptional programs emerge only after the canonical expression is removed, indicating that the unattributed expression recovered by RESCUE captures biologically meaningful signals associated with radial glial and cytoplasmic processes that are not represented in snRNA-seq reference data.

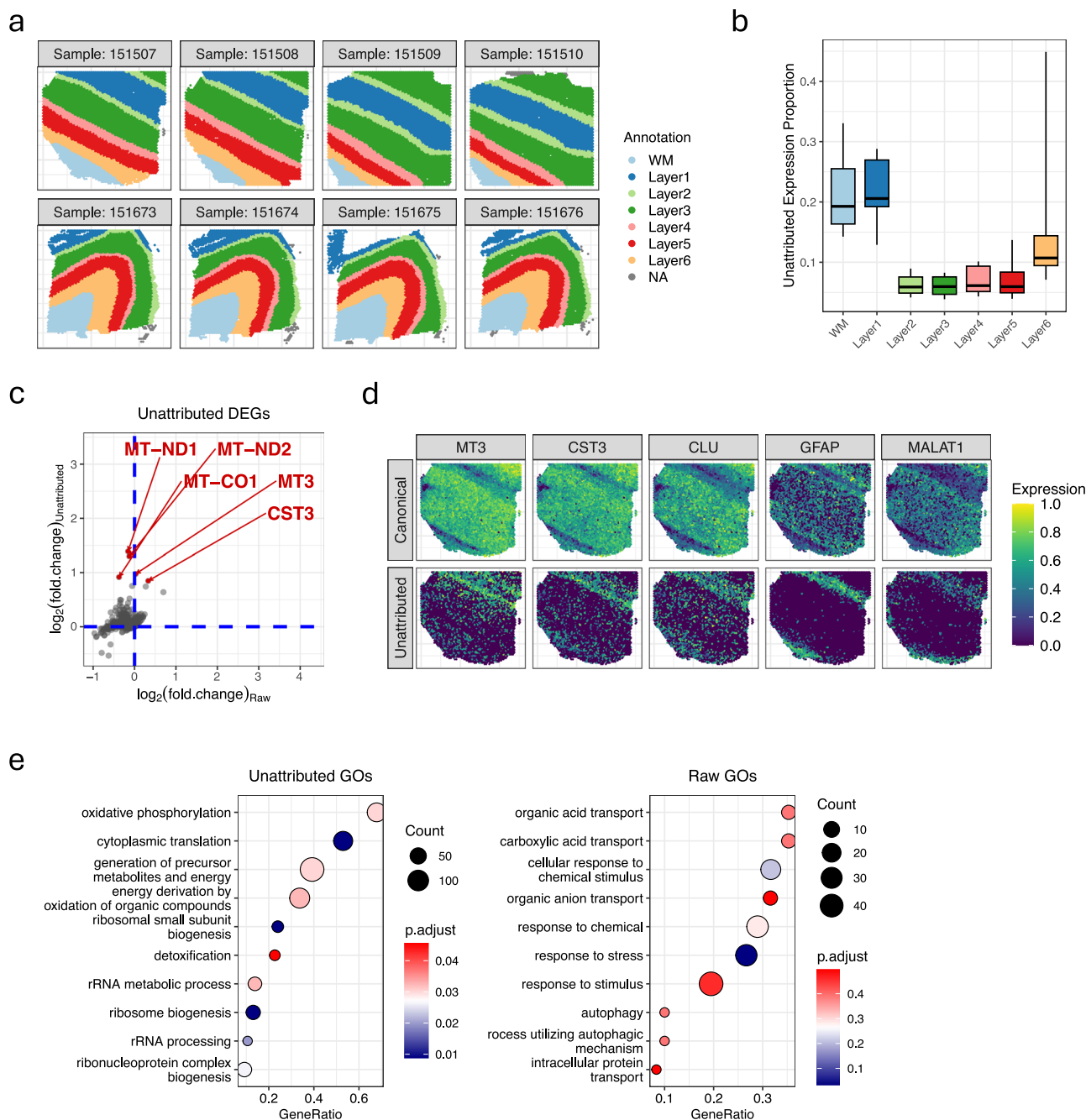


Fig. 5 | Application to the human DLPFC Visium data. a Anatomical annotations of Visium samples from the DLPFC dataset, with cortical layers (Layer 1-6) and white matter (WM) labeled. **b** Proportions of unattributed expression estimated by RESCUE across cortical layers and WM across eight samples⁶⁰ in which all six cortical layers and WM were annotated ($n=8$). Box plots indicate the median (center line), the 25th and 75th percentiles (box bounds), and whiskers extending to the minimum and maximum values. **c** Comparison of log-fold changes for differentially expressed gene (DEG) analysis comparing Layer 1 to all other layers on raw (i.e., observed) expression (x-axis) versus unattributed expression recovered by RESCUE (y-axis). DEG was tested using the R package limma⁶⁵ with two-sided testing and

Benjamini-Hochberg correction for multiple testing. Blue dashed lines indicate zero-fold change. Genes highlighted in red are top Layer 1 markers, detected from unattributed expression. **d** Spatial expression patterns of Layer 1 marker genes derived from unattributed expression, shown separately for canonical and unattributed components. Expression values are min-max normalized separately for each panel. **e** Gene ontology (GO) terms enriched among Layer 1-enriched DEGs based on unattributed (left) and raw (right) expression. Gene set enrichment analysis was performed using the R package clusterProfiler⁶⁹, which uses a weighted Kolmogorov-Smirnov-like test, with Benjamini-Hochberg correction for multiple testing. Source data are provided as a Source Data file.

