MANTIS: Analytics toolkit for spatial metabolomics with matching spatial transcriptomics data

Yu Hao[1], Yeojin Kim[1], Bhavay Aggarwal[1], Saurabh Sinha[1*]

1 The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA

* Corresponding Author
Saurabh Sinha, Email- saurabh.sinha@bme.gatech.edu

ABSTRACT

Motivation: Joint Spatial Metabolomics (SM) and Spatial Transcriptomics (ST) profiling is a powerful approach to fine-mapping of metabolic states associated with tissue function. Current computational tools for analysis of "SM+ST" data focus primarily on alignment and integration of the two modalities, with limited support for probing biological relationships between the two molecular layers.

Results: We present MANTIS, a statistical framework for analyzing co-registered SM+ST profiles at single cell or spot resolution, along with spatial domain or cell type information, to discover metabolite spatial patterns and gene-metabolite relationships. It employs an autocorrelation-preserving permutation strategy to assess statistical significance, yielding calibrated inference under spatial dependence. It disentangles different sources of spatial patterns and correlations, viz., those arising from regional preferences, cell type associations, or other unknown factors. It introduces the use of spatial cross-correlation and spatial partial correlation statistics for quantifying gene-metabolite associations. Across data sets spanning different spatial technologies, tissues and species, MANTIS provides more specific and interpretable discoveries than existing methods through rigorous statistical testing and explicitly modeling confounding structure. To our knowledge, MANTIS is the first toolkit to unify spatial metabolomics, spatial transcriptomics, cell type information and spatial domains within a single framework that emphasizes spatial statistics, hypothesis testing and confounder correction.

Availability and Implementation: Freely available on the web at https://github.com/yuhaotuo/MANTIS.

KEYWORDS

spatial metabolomics; spatial transcriptomics; gene-metabolite association; spatial statistics

INTRODUCTION

The rapidly maturing technology of spatial metabolomics [1], where abundances of many metabolites are measured in a multiplexed manner with spatial resolution, can be transformational for fine-mapping of cellular processes associated with tissue function. For instance, spatial metabolomics (SM) maps of the brain reveal lipid- and metabolite-level variations associated with function in normal [2] as well as disease conditions [3,4]. Recent breakthroughs have enabled SM profiling and spatial transcriptomics (ST) profiling of the same tissue, providing an even richer multi-omics view into the tissue's molecular state [5]. The matched ST profile can help delineate

cell types and spatial domains [6] in the tissue, giving us clarity on the regional selectivity [7] and cell type-specificity of metabolic events and their changes. For instance, Hu et al. [1] used matched SM and spatial proteomics profiling to characterize local metabolite competition of neighboring cells in human lung cancer. Similarly, Vicari et al. [5] analyzed SM+ST profiles of mouse striatum to characterize genes and neuronal subtypes associated with dopamine. The ST profile can also elucidate regulatory mechanisms related to metabolic alterations [8].

The transformative potential of joint metabolomics and transcriptomics profiling cannot be unlocked without the key to integrative multi-omics *analysis*. Spatial metabolomics analysis lacks comprehensive statistical tools for unbiased discovery of metabolic patterns involving cell types and spatial domains. Existing tools for SM data offer a variety of core functionalities such as annotating m/z peaks, dimensionality reduction, clustering and visualization, spatial structure discovery, pairwise metabolite network reconstruction, etc. [2,9,10]. However, these tools operate at the *pixel-level* rather than cell-level, and typically do not incorporate cell type or spatial domain information. Recently introduced tools tackle SM+ST data, addressing cross-modality alignment and cell segmentation [2,11], a necessary step for cell-level analysis. SpaMTP [12] is a state-of-the-art tool for downstream analysis of such cell-level multi-omics data, including gene-metabolite correlation analysis, which can provide important clues into mechanisms associated with metabolic patterns. SpatialMETA [13] uses neural networks for cross-modality (SM, ST) integration and provides several useful downstream analyses of the integrated data, including spatial clustering, markers of clusters, gene-metabolite correlation, etc. However, key analytic goals remain unmet in existing tools for SM+ST data analysis. For instance, current tools [12,13] for discovering gene-metabolite relationships use non-spatial correlation metrics such as the Pearson correlation coefficient; ideally such relationships should be quantified in a spatially aware manner, in order to account for prevalent and diverse spatial covariance structures that may confound the findings. Gene-metabolite correlations may also be confounded by associations with cell types and regions, and careful accounting of these confounders is necessary. Furthermore, correlation scores need to be accompanied by robust statistical testing in order to control false discoveries, an aspect that has not been rigorously addressed in prior work. Finally, while existing tools do provide some support for detecting region- or cell type-associated metabolites, such functions are unable to disentangle spatial variation driven by tissue organization, cell type distributions and other factors, or to assess statistical significance of these effects in a unified framework.

Here, we present a computational toolkit called **MANTIS (M**etabolomics **An**d **T**ranscriptomics **I**n **S**pace) for analysis of paired spatial metabolomics and transcriptomics (SM+ST) data. It provides functions for two broad categories of tasks: (1) discovery of metabolic spatial patterns and (2) identification of gene-metabolite relationships. It exploits cell type annotations of cells and spatial domain delineations of the tissue, if available, to extract biologically interpretable patterns and relationships. (See Table 1.) It employs well-established spatial and non-spatial statistical approaches and a novel strategy to assess statistical significance of spatial patterns. It also disentangles different sources of spatial patterns and correlations, viz., those arising from regional preferences or cell type associations of a metabolite. We present the functions of the toolkit with examples of their application to three data sets where SM profiles are supported by ST profiles of the same or adjacent tissue section. The SM datasets used in this study were generated using different technologies, illustrating the ability of MANTIS to operate across heterogeneous measurement platforms. To our knowledge, MANTIS is the first software toolkit to combine spatial, single-cell metabolomics, transcriptomics, and cell type information with rigorous statistical procedures, revealing metabolite patterns and their gene-level correlates and mechanisms.

RESULTS

Overview of MANTIS toolkit

MANTIS provides a suite of functions for analyzing spatial metabolomics (SM) data on a tissue sample along with co-registered spatial transcriptomics (ST) data for the same or adjacent tissue section. To account for variations in spatial resolution across technologies, we assume that the data set comprises a set of “objects” with a metabolomics profile, a transcriptomics profile and spatial coordinates (**Figure 1**). Objects may be individual cells, or “spots” in certain ST technologies such as 10x Visium. We assume that mismatches in resolution have been addressed and objects are aligned [5,12], for instance through available tools such as SpatialMETA [13] and SpaMTP [12]. We also assume that the ST data have been analyzed via third-party tools to annotate cell types [14] or to deconvolve spots into cell type proportions [15] and to delineate spatial domains of the tissue [14]. Several tools are available for both tasks [14], and studies generating ST data typically provide cell type annotations and often spatial domains (e.g., brain regions [16], tumor clusters and microenvironments [1], etc.) Thus, the input to MANTIS can be described as in **Figure 1A**, as a matrix with rows representing “objects” (cells, spots, etc.), and columns representing information about those objects, e.g., spatial coordinates, metabolite

106 abundances and gene expression, and optionally region and/or cell type annotations. MANTIS
107 includes a set of functions that analyze such a data matrix to report lists of metabolic spatial
108 patterns and gene-metabolite associations (Figure 1), with rigorous significance assessment and
109 handling of confounding factors.

110 We seek to identify metabolites with "interesting" (statistically significant) spatial distributions.
111 Metabolomic mapping studies have pointed to many such metabolic spatial patterns in mouse
112 brain [17], human and primate brains [18], human tonsil and tumor samples [1], etc. Of interest
113 is a metabolite that is (1) abundant in one or a few anatomical/functional regions of the tissue, (2)
114 is expressed in cellular neighborhoods ("niches") marked by specific cell types, or (3) exhibits a
115 statistically unusual spatial distribution even if it does not segregate by regions/cell types [19]
116 (**Figure 1B**). Such patterns may help us link metabolites to regional functions of the tissue [17],
117 reveal metabolic requirements or signaling roles of specific cell types [17,20], or point to
118 previously unrecognized metabolic microenvironments or functional domains within the tissue
119 [2,20-22]. MANTIS provides functions to identify these three classes of metabolite spatial patterns.
120 In addition, MANTIS provides routines to quantify gene-metabolite associations via spatial
121 statistics (Figure 1C), while also accounting for potential confounders such as regions and cell
122 types. We illustrate these functions via application to the mouse brain striatum SM+ST data set
123 of Vicari et al. [5], referred to below as the VICARI data set.

124 **Discovery of region-associated metabolites**

125 A simple and easily interpreted pattern is that of a metabolite preferentially localized to one or
126 more spatial domains ("regions"). Regional preference of metabolites often reflects specialized
127 functions of brain areas, e.g., certain neurotransmitters enriched in regions active in signaling, or
128 energy metabolites concentrated in areas with high metabolic demand [21]. MANTIS provides the
129 '**regional_met**' function to identify such patterns. This function first represents a metabolite's
130 abundances across regions as a probability vector: the "regional abundance profile" for that
131 metabolite. A background profile is obtained similarly by averaging abundances of *all* metabolites
132 in each region. It then quantifies the divergence between the metabolite's profile and background
133 profile using Kullback-Leibler (KL) divergence (**Methods**). This is repeated for every metabolite,
134 and significance threshold is set based on an empirical null distribution obtained by a custom
135 permutation method that creates randomized versions of a metabolite's spatial distribution while
136 preserving autocorrelation, see **Methods**.

