# RESCUE: recovery of idiosyncratic expression patterns in spatial transcriptomics

Young Joo Lee[1], Seokjin Yeo[2], Alex W. Schrader[3], Ian M. Traniello[4], Amy Cash Ahmed[5], Gene E. Robinson[5,6,7], Hee-Sun Han[2,3,5,6*], Sihai Dave Zhao[1,5*]

[1]Department of Statistics, University of Illinois Urbana-Champaign

[2]Department of Bioengineering, University of Illinois Urbana-Champaign

[3]Department of Chemistry, University of Illinois Urbana-Champaign

[4]Lewis-Sigler Institute for Integrative Genomics, Princeton University

[5]Carl R. Woese Institute of Genomic Biology, University of Illinois Urbana-Champaign

[6]Neuroscience Program, University of Illinois Urbana-Champaign

[7]Department of Entomology, University of Illinois Urbana-Champaign

* Corresponding authors: sdzhao@illinois.edu (S.D.Z), hshan@illinois.edu (H.-S.H)

## ABSTRACT

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, they can fail to capture important expression patterns, for example from 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 idiosyncratic spatial expression patterns missed by existing ST analysis methods and enable robust inference even when reference data are 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.

## INTRODUCTION

Profiling cellular expression within its spatial context has become possible with the recent development of spatial transcriptomics (ST) technologies, which measures 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 MERFISH[3], 10x Visium HD[4], or 10x Xenium[5], this is usually done using cell segmentation[6–8]. With low-resolution platforms like 10x Visium[9] 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-seq[22]. Similarly, reference snRNA-seq data, by definition, ignores cytoplasmic expression. Finally, existing methods do not capture extracellular tissue 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 types[23], 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 adipocytes[5], which play crucial roles in metabolic disease and cancer[24], and neutrophils[25], 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 “idiosyncratic” expression. Our purified canonical ST expression component can be used to correct existing cell-type deconvolution methods. Our recovered idiosyncratic ST expression component gives us the new ability to specifically target biology that is distinct from our reference data, as discussed above. The idiosyncratic 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 idiosyncratic 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.

87 **RESULTS**

88 **Overview of RESCUE**

89 RESCUE is predicated on a new model that we propose for integrating ST and reference
90 single-cell data, and can be applied to diverse ST platforms, including MERFISH and
91 Visium, across different organisms. Most existing models[12–16] posit that ST expression at
92 each location is composed of a linear combination of expression patterns present in the
93 reference, where the coefficients of this linear combination are interpreted as the
94 proportions of each reference cell type in that location. However, as described in the
95 introduction, there are numerous sources of ST expression that are left out of this model;
96 we will refer to these as idiosyncratic expression patterns, because they are different from
97 the canonical expression patterns reflected in the reference.

98 To address this issue, we propose a corrected model (**Fig. 1**) that allows the ST
99 profile to have both canonical and idiosyncratic expression components. While it appears
100 minor, this change has important implications. Under this model, results from the statistics
101 literatures[45–47] imply that under certain conditions it is possible to simultaneously recover
102 idiosyncratic expression and estimate the linear combination of reference patterns that
103 constitute canonical expression. The conditions are that only relatively few genes show
104 non-zero, or "sparse" expression, in the idiosyncratic component, which is reasonable
105 given the nature of idiosyncratic expression and the read depth/detection efficiency of ST
106 technologies.

107 Leveraging this insight, we developed RESCUE, an algorithm that simultaneously
108 deconvolves ST data while rescuing spatial idiosyncratic expression. RESCUE solves a
109 regularized robust multivariate regression problem and can be applied to both low- and
110 high-resolution ST data. Unlike existing deconvolution methods[12–16], it is robust to

111 incomplete references, such as snRNA-seq data which lack cytoplasmic expression
112 profiles. The rescued idiosyncratic expression can be analyzed using standard
113 approaches, such as clustering, differentially expressed gene (DEG) analysis, and
114 spatially varying gene (SVG) analysis, to discover new biology hidden from existing ST
115 analyses. We illustrate this in a broad range of examples below. We note that the
116 interpretation of idiosyncratic expression depends on both the tissue context and the
117 choice of reference dataset.
118
119 **Simulations**
120 To evaluate the performance of RESCUE, we conducted a simulation study on synthetic
121 data that mimics real experimental ST data with varying degrees of idiosyncratic
122 expression. As no existing methods are specifically designed to recover idiosyncratic
123 expression, we adapted four widely used deconvolution approaches: non-negative least
124 squares (NNLS), weighted NNLS (WNNLS), RCTD[13], and CARD[14], to recover the
125 idiosyncratic component. For each method, we first estimated cell-type proportions using
126 its standard deconvolution procedure and then recovered the idiosyncratic expression by
127 subtracting the estimated cellular contributions from the observed ST expression.
128 We simulated ST data using real scRNA-seq data from the honey bee brain[48] by
129 introducing synthetic idiosyncratic expression of a small number of genes (Methods).
130 Honey bee brains provide a unique benefit for simulating idiosyncratic expression due to
131 the existence of the neuropil region, where axons and dendrites are spatially segregated
132 from cell bodies. The molecular expression in these regions is largely absent in single-
133 cell sequencing data due to the structural fragility of neurites. During the tissue

134 dissociation step, a necessary step in single-cell sequencing, neurites are mostly sheared
135 off and lost. In contrast, ST measurements performed on intact tissue sections preserve
136 these structures, thus the molecular expression in neuropils as well. Our in-house
137 collected MERFISH data shows that neuropil regions have less than 10% of genes
138 expressed on average, confirming a highly sparse idiosyncratic expression component.
139 We simulated both highly sparse and moderately sparse idiosyncratic expression by
140 varying the idiosyncratic sparsity level (e.g., 0.95 for "highly sparse" and 0.80 for
141 "moderately sparse") and idiosyncratic expression proportion (ranging from 0 to 0.8).

142 We first compared the deconvolution performance with existing methods. **Fig. 2A**
143 shows the accuracy of cell-type proportion estimates across methods, measured by
144 average Lin's concordance correlation coefficient[49] (CCC) (Methods). Expectedly, the
145 performance of standard deconvolution methods deteriorates sharply as idiosyncratic
146 expression level increases. On the other hand, RESCUE maintains high accuracy across
147 all conditions, demonstrating robustness to both high idiosyncratic content and reduced
148 sparsity. Our findings imply that idiosyncratic expression, if not properly accounted for,
149 can introduce systematic biases in cell-type proportion estimates.

150 We also evaluated the quality of idiosyncratic transcriptome recovery (**Fig. 2B**),
151 using relative Frobenius error between the true and estimated idiosyncratic expression
152 matrices (Methods). Overall, RESCUE achieves the lowest recovery error, even under
153 high idiosyncratic expression and low sparsity, while other methods deviate substantially
154 from the ground truth, especially when idiosyncratic signals are dense. Our findings show
155 that idiosyncratic expression, if not properly accounted for, can obscure cell-type
156 deconvolution and other downstream analyses, which we will illustrate below. This has

major implications for deconvolution with incomplete reference data and analyses of complex tissue structures with significant idiosyncratic transcription. RESCUE accurately recovers the hidden transcriptome and cell-type composition with minimal distortion.

**RESCUE accurately recovers the idiosyncratic transcriptome in real tissues**

Using a custom MERFISH platform[50], we generated single molecule-resolution ST data from coronal sections of a honey bee brain, targeting 130 genes relevant to aggression (**Supplementary Table S1**)[51]. 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 Cellpose[6] and Baysor[7], can only identify the boundaries of neuronal bodies, also called somata. Individual axons and dendrites could be stained in 2D cell cultures[52], but in native tissues they likely will not lie in one imaging plane. Thus, traditionally, neurite expression has been either 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. The honey bee brain serves as an ideal test case due to its spatially segregated somata and neuropil regions, allowing precise profiling of neurite expression without interference from somatic expression. Cell soma were segmented using Cellpose[6]. Remarkably, only 73.5% of transcripts were localized within cells, indicating that the idiosyncratic transcriptome accounts for 26.5% of the observed ST expression. These signals, which would have been ignored by standard cell segmentation approaches, highlight the importance of RESCUE in capturing overlooked transcriptomic information.

180 Using honey bee brain scRNA-seq data[48] as a reference, we applied each method
181 to our MERFISH data after dividing the brain into grids of size $20 \times 20\ \mu m^2$. We first used
182 RESCUE and the other methods to estimate the proportion of canonical expression
183 relative to the total expression in each of our grids. Leveraging the spatial segregation of
184 the honey bee brain, we then obtained a ground truth proportion of transcripts in each
185 grid that are localized inside cell bodies via cell segmentation. RESCUE's estimates
186 closely matched those from cell segmentation (**Fig. 3A**) and were more accurate than
187 estimates from other methods based on two metrics: Lin's CCC and RMSE between the
188 segmentation-based and estimated proportions (**Fig. 3B**, Methods). Note that, in this
189 example, RESCUE solely relies on reference scRNA-seq data and does not rely on cell
190 segmentation.