Discussion

RESCUE is designed to acknowledge and recover the unattributed transcriptome, which encompasses gene expression patterns not captured by cell segmentation or deconvolution approaches in ST data analysis. These unattributed signals, though typically overlooked, can

carry substantial functional and biological significance. While the field has developed numerous transcriptomic profiling methods, such as scRNA-seq for isolating gene expression at the single-cell level, Riboseq for ribosome-associated transcripts, and CLIP-seq for binding-protein interactions, each of these technologies inevitably captures

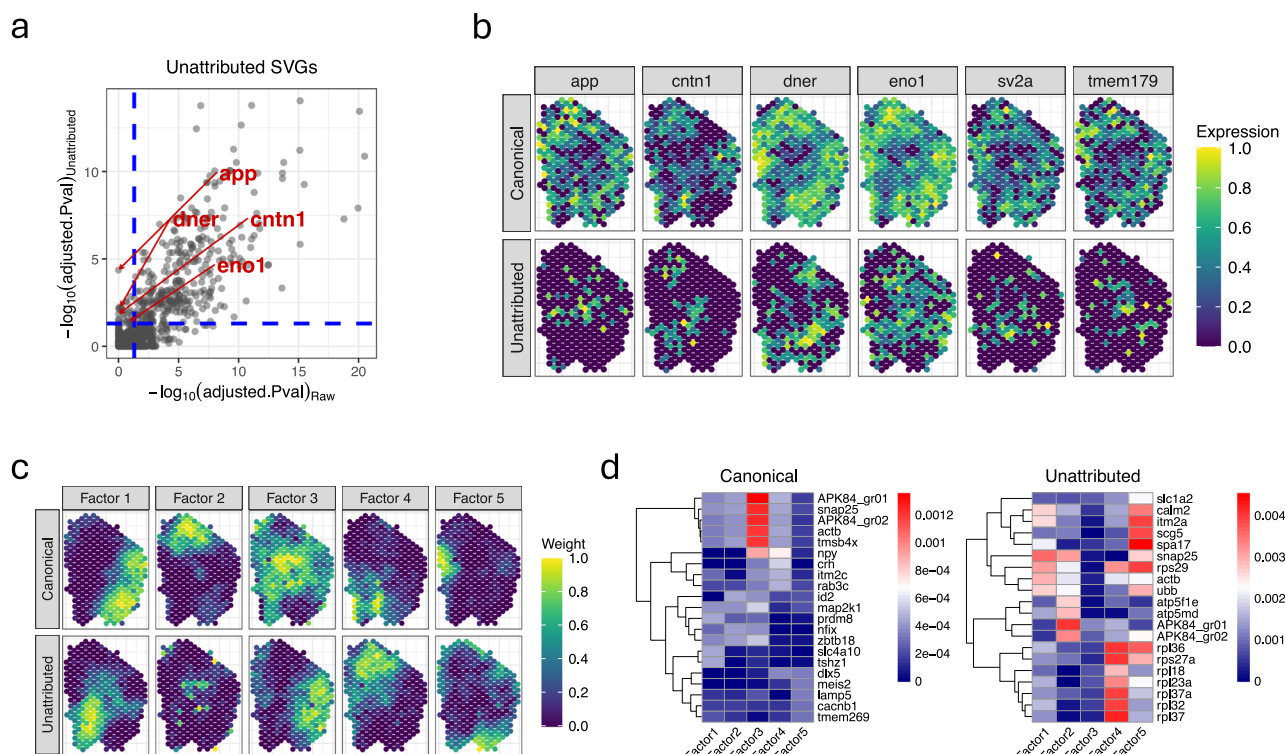


Fig. 6 | Application to the cichlid fish telencephalon Visium data. **a** Comparison of adjusted p-values for spatially variable gene (SVG) detection on raw (i.e., observed) expression (x-axis) versus unattributed expression recovered by RESCUE (y-axis). Blue dashed lines indicate the significance threshold at adjusted $p = 0.05$. Genes labeled in red are SVGs uniquely detected from the unattributed expression. SVG was tested using the R package SPARK-X⁷², which uses a non-parametric covariance test with Benjamini-Yekutieli correction for multiple testing. **b** Spatial

expression patterns of unattributed-only SVGs, shown separately for canonical and unattributed components. Expression values are min-max normalized separately for each panel. **c** Non-negative matrix factorization (NMF) of the canonical and unattributed expression component inferred by RESCUE. Weight values are min-max normalized separately for each panel. **d** Heatmaps of genes with high loadings in canonical (left) and unattributed (right) factors. Source data are provided as a Source Data file.

only a fraction of the total transcriptome. RESCUE provides a systematic framework to quantify and analyze what is missing: the molecular signals outside the scope of reference datasets.

By formulating the problem as a robust sparse recovery framework, RESCUE decomposes observed ST data into a reference-aligned canonical component and a biologically meaningful unattributed component. RESCUE focuses on characterizing what is not explained by reference profiles and introduces a broader conceptual innovation: computational negative selection. Rather than focusing solely on what to include (as in positive selection), we define a reference component to subtract out and focus on what remains. In this work, we used scRNA-seq and snRNA-seq data as reference profiles and focused primarily on brain tissue, but the concept is general and can be applied to a wide range of other settings, for example, muscle tissue, where multinucleated cells make segmentation challenging and single-cell reference data difficult to interpret. The choice of reference therefore requires careful design and quality control.

One striking finding emerging from our brain tissue analyses is the substantial amount of expression (26.5% of total expression) found in the unattributed transcriptome of the honey bee brain MERFISH data. While localization of mRNAs to neurites is a well-studied phenomenon^{26–30}, our results, along with the estimates for the adult mouse brain from Ament and Poulpoulos²², additionally suggest that these local expression patterns may be different from those found in cell soma. RESCUE enables a reexamination of existing datasets to more carefully study this apparently divergent expression. Despite these promising biological insights revealed by RESCUE, reliance on computational inference alone may have potential limitations, underscoring the value of experimental validation when feasible. Orthogonal

approaches, such as in situ hybridization, could further strengthen and refine these findings; however, such validation is often impractical in tissues where cell bodies and non-cell-body structures are spatially intermingled and difficult to segment. These challenges highlight the importance of discovery-oriented computational frameworks like RESCUE, which are designed to uncover biologically meaningful patterns and guide interpretation in settings where experimental confirmation is difficult.

There are several potential directions for future work. First, in cases where no suitable reference is available, extensions of RESCUE to reference-free or unsupervised frameworks may be possible. For example, methods based on robust principal component analysis (PCA)^{78,79} could be adapted to extract unattributed structure without predefined reference matrices. Second, applying RESCUE to high-resolution ST data requires creating a predefined grid, and further research is required to develop a systematic method to determine the optimal grid size. Third, while the gene-level sparsity assumption underlying RESCUE is biologically plausible in many settings, there may exist scenarios in which the unattributed component is not sparse but instead exhibits alternative structure, such as low-rank or low-rank plus sparse residuals. Extending RESCUE to accommodate such residual regimes represents an important direction for future research. Despite these remaining challenges, we have already shown that RESCUE can still provide meaningful analyses of unattributed expression.