137 Using the regional_met function on the VICARI data set, which includes 9 delineated spatial
138 domains (regions in the mouse striatum) (**Figure 2A**) we obtained a list of 66 significant metabolic
139 spatial patterns (FDR < 0.1, **Figure 2D**, **Supplementary Table S1**) out of the 500 highly variable
140 metabolites examined. **Figures 2B and 2C** show two such metabolites – m/z 725.551, which is
141 enriched in the ventrolateral striatum (VL), and m/z 822.614, which is relatively abundant in the
142 corpus callosum (CC) and anterior commissure (ACO). The complete set of region-associated
143 metabolites spans a small set of patterns (**Figure 2E, Supplementary Figure S1**), several of
144 which are similar to cell types' distributions (**Supplementary Figure S2**), suggesting that
145 metabolite-cell type associations may underlie their regional preferences (or vice versa).

146 *Comparison to existing tools:* Available tools such as SpatialMETA and SpaMTP provide a similar
147 functionality of identifying "marker" metabolites, by testing if a metabolite's abundance is enriched
148 in a region compared to other regions, using a between-group difference test [13]. Our approach
149 is different in that (a) it tests for non-randomness of the overall regional distribution of a metabolite
150 rather than enrichment in each region separately, thus avoiding a multiple testing burden, and (b)
151 performs a statistical test that controls for the spatial autocorrelation of the metabolite, which is
152 not addressed in current tools. We show in **Supplementary Table S2** that the "one region vs.
153 others" test reports a very large number of significant associations. For instance, a Wilcoxon test
154 finds 2,886 metabolite-region associations at FDR < 0.05 (64% of 4,500 pairs tested). In contrast,
155 MANTIS performs only 500 tests (one per metabolite) and found 66 significant region-associated
156 metabolites at FDR < 0.1. Though there is no "ground truth" to objectively establish their relative
157 error rates, we interpret these findings as suggesting that MANTIS reveals a more conservative
158 set of regional associations that is likely to have a lower false positive rate.

159 **Discovery of cell type-associated metabolites**

160 There have been several reports of specific metabolites being enriched in certain cell types and
161 in tissue regions with specific cell type compositions [1,17,20]. These associations may be specific
162 to a region, e.g., elevated glycogen in astrocytes in the hippocampus and cerebellar cortex [23],
163 or transcend regional boundaries, e.g., lactate predominantly released by astrocytes to fuel
164 neuronal activity [24]. Findings of metabolite-cell type associations can point to instances of
165 metabolic reprogramming in certain disease contexts.

166 MANTIS provides the '**celltype_met**' function to discover cell type-associated metabolites. It tests
167 for spatial covariance between a cell type's presence and a metabolite's abundance, using the
168 Spatial Cross-Correlation Index (SCI) [25] (**Methods**), which is closely related to bivariate Moran's

169 I [26]. For spot-level data, cell type presence is assumed to be obtained from deconvolution tools
170 while for cell-level data a Boolean indicator of cell type annotation is used. A permutation-based
171 empirical null distribution that controls for metabolite spatial autocorrelations is used to set a
172 significance threshold on the SCI score (**Methods**).

173 Using this function on the VICARI data set, which contains 48 different cell types (**Figure 3A**), we
174 obtained 1,518 significant cell type-metabolite associations (FDR<0.05) (**Supplementary Table**
175 **S3**), with the most associations being for oligodendrocytes (OLIG), Schwann Cells (SCHW) and
176 Satellite glia (SATG) (**Figure 3B**). **Figure 3C** depicts the strong spatial correlation between cell
177 type Cerebellum neurons and metabolite 172.001 m/z (SCI score 0.474). Since certain cell types
178 are statistically segregated across spatial domains, we expected and confirmed that many of the
179 discovered cell type-metabolite associations are also region-associated metabolites (**Figure 3D**).

180 *Comparison to existing tools:* To our knowledge, existing tools do not explicitly test for cell type
181 associations of metabolites. A possible baseline approach is to use Pearson correlation or
182 Wilcoxon test depending on whether cell type information is quantitative (proportions) or binary.
183 Such an approach differs from MANTIS in two ways: the score is calculated in a spatially agnostic
184 manner, and the significance test does not account for spatial autocorrelations. We evaluated this
185 baseline method (**Supplementary Table S4**) and found 11,869 significant cell type-metabolite
186 associations at FDR < 0.05, ~49.4% of 24,000 tested pairs. In the absence of a reliable "ground
187 truth" we are unable to perform an objective comparison of error rates to that of MANTIS, but the
188 far larger number of associations reported (at the same FDR threshold) suggests a weak null
189 hypothesis used in the non-spatial baseline. **Supplementary Figure S3** shows an example
190 association that is designated significant by the baseline method but not by MANTIS.

191 **Discovery of spatially variable metabolites**

192 A metabolite's spatial pattern may not necessarily segregate by spatial domains or cell type niches
193 but still be relevant to tissue function. We drew inspiration from the literature on "spatially varying
194 genes" (SVGs) [27,28] to search for such patterns in SM data. These tools test if spatially proximal
195 cells tend to exhibit similar levels of the molecule of interest. The '**spatvar_met**' function in
196 MANTIS provides a similar functionality – detection of spatially variable metabolites (SVMs) –
197 using the Moran's I score [29] . We recognized that a region-associated metabolite is likely to be
198 also detected as a SVM because being enriched in a region implies clustering of cells/spots where
199 the metabolite is found. Similarly, cell type-associated metabolites are also likely to be detected
200 as a SVM. Thus, the spatvar_met function corrects for the confounding effects of spatial domains

201 (or cell type) by first regressing metabolite abundance against spatial domain (or cell type)
202 information and applying the Moran's I score to the residuals (**Figure 3E, Methods**). A metabolite
203 that retains a high Moran's I after both adjustments is designated as a spatially variable metabolite.

204 Using this function on the VICARI data, we discovered 73 spatially variable metabolites (out of
205 500), based on a Moran's I threshold of 0.2 (**Supplementary Table S5**). **Figure 3F** shows two
206 distinct examples of such patterns: 390.166 m/z displays a broad gradient emanating from the
207 dorsolateral injection site (dopamine delivered from the upper-left), consistent with diffusion from
208 that source, whereas 838.630 m/z exhibits discrete focal hotspots that do not align well with major
209 domains or local cell-type composition.

210 The spatvar_met routine includes a user-tunable parameter called the length scale (l) that
211 influences the type of pattern reported. Smaller l increases sensitivity to local structure, while
212 larger l accommodates larger-scale patterns. **Supplementary Table S6** shows a smooth,
213 monotonic decrease in the number of SVMs identified with increasing l, with **Figures 3G-3I**
214 showing how extreme values of the length scale can reveal complementary spatial patterns. The
215 analyses above (Figures 3E,F) were performed with an intermediate value of $l^2 = 10d_{min}^2$ ($d_{min}$is
216 distance between spots).

217 *Comparison to existing tools:* The Moran's I functionality is also provided as part of the
218 SpatialMETA and SpaMTP tools. However, these implementations do not account for the
219 confounding effects of spatial domains and cell types. To underscore the consequences of this
220 difference, we show (**Supplementary Table S5)** that of the 181 (36% of all) metabolites with
221 Moran's I ≥ 0.2, only 73 retain that level of autocorrelation after accounting for confounders. Thus,
222 the spatvar_met function avoids redundancy and mutual confounding among different types of
223 metabolic spatial patterns. We also tested the SVG detection tool SpatialDE [27] on this data set,
224 identifying all 500 metabolites as significant (FDR < 0.02), most of which can be explained by
225 spatial domain and/or cell type associations, which are not corrected for by that tool
226 (**Supplementary Table S7**).

227 **Discovering gene-metabolite associations**

228 The routines introduced above primarily analyze SM data and reveal metabolite-associated
229 patterns, using ST data in a supporting role. We now outline the more direct form of cross-omics
230 functionality provided by MANTIS –identifying gene-metabolite pairs with spatially correlated
231 expression. The "**compute_genemet_sci**" function calculates SCI score for every gene-

metabolite pair. This score is high if the gene and metabolite exhibit co-expression in the same spots and/or in proximal spots (**Methods**).

We used the compute_genemet_sci function on the VICARI data set, focusing on 1000 highly variable genes and 500 highly variable metabolites (**Supplementary Table S8**, **Figure 4A**), noting extreme values in excess of 0.6. Visualizations of two sample pairs clearly depicts the high gene-metabolite covariation (**Figure 4B, Supplementary Figure S4**). SCI scores are assessed for statistical significance using an empirical null distribution derived from a randomized data set that preserves the metabolite's spatial autocorrelation (**Supplementary Figure S5**), and we noted that 12.3% of all gene-metabolite pairs tested were significant at an FDR < 0.05 (Supplementary Table S8).

A gene and metabolite that are expressed specifically in the same cell type will tend to exhibit high SCI. Furthermore, if there are many genes and metabolites specific to that cell type, then all pairs thereof will be significantly associated. It is more parsimonious to separate out such cases and report them as sets of genes and metabolites associated with that cell type, rather than as directly associated gene-metabolite pairs. Thus, MANTIS provides the "**compute_SPC-CT**" function to report gene-metabolite spatial correlations after removing the confounding effect of cell type information. It uses a statistical score that we call "Spatial Partial Correlation" (SPC), where the gene and metabolite levels are first fit separately to cell type and the residuals of the fit are subjected to SCI calculation (**Methods**).

Using the "compute_SPC-CT" function, we obtained a revised compendium of gene-metabolite associations (**Supplementary Table S9, Supplementary Figure S6**), with 6.8% of the pairs significant at adjusted p-value < 0.05. This confirms that the majority of the significant pairs identified initially (Figure 4A) were likely a result of cell type-specific expression. Two of the strongest gene-metabolite spatial associations observed after accounting for cell type information are shown in **Figure 4C** and **Supplementary Figure S7**. In Figure 4C, the *Raph1*–848.540 m/z pair shows a closely matched spatial pattern across the tissue; this agreement remains after residualizing cell type effects (see **Supplementary Figure S8**). A secondary product of the compute_SPC-CT function is a listing of gene-metabolite pairs whose strong statistical association is primarily driven by their link to cell types (**Supplementary Table S10**). An example of such a pair is shown in **Supplementary Figure S9**.