191 Finally, we evaluated how well each method recovered the gene expression
192 specifically within the somatic regions (**Fig. 3C**). We calculated the Lin's CCC between
193 the true and estimated gene expression within the grids located in somata. Again,
194 RESCUE significantly outperformed other methods, accurately recovering canonical as
195 well as idiosyncratic expression. Collectively, these results validate that RESCUE is
196 highly effective in partitioning the full transcriptome into its somatic and neuropil
197 components, capturing valuable expression signals that would otherwise be lost using
198 conventional approaches.

199

200 **RESCUE enables accurate deconvolution of ST data with incomplete reference**

201 Here, we illustrate how RESCUE's estimated canonical expression component can be
202 used to achieve accurate deconvolution of ST data when key cell types are missing from

203 the reference, a challenge for all existing methods that has been noted in a recent study[23].
204 We analyzed paired single-cell FFPE-seq (scFFPE-seq) and 10x Visium ST data from
205 human breast cancer tissue[5]. In this ST dataset, adipocytes are manually annotated in
206 approximately 10% of spatial spots (**Fig. 4A**). However, adipocytes are absent from the
207 scFFPE-seq reference, likely due to their high lipid content interfering with fixation and
208 processing. These cells are prone to floating, rupturing, or sticking to plastic surfaces
209 during tissue dissociation, which hinders their recovery in single-cell sequencing
210 protocols[5,17]. As a result, reference-based methods that assume the observed expression
211 is fully explained by the available reference cell types tend to misattribute adipocyte-
212 derived transcripts to other cell types.

213 In adipocyte-rich regions, all four existing methods (NNLS, WNNLS, RCTD, and
214 CARD) fail to account for the missing adipocyte transcriptome and instead misattribute
215 the signal to other reference cell types, such as stromal and endothelial cells (**Fig. 4C**).
216 This systematic misassignment likely arises from transcriptional similarity, as shown by
217 the strong correlation between adipocyte and stromal/endothelia profiles in pseudo-bulk
218 measurements from Visium and scFFPE-seq (**Fig. 4B**). On the other hand, RESCUE
219 accommodates idiosyncratic expression and avoids forcing unassigned signal into
220 existing reference cell types, which results in more accurate and biologically plausible
221 estimates. As shown in **Fig. 4D**, RESCUE's estimated proportions closely align with the
222 scFFPE-seq composition, while other methods exhibit significant distortions. Together,
223 these results demonstrate that RESCUE improves deconvolution accuracy in settings
224 with incomplete references by preventing the systematic misassignment of idiosyncratic
225 transcriptomes to incorrect cell types.

226

227 **RESCUE reveals idiosyncratic biological processes in diverse tissues**

228 In many neuroscience experiments, cell-level information is often characterized using
229 snRNA-seq[53,54], which does not capture cytoplasmic or neurite expression. Here, we
230 illustrate how RESCUE can uncover unique biological processes within the idiosyncratic
231 expression of the human dorsolateral prefrontal cortex (DLPFC) and cichlid fish
232 telencephalon, using snRNA-seq as a reference.

233

234 **Human DLPFC.** Maynard et al.[55] profiled the DLPFC of neurotypical adult donors using
235 paired snRNA-seq and 10x Visium ST. We applied RESCUE to eight samples in which
236 all six cortical layers and the white matter (WM) were annotated, to recover idiosyncratic
237 expression patterns not represented in the snRNA-seq reference (**Fig 5A**). As shown in
238 **Fig. 5B**, RESCUE reveals that Layer 1 and the WM contain substantially higher levels of
239 idiosyncratic expression compared to other cortical layers. This is consistent with the fact
240 that Layer 1 has sparser cell densities and contains neuropil-rich regions[55], and that WM
241 contains a substantial amount of axons and myelin[56], whose transcriptomic content is
242 mostly cytoplasmic and thus is absent from the reference snRNA-seq data.

243 RESCUE enables us to study biology specific to the idiosyncratic transcriptome of
244 Layer 1. We first identified DEGs by comparing Layer 1 to all other layers using either the
245 raw observed ST expression or the idiosyncratic expression recovered by RESCUE.
246 These two analyses gave substantially different results (**Fig. 5C**). For example, the top
247 five genes most enriched in Layer 1 based on idiosyncratic expression had either fold-

248 changes in the opposite direction (*MT-ND1*, *MT-ND2*, and *MT-CO1*) or negligible fold-
249 changes (*MT3* and *CST3*) in the differential expression analysis using raw data.
250 To explore the underlying biological processes, we performed gene ontology (GO)
251 enrichment analysis separately for DEGs derived from the idiosyncratic expression and
252 from the raw ST data (**Fig. 5D**). Idiosyncratic GO terms include mitochondrial processes:
253 "oxidative phosphorylation", "energy derivation by oxidation of organic compounds", and
254 "generation of precursor metabolites and energy". In contrast, no mitochondria-related
255 pathways reached statistical significance in the analysis based on observed ST
256 expression; only the broad term "response to stress" passed the FDR < 0.05 threshold.
257 These results align with the elevated metabolic demands of a neurite-enriched region like
258 Layer 1, which is densely populated with synapses and reliant on local energy production
259 to support axonal and dendritic activity[57,58]. Collectively, these findings demonstrate that
260 isolating idiosyncratic expression can uncover previously obscured yet functionally
261 important gene expression programs in Layer 1 of the human DLPFC.
262
263 **Cichlid Fish Telencephalon.** Johnson et al.[59] collected paired snRNA-seq and Visium
264 ST data from the dorsal telencephalon of individual cichlid fishes exhibiting distinct social
265 behaviors. As shown in **Fig. 6A**, anatomical annotations delineate multiple subdivisions
266 of the dorsal pallium, structured around radial glial (RG) zones that extend long
267 projections from the ventricular zone into superficial pallial regions[60,61]. These RG
268 projections span nearly the entire depth of the pallium and may play a role in shaping
269 spatial gene expression profiles across distinct layers and compartments, but they do not
270 survive tissue dissociation and thus their transcriptomes are not captured by snRNA-seq.

We applied RESCUE to this dataset to recover idiosyncratic transcriptomic signals using the snRNA-seq data as a reference. To assess spatial structure in the rescued transcriptome, we performed SVG detection using SPARK-X[62] on both the observed ST expression and the idiosyncratic expression. As shown in **Fig. 6B**, a substantial number of SVGs are uniquely identified in the idiosyncratic transcriptome. In **Fig. 6C**, we show the idiosyncratic 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[63,64], *cntn1* (Contactin-1) is a GPI-anchored adhesion molecule essential for axon guidance and neurite outgrowth[65], *dner* mediates neuron-glia interactions that are critical for astrocyte differentiation[66], and *eno1* ($\alpha$-enolase) promotes neuritogenesis and contributes to structural remodeling in the developing nervous system[67]. Indeed, Johnson et al.[59] 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.

## DISCUSSION

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

295 significance. While the field has developed numerous transcriptomic profiling methods,
296 such as scRNA-seq for isolating gene expression at the single-cell level, Ribo-seq for
297 ribosome-associated transcripts, and CLIP-seq for binding-protein interactions, each of
298 these technologies inevitably captures only a fraction of the total transcriptome. RESCUE
299 provides a systematic framework to quantify and analyze what is missing: the molecular
300 signals outside the scope of reference datasets.

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

313 One striking finding emerging from our brain tissue analyses is the substantial
314 amount of expression (26.5% of total expression) found in the idiosyncratic transcriptome
315 of the honey bee brain MERFISH data. While localization of mRNAs to neurites is a well-
316 studied phenomenon[26–30], our results, along with the estimates for the adult mouse brain
317 from Ament and Poulpoulos[22], 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.

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[68,69] could be adapted to extract idiosyncratic 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. Despite these remaining challenges, we have already shown that RESCUE can still provide meaningful analyses of spatial idiosyncratic 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 “idiosyncratic 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 $Y \in \mathbb{R}_{\geq 0}^{G \times J}$ of $G$ genes in $J$ grids as

$$Y = X\alpha + E, \tag{1}$$

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

Instead, we further decompose the error $E$, motivated by the robust regression framework[45–47]. Specifically, we consider the multivariate mean-shift model

$$Y = X\alpha + R + \mathcal{E}, \tag{2}$$

where $X\alpha$ 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, $R \in \mathbb{R}_{\geq 0}^{G \times J}$ is an idiosyncratic term that allows for outliers that cannot be captured by the reference data, and $\mathcal{E} \in \mathbb{R}^{G \times J}$ is a noise matrix that allows for residual variability between the ST sample and the reference data. Our key assumption is that $R$ is sparse. This is necessary for identifiability and is also reasonable because we expect relatively few

genes to exhibit idiosyncratic transcriptome within each grid. We also assume that both $\alpha$ and $R$ are non-negative, for biological interpretation.