Finally, RESCUE reminds us that the biology we wish to study and the measurements we are able to make are not always one and the same. For instance, we typically accept single-cell reference data as truly representing the expression of the entire single cell, but in many cases it does not, as we have illustrated above. Any measurement

technology has inherent biases. As a result, recognizing, acknowledging and properly interpreting what is missed, in other words, the unattributed expression, is important for a comprehensive understanding of a system.

Methods

RESCUE model

We assume that our ST data profiles G genes in J grids across the tissue section. In low-resolution ST technologies, these grids are predefined as multicellular spots, whereas in higher-resolution ST technologies, these can be constructed at any desired size, which allows inference at different biological scales. To introduce the statistical model behind RESCUE, we begin by reviewing existing linear deconvolution methods^{12–16}. These methods decompose the ST expression matrix $\mathbf{Y} \in \mathbb{R}_{\geq 0}^{G \times J}$ of genes in J grids as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{E} \quad (1)$$

where $\mathbf{X} \in \mathbb{R}_{\geq 0}^{G \times K}$ is a matrix of K cell-type expression reference profiles, $\boldsymbol{\alpha} \in \mathbb{R}_{\geq 0}^{K \times J}$ is a matrix of inferred cell-type proportions within each grid, and $\mathbf{E} \in \mathbb{R}^{G \times J}$ is assumed to be a mean-zero noise matrix. The issue is that (1) implies that expectation of ST expression is $\mathbb{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\alpha}$, which ignores unattributed transcriptome.

Instead, we further decompose the error $\mathbf{E}$, motivated by the robust regression framework^{45–47}. Specifically, we consider the multivariate mean-shift model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{R} + \boldsymbol{\varepsilon} \quad (2)$$

where $\mathbf{X}\boldsymbol{\alpha}$ represents the expected expression values of each gene in each grid due to the presence of cells whose types are represented in the reference sc/snRNA-seq data, $\mathbf{R} \in \mathbb{R}_{\geq 0}^{G \times J}$ is an unattributed expression that allows for outliers that cannot be captured by the reference data, and $\boldsymbol{\varepsilon} \in \mathbb{R}^{G \times J}$ is a noise matrix that allows for residual variability between the ST sample and the reference data.

Following the mean-shift model in the robust regression framework, we assume that the unattributed component $\mathbf{R}$ is sparse. Specifically, the sparsity assumption reflects the expectation that only a subset of genes exhibits unattributed expression within each grid. This assumption is required for identifiability of the decomposition and is biologically plausible, as we expect relatively few genes to exhibit unattributed expression within each grid. Consistent with this assumption, we observe empirically that unattributed expression in neuropil regions of the honey bee brain is highly sparse, with 92% of genes showing no unattributed expression on average. Moreover, even in extracellular environments, ECM proteins typically involve highly sparse, restricted gene sets, with only 1.3–1.5% of genes showing nonzero expression, rather than transcriptome-wide activation⁸⁰.

Optimization problem

RESCUE aims to simultaneously recover the sparse non-negative matrix $\mathbf{R}$ and non-negative cell-type proportions $\boldsymbol{\alpha}$ in our model (2). Specifically, we consider the following optimization problem:

$$\min_{\boldsymbol{\alpha}, \mathbf{R} \geq 0} \|\mathbf{W} \odot \mathbf{R}\|_1 \quad \text{subject to} \quad \|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{R}\|_F^2 \leq \lambda \quad (3)$$

where $\mathbf{W}$ is a weight matrix, $\odot$ denotes the Hadamard product, $\|\cdot\|_F$ denotes the Frobenius norm, and $\lambda > 0$ is a tuning parameter. We set each entry of the weight matrix as $W_{gj} = 1/\sqrt{\text{Var}(R_{gj})}$ to account for the fact that different genes have different variances across different locations. These weights will be estimated through an iterative variance estimation procedure. The ℓ_1 objective function enforces sparsity of our estimated $\mathbf{R}$, while the constraint ensures that the residual variability $\boldsymbol{\varepsilon}$ is not too large, in other words, that the ST data do not deviate too much from the reference sc/snRNA-seq data.

In practice, we solve the Lagrangian form of this problem:

$$\min_{\boldsymbol{\alpha}, \mathbf{R} \geq 0} f(\boldsymbol{\alpha}, \mathbf{R}) = \|\mathbf{W} \odot \mathbf{R}\|_1 + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{R}\|_F^2 \quad (4)$$

where $\mu > 0$ is a tuning parameter. We apply an efficient alternating minimization to solve two convex sub-problems in each iteration:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha} \geq 0} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{R}\|_F^2 \quad (5)$$

$$\hat{\mathbf{R}} = \arg\min_{\mathbf{R} \geq 0} \|\mathbf{W} \odot \mathbf{R}\|_1 + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}} - \mathbf{R}\|_F^2 = S_+ \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}}, \frac{\mathbf{W}}{\mu} \right) \quad (6)$$

where $S_+(x, \tau) = \max(x - \tau, 0)$ denotes the entry-wise non-negative soft-thresholding function.