Another potential confounder for gene-metabolite associations is spatial domains (SD) in the tissue -- a gene and metabolite localized in the same SD are likely to exhibit a high SCI. MANTIS

264 provides the function '**compute_SPC-SD**' to report gene-metabolite associations not explainable
265 by co-localization in a SD **(Methods**) – it first removes SD-related signals from the gene's and
266 metabolite's expression distribution and then calculates SCI on the residuals. Applying this
267 function yielded 9.4% of the tested gene-metabolite pairs as significantly associated at adjusted
268 p-value < 0.05 (**Supplementary Table S11**). Intersecting these pairs with those from
269 **Supplementary Table S9**, we obtained a list of 11,890 (2.4% of the 500,000 tested) gene-
270 metabolite pairs whose associations are not attributable to co-expression in certain cell types or
271 spatial domains (**Supplementary Table S12, Figure 4D**).

272 *Comparison with state-of-the-art tools:* The only available functionality for detecting gene-
273 metabolite relationships in SM+ST data is the correlation routine in SpaMTP and SpatialMETA
274 based on Pearson correlation coefficient (PCC). To assess the difference between this and the
275 SCI-based approach of MANTIS, we calculated PCC between every gene-metabolite pair,
276 observing ~385K pairs (77% of the 500K tested) to be significant at FDR < 0.05 (**Supplementary**
277 **Table S13**). This is a far greater number than the 61,552 (12.3% of 500K tested) pairs reported
278 by the compute_genemet_sci routine (above). The wide gap is attributable to the use of a spatial
279 cross-correlation score (SCI) in place of the standard correlation, and statistical testing using a
280 specially constructed null distribution. We present representative examples to illustrate the impact
281 of these two methodological choices, with **Supplementary Figure S10** showing a gene-
282 metabolite pair that is significant by PCC (FDR < 0.05) but not by SCI and **Supplementary Figure**
283 **S11** shows a gene-metabolite pair where the SCI is significant (FDR < 0.05) if tested through
284 standard randomization but not via autocorrelation-preserving randomization (FDR > 0.1). Visual
285 inspection of the feature maps reveals no clear spatial covariation in either example. These
286 analyses and examples show that the compute_genemet_sci provides a more conservative test
287 of gene-metabolite spatial covariation compared to the standard correlation approach.

288 **Applications to human multi-omics data sets**

289 We applied MANTIS to a published human lung-cancer multi-omics dataset [1] that contains 190
290 metabolites and 22 proteomic markers measured at single-cell resolution across ~920 cells
291 (**Figure 5A**). We did not have ready access to cell type or regional annotations, so we directly
292 used the "compute_genemet_sci" function, treating protein abundances as a measure of gene
293 expression. This revealed 547 significant protein-metabolite pairs (FDR < 0.05), i.e., ~13% of
294 4180 tested pairs (**Figure 5B**, **Supplementary Table S14**). T-cell markers CD45RA and CD45RO
295 were among the most metabolite-associated proteins, suggesting that immune infiltration and
296 activation states are closely tied to spatial metabolic structure [30,31]. E-cadherin, an epithelial

297 cell marker, is also highly metabolite-associated, potentially reflecting the distinct metabolic
298 programs linked to EMT and/or tumor-stroma boundaries [32,33]. The many associations of
299 Histone H3 may reflect simply a higher metabolite abundance in cell-dense regions or a more
300 specific link to chromatin state. In **Figures 5C,D**, we see that the metabolite glycine (m/z 74.0)
301 has significantly higher expression in tumor regions, defined by ECadherin and PanKeratin
302 positive cells. This is consistent with established mechanisms whereby tumor cells up-regulate
303 glycolysis and one-carbon (serine/glycine) metabolism to support rapid growth, nucleotide and
304 protein synthesis, resulting in elevated glycine accumulation in tumor zones [34]. As a second
305 example, CD45RA (T cell marker) expression positively correlates with ceramide-phosphocholine
306 (m/z 101.961) abundance (**Figures 5E,F**). This may reflect local remodeling of sphingolipids in
307 immune infiltration, consistent with known roles of sphingolipids their ceramide products in T cell
308 activation and maintenance [35].

309 We next applied MANTIS to a human striatum spatial multi-omics sample [5] of postmortem tissue
310 from a 94-year-old Parkinson's disease patient, examining associations between the top 1000
311 highly variable genes and all 1538 detected metabolites (**Figure 5G**). The "compute_genemet_sci"
312 function revealed extensive gene-metabolite associations, with 25% of the tested pairs being
313 significant at FDR < 0.05 (**Supplementary Table S15**). Many of these associations involved the
314 metabolite dopamine (m/z 674.281, **Figure 5I**), including a strong negative spatial correlation with
315 *Mbp* (Myelin Basic Protein) and *Plp1* (Proteolipid Protein 1) (**Figure 5H**), which are major myelin-
316 sheath proteins produced by oligodendrocytes and are enriched in white matter but largely absent
317 in grey matter. Regions with high dopamine signal correspond to the caudate nucleus, a gray-
318 matter region rich in neurons, whereas regions with low dopamine signal show white-matter
319 characteristics with higher *Mbp* and *Plp1* expression. This spatial segregation likely explains the
320 observed negative correlation, which did not remain significant (SCI = 0.08) after regressing out
321 the spatial domain.

322 To test whether other reported associations were driven by spatial domains, we used the
323 "compute_SPC-SD" function that accounts for spatial domains Caudate nucleus and Claustrum
324 (**Figure 5K**). The SCI values fell sharply (**Figure 5L**) and only 5,079 (0.3%) of gene–metabolite
325 pairs remained significant (FDR < 0.05) (**Supplementary Table S16**), indicating that most of the
326 originally observed co-localizations are attributable to spatial domains. **Figure 5J** provides an
327 overview of the gene-metabolite pairs that remain significant, showing that dopamine (m/z =
328 674.2805) has among the most number of associations even after accounting for domain effects.
329 We noticed a few dense biclusters (modules) of gene–metabolite pairs in the map, e.g., one

330 shown by green box in Fig. 5J. **Figures 5M,N** visualize a representative pair from this module
331 (*Cntn2* (Contactin-2) paired with GABA-H2O (m/z 353.1652), FDR = 0.031), showing a
332 significantly negative spatial correlation in the top right rather than following the spatial domain
333 pattern of Fig. 5K. *Cntn2* is associated with myelinated axons [36] and oligodendroglial structures,
334 whereas GABA-H2O reflects GABAergic neuronal activity [37]. Because fiber-enriched,
335 oligodendroglial compartments and GABAergic neuron–rich compartments rarely coincide in the
336 same tissue locations, *Cntn2* and GABA-H2O exhibit an inverse spatial relationship.

337

338 **DISCUSSION**

339 There is rapidly growing interest in tools for analysis of spatial metabolomics (SM) data [2-4],
340 especially those that analyze matched SM and ST data on the same biological sample [1,8,11-
341 13,38]. MANTIS is a computational toolkit for analysis of such co-registered SM+ST data sets at
342 cell/spot resolution, along with cell type annotations and regional delineations, if available. Its
343 main functionalities include identifying metabolites with statistically interesting spatial patterns and
344 discovering gene-metabolite relationships via correlation analysis.

345 MANTIS adopts new statistical approaches designed to handle confounders and control false
346 positive errors. For example, it uses spatial cross-correlation to discover gene-metabolite
347 relationships, in contrast to standard correlation measures used by SpaMTP and SpatialMETA.
348 Prior work makes a compelling case for use of spatial statistics when quantifying associations in
349 spatial data, omics [39] and otherwise [40], and our empirical comparisons show that a large
350 number of associations reported by non-spatial analysis are not significant when using spatial
351 scores (Supplementary Table S17). Furthermore, MANTIS' special method for significance testing,
352 based on autocorrelation-preserving permutations, distinguishes it from simple correlation
353 measures implemented in existing tools, and provides a stringent control on false positive errors.

354 Another important methodological contribution of MANTIS is its handling of potential confounders
355 when testing associations. A gene and metabolite may be spatially co-expressed because they
356 are individually associated with the same spatial region or cell types. Our decision was to report
357 such associations separately from those pairs whose spatial covariance cannot be explained by
358 simple regional or cell type associations of either molecule. Thus, the "compute_SPC-CT" function
359 reports significant gene-metabolite pairs after removing the effect of spatial domains or cell types
360 on a pair's association. Similar handling of confounders is also featured in MANTIS the function

361 “spatvar_met” for discovering “spatially variable metabolites” (SVM). This function may be
362 employed with varying length scale parameter l to reveal patterns at different spatial scales.

363 The goals and functionalities of MANTIS are different from an emerging genre of tools such as
364 MISO [1,13,41] that combine single-cell metabolomics and gene expression data to obtain a
365 unified multi-omic representation of each cell, which is then used to discover spatial
366 neighborhoods with characteristic molecular signatures. However, these tools do not directly
367 address the explicit recovery of gene-metabolite relationships in a spatial omics setting.

368 There are several methodological aspects of the MANTIS toolkit that need future research and
369 improvement. For instance, the SCI score is better suited for spot-level data where each spot is
370 assigned cell type proportions (real valued), but may not be ideal when cell type annotations are
371 categorical or binary. For such situations, a spatial covariance test tailored for discrete variables,
372 such as the Proximal Pairs test [42], may be more appropriate. Another direction of future work
373 will be to improve how spatial autocorrelation is accounted for when quantifying bivariate spatial
374 correlation. Currently this is addressed by estimating an empirical null distribution of the SCI score
375 that conditions on metabolite autocorrelation levels, but future versions may explore a modified
376 version of the SCI score that is normalized by the autocorrelation levels of both variables.