**Estimation algorithm**

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

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

where $W$ 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 $W_{gj} = 1/\sqrt{Var(R_{gj})}$ to account for the fact that different genes have different variances across different locations. The $\ell_1$ objective function enforces sparsity of our estimated $R$, while the constraint ensures that the residual variability $\mathcal{E}$ 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:

$$\min_{\alpha, R \geq 0} f(\alpha, R) = \|W \odot R\|_1 + \frac{\mu}{2}\|Y - R - X\alpha\|_F^2, \tag{3}$$

where $\mu > 0$ is a tuning parameter. We apply an efficient alternating minimization to solve two convex sub-problems in each iteration. Specifically, our proposed algorithm is a multivariate extension of the soft-IPOD[45], incorporating non-negative constraints and the weight matrix:

$$\hat{\alpha} = \underset{\alpha \geq 0}{\operatorname{argmin}} \|Y - R - X\alpha\|_F^2, \tag{4}$$

$$\hat{R} = \underset{R \geq 0}{\operatorname{argmin}} \|W \odot R\|_1 + + \frac{\mu}{2}\|Y - R - X\alpha\|_F^2 = S_+\left(Y - X\alpha, \frac{W}{\mu}\right), \tag{5}$$

where $S_+(x,\tau) = \max(x-\tau, 0)$ denotes the entry-wise non-negative soft-thresholding function. In the presence of multiple outliers, it is important to address masking and swamping effects. Our proposed weight plays a similar role as the adaptive weight used in Kong et al.[47], but it can be iteratively estimated within the algorithm, not requiring a robust initial estimator.

To estimate $W_{gj}$, we iteratively estimate the variance of $R_{gj}$. In ST deconvolution models, it is common to model the observed gene expression count $\tilde{Y}_{gj}$ under a Poisson assumption[13,16]:

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

where $N_j = \sum_g \tilde{Y}_{gj}$ represents the total transcript count, or number of UMIs for grid $j$, and $\lambda_{gj}$ 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

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

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 idiosyncratic expression for grid $j$, respectively, and $(\lambda_{gj}^L, \lambda_{gj}^R)$ denote the expected expression rates of these components. We can write $N_j^L = \pi_j N_j$ and $N_j^R = (1-\pi_j)N_j$, where $\pi_j \in [0,1]$ denotes the proportion of canonical 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 idiosyncratic component as

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

where the approximation follows from $\widehat{\mathrm{Var}}(\tilde{R}_{gj}) = \widehat{N_j^R \lambda_{gj}^R}$ and $\tilde{R}_{gj} = R_{gj}/10^6 \times N_j$. Here, the term $\widehat{N_j^R \lambda_{gj}^R}$ 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 $\delta$ is a small positive constant.

**Tuning parameter selection**

RESCUE aims to recover idiosyncratic transcriptome by balancing the sparsity of $R$, representing the number of outliers rejected, against the level of residual variability $\mathcal{E}$. Following the approaches in robust regression and robust PCA[45,68], we set $\mu = GJ/C\|Y\|_1$ and instead focus on finding the optimal $C$. However, selecting the optimal tuning parameter can be challenging in the presence of outliers[69,70].

To address this, we adopt a surrogate matching approach[71]. The idea is to choose the optimal tuning parameter $C$ using a surrogate ST dataset with "ground truth" canonical and idiosyncratic transcriptomes. To construct this surrogate ST dataset, we apply NNLS to the observed ST data to obtain rough estimates for $\alpha$ and $R$:

$$\hat{R}_{NNLS} = (Y - X\hat{\alpha}_{NNLS})_+, \quad \text{where} \quad \hat{\alpha}_{NNLS} = \underset{\alpha \geq 0}{\operatorname{argmin}} \|Y - X\alpha\|_F^2, \tag{7}$$

where $(x)_+ = \max(x, 0)$ denotes the zero-thresholding function. We then simulate a surrogate ST data $\tilde{Y}_{\mathrm{sim}}$ from a Poisson distribution with a rescaled expected expression rate, preserving the overall signal-to-noise characteristics observed in the real data. RESCUE is applied to $\tilde{Y}_{\mathrm{sim}}$ across a range of $\mu$ values, and the optimal $C$ is selected as the value that minimizes the recovery error of the canonical and idiosyncratic components in the surrogate dataset.

423

424 **Platform effect normalization**

425 To mitigate cross-platform effects between ST data and sc/snRNA-seq reference data,
426 we apply platform effect normalization. This step is essential because differences in
427 experimental platforms can introduce systematic shifts in gene expression levels that
428 confound downstream analysis. Without proper alignment of scale, these differences may
429 lead to biased estimation of cell-type contributions or idiosyncratic signals. We adopt the
430 normalization strategy from RCTD[13], which rescales each gene in the reference data to
431 match the overall expression level of the corresponding gene in the ST data. Specifically,
432 we first compute pseudo-bulk measurements for both ST and reference datasets and
433 then apply a gene-wise log-scale correction to harmonize them.

434

435 **Data normalization and reference matrix**

436 To account for technical variability and the effect of library size, we adopt CPM
437 normalization but fit RESCUE in the linear space. We avoid log-transformation for two
438 main reasons. First, log-transformed input tends to underestimate true signals in gene
439 expression deconvolution[72]. Second, log-transformation requires adding pseudo-counts,
440 which can introduce significant artificial distortion between zeros and non-zeros[73,74].
441 Instead, we work in the linear scale, employing an iteratively reweighted algorithm to
442 stabilize variance adaptively. This approach allows the model to adjust for varying data
443 scales and minimizes the influence of outliers without artificially altering data distributions.
444 The normalized reference matrix is then used to construct average cell-type expression.

445

446 **Data generating process for simulation studies**

447 For a total of $J = 2{,}500$ grids, we first generated the total UMI count for grid $j$ from a
448 negative binomial distribution $N_j \sim \mathrm{NB}(\mathrm{mean} = 10{,}000, \mathrm{size} = 2)$. We generated the raw
449 count for each gene in each grid as $\tilde{Y}_{gj} = \tilde{L}_{gj} + \tilde{R}_{gj}$, where $\tilde{R}_{gj}$ denotes the idiosyncratic
450 expression and $\tilde{L}_{gj}$ represents the canonical expression that could have been captured
451 by sc/snRNA-seq reference data.

452 To generate $\tilde{L}_{gj}$, we used the real scRNA-seq data from honey bee brains[48] as the
453 source of canonical expression profiles. We selected the top $G = 500$ number of highly
454 variable genes and constructed the reference matrix $X$ using $K = 20$ number of cell types
455 annotated via Louvain clustering in Seurat[75]. Latent mixing proportions were drawn as
456 $\beta_{kj} \sim \mathrm{Beta}(1,3)$ and normalized as $\alpha_{kj} = \beta_{kj} / \sum_k \beta_{kj}$, with at most 10 cell types per spot.
457 Defining $\lambda^L_{gj} = (X\alpha)_{gj}$, we generated the cell-type-driven canonical expression as

458 $$\tilde{L}_{gj} \sim \mathrm{Multinomial}\left(\pi_j N_j, P_{gj} = \frac{\lambda^L_{gj}}{\sum_g \lambda^L_{gj}}\right).$$

459 To generate sparse idiosyncratic expression, we sampled

460 $$\tilde{R}_{gj} \sim \mathrm{Multinomial}\left((1-\pi_j)N_j, Q_{gj} = \frac{\lambda^R_{gj}}{\sum_g \lambda^R_{gj}}\right),$$

461 where $\lambda^R_{gj} \sim \mathrm{Gamma}(1, 1/3)$ with probability $(1-s)$ and $\lambda^R_{gj} = 0$ with probability $s$. Here,
462 we introduced a sparsity parameter $s$, which makes $(s \times 100)$ percentage of intensity
463 parameter $\lambda^R_{gj}$ zero-inflated. Under this setting, we generated $\tilde{L}_{gj}$ and $\tilde{R}_{gj}$ to satisfy and
464 $\sum_g \left(\tilde{L}_{gj} + \tilde{R}_{gj}\right) = N_j$. In our simulations, we considered $\pi_j \in \{0, 0.2, 0.4, 0.6, 0.8\}$ and $s \in$
465 $\{0.8, 0.85, 0.9, 0.95\}$ to control the proportion of canonical expression and the sparsity level
466 of idiosyncratic transcriptome, respectively.

## Evaluation metrics for simulation studies

To evaluate the robust estimation of cell-type proportions, we measured Lin's CCC between the true and estimated proportion of the $k$th cell type, defined as

$$\mathrm{CCC}(\alpha_k, \hat{\alpha}_k) = \frac{2\mathrm{Cov}(\alpha_k, \hat{\alpha}_k)}{\sigma_{\alpha_k}^2 + \sigma_{\hat{\alpha}_k}^2 + \left(\mu_{\alpha_k} - \mu_{\hat{\alpha}_k}\right)^2},$$

where $\mu_{\alpha_k}$ 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 points[71]. The concordance increases as the CCC value gets closer to 1. We took the average of Lin's CCC across all cell types to compare the cell-type deconvolution performance.