Estimation of adaptive weights

Kong et al.⁴⁷ proposed adaptive lasso-type weights to reduce masking and swamping effects in the presence of multiple outliers. Our weighting scheme plays a similar role but differs in that the weights are learned iteratively through variance estimation of the unattributed component. In ST deconvolution models, the observed count $\tilde{Y}_{gj}$ is often modeled under a Poisson assumption^{13,16}:

$$\tilde{Y}_{gj} | \lambda_{gj} \sim \text{Poisson}(N_j \lambda_{gj}) \quad (7)$$

where $N_j = \sum_g \tilde{Y}_{gj}$ represents the total transcript count, or number of UMIs for grid j , and λ_{gj} denotes the expected expression rate of gene g at grid j . Rather than directly modeling $\tilde{Y}_{gj}$, we assume that it can be decomposed as $\tilde{Y}_{gj} = \tilde{L}_{gj} + \tilde{R}_{gj}$, where

$$\tilde{L}_{gj} | \lambda_{gj}^L \sim \text{Poisson}(N_j^L \lambda_{gj}^L) \quad \text{and} \quad \tilde{R}_{gj} | \lambda_{gj}^R \sim \text{Poisson}(N_j^R \lambda_{gj}^R) \quad (8)$$

Here, $N_j^L = \sum_g \tilde{L}_{gj}$ and $N_j^R = \sum_g \tilde{R}_{gj}$ represent the total transcript counts of the canonical and unattributed expression for grid j , respectively, and $(\lambda_{gj}^L, \lambda_{gj}^R)$ denote the expected expression rates of these components. We can write $N_j^L = (1 - \pi_j)N_j$ and $N_j^R = \pi_j N_j$, where $\pi_j \in [0, 1]$ denotes the proportion of unattributed expression for grid j .

When using counts per million (CPM) normalization, i.e., $Y_{gj} = \tilde{Y}_{gj}/N_j \times 10^6$, we can estimate the variance of unattributed component as

$$\text{Var}(R_{gj}) = \text{Var}\left(\frac{\tilde{R}_{gj}}{N_j} \times 10^6\right) = \left(\frac{10^6}{N_j}\right)^2 \times \text{Var}(\tilde{R}_{gj}) \approx \frac{10^6}{N_j} \times R_{gj} \quad (9)$$

where the approximation follows from $\widehat{\text{Var}}(\tilde{R}_{gj}) = N_j^R \lambda_{gj}^R$ and $\tilde{R}_{gj} = R_{gj}/10^6 \times N_j$. Here, the term $N_j^R \lambda_{gj}^R$ can be iteratively updated using the plug-in estimator $\tilde{R}_{gj} \approx N_j^R \lambda_{gj}^R$. To avoid division by zero, we set $W_{gj} = 1/\max\left(\sqrt{\frac{10^6}{N_j} \times R_{gj}}, \delta\right)$, where δ is a small positive constant. We summarize the RESCUE algorithm in Supplementary Algorithm 1.

Tuning parameter selection

RESCUE aims to recover unattributed expression by balancing the sparsity of $\mathbf{R}$, which reflects the number of rejected outliers, against the level of residual variability $\boldsymbol{\varepsilon}$. Following standard practices in robust regression and robust PCA^{45,78}, we set $\mu = GJ/c\|\mathbf{Y}\|_1$ and focus on selecting the optimal value of the tuning parameter c . However, choosing the optimal tuning parameter is challenging in the presence of outliers^{79,81}.

To select the optimal tuning parameter, we adopt a data-driven strategy recently proposed for related deconvolution problems in bulk

RNA-seq analysis⁸². The key idea is to generate simulated data that closely resemble the observed ST data and use this for evaluation. Specifically, we use a surrogate ST dataset with ground truth proportions of canonical and unattributed expression. To construct this surrogate ST dataset, we first apply the non-negative least squares (NNLS) to the observed ST data to obtain initial estimates for α and $\mathbf{R}$:

$$\hat{\mathbf{R}}_{\text{NNLS}} = (\mathbf{Y} - \mathbf{X}\hat{\alpha}_{\text{NNLS}})_+, \quad \text{where} \quad \hat{\alpha}_{\text{NNLS}} = \operatorname{argmin}_{\alpha \geq 0} \|\mathbf{Y} - \mathbf{X}\alpha\|_F^2 \quad (10)$$

where $(x)_+ = \max(x, 0)$ denotes the zero-thresholding function. We then simulate a surrogate ST data $\tilde{\mathbf{Y}}_{\text{sim}}$ from a Poisson distribution with rescaled expected expression rates, preserving the overall signal-to-noise characteristics observed in the real data. RESCUE is applied to $\tilde{\mathbf{Y}}_{\text{sim}}$ across a range of μ values, and the optimal c is selected as the value that minimizes the recovery error of the canonical and unattributed components in the surrogate dataset. We summarize the tuning procedure in Supplementary Algorithm 2.

Platform effect normalization

To mitigate cross-platform effects between ST data and sc/snRNA-seq reference data, we apply platform effect normalization. This step is essential because differences in experimental platforms can introduce systematic shifts in gene expression levels that confound downstream analysis. Without proper alignment of scale, these differences may lead to biased estimation of cell-type contributions or unattributed signals. We adopt the normalization strategy from RCTD¹³, which rescales each gene in the reference data to match the overall expression level of the corresponding gene in the ST data. Specifically, we first compute pseudo-bulk measurements for both ST and reference datasets and then apply a gene-wise log-scale correction to harmonize them.

Data normalization and reference matrix

To account for technical variability and the effect of library size, we adopt CPM normalization but fit RESCUE in the linear space. We avoid log-transformation for two main reasons. First, log-transformed input tends to underestimate true signals in gene expression deconvolution⁸³. Second, log-transformation requires adding pseudo-counts, which can introduce significant artificial distortion between zeros and non-zeros^{84,85}. Instead, we work in the linear scale, employing an iteratively reweighted algorithm (Supplementary Algorithm 1) to stabilize variance adaptively. This approach allows the model to adjust for varying data scales and minimizes the influence of outliers without artificially altering data distributions. The normalized reference matrix is then used to construct average cell-type expression.

Reference dependence of unattributed expression

RESCUE is a computational negative selection method which relies on the choice of reference, and interpretation of the recovered unattributed expression therefore requires consideration of both biological context and reference completeness. To further assess the impact of reference selection within the same tissue, we extended the honey bee brain analysis using snRNA-seq⁸⁶ and segmented MERFISH cells as alternative reference profiles (Supplementary Fig. 5). Using segmented MERFISH cells produced unattributed expression patterns highly consistent with those obtained using scRNA-seq, as both references capture cellular transcripts beyond the nucleus and therefore explain similar fractions of the ST signal. In contrast, the snRNA-seq reference yielded higher levels of unattributed expression on average, reflecting its enrichment for nuclear RNA and underrepresentation of cytoplasmic and neuritic transcripts. These results indicate that variation in the recovered unattributed component reflects biologically meaningful differences in reference coverage rather than technical artifacts, and that interpretation of unattributed expression should be guided

by the biological properties of both the tissue and the reference dataset.