377 MANTIS is intended to add to the developing ecosystem of tools for SM+ST data analysis. These
378 tools offer several useful functionalities, including multi-omics data alignment and mapping,
379 clustering and discovery of markers, and pathway analysis. Our goal was to improve the rigor and
380 precision of spatial pattern discovery and association analysis, and we believe MANTIS presents
381 an important step forward in this direction.

382

383 **METHODS**

384 **Data source and preprocessing**

385 We used the dataset from Vicari et al. [5], which provides matched ST (Visium) and SM (MALDI-
386 MSI) data from mouse and human striatum. Gene expression and metabolite abundance
387 measurements are already aligned to the same coordinate system, at the resolution of Visium
388 spots. Also included are annotations of spatial domains and cell type proportions (obtained using
389 Stereoscope [15]) at each spot. We analyzed one mouse striatum tissue section (V11L12-038),
390 with data on 19,126 genes and 2,754 metabolites across 2,493 spatial spots. Raw transcript
391 counts were normalized by library size and log-transformed. We applied MAGIC [43] to reduce

technical noise. For each gene, observations outside the interval $[Q_1 - 1.5\,\mathrm{IQR},\ Q_3 + 1.5\,\mathrm{IQR}]$ were classified as outliers and corrected by local averaging. For metabolomics data, intensities were median-normalized, log-transformed, and corrected for outliers in the same way. Then, we selected the top 1,000 highly variable genes and 500 highly variable metabolites (HVMs) for downstream analysis. The human sample (postmortem striatal tissue from a Parkinson's disease patient) includes 3,962 spots, 14,826 genes and 1,538 metabolites. Preprocessing was as above, and we selected 1000 highly variable genes (HVGs) and all 1538 metabolites for analysis. We also analyzed a lung cancer dataset from Hu et al. [1], where levels of 190 metabolites and 22 proteins (cell type markers) are measured in 920 cells. We used the gene–expression matrix accompanying the publication without further normalization.

Regional (spatial domain) enrichment analysis: regional_met function

KL divergence. To identify metabolites enriched in specific spatial domains (SD), we constructed regional abundance profiles for each metabolite. For metabolite m, we first computed its mean abundance in region r, and normalized these across regions to form a probability vector $\{p_{rm}\}_r$. A background profile q_r was similarly defined by averaging across all metabolites. We then quantified the divergence between each metabolite's profile p_m and the background profile q using Kullback–Leibler (KL) divergence: $D_{\mathrm{KL}} = \sum_{r=1}^{k} p_{rm} \log \frac{p_{rm}}{q_r}$, and the contribution of each region was given by the term $p_{rm} \log \frac{p_{rm}}{q_r}$.

Significance of KL divergence score. To construct the null model, we generate one "null spatial map" X_m^{null} for each metabolite m – a randomized version of the metabolite's spatial distribution where the marginal distribution as well as spatial autocorrelation are preserved – using the sampling procedure described below. We compute the KL divergence score based on X_m^{null} as for the real data. Collecting these scores for all metabolites we obtain an empirical null distribution $\{KL(X_m^{null})\}_m$. We then fit an analytical distribution to this empirical null, through maximum likelihood. For this, we screen a panel of candidate distribution families, fitted using their standard parameterizations as implemented in scipy.stats and choose the best fit. We evaluate significance for each metabolite m by comparing its observed KL divergence to the fitted CDF and obtaining one-sided p-value, which is then corrected for multiple testing across metabolites with Benjamini–Hochberg FDR procedure.

Simulated-annealing (SA) sampler of null spatial maps

We generate null spatial map of a metabolite by repeated swaps of its abundance between pairs of objects (spots/cells), which preserves the marginal distribution. The procedure strives to retain the autocorrelation level of the original spatial map, as follows. A weighted spatial graph G is first built from object coordinates, with adjacent objects being connected by edges. Let i and j index two nodes (objects) in the graph, we write $(i,j) \in E$ if the two spots are neighbors (Euclidean distance $d_{ij} < l$), l being a user-specified length scale with default value $l = 4d_{min}$, where d_{min} is the shortest distance between spots/cells. Edge (i,j) is assigned a weight w_{ij} as follows:

$$w_{ij} = \begin{cases} 1 & d_{ij} \leq l \\ e^{\left(\frac{-d_{ij}^2}{2l^2}\right)} & d_{ij} \geq l \end{cases}$$

Let $X_m = \{X_{im}\}_i$ denote the spatial map of a metabolite m across all spots i. Let $M(X)$ be the autocorrelation of any map X, i.e., $M(X) = \sum_{(i,j)\in E} w_{ij}\, X_i\, X_j$. The sampler explores randomized maps $X_m^r = \{X_{im}^r\}_i$ such that the autocorrelation $M(X_m^r)$ remains close to the autocorrelation $M_0 = M(X_m)$ of the original map X_m. In each iteration, the sampler evaluates a proposed swap of metabolite levels between two randomly chosen nodes i,j. Denoting the current spatial map by X^{curr} and the autocorrelation of the update map after the swap by M_{new}, the proposal is accepted with probability $\alpha = \min[\,1, \exp(-(\|M_{new} - M_0\|_1 - \|M(X^{curr}) - M_0\|_1)/T)]$, where T is the current "temperature" of the SA sampler.

See Supplementary Methods S1 for additional details of implementation of the SA sampler of randomly permuted spatial maps.

Outputs: Each SA run returns the final permutation π of node indices obtained by the sequence of swaps performed.

Spatial Cross-Correlation Index (SCI)

Definition. For vectors $\boldsymbol{x}, \boldsymbol{y} \in R^N$ and a spatial weight matrix $W = [w_{ij}]$, we define the Spatial Cross-Correlation Index (SCI) as a bivariate Moran-type correlation: $\mathrm{SCI}(x,y) = \frac{1}{N}\, \tilde{x}^T W\, \tilde{y}$, where $\tilde{x}_i = \frac{x_i - \bar{x}}{s_x}$, $\tilde{y}_i = \frac{y_i - \bar{y}}{s_y}$, and N is the number of objects (spots/cells). Here, $\bar{x}$, $\bar{y}$ denote sample means and s_x, s_y are sample standard deviations.

Uses of SCI: In all uses of SCI, y is the metabolite intensity. For celltype_met, x is cell-type proportion (for spot-based data) or a binary indicator of cell type (for single-cell data). For "compute_genemet_sci", x is gene expression.

Weight matrix W is defined as:

$$w_{ij} = \begin{cases} 1, & i = j \\ e^{\left(-d_{ij}^2/(2l^2)\right)}, & i \neq j \end{cases}.$$

We row-normalize the matrix, i.e., $\sum_j w_{ij} = 1$. The default value of l is d_{min}.

Significance testing. We use sampler-based empirical nulls (described above) to control for metabolite spatial autocorrelation when assessing SCI significance. (See **Supplementary Methods S2**.)

Detection of spatially variable metabolites (SVM)

Moran's I: We quantify spatial autocorrelation of a metabolite's distribution $\{X_{im}\}_i$ using a revised Moran's I. Let W= $[w_{ij}]$ with $w_{ii} = 0$ and $w_{ij} = e^{\left(-d_{ij}^2/(2l^2)\right)}$ for $i \neq j$. We row-normalize W such that $\sum_{j=1}^{n} w_{ij} = 1$ for all i. Under this row-normalized formulation, Moran's I for metabolite m is given by $I(m) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(X_{im}-\overline{X_m})(X_{jm}-\overline{X_m})}{\sum_{i=1}^{n}(X_{im}-\overline{X_m})^2}$.

Control for cell type associations: We first regress the metabolite abundance against cell type. Specifically, we fit a LASSO model to $\{X_{im}\}_i$ using cell type(s) of object i as covariates and select a sparse subset of cell types informative of X_{im}. Then we refit an OLS model on the selected cell type covariates and calculate Moran's I as above on the residuals $r_{im}^{(CT)}$ of the metabolite, to obtain $I^{CT}(m)$.

Control for spatial domain association. We encode each objefffct's region (spatial domain) annotation with a one-hot feature vector, which is used to predict metabolite abundance via an OLS model and obtain residuals $r_{im}^{(SD)}$, on which we then calculate Moran's I as above to obtain $I^{SD}(m)$.

We designate a metabolite as SVM if $\min(I^{CT}(m), I^{SD}(m)) \geq \tau$ for some threshold τ.

Spatial Partial Correlation (SCI with adjustment for cell type or spatial domain)

Adjustment for cell type/spatial domain. For both the gene expression matrix and metabolite abundance matrix, we apply the same adjustments as above: spatial domains by OLS and cell types by LASSO with post-selection OLS refit. This yields residuals $r_{im}^{(SD)}$, $r_{ig}^{(SD)}$ of metabolite abundance and gene expression respectively after accounting for effects of spatial domains, and

similarly the residuals $r_{im}^{(CT)}$, $r_{ig}^{(CT)}$ after accounting for cell types. We define "spatial partial correlation" (SPC) as the SCI score between residuals, i.e. SPC-CT$(g,m) = SCI(\{r_{im}^{(CT)}\}_i, \{r_{ig}^{(CT)}\}_i)$ and SPC-SD$(g,m) = SCI(\{r_{im}^{(SD)}\}_i, \{r_{ig}^{(SD)}\}_i)$

Null model and statistical testing of SPC: We use the SA sampler to obtain a randomized version X^r of each metabolite's spatial map X, and the full collection of randomized spatial maps is then subjected to SPC-SD or SPC-CT calculations, thus resulting in null distributions and empirical p-value (as well as BH-FDR q-value) calculations as above.

KEY POINTS

- MANTIS is an analytics toolkit for paired spatial metabolomics + spatial transcriptomics data, supporting discovery of metabolic spatial patterns and gene-metabolite relationships.
- MANTIS quantifies gene-metabolite spatial associations using a spatially aware correlation score and performs autocorrelation-preserving permutation testing to provide rigorous statistical significance and reduce false positives.
- MANTIS disentangles confounding effects of spatial domains and cell-type composition using spatial partial correlation, separating domain/cell-type-driven co-localization from more specific associations.