To evaluate the accuracy of idiosyncratic transcriptome recovery, we measured relative Frobenius error, defined as

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

where $R$ and $\hat{R}$ denote the true and estimated idiosyncratic transcriptomes.

## Evaluation metrics for honey bee brain MERFISH data

We measured Lin's CCC and the root mean squared error (RMSE) between the segmentation-based (true) and estimated idiosyncratic proportions, defined as follows:

$$\mathrm{CCC}(\rho, \hat{\rho}) = \frac{2\mathrm{Cov}(\rho, \hat{\rho})}{\sigma_\rho^2 + \sigma_{\hat{\rho}}^2 + \left(\mu_\rho - \mu_{\hat{\rho}}\right)^2} \quad \text{and} \quad \mathrm{RMSE}(\rho, \hat{\rho}) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left(\rho_j - \hat{\rho}_j\right)^2},$$

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

## DATA AVAILABILITY

Human breast cancer dataset[5]: 10x Visium and scFFPE-seq data are available via GEO accession GSE243280. Human DLPFC dataset[55]: 10x Visium and snRNA-seq data are available at https://github.com/LieberInstitute/spatialLIBD. Cichlid fish brain dataset[59]: 10x Visium and snRNA-seq data are available via GEO accession GSE217619. Honey bee brain scRNA-seq data[48] is available at https://doi.org/10.6084/m9.figshare.16832518.

## CODE AVAILABILITY

RESCUE is available as an R package at https://github.com/brunoyjlee/RESCUE.

## 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 Y.J.L., S.Y., H.-S.H. and S.D.Z., R35GM147420 to A.W.S and H.-S.H. and R01AT013189 to S.D.Z.).

## REFERENCES


1. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
2. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022).
3. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* (2015) doi:10.1126/science.aaa6090.
4. 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**, (2023).
5. Janesick, A. *et al.* High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).
6. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
7. Petukhov, V. *et al.* Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* **40**, 345–354 (2022).
8. Polański, K. *et al.* Bin2cell reconstructs cells from high resolution Visium HD data. *Bioinformatics* **40**, btae546 (2024).
9. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

10. Rodriques, 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., Cole, E., Petukhov, V. & Kharchenko, P. V. Impact of Segmentation Errors in Analysis of Spatial Transcriptomics Data. 2025.01.02.631135 Preprint at https://doi.org/10.1101/2025.01.02.631135 (2025).

548 19. Nguyen, H., Nguyen, H., Tran, D., Draghici, S. & Nguyen, T. Fourteen years of
549 cellular deconvolution: methodology, applications, technical evaluation and
550 outstanding challenges. *Nucleic Acids Res.* **52**, 4761–4783 (2024).
551 20. Garmire, L. X. *et al.* Challenges and perspectives in computational deconvolution of
552 genomics data. *Nat. Methods* **21**, 391–400 (2024).
553 21. Gaspard-Boulinc, L. C., Gortana, L., Walter, T., Barillot, E. & Cavalli, F. M. G. Cell-
554 type deconvolution methods for spatial transcriptomics. *Nat. Rev. Genet.* 1–19 (2025)
555 doi:10.1038/s41576-025-00845-y.
556 22. Ament, S. A. & Poulopoulos, A. The brain's dark transcriptome: Sequencing RNA in
557 distal compartments of neurons and glia. *Curr. Opin. Neurobiol.* **81**, 102725 (2023).
558 23. Ivich, A. *et al.* Missing cell types in single-cell references impact deconvolution of
559 bulk data but are detectable. *Genome Biol.* **26**, 86 (2025).
560 24. Han, J. *et al.* Single-cell sequencing unveils key contributions of immune cell
561 populations in cancer-associated adipose wasting. *Cell Discov.* **8**, 122 (2022).
562 25. Fetit, R. *et al.* Characterizing Neutrophil Subtypes in Cancer Using scRNA
563 Sequencing Demonstrates the Importance of IL1β/CXCR2 Axis in Generation of
564 Metastasis-specific Neutrophils. *Cancer Res. Commun.* **4**, 588–606 (2024).
565 26. Sutton, M. A. *et al.* Miniature Neurotransmission Stabilizes Synaptic Function via
566 Tonic Suppression of Local Dendritic Protein Synthesis. *Cell* **125**, 785–799 (2006).
567 27. Cajigas, I. J. *et al.* The Local Transcriptome in the Synaptic Neuropil Revealed by
568 Deep Sequencing and High-Resolution Imaging. *Neuron* **74**, 453–466 (2012).

569 28. Hafner, A.-S., Donlin-Asp, P. G., Leitch, B., Herzog, E. & Schuman, E. M. Local
570 protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic
571 compartments. *Science* **364**, eaau3644 (2019).
572 29. Glock, C. *et al*. The translatome of neuronal cell bodies, dendrites, and axons. *Proc*.
573 *Natl*. *Acad*. *Sci*. **118**, e2113929118 (2021).
574 30. von Kügelgen, N. & Chekulaeva, M. Conservation of a core neurite transcriptome
575 across neuronal types and species. *WIREs RNA* **11**, e1590 (2020).
576 31. Lee, K.-W., Kim, H.-J., Sung, J.-J., Park, K.-S. & Kim, M. Defective neurite outgrowth
577 in aphidicolin/cAMP-induced motor neurons expressing mutant Cu/Zn superoxide
578 dismutase. *Int*. *J*. *Dev*. *Neurosci*. **20**, 521–526 (2002).
579 32. Markmiller, S. *et al*. Persistent mRNA localization defects and cell death in ALS
580 neurons caused by transient cellular stress. *Cell Rep*. **36**, 109685 (2021).
581 33. Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time:
582 visualizing and understanding mRNA localization. *Nat*. *Rev*. *Mol*. *Cell Biol*. **16**, 95–109
583 (2015).
584 34. Gandalovičová, A., Vomastek, T., Rosel, D. & Brábek, J. Cell polarity signaling in the
585 plasticity of cancer cell invasiveness. *Oncotarget* **7**, 25022–25049 (2016).
586 35. Backlund, M. *et al*. Plasticity of nuclear and cytoplasmic stress responses of RNA-
587 binding proteins. *Nucleic Acids Res*. **48**, 4725–4740 (2020).
588 36. Apicco, D. J. *et al*. Reducing the RNA binding protein TIA1 protects against tau-
589 mediated neurodegeneration in vivo. *Nat*. *Neurosci*. **21**, 72–80 (2018).

590 37. Wolozin, B. & Ivanov, P. Stress granules and neurodegeneration. *Nat. Rev.*
591 *Neurosci.* **20**, 649–666 (2019).
592 38. Ridder, K. *et al.* Extracellular vesicle-mediated transfer of functional RNA in the
593 tumor microenvironment. *OncoImmunology* **4**, e1008371 (2015).
594 39. Li, I. & Nabet, B. Y. Exosomes in the tumor microenvironment as mediators of
595 cancer therapy resistance. *Mol. Cancer* **18**, 32 (2019).
596 40. O'Brien, K., Breyne, K., Ughetto, S., Laurent, L. C. & Breakefield, X. O. RNA delivery
597 by extracellular vesicles in mammalian cells and its applications. *Nat. Rev. Mol. Cell*
598 *Biol.* **21**, 585–606 (2020).
599 41. Landry, F. J. & Findlay, S. R. Purification of human basophils by negative selection. *J.*
600 *Immunol. Methods* **63**, 329–336 (1983).
601 42. Hansel, T. T. *et al.* Purification of human blood eosinophils by negative selection
602 using immunomagnetic beads. *J. Immunol. Methods* **122**, 97–103 (1989).
603 43. Hansel, T. T. *et al.* An improved immunomagnetic procedure for the isolation of
604 highly purified human blood eosinophils. *J. Immunol. Methods* **145**, 105–110 (1991).
605 44. Munoz, N. M. & Leff, A. R. Highly purified selective isolation of eosinophils from
606 human peripheral blood by negative immunomagnetic selection. *Nat. Protoc.* **1**, 2613–
607 2620 (2006).
608 45. She, Y. & Owen, A. B. Outlier Detection Using Nonconvex Penalized Regression. *J.*
609 *Am. Stat. Assoc.* **106**, 626–639 (2011).
610 46. Lee, Y., MacEachern, S. N. & Jung, Y. Regularization of Case-Specific Parameters for
611 Robustness and Efficiency. *Stat. Sci.* **27**, 350–372 (2012).