Choice of grid resolution for honey bee brain MERFISH data

For high-resolution ST platforms such as MERFISH, RESCUE operates on spatially aggregated expression profiles defined over user-specified grids. The choice of grid size reflects a trade-off between spatial resolution and sparsity: smaller grids preserve fine spatial structure but increase sparsity, whereas larger grids reduce sparsity at the cost of spatial detail. As a practical guideline, we recommend selecting a grid size such that the average total transcript count per grid is comparable to the typical total transcript count per cell. In the honey bee MERFISH analysis, we used a grid size of $20 \times 20 \mu\text{m}^2$, which approximates cell-scale resolution. To assess sensitivity to grid resolution, we varied grid size across a range of spatial scales and found that RESCUE produces consistent results across a broad range of grid resolutions (Supplementary Fig. 6).

Subtraction-based estimation of unattributed expression for existing methods

Most existing ST deconvolution methods are designed to estimate cell-type proportions, rather than to recover expression levels directly. Because these methods do not explicitly model unattributed expression, we adopted a subtraction-based framework, following a strategy used in related deconvolution settings for bulk RNA-seq data²³. Specifically, we first estimated cell-type proportions using each deconvolution method and then recovered the unattributed expression by subtracting the estimated cellular contributions from the observed expression, followed by zero-thresholding to avoid negative values. Under this framework, the unattributed expression is defined as the portion of the observed ST expression that cannot be explained by the reference-based cellular expression implied by each method's estimated cell-type proportions.

Data generating process for simulation studies

For a total of $J = 2,500$ grids, we first generated the total UMI count for grid j from a negative binomial distribution $N_j \sim \text{NB}(\text{mean} = M, \text{size} = 1/\sigma)$. We generated the raw count for each gene in each grid as $\tilde{Y}_{gj} = \tilde{L}_{gj} + \tilde{R}_{gj}$, where $\tilde{R}_{gj}$ denotes the unattributed expression and $\tilde{L}_{gj}$ represents the canonical expression that could have been captured by sc/snRNA-seq reference data.

To generate $\tilde{L}_{gj}$, we used the real scRNA-seq data from honey bee brains⁴⁸ as the source of canonical expression profiles. We selected the top $G = 500$ number of highly variable genes and constructed the reference matrix $\mathbf{X}$ using $K = 20$ number of cell types annotated via Louvain clustering in Seurat⁸⁷. Latent mixing proportions were drawn as $\beta_{kj} \sim \text{Beta}(1, 3)$ and normalized as $\alpha_{kj} = \beta_{kj} / \sum_k \beta_{kj}$, with at most 10 cell types per spot. Defining $\lambda_{gj}^{\mathbf{L}} = (\mathbf{X}\alpha)_{gj}$, we generated the cell-type-driven canonical expression as

$$\tilde{L}_{gj} \sim \text{Multinomial}\left(\left(1 - \pi_j\right)N_j, P_{gj} = \frac{\lambda_{gj}^{\mathbf{L}}}{\sum_g \lambda_{gj}^{\mathbf{L}}}\right) \quad (11)$$

To generate sparse unattributed expression, we sampled

$$\tilde{R}_{gj} \sim \text{Multinomial}\left(\pi_j N_j, Q_{gj} = \frac{\lambda_{gj}^{\mathbf{R}}}{\sum_g \lambda_{gj}^{\mathbf{R}}}\right) \quad (12)$$

where $\lambda_{gj}^{\mathbf{R}} \sim \text{Gamma}(1, 1/3)$ with probability $(1 - s)$ and $\lambda_{gj}^{\mathbf{R}} = 0$ with probability s . Here, we introduced a sparsity parameter s , which makes $(s \times 100)$ percentage of intensity parameter $\lambda_{gj}^{\mathbf{R}}$ zero-inflated. Under this setting, we can generate $\tilde{L}_{gj}$ and $\tilde{R}_{gj}$ that satisfy $\sum_g (\tilde{L}_{gj} + \tilde{R}_{gj}) = N_j$. In our simulations, we fixed $M = 10,000$ and $\sigma = 0.5$, while varying $\pi_j \in \{0, 0.2, 0.4, 0.6, 0.8\}$ and $s \in \{0.8, 0.85, 0.9, 0.95\}$ to control the proportion of canonical expression and the sparsity level of

unattributed transcriptome, respectively (Fig. 2). We then examined the robustness of RESCUE under additional simulation settings by varying the average total UMI count per spot (M), as well as varying noise levels by adjusting library size variability across spots (σ) (Supplementary Fig. 1).

Evaluation metrics for cell-type deconvolution in simulation studies

To evaluate the accuracy of cell-type deconvolution, we considered three complementary metrics: Lin's CCC, RMSE, and JSD. We first measured Lin's CCC between the true and estimated proportion of the k th cell type, defined as

$$\text{CCC}(\alpha_k, \hat{\alpha}_k) = \frac{2\text{Cov}(\alpha_k, \hat{\alpha}_k)}{\sigma_{\alpha_k}^2 + \sigma_{\hat{\alpha}_k}^2 + (\mu_{\alpha_k} - \mu_{\hat{\alpha}_k})^2} \quad (13)$$

where μ_{α_k} and $\mu_{\hat{\alpha}_k}$ are the means of the two variables, and $\sigma_{\alpha_k}^2$ and $\sigma_{\hat{\alpha}_k}^2$ denote the corresponding variances. We used Lin's CCC because it captures shifts in both location and scale, making it robust to high leverage points⁸². The concordance increases as the CCC value gets closer to 1. We report the average CCC across all cell types.

To quantify estimation error at the spatial location level, we computed the RMSE between the true and estimated cell-type proportion vectors. For spatial location j , RMSE is defined as

$$\text{RMSE}_j = \sqrt{\frac{1}{K} \sum_{k=1}^K (\alpha_{kj} - \hat{\alpha}_{kj})^2} \quad (14)$$

where K is the number of cell types, and α_{kj} and $\hat{\alpha}_{kj}$ denote the true and estimated proportions of cell type k at location j , respectively. We report the average RMSE across all spatial locations, with smaller values indicating better performance.