ACKNOWLEDGMENTS

We thank Dr. Ahmet Coskun and Efe Ozturk for helpful discussions and sharing data.

DATA AVAILABILITY

The datasets used in this study were obtained from publicly available sources and are listed in Supplementary Table S18. MANTIS is implemented in Python and distributed as open-source software at https://github.com/yuhaotuo/MANTIS

.

REFERENCES

1. Hu T, Allam M, Cai S, et al. Single-cell spatial metabolomics with cell-type specific protein profiling for tissue systems biology. *Nat Commun*. Dec 13 2023;14(1):8260. doi:10.1038/s41467-023-43917-5
2. Clarke HA, Ma X, Shedlock CJ, et al. Spatial mapping of the brain metabolome lipidome and glycome. *Nat Commun*. May 12 2025;16(1):4373. doi:10.1038/s41467-025-59487-7
3. Panda P, Ferreira CR, Cooper BR, Schaser AJ, Aryal UK. Multiplatform Lipid Analysis of the Brain of Aging Mice by Mass Spectrometry. *J Proteome Res*. Mar 7 2025;24(3):1077–1091. doi:10.1021/acs.jproteome.4c00688
4. Michno W, Bowman A, Jha D, et al. Spatial Neurolipidomics at the Single Amyloid-beta Plaque Level in Postmortem Human Alzheimer's Disease Brain. *ACS Chem Neurosci*. Feb 21 2024;15(4):877–888. doi:10.1021/acschemneuro.4c00006
5. Vicari M, Mirzazadeh R, Nilsson A, et al. Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nature Biotechnology*. 2024/07/01 2024;42(7):1046–1050. doi:10.1038/s41587-023-01937-y
6. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications*. 2022/04/01 2022;13(1):1739. doi:10.1038/s41467-022-29439-6
7. Wang L, Cui CY, Lee CT, et al. Spatial transcriptomics of the aging mouse brain reveals origins of inflammation in the white matter. *Nat Commun*. Apr 4 2025;16(1):3231. doi:10.1038/s41467-025-58466-2
8. Ji L, Wang D, Zhuo G, et al. Spatial Metabolomics and Transcriptomics Reveal Metabolic Reprogramming and Cellular Interactions in Nasopharyngeal Carcinoma with High PD-1 Expression and Therapeutic Response. *Theranostics*. 2025;15(7):3035–3054. doi:10.7150/thno.102822
9. Li H, Li D, Ledru N, et al. Transcriptomic, epigenomic, and spatial metabolomic cell profiling redefines regional human kidney anatomy. *Cell Metabolism*. 2024;36(5):1105–1125.e10. doi:10.1016/j.cmet.2024.02.015
10. Bemis KA, Föll MC, Guo D, Lakkimsetty SS, Vitek O. Cardinal v.3: a versatile open-source software for mass spectrometry imaging analysis. *Nature Methods*. 2023/12/01 2023;20(12):1883–1886. doi:10.1038/s41592-023-02070-z
11. Williams EC, Franzén L, Lindvall MO, et al. Spatially resolved integrative analysis of transcriptomic and metabolomic changes in tissue injury studies. *bioRxiv*. 2025:2025.02.26.640381. doi:10.1101/2025.02.26.640381
12. Causer A, Lu T, Fitzgerald C, et al. SpaMTP: Integrative Statistical Analysis and Visualisation of Spatial Metabolomics and Transcriptomics data. *bioRxiv*. 2024:2024.10.31.621429. doi:10.1101/2024.10.31.621429
13. Tian R, Xue Z, Chen Y, et al. Integrating cross-sample and cross-modal data for spatial transcriptomics and metabolomics with SpatialMETA. *Nature Communications*. 2025/10/06 2025;16(1):8855. doi:10.1038/s41467-025-63915-z

14. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*. 2024/02/01 2024;42(2):293–304. doi:10.1038/s41587-023-01767-y

15. Andersson A, Bergenstråhle J, Asp M, et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Communications Biology*. 2020/10/09 2020;3(1):565. doi:10.1038/s42003-020-01247-y

16. Maynard KR, Collado-Torres L, Weber LM, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience*. 2021/03/01 2021;24(3):425–436. doi:10.1038/s41593-020-00787-0

17. Ivanisevic J, Epstein AA, Kurczy ME, et al. Brain region mapping using global metabolomics. *Chem Biol*. Nov 20 2014;21(11):1575–84. doi:10.1016/j.chembiol.2014.09.016

18. Osetrova M, Tkachev A, Mair W, et al. Lipidome atlas of the adult human brain. *Nat Commun*. May 25 2024;15(1):4455. doi:10.1038/s41467-024-48734-y

19. Lee CY, Soliman H, Geraghty BJ, et al. Lactate topography of the human brain using hyperpolarized (13)C-MRI. *Neuroimage*. Jan 1 2020;204:116202. doi:10.1016/j.neuroimage.2019.116202

20. Wang G, Heijs B, Kostidis S, et al. Analyzing cell-type-specific dynamics of metabolism in kidney repair. *Nat Metab*. Sep 2022;4(9):1109–1118. doi:10.1038/s42255-022-00615-8

21. Ding J, Ji J, Rabow Z, et al. A metabolome atlas of the aging mouse brain. *Nature Communications*. 2021/10/15 2021;12(1):6021. doi:10.1038/s41467-021-26310-y

22. Yang Q, Cai Y, Guo S, Wang Z, Qiu S, Zhang A. Localization analysis of metabolites from complex biological samples-recent analytical technique of mass spectrometry imaging. *Front Mol Biosci*. 2023;10:1169449. doi:10.3389/fmolb.2023.1169449

23. Oe Y, Baba O, Ashida H, Nakamura KC, Hirase H. Glycogen distribution in the microwave-fixed mouse brain reveals heterogeneous astrocytic patterns. *Glia*. Sep 2016;64(9):1532–45. doi:10.1002/glia.23020

24. Zhang YM, Qi YB, Gao YN, et al. Astrocyte metabolism and signaling pathways in the CNS. *Front Neurosci*. 2023;17:1217451. doi:10.3389/fnins.2023.1217451

25. Chen Y. A New Methodology of Spatial Cross-Correlation Analysis. *PLOS ONE*. 2015;10(5):e0126158. doi:10.1371/journal.pone.0126158

26. Lee S-I. Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *Journal of Geographical Systems*. 2001/12/01 2001;3(4):369–385. doi:10.1007/s101090100064

27. Svensson V, Teichmann SA, Stegle O. SpatialDE: identification of spatially variable genes. *Nat Methods*. May 2018;15(5):343–346. doi:10.1038/nmeth.4636

28. 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*. Jun 21 2021;22(1):184. doi:10.1186/s13059-021-02404-0

29. MORAN PAP. NOTES ON CONTINUOUS STOCHASTIC PHENOMENA. *Biometrika*. 1950;37(1-2):17–23. doi:10.1093/biomet/37.1-2.17

30. Munford H, Dimeloe S. Intrinsic and Extrinsic Determinants of T Cell Metabolism in Health and Disease. Review. *Frontiers in Molecular Biosciences*. 2019–October–25 2019;Volume 6 - 2019doi:10.3389/fmolb.2019.00118

31. Corrado M, Pearce EL. Targeting memory T cell metabolism to improve immunity. *The Journal of Clinical Investigation*. 01/04/ 2022;132(1)doi:10.1172/JCI148546

32. Sciacovelli M, Frezza C. Metabolic reprogramming and epithelial-to-mesenchymal transition in cancer. *Febs j*. Oct 2017;284(19):3132–3144. doi:10.1111/febs.14090

33. Wei Q, Qian Y, Yu J, Wong CC. Metabolic rewiring in the promotion of cancer metastasis: mechanisms and therapeutic implications. *Oncogene*. 2020/09/01 2020;39(39):6139–6156. doi:10.1038/s41388-020-01432-7

34. Sun W, Liu R, Gao X, et al. Targeting serine-glycine-one-carbon metabolism as a vulnerability in cancers. *Biomarker Research*. 2023/05/05 2023;11(1):48. doi:10.1186/s40364-023-00487-4

35. Hannun YA, Obeid LM. Principles of bioactive lipid signalling: lessons from sphingolipids. *Nat Rev Mol Cell Biol*. Feb 2008;9(2):139–50. doi:10.1038/nrm2329

36. Zoupi L, Savvaki M, Kalemaki K, Kalafatakis I, Sidiropoulou K, Karagogeos D. The function of contactin-2/TAG-1 in oligodendrocytes in health and demyelinating pathology. *Glia*. Mar 2018;66(3):576–591. doi:10.1002/glia.23266

37. Tepper JM, Tecuapetla F, Koos T, Ibanez-Sandoval O. Heterogeneity and Diversity of Striatal GABAergic Interneurons. Review. *Frontiers in Neuroanatomy*. 2010–December–29 2010;Volume 4 - 2010doi:10.3389/fnana.2010.00150

38. Kriel J, Moffet JJD, Lu T, et al. An integrative spatial multi-omic workflow for unified analysis of tumor tissue. *bioRxiv*. 2024:2024.10.15.618574. doi:10.1101/2024.10.15.618574

39. Li Z, Wang T, Liu P, Huang Y. SpatialDM for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns. *Nature Communications*. 2023/07/06 2023;14(1):3995. doi:10.1038/s41467-023-39608-w

40. Chen JG, Chávez-Fuentes JC, O'Brien M, et al. Giotto Suite: a multiscale and technology-agnostic spatial multiomics analysis ecosystem. *Nature Methods*. 2025/10/01 2025;22(10):2052–2064. doi:10.1038/s41592-025-02817-w

41. Coleman K, Schroeder A, Loth M, et al. Resolving tissue complexity by multimodal spatial omics modeling with MISO. *Nature Methods*. 2025/03/01 2025;22(3):530–538. doi:10.1038/s41592-024-02574-2

42. Kumar A, Schrader AW, Aggarwal B, et al. Intracellular spatial transcriptomic analysis toolkit (InSTAnT). *Nature Communications*. 2024/09/06 2024;15(1):7794. doi:10.1038/s41467-024-49457-w

43. van Dijk D, Sharma R, Nainys J, et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell*. 2018;174(3):716–729.e27. doi:10.1016/j.cell.2018.05.061

Figure Legends

Figure 1. Overview of MANTIS. (A) MANTIS analyzes spatial multi-omics data profiling metabolite intensities and gene expression levels for each cell or spot in a tissue section. The input is assumed to be metabolomics profiles and transcriptomics profiles of cells/spots along with their spatial locations. Optionally, the cells/spots may be annotated with their spatial domains (outlined by dashed lines) and/or cell types (colored circles), to be used to decompose sources of spatial variation. **(B)** One set of routines in MANTIS detect metabolites with interesting spatial patterns, including associations with spatial domains (regions) or cell types, as well as non-random spatial patterns not easily attributed to regions or cell types. **(C)** Another set of routines report gene-metabolite pairs with correlated spatial patterns (gene shown in left panel, metabolite in right panel).