612 47. Kong, D., Bondell, H. D. & Wu, Y. Fully Efficient Robust Estimation, Outlier Detection
613 and Variable Selection Via Penalized Regression. *Stat. Sin.* **28**, 1031–1052 (2018).
614 48. Traniello, I. M. *et al.* Single-cell dissection of aggression in honeybee colonies. *Nat.*
615 *Ecol. Evol.* **7**, 1232–1244 (2023).
616 49. Lin, L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility.
617 *Biometrics* **45**, 255–268 (1989).
618 50. Kumar, A. *et al.* Intracellular spatial transcriptomic analysis toolkit (InSTAnT). *Nat.*
619 *Commun.* **15**, 7794 (2024).
620 51. Yeo, S., Schrader, A. W., Lee, J., Asadian, M. & Han, H.-S. Spot-Based Global
621 Registration for Subpixel Stitching of Single-Molecule Resolution Images for Tissue-
622 Scale Spatial Transcriptomics. *Anal. Chem.* **96**, 6517–6522 (2024).
623 52. Wang, G. *et al.* Spatial organization of the transcriptome in individual neurons.
624 2020.12.07.414060 Preprint at https://doi.org/10.1101/2020.12.07.414060 (2020).
625 53. Lake, B. B. *et al.* A comparative strategy for single-nucleus and single-cell
626 transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA.
627 *Sci. Rep.* **7**, 6031 (2017).
628 54. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat.*
629 *Methods* **14**, 955–958 (2017).
630 55. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human
631 dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).
632 56. DeFelipe, J. *et al.* Cortical white matter: beyond the pale remarks, main conclusions
633 and discussion. *Front. Neuroanat.* **4**, 1692 (2010).

634 57. Overly, C. C., Rieff, H. I. & Hollenbeck, P. J. Organelle motility and metabolism in
635 axons vs dendrites of cultured hippocampal neurons. *J. Cell Sci.* **109**, 971–980 (1996).
636 58. Harris, J. J., Jolivet, R. & Attwell, D. Synaptic Energy Use and Supply. *Neuron* **75**,
637 762–777 (2012).
638 59. Johnson, Z. V. *et al.* Cellular profiling of a recently-evolved social behavior in cichlid
639 fishes. *Nat. Commun.* **14**, 4891 (2023).
640 60. Than-Trong, E. & Bally-Cuif, L. Radial glia and neural progenitors in the adult
641 zebrafish central nervous system. *Glia* **63**, 1406–1428 (2015).
642 61. Miranda-Negrón, Y. & García-Arrarás, J. E. Radial glia and radial glia-like cells: Their
643 role in neurogenesis and regeneration. *Front. Neurosci.* **16**, (2022).
644 62. Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and
645 robust detection of spatial expression patterns for large spatial transcriptomic studies.
646 *Genome Biol.* **22**, 184 (2021).
647 63. Abramsson, A. *et al.* The zebrafish amyloid precursor protein-b is required for motor
648 neuron guidance and synapse formation. *Dev. Biol.* **381**, 377–388 (2013).
649 64. Newman, M., Ebrahimie, E. & Lardelli, M. Using the zebrafish model for Alzheimer's
650 disease research. *Front. Genet.* **5**, (2014).
651 65. Haenisch, C. *et al.* The neuronal growth and regeneration associated Cntn1
652 (F3/F11/Contactin) gene is duplicated in fish: expression during development and retinal
653 axon regeneration. *Mol. Cell. Neurosci.* **28**, 361–374 (2005).
654 66. Eiraku, M. *et al.* DNER acts as a neuron-specific Notch ligand during Bergmann glial
655 development. *Nat. Neurosci.* **8**, 873–880 (2005).

656 67. Díaz-Ramos, À., Roig-Borrellas, A., García-Melero, A. & López-Alemany, R. α-
657 Enolase, a Multifunctional Protein: Its Role on Pathophysiological Situations. *BioMed*
658 *Res. Int.* **2012**, 156795 (2012).
659 68. Zhou, Z., Li, X., Wright, J., Candès, E. & Ma, Y. Stable Principal Component Pursuit.
660 in *2010 IEEE International Symposium on Information Theory* 1518–1522 (2010).
661 doi:10.1109/ISIT.2010.5513535.
662 69. Mateos, G. & Giannakis, G. B. Robust PCA as Bilinear Decomposition With Outlier-
663 Sparsity Regularization. *Trans Sig Proc* **60**, 5176–5190 (2012).
664 70. Giannakis, G. B., Mateos, G., Farahmand, S., Kekatos, V. & Zhu, H. USPACOR:
665 Universal sparsity-controlling outlier rejection. in *2011 IEEE International Conference on*
666 *Acoustics, Speech and Signal Processing (ICASSP)* 1952–1955 (2011).
667 doi:10.1109/ICASSP.2011.5946891.
668 71. Huang, P., Cai, M., Lu, X., McKennan, C. & Wang, J. Accurate estimation of rare cell-
669 type fractions from tissue omics data via hierarchical deconvolution. *Ann. Appl. Stat.* **18**,
670 1178–1194 (2024).
671 72. Zhong, Y. & Liu, Z. Gene expression deconvolution in linear space. *Nat. Methods* **9**,
672 8–9 (2012).
673 73. Hicks, S. C., Townes, F. W., Teng, M. & Irizarry, R. A. Missing data and technical
674 variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562–578 (2018).
675 74. Townes, F. W., Hicks, S. C., Aryee, M. J. & Irizarry, R. A. Feature selection and
676 dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome*
677 *Biol.* **20**, 295 (2019).

75. 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).

683 **Figure Legends**

684 **Figure 1. Schematic overview of RESCUE.** Overview of RESCUE. RESCUE is a
685 computational negative selection method designed to recover the idiosyncratic
686 transcriptome. It takes as input ST data to be decomposed, along with a reference matrix
687 constructed from sc/snRNA-seq data from the same tissue. With these two data inputs,
688 RESCUE performs robust sparse recovery of the idiosyncratic transcriptome through a
689 regularized multivariate regression problem. To account for gene variability across
690 different locations, RESCUE iteratively updates a weight matrix within its alternating
691 minimization algorithm. RESCUE partitions the full transcriptome profile into the canonical
692 and idiosyncratic expression, which can be further used for downstream analysis, such
693 as DEG and SVG analyses. Created with BioRender.com.

694

695 **Figure 2. Simulation results of deconvolution and idiosyncratic transcriptome**
696 **recovery. A.** Comparisons of cell-type deconvolution performance using the average
697 Lin's CCC across cell types. **B.** Comparisons of idiosyncratic transcriptome recovery
698 performance using the relative Frobenius error. The proportion of idiosyncratic expression
699 ranges from 0% to 80% and the idiosyncratic sparsity level ranges from 80% to 95%.

700

701 **Figure 3. Application to the honey bee brain MERFISH data. A.** Comparison of
702 segmentation-based ("Cell Mask") and estimated somatic proportions. Cell mask ratio is
703 computed as the proportion of MERFISH spots in each grid lying within a cell. **B.** Bar plots
704 of Lin's CCC and RMSE between segmentation-based and estimated idiosyncratic
705 proportions. For each grid, idiosyncratic proportions are calculated as 1 – (proportion of

706 grid's somatic expression). **C.** Comparison of segmented and estimated transcript counts
707 for each grid in cell body regions where the cell mask ratio is 1. Grey dashed lines indicate
708 $y = x$ and blue solid lines represent the smoothed line from estimates.
709
710 **Figure 4. Application to the human breast cancer Visium data. A.** Manual annotation
711 of Visium ST data from human breast cancer tissue. **B.** Heatmap of gene expression
712 correlations between pseudo-bulk Visium and scFFPE-seq cell types. Red box highlights
713 high similarity between adipocytes and other cell types including stromal cells. **C.**
714 Estimated cell-type proportions in adipocyte regions across methods. Existing methods
715 overestimate stromal and endothelial populations due to the absence of adipocytes in the
716 reference. RESCUE avoids such misattributions by modeling idiosyncratic expression. **D.**
717 Comparison of estimated versus reference cell-type proportions across all methods.
718
719 **Figure 5. Application to the human DLPFC Visium data. A.** Anatomical annotations of
720 10x Visium samples from the DLPFC dataset. **B.** Idiosyncratic expression proportions in
721 each layer estimated by RESCUE. RESCUE estimates higher idiosyncratic proportions
722 in Layer 1 and WM, consistent with their enrichment in neuropil and myelinated axons,
723 respectively. **C.** Differential expression analysis comparing Layer 1 to other layers using
724 observed ST expression ("raw") versus idiosyncratic expression. Each point represents a
725 gene, with log-fold changes from raw expression (x-axis) versus idiosyncratic expression
726 (y-axis). Blue dashed lines indicate the thresholds at log-fold changes = 0. Red-labeled
727 genes are top markers of Layer 1, detected from the idiosyncratic expression. **D.** Gene