Finally, we computed the JSD between the true and estimated cell-type proportion distributions at each spatial location. For spatial location j , JSD is defined as

$$\text{JSD}_j = \frac{1}{2} \text{KL}(\alpha_j \| \mathbf{m}_j) + \frac{1}{2} \text{KL}(\hat{\alpha}_j \| \mathbf{m}_j), \quad \mathbf{m}_j = \frac{1}{2} (\alpha_j + \hat{\alpha}_j) \quad (15)$$

where $\text{KL}(\cdot \| \cdot)$ denotes the Kullback-Leibler divergence. We report the average JSD across all spatial locations, with smaller values indicating better performance.

Evaluation metrics for unattributed expression recovery in simulation studies

To evaluate the accuracy of unattributed expression recovery, we measured the relative Frobenius error, defined as

$$(\text{Relative Frobenius Error}) = \frac{\|\mathbf{R} - \hat{\mathbf{R}}\|_F}{\|\mathbf{R}\|_F} \quad (16)$$

where $\mathbf{R}$ and $\hat{\mathbf{R}}$ denote the true and estimated unattributed expression matrices, respectively.

Evaluation metrics for honey bee brain MERFISH data

We measured Lin's CCC and the RMSE between the segmentation-based (true) and estimated unattributed expression proportions,

defined as follows:

$$\text{CCC}(\rho, \hat{\rho}) = \frac{2\text{Cov}(\rho, \hat{\rho})}{\sigma_{\rho}^2 + \sigma_{\hat{\rho}}^2 + (\mu_{\rho} - \mu_{\hat{\rho}})^2} \quad \text{and} \quad (17)$$

$$\text{RMSE}(\rho, \hat{\rho}) = \sqrt{\frac{1}{J} \sum_{j=1}^J (\rho_j - \hat{\rho}_j)^2}$$

where $\rho_j = \sum_g \tilde{R}_{gj} / \sum_g \tilde{Y}_{gj}$ denotes the true unattributed expression proportion in grid j , and $\hat{\rho}_j$ denotes its estimated value. Here, μ_{ρ} and $\mu_{\hat{\rho}}$ are the means of two variables, and σ_{ρ}^2 and $\sigma_{\hat{\rho}}^2$ denote the corresponding variances.

Benchmarking computational runtime

We evaluated the computational runtime of each method across ST datasets of varying sizes (Supplementary Fig. 7). Across all benchmarks, RESCUE completed within practical time limits for contemporary ST datasets, demonstrating scalability to datasets with hundreds to thousands of spatial locations and genes. All analyses were performed using CPU-only computation on a standard laptop (Apple M3 Pro with 36GB of memory), without GPUs or high-performance computing clusters.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The honey bee brain MERFISH data used in this study is deposited in our GitHub repository: <https://github.com/brunoyjlee/RESCUE/tree/main/data>. The honey bee brain scRNA-seq data⁴⁸ is available at <https://doi.org/10.6084/m9.figshare.16832518>. The human breast cancer dataset⁵ is available via GEO with accession ID [GSE243280](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE243280). The human DLPFC dataset⁶⁰ is available at <https://github.com/LieberInstitute/spatialLIBD>. The cichlid fish brain dataset⁶¹ is available via GEO with accession ID [GSE217619](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE217619). Source data are provided with this paper.

Code availability

RESCUE is available as an R package at <https://github.com/brunoyjlee/RESCUE>. Codes for the version of RESCUE used in this paper are also deposited at Zenodo⁸⁸ (<https://doi.org/10.5281/zenodo.18965449>).

References

- Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
- Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022).
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* <https://doi.org/10.1126/science.aaa6090> (2015).
- Nagendran, M. et al. 1457 Visium HD enables spatially resolved, single-cell scale resolution mapping of FFPE human breast cancer tissue. *J. Immunother. Cancer* **11**, A1620–A1620 (2023).
- Janesick, A. et al. High resolution mapping of the tumor micro-environment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).
- Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
- Petukhov, V. et al. Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* **40**, 345–354 (2022).
- Polański, K. et al. Bin2cell reconstructs cells from high resolution Visium HD data. *Bioinformatics* **40**, btac546 (2024).
- Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