Figure 2. Region-associated metabolites in mouse striatum SM+ST data (VICARI). (A) Spatial domains of the striatum profiled in the data set. ACB: Nucleus accumbens, CP: Caudate putamen, CTX: Cortex, LSX: Lateral septal complex, OT: Olfactory tubercle, PAL: Pallidum, VL: Ventral striatum, aco: anterior commissure, cc: corpus callosum. **(B, C)** Examples of region-associated metabolites discovered by MANTIS. Left subpanels show metabolite abundance maps; right subpanels show the metabolite's regional abundance profile and its enrichment (log fold change relative to average of all metabolites) in specific domains. m/z 725.551 is enriched in VL (B), while m/z 822.614 shows preferential abundance in cc/aco (C). **(D)** Distribution of KL-divergence scores of metabolites, representing their statistical association with spatial domains. The observed distribution over 500 metabolites in the VICARI data (blue) is contrasted with an empirical null distribution (gray) constructed with a randomization procedure that preserves spatial autocorrelations. **(E)** Heatmap showing regional abundance profiles of select metabolites. Color intensities represent abundance of a metabolite (column) in each region (row). Shown metabolites were selected to illustrate major patterns of region-associations seen across the complete set of significant region-associated metabolites.

Figure 3. Cell-type-associated metabolites and spatially variable metabolites in VICARI data. (A) Spatial distribution of eight most frequent cell types in the mouse striatum section. TEGLU (Telencephalic Glutamatergic Neurons), MSN2 (Medium Spiny Neurons, Subtype 2), PSPEP (Pallidal/Striatal Peptidergic Neurons), OLIG (Oligodendrocytes), MSN4 (Medium Spiny Neurons, Subtype 4), MSN3 (Medium Spiny Neurons, Subtype 3), DE/ME INH (Diencephalic/Mesencephalic Inhibitory Neurons), EPEN (Ependymal Cells), and Others (other cell-types). **(B)** Counts of significant cell-type-metabolite associations reported by MANTIS, for each cell type; OLIG, SCHW, and SATG show the largest numbers of associations. **(C)** Example of significant cell-type-metabolite association: Cerebellum neurons proportion (left) and metabolite m/z 172.001 (right) show a high SCI score of 0.474. **(D)** Cell-type-associated metabolites often include region-associated metabolites, as cell types themselves exhibit regional preferences. For each cell type, the bar shows the number of metabolites associated with that cell type and the number of such metabolites that are also region-associated. The dashed line

represents the total number of region-associated metabolites (n = 66). **(E)** Effect of confounder adjustment for spatial autocorrelation statistics (Moran's I) used to detect spatially variable metabolites (SVMs). Scatter plots compare Moran's I of each metabolite's spatial distribution (unadjusted, x-axis) vs adjusted values after regressing out region (left) or cell-type (right). **(F)** Examples of metabolites retaining spatial autocorrelation signal (high Moran's I) after adjustment for region and cell type; the unadjusted vs adjusted scores for these two examples are shown in panel E. **(G)** Effect of the "length scale" parameter used in Moran's I. Scatter plot compares Moran's I at two extreme spatial scales ($l = d_{min}$ and $l = 100d_{min}$), and two examples, m/z = 866.648, m/z = 399.145 and m/z = 196.096, are highlighted to illustrate opposite trends. **(H)** Feature map of m/z = 866.637, which exhibits strong fine-grained boundaries and high Moran's I at small l but loses spatial autocorrelation at larger l. **(I-J)** Feature maps of m/z = 399.145 and m/z = 196.096, which display a broad gradient that becomes more obvious at large l.

Figure 4. Gene-metabolite spatial covariation

(A) Spatial Correlation Index (SCI) scores for all 500,000 gene-metabolite pairs in the VICARI data set (blue), compared to empirical null distribution (gray) derived by a randomization procedure that preserves autocorrelations. **(B)** Example of significant gene-metabolite spatial associations (SCI q-value 0.002). Shown are gene expression of Scn3b (left) and abundance of metabolite 769.552 m/z (right) in the striatum tissue section. **(C)** Example of gene-metabolite pair with significant spatial association (SCI q-value 0.003) even after accounting for their cell type-associations. (SCI is calculated after regressing out the effect of cell type from both variables.) **(D)** Example of a gene–metabolite pair *Cltc* and 195.092 m/z that remains significantly spatially associated even after adjusting for cell-type composition and spatial domain.

Figure 5. Gene-metabolite spatial covariation in human datasets

(A) Schematic of the single cell SM + spatial proteomics (SP) human lung cancer sample (920 cells, 22 proteins, 180 metabolites). **(B)** Number of significant protein–metabolite associations identified for each protein. **(C)** Spatial expression patterns of the tumor-associated proteins PanKeratin and E-Cadherin. **(D)** Spatial abundance of glycine (m/z 74.0135), which colocalizes with epithelial tumor marker proteins. **(E)** Spatial expression of the immune T-cell marker CD45RA. **(F)** Spatial abundance of ceramide phosphocholine (m/z 101.961), which shows strong correspondence with CD45RA. **(G)** Schematic of the ST+SM human striatum sample (3962 spots, 1000 HVGs, 1538 metabolites). **(H)** Spatial expression of oligodendrocyte-associated genes *Mbp* and *Plp1*. **(I)** Spatial distribution of dopamine (m/z 674.2805), which is strongly enriched in regions expressing *Mbp*/*Plp1*. **(J)** Heatmap of significant gene–metabolite associations identified by SPC–SD, highlighting broad associations involving dopamine as well as a more localized association involving *Cntn2* and GABA·H_2O (m/z 532.096). The green box marks the bicluster from which the *Cntn2*–m/z 532.096 example was selected. (K) Two spatial domains in this sample. Cd: caudate, Cl: claustrum. **(L)** Comparison of SCI and SPC–SD scores for all gene–metabolite pairs. **(M)** Spatial expression of *Cntn2*. **(N)** Spatial abundance of m/z 532.096, showing colocalization with *Cntn2*.

Figure 1

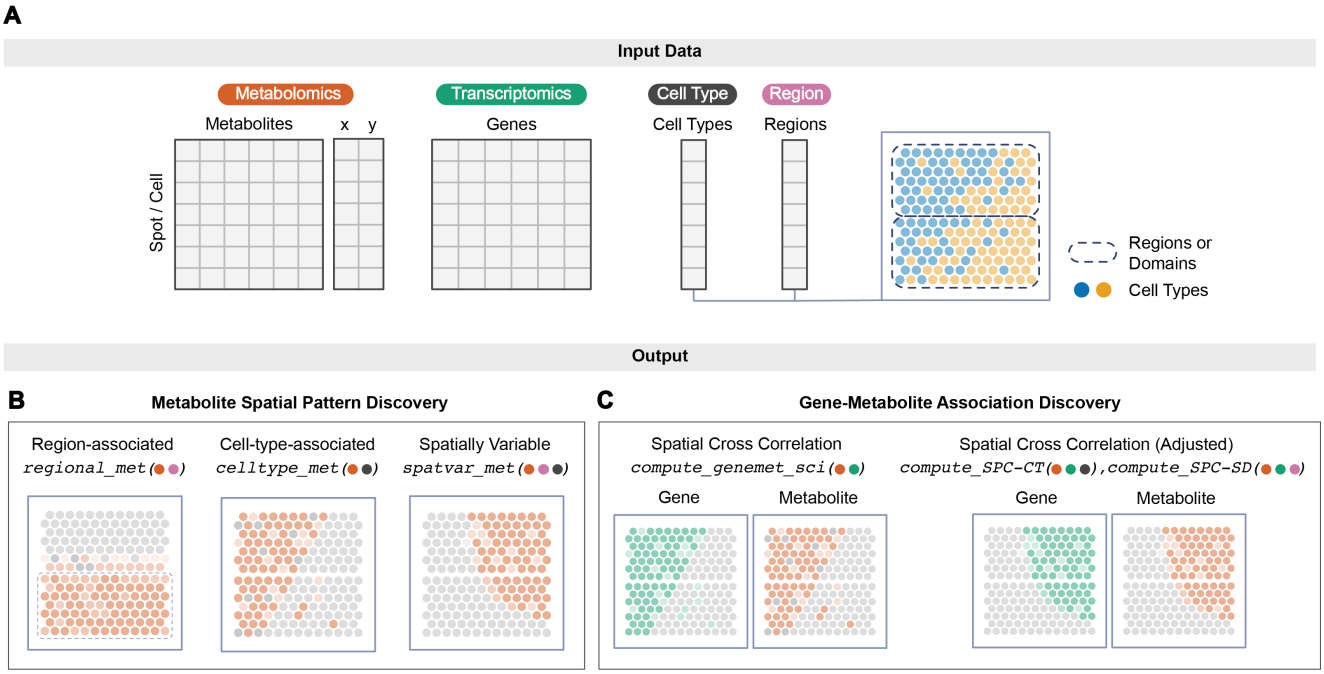

A
Input Data
Metabolomics
Metabolites
x y
Spot / Cell
Transcriptomics
Genes
Cell Type
Cell Types
Region
Regions
Regions or Domains
Cell Types
Output
B
Metabolite Spatial Pattern Discovery
Region-associated
regional_met()
Cell-type-associated
celltype_met()
Spatially Variable
spatvar_met()
C
Gene-Metabolite Association Discovery
Spatial Cross Correlation
compute_genemet_sci()
Gene
Metabolite
Spatial Cross Correlation (Adjusted)
compute_SPC-CT(),compute_SPC-SD()
Gene
Metabolite