728 ontology (GO) enrichment analysis results of Layer 1-enriched differentially expressed
729 genes (DEGs) based on idiosyncratic (left) and raw (right) expression.
730
731 **Figure 6. Application to the cichlid fish telencephalon Visium data. A.** Schematic of
732 radial glial (RG) anatomy (left; Johnson et al.[59]) and regional annotation of Visium ST data
733 from the dorsal telencephalon of cichlid fish (right). **B.** Comparison of adjusted p-values
734 for SVG detection using raw (x-axis) versus idiosyncratic (y-axis) expression. Blue
735 dashed lines indicate the significance threshold at adjusted p-values = 0.05. Red-labeled
736 genes are examples of SVGs detected from idiosyncratic expression but not detectable
737 from raw data. **C.** Spatial expression maps of idiosyncratic expression for unique SVGs.
738

**Figure 1.**

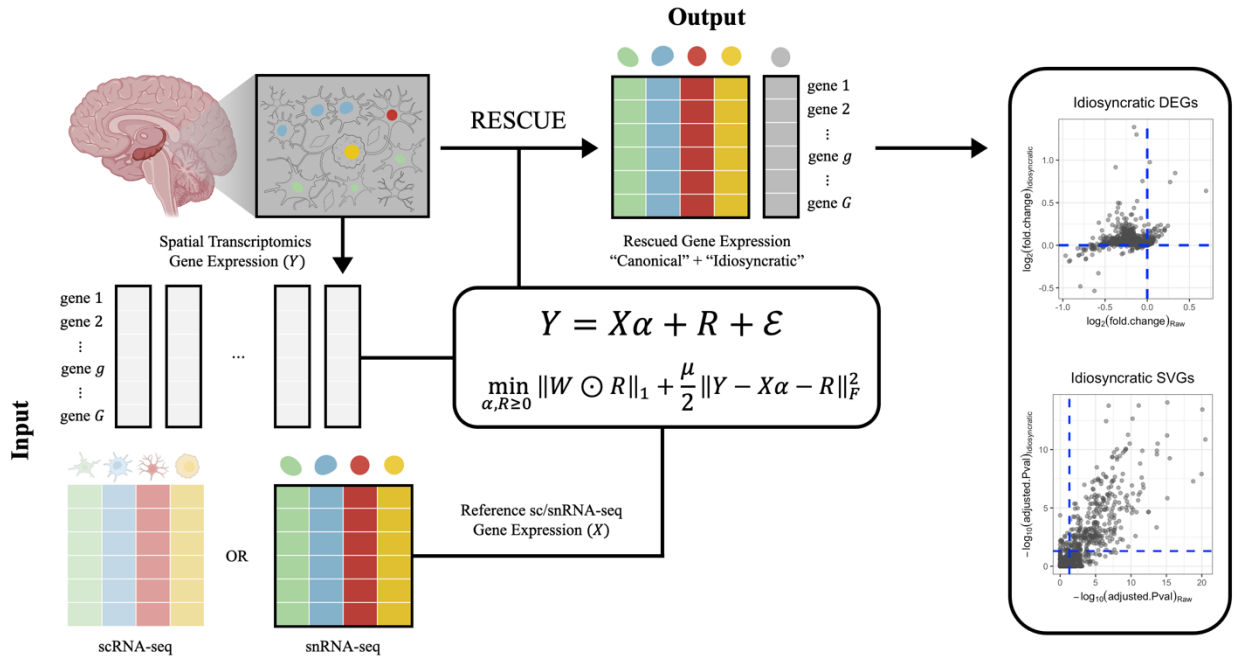



**Figure 2.**

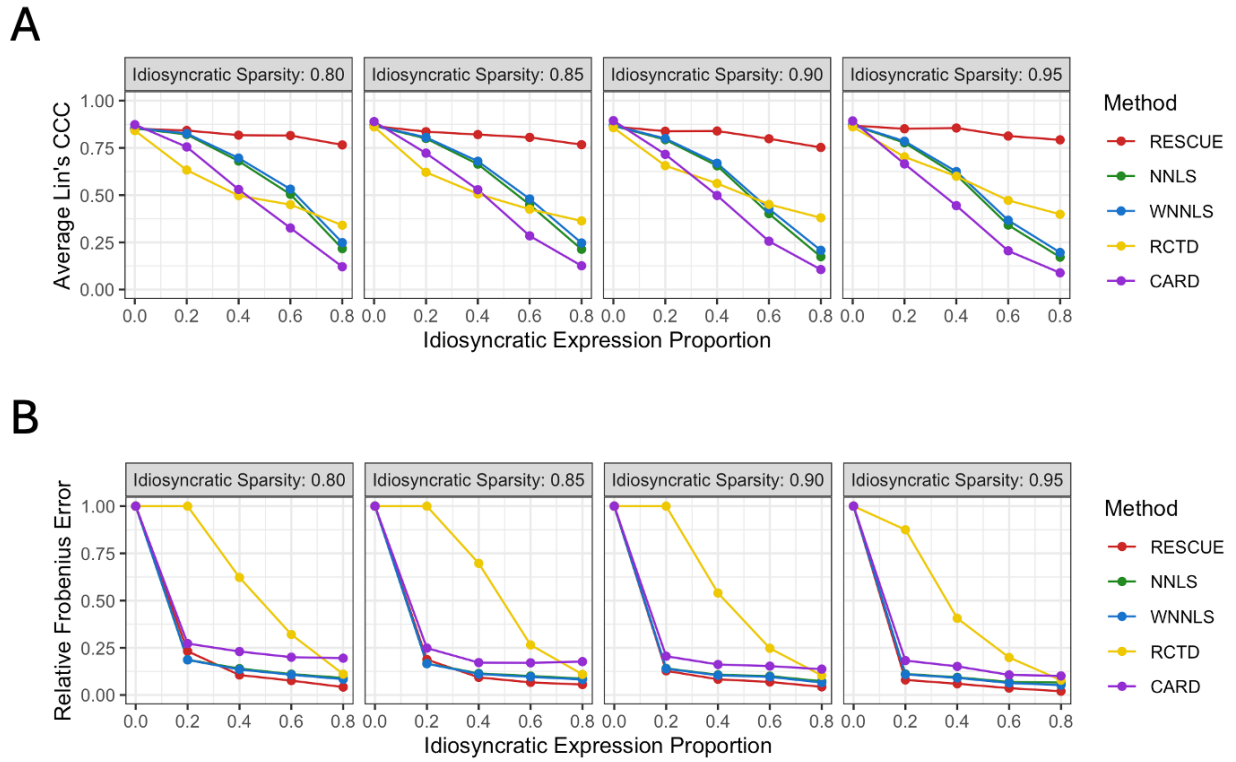


**Figure 3.**

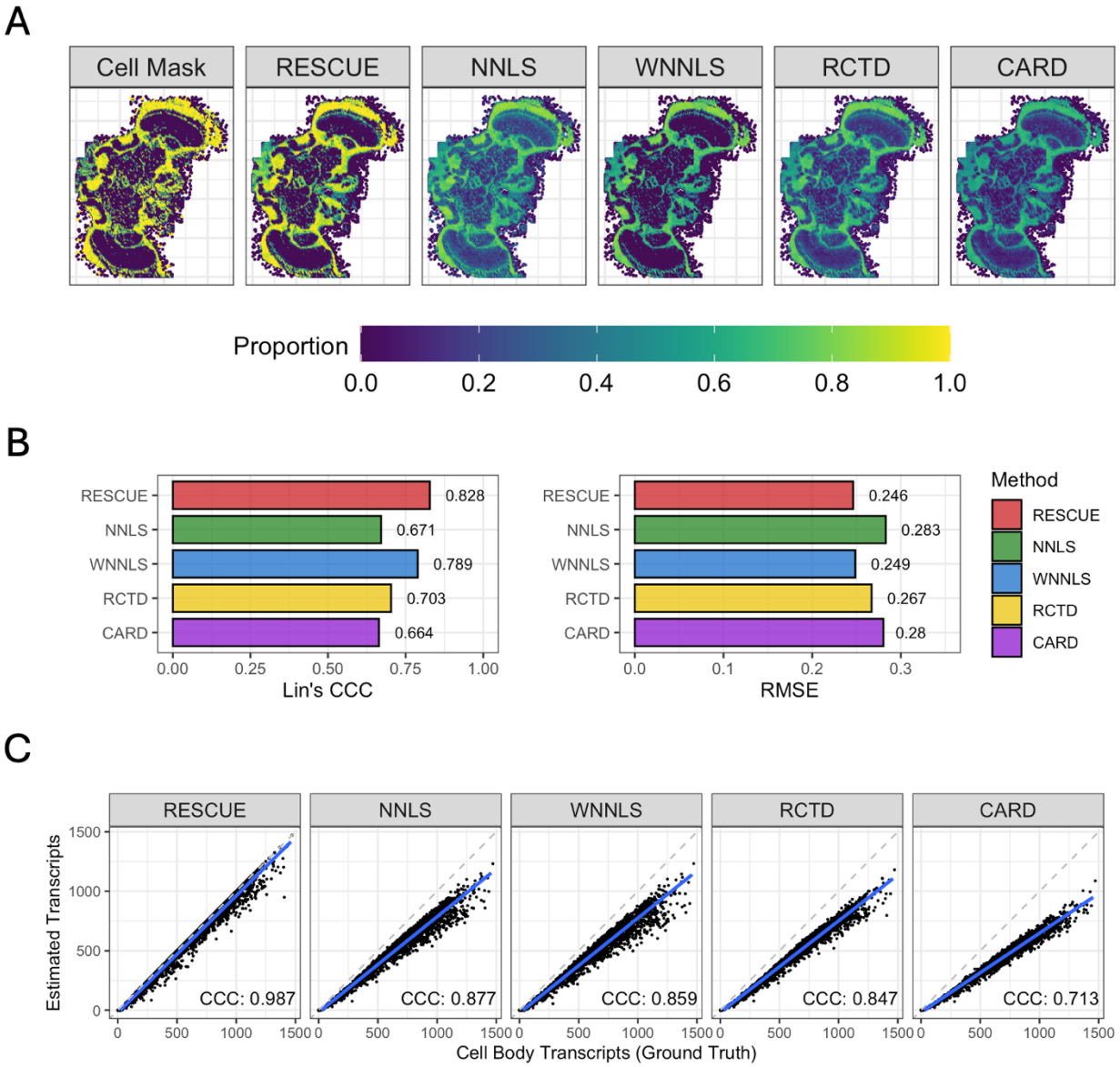

A
Cell Mask
RESCUE
NNLS
WNNLS
RCTD
CARD
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
B
RESCUE
NNLS
WNNLS
RCTD
CARD
0.828
0.671
0.789
0.703
0.664
0.00
0.25
0.50
0.75
1.00
Lin's CCC
0.246
0.283
0.249
0.267
0.28
0.0
0.1
0.2
0.3
RMSE
Method
RESCUE
NNLS
WNNLS
RCTD
CARD
C
Estimated Transcripts
Cell Body Transcripts (Ground Truth)
CCC: 0.987
CCC: 0.877
CCC: 0.859
CCC: 0.847
CCC: 0.713
0
500