10. Rodrigues, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
11. Stickels, R. R. et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
12. 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.* **49**, e50 (2021).
13. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).
14. Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat. Biotechnol.* **40**, 1349–1359 (2022).
15. Zhou, Z., Zhong, Y., Zhang, Z. & Ren, X. Spatial transcriptomics deconvolution at single-cell resolution using Redeconve. *Nat. Commun.* **14**, 7930 (2023).
16. Liu, Z., Wu, D., Zhai, W. & Ma, L. SONAR enables cell type deconvolution with spatially weighted Poisson-Gamma model for spatial transcriptomics. *Nat. Commun.* **14**, 4727 (2023).
17. Si, Y. et al. FICTURE: scalable segmentation-free analysis of submicron-resolution spatial transcriptomics. *Nat. Methods* **21**, 1843–1854 (2024).
18. Mitchel, J., Gao, T., Petukhov, V., Cole, E. & Kharchenko, P. V. Impact of segmentation errors in analysis of spatial transcriptomics. *Nat. Genet.* **58**, 434–444 (2026).
19. Nguyen, H., Nguyen, H., Tran, D., Draghici, S. & Nguyen, T. Fourteen years of cellular deconvolution: methodology, applications, technical evaluation and outstanding challenges. *Nucleic Acids Res.* **52**, 4761–4783 (2024).
20. Garmire, L. X. et al. Challenges and perspectives in computational deconvolution of genomics data. *Nat. Methods* **21**, 391–400 (2024).
21. Gaspard-Boulinc, L. C., Gortana, L., Walter, T., Barillot, E. & Cavalli, F. M. G. Cell-type deconvolution methods for spatial transcriptomics. *Nat. Rev. Genet.* 1–19 <https://doi.org/10.1038/s41576-025-00845-y> (2015).
22. Ament, S. A. & Pouloupoulos, A. The brain's dark transcriptome: sequencing RNA in distal compartments of neurons and glia. *Curr. Opin. Neurobiol.* **81**, 102725 (2023).
23. Ivich, A. et al. Missing cell types in single-cell references impact deconvolution of bulk data but are detectable. *Genome Biol.* **26**, 86 (2025).
24. Han, J. et al. Single-cell sequencing unveils key contributions of immune cell populations in cancer-associated adipose wasting. *Cell Discov.* **8**, 122 (2022).
25. Fetit, R. et al. Characterizing neutrophil subtypes in cancer using scRNA sequencing demonstrates the importance of IL1 β /CXCR2 axis in generation of metastasis-specific neutrophils. *Cancer Res. Commun.* **4**, 588–606 (2024).
26. Sutton, M. A. et al. Miniature neurotransmission stabilizes synaptic function via tonic suppression of local dendritic protein synthesis. *Cell* **125**, 785–799 (2006).
27. Cajigas, I. J. et al. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron* **74**, 453–466 (2012).
28. Hafner, A.-S., Donlin-Asp, P. G., Leitch, B., Herzog, E. & Schuman, E. M. Local protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments. *Science* **364**, eaau3644 (2019).
29. Glock, C. et al. The translome of neuronal cell bodies, dendrites, and axons. *Proc. Natl. Acad. Sci. USA* **118**, e2113929118 (2021).
30. von Kügelgen, N. & Chekulaeva, M. Conservation of a core neurite transcriptome across neuronal types and species. *WIREs RNA* **11**, e1590 (2020).
31. Lee, K.-W., Kim, H.-J., Sung, J.-J., Park, K.-S. & Kim, M. Defective neurite outgrowth in aphidicolin/cAMP-induced motor neurons expressing mutant Cu/Zn superoxide dismutase. *Int. J. Dev. Neurosci.* **20**, 521–526 (2002).
32. Markmiller, S. et al. Persistent mRNA localization defects and cell death in ALS neurons caused by transient cellular stress. *Cell Rep.* **36**, 109685 (2021).
33. Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (2015).
34. Gandalovičová, A., Vomastek, T., Rosel, D. & Brábek, J. Cell polarity signaling in the plasticity of cancer cell invasiveness. *Oncotarget* **7**, 25022–25049 (2016).
35. Backlund, M. et al. Plasticity of nuclear and cytoplasmic stress responses of RNA-binding proteins. *Nucleic Acids Res.* **48**, 4725–4740 (2020).
36. Apicco, D. J. et al. Reducing the RNA binding protein TIA1 protects against tau-mediated neurodegeneration in vivo. *Nat. Neurosci.* **21**, 72–80 (2018).
37. Wolozin, B. & Ivanov, P. Stress granules and neurodegeneration. *Nat. Rev. Neurosci.* **20**, 649–666 (2019).
38. Ridder, K. et al. Extracellular vesicle-mediated transfer of functional RNA in the tumor microenvironment. *Oncotarget* **4**, e1008371 (2015).
39. Li, I. & Nabet, B. Y. Exosomes in the tumor microenvironment as mediators of cancer therapy resistance. *Mol. Cancer* **18**, 32 (2019).
40. O'Brien, K., Breyne, K., Ughetto, S., Laurent, L. C. & Breakefield, X. O. RNA delivery by extracellular vesicles in mammalian cells and its applications. *Nat. Rev. Mol. Cell Biol.* **21**, 585–606 (2020).
41. Landry, F. J. & Findlay, S. R. Purification of human basophils by negative selection. *J. Immunol. Methods* **63**, 329–336 (1983).
42. Hansel, T. T. et al. Purification of human blood eosinophils by negative selection using immunomagnetic beads. *J. Immunol. Methods* **122**, 97–103 (1989).
43. Hansel, T. T. et al. An improved immunomagnetic procedure for the isolation of highly purified human blood eosinophils. *J. Immunol. Methods* **145**, 105–110 (1991).
44. Munoz, N. M. & Leff, A. R. Highly purified selective isolation of eosinophils from human peripheral blood by negative immunomagnetic selection. *Nat. Protoc.* **1**, 2613–2620 (2006).
45. She, Y. & Owen, A. B. Outlier detection using nonconvex penalized regression. *J. Am. Stat. Assoc.* **106**, 626–639 (2011).
46. Lee, Y., MacEachern, S. N. & Jung, Y. Regularization of case-specific parameters for robustness and efficiency. *Stat. Sci.* **27**, 350–372 (2012).
47. Kong, D., Bondell, H. D. & Wu, Y. Fully efficient robust estimation, outlier detection and variable selection via penalized regression. *Stat. Sin.* **28**, 1031–1052 (2018).
48. Traniello, I. M. et al. Single-cell dissection of aggression in honey-bee colonies. *Nat. Ecol. Evol.* **7**, 1232–1244 (2023).
49. Lin, L. I.-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255–268 (1989).
50. Kumar, A. et al. Intracellular spatial transcriptomic analysis toolkit (InSTAnT). *Nat. Commun.* **15**, 7794 (2024).
51. Kim, Y. et al. SpaceExpress enables differential spatial transcriptomics with natural coordinate systems. 2024.12.19.628720 Preprint at <https://doi.org/10.1101/2024.12.19.628720> (2025).
52. Yeo, S., Schrader, A. W., Lee, J., Asadian, M. & Han, H.-S. Spot-based global registration for subpixel stitching of single-molecule resolution images for tissue-scale spatial transcriptomics. *Anal. Chem.* **96**, 6517–6522 (2024).