Figure 2

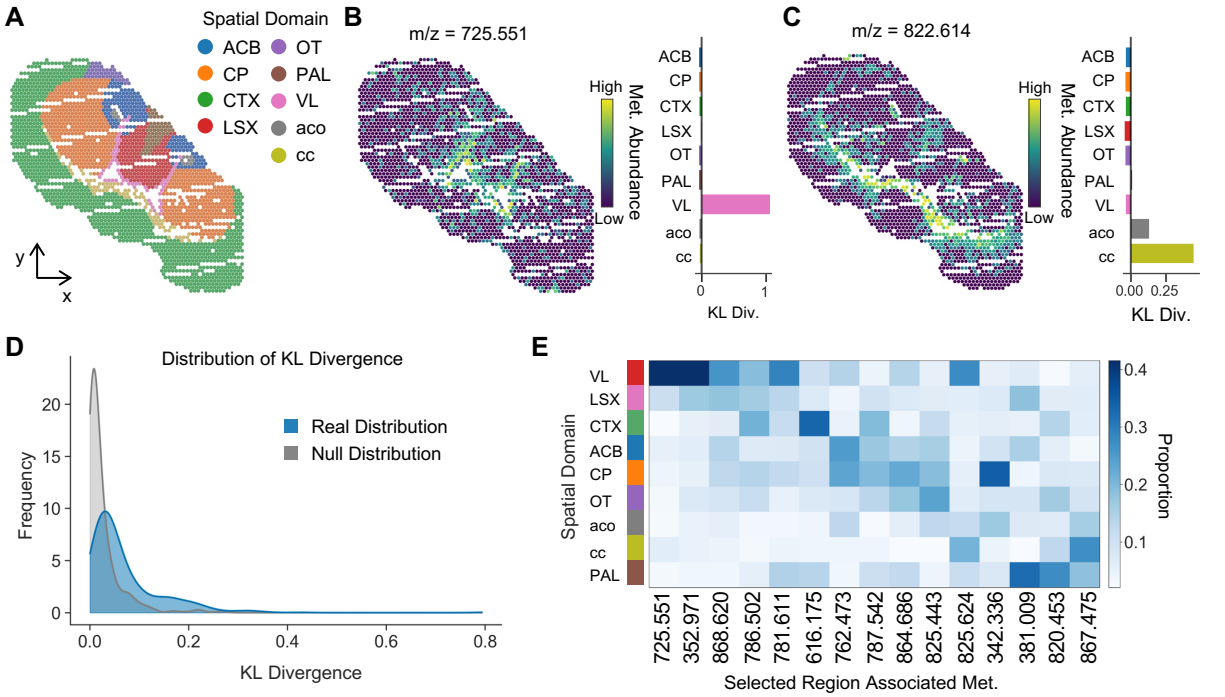

A
Spatial Domain
ACB
CP
CTX
LSX
OT
PAL
VL
aco
cc
y
x
B
m/z = 725.551
High
Low
Met. Abundance
ACB
CP
CTX
LSX
OT
PAL
VL
aco
cc
0
1
KL Div.
C
m/z = 822.614
High
Low
Met. Abundance
ACB
CP
CTX
LSX
OT
PAL
VL
aco
cc
0.00
0.25
KL Div.
D
Distribution of KL Divergence
Real Distribution
Null Distribution
Frequency
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
KL Divergence
E
Spatial Domain
VL
LSX
CTX
ACB
CP
OT
aco
cc
PAL
725.551
352.971
868.620
786.502
781.611
616.175
762.473
787.542
864.686
825.443
825.624
342.336
381.009
820.453
867.475
Selected Region Associated Met.
0.4
0.3
0.2
0.1
Proportion

Figure 3

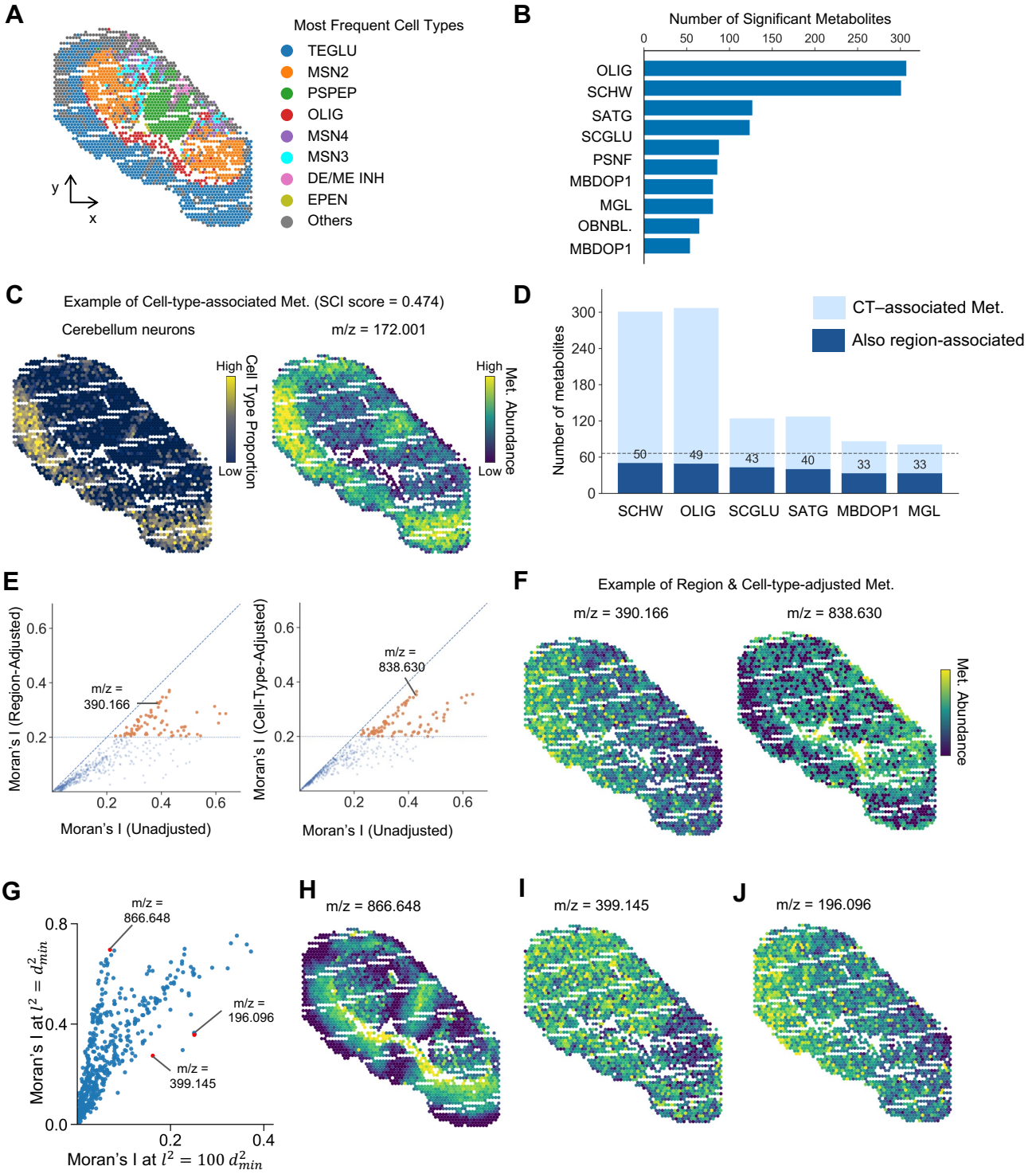

A
Most Frequent Cell Types
TEGLU
MSN2
PSPEP
OLIG
MSN4
MSN3
DE/ME INH
EPEN
Others
y
x
B
Number of Significant Metabolites
0
50
100
150
200
250
300
OLIG
SCHW
SATG
SCGLU
PSNF
MBDOP1
MGL
OBNBL.
MBDOP1
C
Example of Cell-type-associated Met. (SCI score = 0.474)
Cerebellum neurons
m/z = 172.001
High
Low
Cell Type Proportion
High
Low
Met. Abundance
D
Number of metabolites
300
240
180
120
60
0
CT–associated Met.
Also region-associated
50
49
43
40
33
33
SCHW
OLIG
SCGLU
SATG
MBDOP1
MGL
E
Moran's I (Region-Adjusted)
Moran's I (Unadjusted)
m/z = 390.166
Moran's I (Cell-Type-Adjusted)
Moran's I (Unadjusted)
m/z = 838.630
0.2
0.4
0.6
F
Example of Region & Cell-type-adjusted Met.
m/z = 390.166
m/z = 838.630
Met. Abundance
G
Moran's I at $l^2 = d^2_{min}$
Moran's I at $l^2 = 100\ d^2_{min}$
m/z = 866.648
m/z = 196.096
m/z = 399.145
0.0
0.4
0.8
0.2
H
m/z = 866.648
I
m/z = 399.145
J
m/z = 196.096