1000
1500

**Figure 4.**

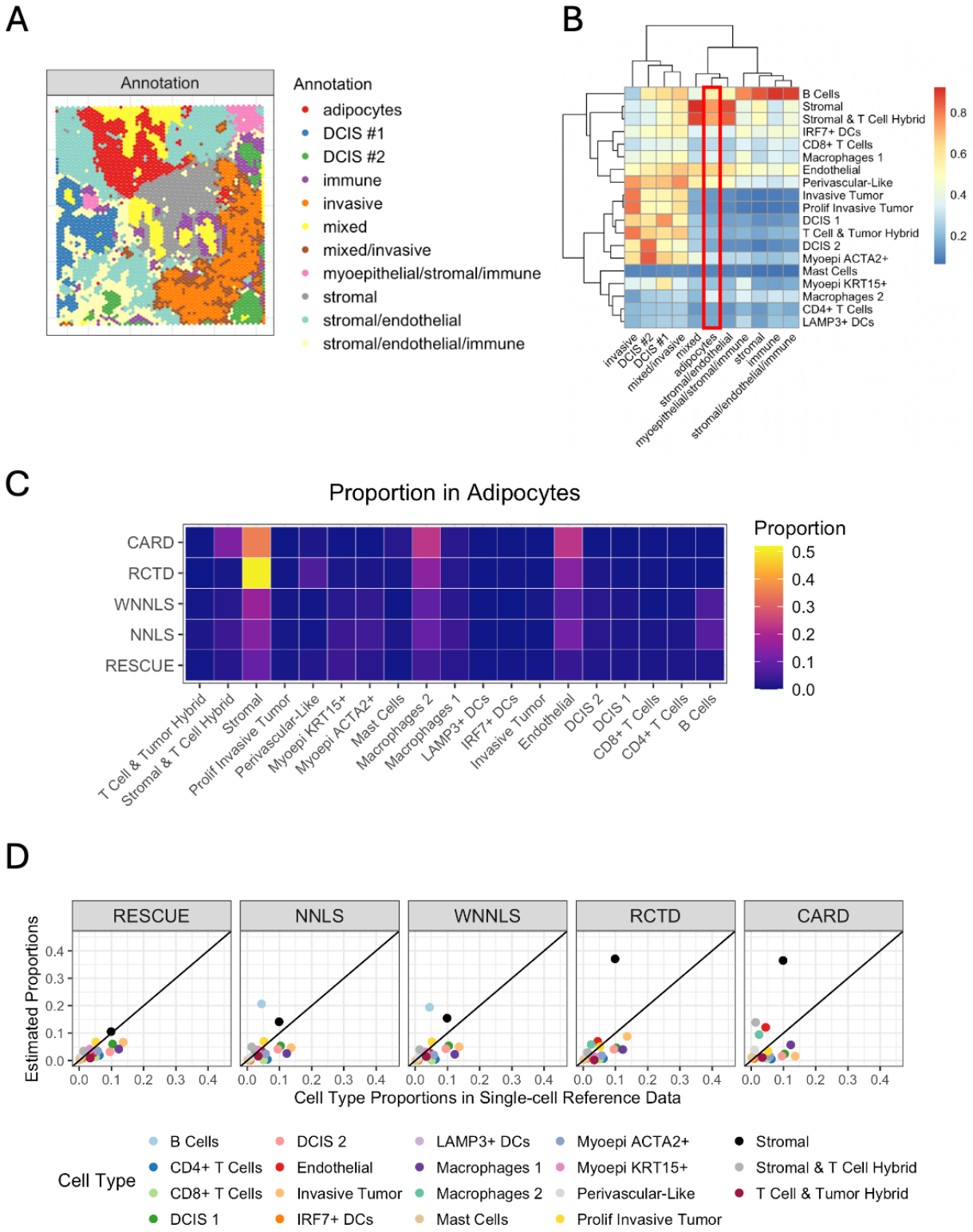

A
Annotation
Annotation
adipocytes
DCIS #1
DCIS #2
immune
invasive
mixed
mixed/invasive
myoepithelial/stromal/immune
stromal
stromal/endothelial
stromal/endothelial/immune
B
B Cells
Stromal
Stromal & T Cell Hybrid
IRF7+ DCs
CD8+ T Cells
Macrophages 1
Endothelial
Perivascular-Like
Invasive Tumor
Prolif Invasive Tumor
DCIS 1
T Cell & Tumor Hybrid
DCIS 2
Myoepi ACTA2+
Mast Cells
Myoepi KRT15+
Macrophages 2
CD4+ T Cells
LAMP3+ DCs
invasive
DCIS #2
DCIS #1
mixed/invasive
mixed
adipocytes
stromal/endothelial
myoepithelial/stromal/immune
stromal
immune
stromal/endothelial/immune
0.8
0.6
0.4
0.2
C
Proportion in Adipocytes
CARD
RCTD
WNNLS
NNLS
RESCUE
T Cell & Tumor Hybrid
Stromal & T Cell Hybrid
Stromal
Prolif Invasive Tumor
Perivascular-Like
Myoepi KRT15+
Myoepi ACTA2+
Mast Cells
Macrophages 2
Macrophages 1
LAMP3+ DCs
IRF7+ DCs
Invasive Tumor
Endothelial
DCIS 2
DCIS 1
CD8+ T Cells
CD4+ T Cells
B Cells
Proportion
0.5
0.4
0.3
0.2
0.1
0.0
D
RESCUE
NNLS
WNNLS
RCTD
CARD
Estimated Proportions
0.0
0.1
0.2
0.3
0.4
Cell Type Proportions in Single-cell Reference Data
Cell Type
B Cells
CD4+ T Cells
CD8+ T Cells
DCIS 1
DCIS 2
Endothelial
Invasive Tumor
IRF7+ DCs
LAMP3+ DCs
Macrophages 1
Macrophages 2
Mast Cells
Myoepi ACTA2+
Myoepi KRT15+
Perivascular-Like
Prolif Invasive Tumor
Stromal
Stromal & T Cell Hybrid
T Cell & Tumor Hybrid