53. Wang, G. et al. Spatial organization of the transcriptome in individual neurons. 2020.12.07.414060 Preprint at <https://doi.org/10.1101/2020.12.07.414060> (2020).
54. Moffitt, J. R. et al. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc. Natl. Acad. Sci. USA* **113**, 14456–14461 (2016).
55. Wang, G., Moffitt, J. R. & Zhuang, X. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep.* **8**, 4847 (2018).
56. Khaledian, B., Thibes, L. & Shimono, Y. Adipocyte regulation of cancer stem cells. *Cancer Sci.* **114**, 4134–4144 (2023).
57. Hu, X. et al. Decorin-mediated suppression of tumorigenesis, invasion, and metastasis in inflammatory breast cancer. *Commun. Biol.* **4**, 72 (2021).
58. Cai, R. et al. Primary breast tumor induced extracellular matrix remodeling in premetastatic lungs. *Sci. Rep.* **13**, 18566 (2023).
59. de Freitas Oliveira-Tore, C. et al. Non-canonical extracellular complement pathways and the complosome paradigm in cancer: a scoping review. *Front. Immunol.* **16**, 1519465 (2025).
60. Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).
61. Johnson, Z. V. et al. Cellular profiling of a recently-evolved social behavior in cichlid fishes. *Nat. Commun.* **14**, 4891 (2023).
62. Lake, B. B. et al. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 6031 (2017).
63. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
64. DeFelipe, J. et al. Cortical white matter: beyond the pale remarks, main conclusions and discussion. *Front. Neuroanat.* **4**, 1692 (2010).
65. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
66. Sakers, K. et al. Astrocytes locally translate transcripts in their peripheral processes. *Proc. Natl. Acad. Sci. USA* **114**, E3830–E3838 (2017).
67. Jiang, T. et al. The lncRNA MALAT1/miR-30/Spastin Axis Regulates Hippocampal Neurite Outgrowth. *Front. Cell. Neurosci.* **14**, 555747 (2020).
68. Overly, C. C., Rieff, H. I. & Hollenbeck, P. J. Organelle motility and metabolism in axons vs dendrites of cultured hippocampal neurons. *J. Cell Sci.* **109**, 971–980 (1996).
69. Harris, J. J., Jolivet, R. & Attwell, D. Synaptic energy use and supply. *Neuron* **75**, 762–777 (2012).
70. Than-Trong, E. & Bally-Cuif, L. Radial glia and neural progenitors in the adult zebrafish central nervous system. *Glia* **63**, 1406–1428 (2015).
71. Miranda-Negrón, Y. & García-Arrarás, J. E. Radial glia and radial glia-like cells: Their role in neurogenesis and regeneration. *Front. Neurosci.* **16**, 1006037 (2022).
72. Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biol.* **22**, 184 (2021).
73. Abramsson, A. et al. The zebrafish amyloid precursor protein-b is required for motor neuron guidance and synapse formation. *Dev. Biol.* **381**, 377–388 (2013).
74. Newman, M., Ebrahimie, E. & Lardelli, M. Using the zebrafish model for Alzheimer's disease research. *Front. Genet.* **5**, 189 (2014).
75. Haenisch, C. et al. The neuronal growth and regeneration associated Cntn1 (F3/F11/Contactin) gene is duplicated in fish: expression during development and retinal axon regeneration. *Mol. Cell. Neurosci.* **28**, 361–374 (2005).
76. Eiraku, M. et al. DNER acts as a neuron-specific Notch ligand during Bergmann glial development. *Nat. Neurosci.* **8**, 873–880 (2005).
77. Díaz-Ramos, À., Roig-Borrellas, A., García-Melero, A. & López-Alemany, R. α -Enolase, a multifunctional protein: its role on pathophysiological situations. *BioMed. Res. Int.* **2012**, 156795 (2012).
78. Zhou, Z., Li, X., Wright, J., Candès, E. & Ma, Y. Stable principal component pursuit. in *2010 IEEE International Symposium on Information Theory* 1518–1522 <https://doi.org/10.1109/ISIT.2010.5513535> (2010).
79. Mateos, G. & Giannakis, G. B. Robust PCA as bilinear decomposition with outlier-sparsity regularization. *Trans. Sig. Proc.* **60**, 5176–5190 (2012).
80. Manabe, R. et al. Transcriptome-based systematic identification of extracellular matrix proteins. *Proc. Natl. Acad. Sci. USA* **105**, 12849–12854 (2008).
81. Giannakis, G. B., Mateos, G., Farahmand, S., Kekatos, V. & Zhu, H. USPACOR: Universal sparsity-controlling outlier rejection. in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1952–1955 <https://doi.org/10.1109/ICASSP.2011.5946891> (2011).
82. Huang, P., Cai, M., Lu, X., McKenna, C. & Wang, J. Accurate estimation of rare cell-type fractions from tissue omics data via hierarchical deconvolution. *Ann. Appl. Stat.* **18**, 1178–1194 (2024).
83. Zhong, Y. & Liu, Z. Gene expression deconvolution in linear space. *Nat. Methods* **9**, 8–9 (2012).
84. Hicks, S. C., Townes, F. W., Teng, M. & Irizarry, R. A. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562–578 (2018).
85. Townes, F. W., Hicks, S. C., Aryee, M. J. & Irizarry, R. A. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biol.* **20**, 295 (2019).
86. Mu, X. et al. Single-nucleus and spatial transcriptomics identify brain landscape of gene regulatory networks associated with behavioral maturation in honeybees. *Nat. Commun.* **16**, 3343 (2025).
87. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
88. Lee, Y. J. et al. RESCUE: recovery of unattributed expression patterns in spatial transcriptomics. *Zenodo* <https://doi.org/10.5281/zenodo.18965449>.
89. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS A J. Integr. Biol.* **16**, 284–287 (2012).

Acknowledgements

We thank Dr. Tina Barbasch on valuable discussions on the analysis of fish brain data. Funding: This work was supported by the National Institutes of Health (R21HG013180 to H.-S.H. and S.D.Z., R35GM147420 to H.-S.H. and R01AT013189 to S.D.Z.).

Author contributions

Y.J.L., H.-S.H., and S.D.Z. conceptualized the problem and developed the methodology. Y.J.L. performed the numerical studies, interpreted the results, and developed the software. S.Y. and A.W.S. performed data processing and cell segmentation on honey bee MERFISH data. J.L. and M.A. generated honey bee MERFISH dataset. J.L. and I.M.T. selected the gene panel for honey bee MERFISH experiments. I.M.T. and A.C.A. collected honey bee brains. Y.J.L., H.-S.H., and S.D.Z. wrote the manuscript with feedback from G.E.R. H.-S.H. and S.D.Z. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-71720-5>.

Correspondence and requests for materials should be addressed to Hee-Sun Han or Sihai Dave Zhao.

Peer review information *Nature Communications* thanks Guoqiang Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026