Figure 4

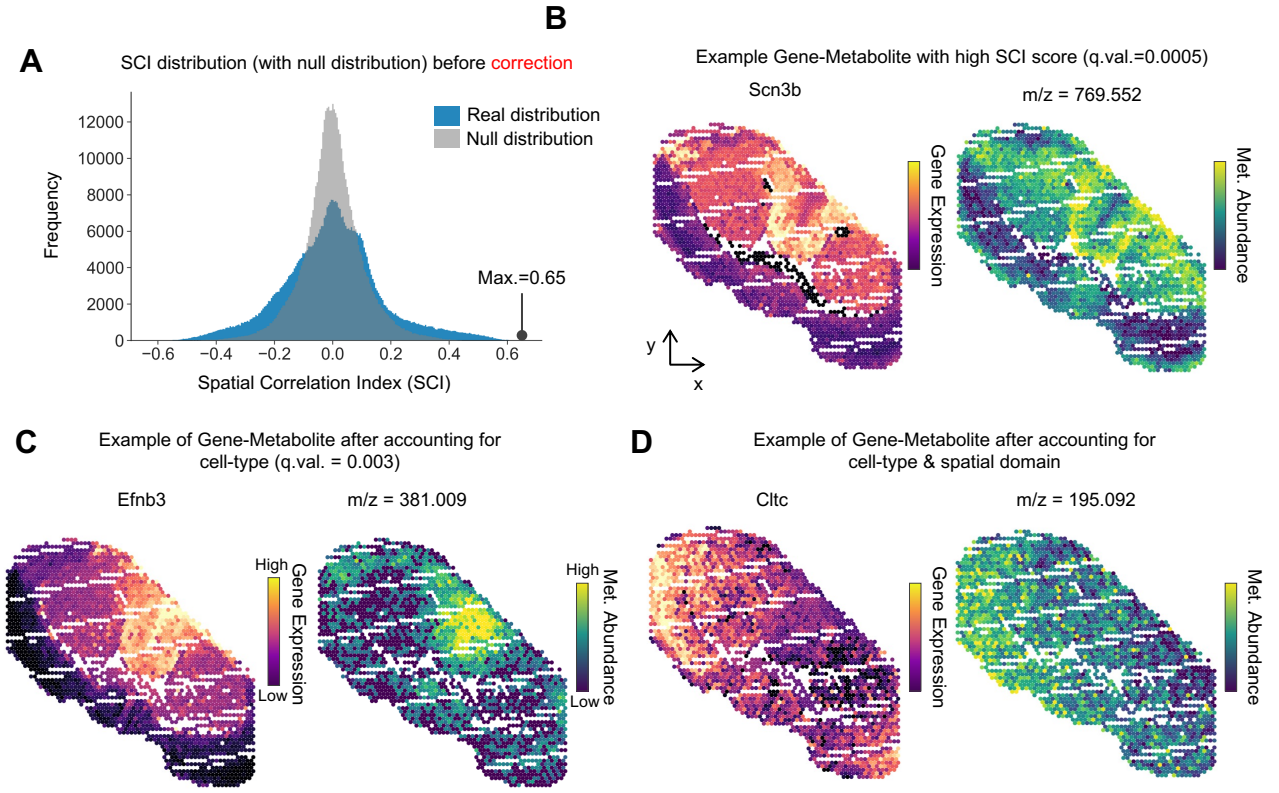
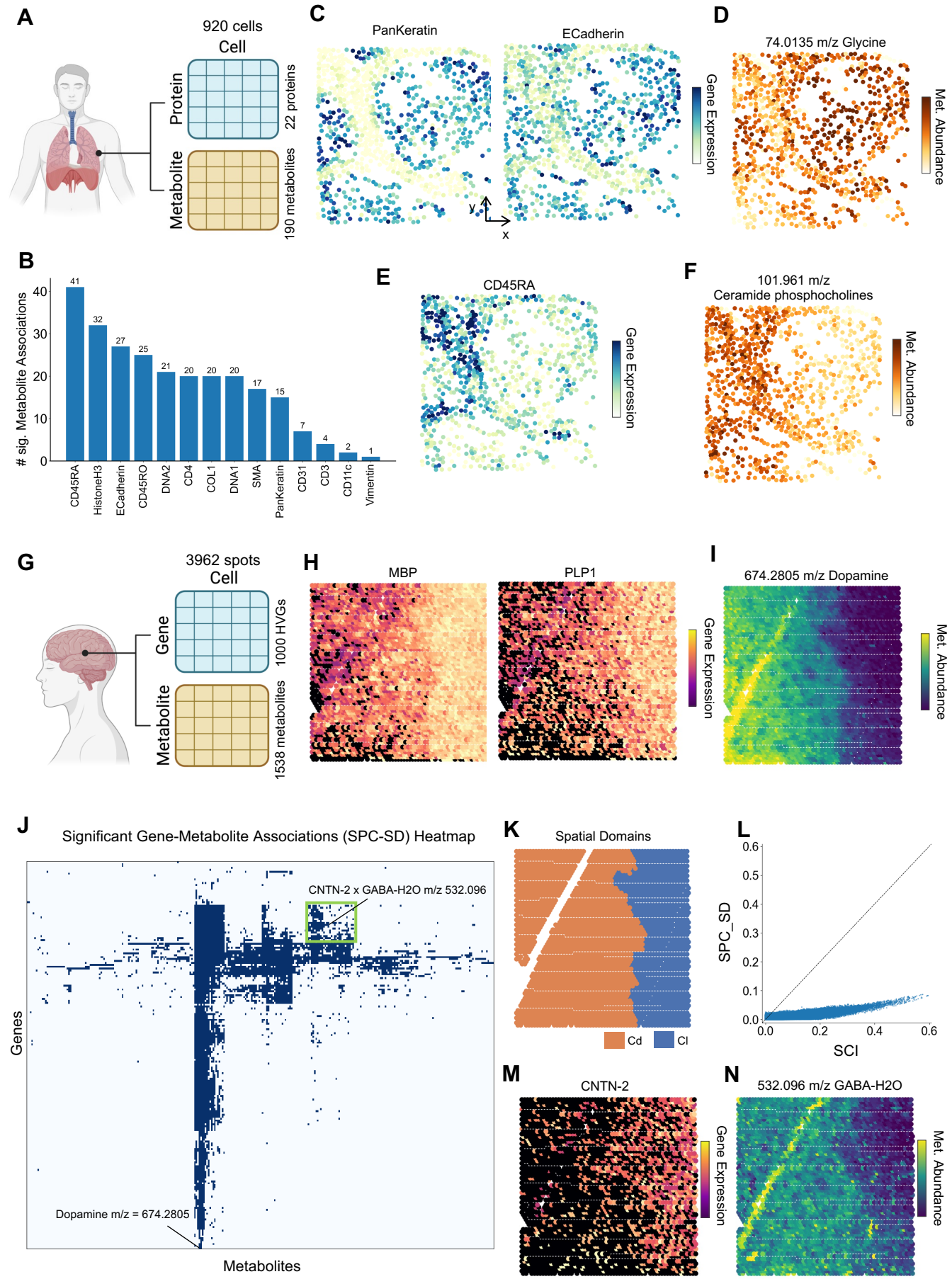

A
SCI distribution (with null distribution) before correction
Real distribution
Null distribution
Frequency
12000
10000
8000
6000
4000
2000
0
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
Max.=0.65
Spatial Correlation Index (SCI)
B
Example Gene-Metabolite with high SCI score (q.val.=0.0005)
Scn3b
m/z = 769.552
Gene Expression
Met. Abundance
y
x
C
Example of Gene-Metabolite after accounting for cell-type (q.val. = 0.003)
Efnb3
m/z = 381.009
High
Low
Gene Expression
High
Low
Met. Abundance
D
Example of Gene-Metabolite after accounting for cell-type & spatial domain
Cltc
m/z = 195.092
Gene Expression
Met. Abundance

Figure 5


A
920 cells
Cell
Protein
22 proteins
Metabolite
190 metabolites
C
PanKeratin
ECadherin
Gene Expression
y
x
D
74.0135 m/z Glycine
Met. Abundance
B
sig. Metabolite Associations
41
32
27
25
21
20
20
20
17
15
7
4
2
1
CD45RA
HistoneH3
ECadherin
CD45RO
DNA2
CD4
COL1
DNA1
SMA
PanKeratin
CD31
CD3
CD11c
Vimentin
E
CD45RA
Gene Expression
F
101.961 m/z
Ceramide phosphocholines
Met. Abundance
G
3962 spots
Cell
Gene
1000 HVGs
Metabolite
1538 metabolites
H
MBP
PLP1
Gene Expression
I
674.2805 m/z Dopamine
Met. Abundance
J
Significant Gene-Metabolite Associations (SPC-SD) Heatmap
CNTN-2 x GABA-H2O m/z 532.096
Genes
Dopamine m/z = 674.2805
Metabolites
K
Spatial Domains
Cd
Cl
L
SPC_SD
SCI
M
CNTN-2
Gene Expression
N
532.096 m/z GABA-H2O
Met. Abundance

	Feature	MANTIS	SpatialMETA	SpaMTP
Metabolite Analysis	Cell Type-associated metabolite discovery	✓	✓	
	Region-associated metabolite discovery	✓	✓	✓
	Spatially variable metabolite discovery adjusted for cell type or region	✓		
Metabolite-gene association	Statistical method	Spatial correlation	PCC	
	Adjusted for cell type- or region-associations	✓		
Spatially aware null model for significance testing (autocorrelation-controlled)		✓		
Metabolite annotation			✓	✓
Alignment			✓	✓
Clustering			✓	✓