**Figure 5.**

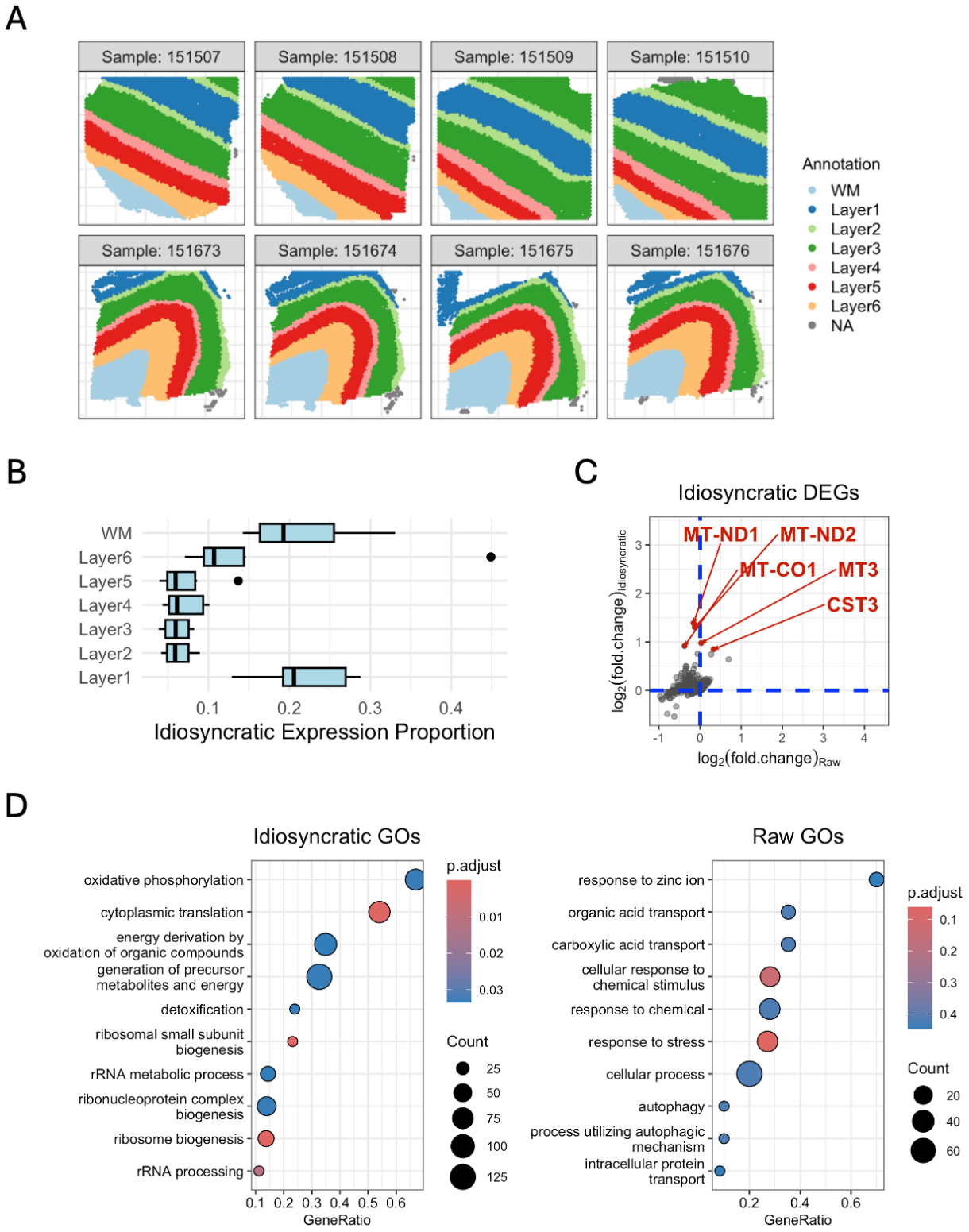
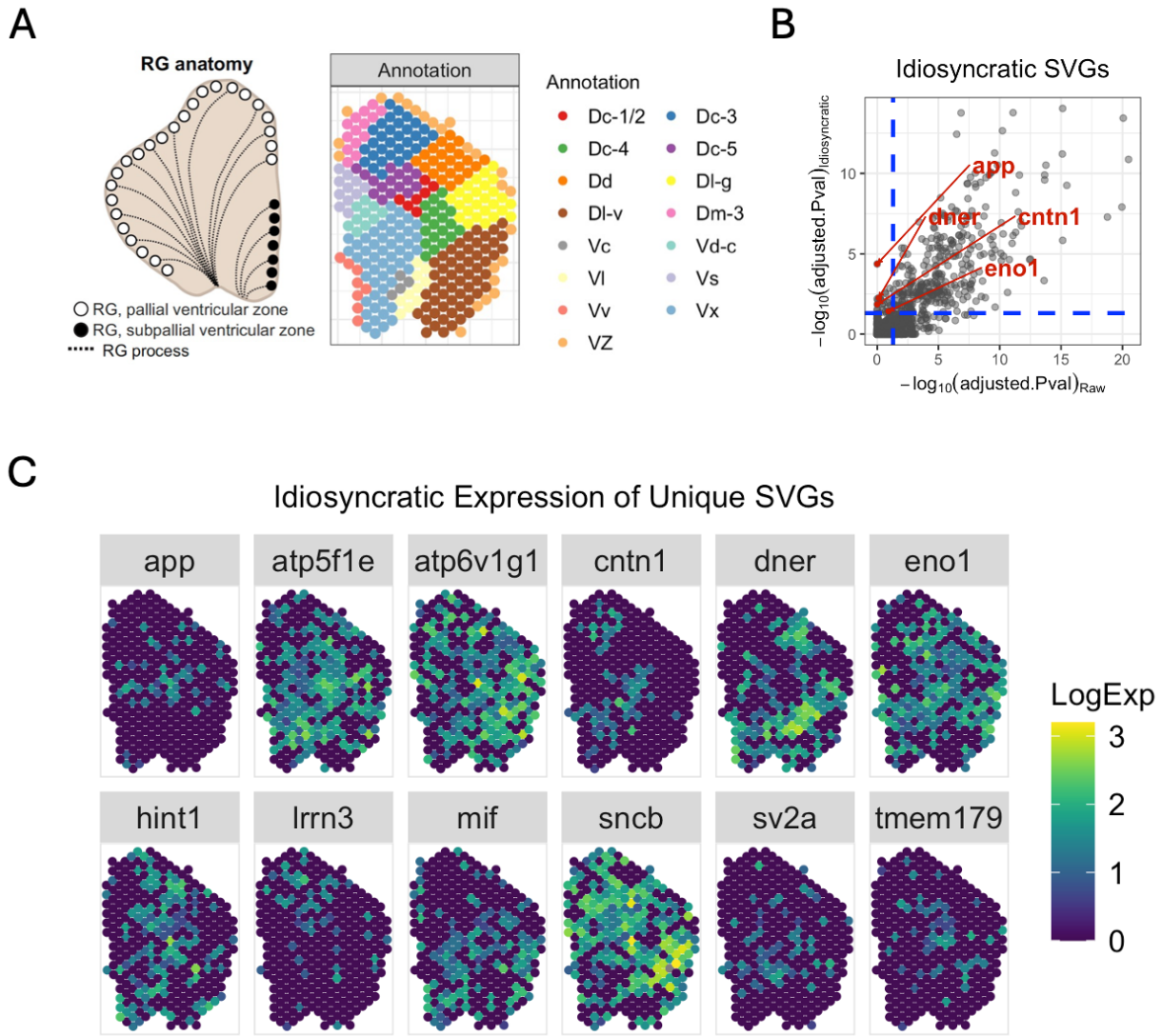

A
Sample: 151507
Sample: 151508
Sample: 151509
Sample: 151510
Sample: 151673
Sample: 151674
Sample: 151675
Sample: 151676
Annotation
WM
Layer1
Layer2
Layer3
Layer4
Layer5
Layer6
NA
B
WM
Layer6
Layer5
Layer4
Layer3
Layer2
Layer1
0.1
0.2
0.3
0.4
Idiosyncratic Expression Proportion
C
Idiosyncratic DEGs
MT-ND1
MT-ND2
MT-CO1
MT3
CST3
log2(fold.change)Idiosyncratic
log2(fold.change)Raw
D
Idiosyncratic GOs
oxidative phosphorylation
cytoplasmic translation
energy derivation by oxidation of organic compounds
generation of precursor metabolites and energy
detoxification
ribosomal small subunit biogenesis
rRNA metabolic process
ribonucleoprotein complex biogenesis
ribosome biogenesis
rRNA processing
GeneRatio
p.adjust
Count
Raw GOs
response to zinc ion
organic acid transport
carboxylic acid transport
cellular response to chemical stimulus
response to chemical
response to stress
cellular process
autophagy
process utilizing autophagic mechanism
intracellular protein transport
GeneRatio
p.adjust
Count

**Figure 6.**


A
RG anatomy
RG, pallial ventricular zone
RG, subpallial ventricular zone
RG process
Annotation
Annotation
Dc-1/2
Dc-3
Dc-4
Dc-5
Dd
Dl-g
Dl-v
Dm-3
Vc
Vd-c
Vl
Vs
Vv
Vx
VZ
B
Idiosyncratic SVGs
app
dner
cntn1
eno1
−log10(adjusted.Pval)Idiosyncratic
−log10(adjusted.Pval)Raw
C
Idiosyncratic Expression of Unique SVGs
app
atp5f1e
atp6v1g1
cntn1
dner
eno1
hint1
lrrn3
mif
sncb
sv2a
tmem179
LogExp
3
2
1
0