Characterizing cell-type spatial relationships across length scales in spatially resolved omics data

Rafael dos Santos Peixoto[1,2,*], Brendan F. Miller[1,2,*], Maigan A. Brusko[3], Lyla Atta[1,2], Manjari Anant[1,4], Mark A. Atkinson[3], Todd M. Brusko[3], Clive H. Wasserfall[3], Jean Fan[1,2]

[1] Center for Computational Biology, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21211

[2] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218,

[3] Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, FL 32611

[4] Department of Neuroscience, Johns Hopkins University, Baltimore, MD 21218

* These authors contributed equally

Correspondence should be addressed to:

Jean Fan (jeanfan@jhu.edu)

Key words: spatial omics, single cell, colocalization, spatial relationships, spatial subsetting

Abstract

Spatially resolved omics technologies provide molecular profiling of cells while preserving their organization within tissues, allowing for the evaluation of cell-type spatial relationships. We developed CRAWDAD to quantify cell-type spatial relationships across length scales. We highlight the utility of such multi-scale characterization on simulated data, recapitulate expected cell-type spatial relationships in tissues such as the mouse brain and embryo, and delineate functionally relevant spatial-defined cell-type subsets in the human spleen.

Main

Spatially resolved omics technologies enable the identification of distinct cell types while preserving cellular spatial organization within tissues, providing an opportunity to evaluate cell-type organizational relationships and advance our understanding of the association between cell-type organization and tissue function. Cell-type colocalization, defined as which cell types are spatially near each other, is critical for our understanding of healthy tissue function[1] as well as aberrations in disease[2]. Such cell-type colocalizations can occur at different length scales, with some cell types colocalizing to engage in paracrine signaling and other close-range interactions at a fine, micrometer length scale[3]; others colocalizing into distinct microenvironments and functional tissue units at a more meso-scale[1]; while others colocalizing into anatomical structures at a more macro-scale (Fig 1a).

To quantitatively evaluate such cell-type spatial relationships across different length scales, we present CRAWDAD, Cell-type Relationship Analysis Workflow Done Across Distances (available at https://github.com/JEFworks-Lab/CRAWDAD and as Supplementary Software). Briefly, given cell positions and their cell-type annotations, CRAWDAD evaluates the statistical enrichment of each cell-type within a spatial neighborhood around cells of a reference cell-type compared to a series of empirical null backgrounds created by shuffling cell-type labels within spatial grids of varying sizes in order to distinguish between cell-type spatial relationships that may occur at different length scales (Online Methods). CRAWDAD uses a binomial proportion testing framework to derive empirical p-values that are corrected for multiple testing and converted to Z-scores to distinguish between significant cell-type enrichments versus depletions. CRAWDAD is amenable to any spatially resolved omics data for

52 which cell positions and cell-type annotations can be obtained. We demonstrate CRAWDAD on
53 simulated as well as real spatial omics data for tissues assayed by seqFISH[4], Slide-Seq[5], and
54 CODEX[6].
55
56 First, to highlight the utility of evaluating cell-type spatial relationships at multiple length scales,
57 we simulated 8,000 cells representing four cell types in a 1000µm-by-1000µm tissue sample
58 (Online Methods). In this simulated spatial dataset, cell-types B and C are intermixed with one
59 another, while cell-type A forms a distinct structure around cell-types B and C, that further
60 isolates them from cell-type D (Fig 1b). Applying CRAWDAD to characterize the spatial
61 relationships between each cell-type pair at multiple length scales (Online Methods), we found
62 that the neighborhood around reference cell-type C is significantly enriched with cells of cell-
63 type B from a fine length scale of ~300µm (Fig 1c-d). Likewise, we observed that the
64 neighborhood around reference cell-type B is significantly enriched with cells of cell-type C
65 from a similarly fine length scale (Fig 1c, Supp Fig 1a). Given this mutual enrichment
66 relationship, we interpret cell-types B and C to be spatially colocalized from this fine length
67 scale. In contrast, we noted cell-types A and B to be spatially colocalized from a comparably
68 coarser length scale of ~600µm (Fig 1c, e, Supp Fig 1b). Notably, an evaluation of spatial
69 colocalization that only considers the whole tissue scale may miss these distinctions regarding
70 length scales (Supp Fig 1c). Further, by comparing against locally defined backgrounds,
71 CRAWDAD's quantified cell-type spatial relationships do not change as the global background
72 is broadened to a larger tissue section inclusive of other cell-types in contrast to other methods
73 (Supp Fig 1d-f). In this manner, CRAWDAD can identify significant spatial relationships among
74 cell-type pairs that occur at different length scales.

75

76 We next applied CRAWDAD to evaluate cell-type spatial relationships in spatial omics datasets
77 of real tissues. First, we analyzed a spatial transcriptomics dataset obtained from a 4000µm-by-
78 4000µm section of a highly spatially organized tissue, the mouse cerebellum, assayed by Slide-
79 Seq[7] (Fig 2a). As Slide-Seq uses barcoded beads to profile the gene expression within tissues in a
80 spatially resolved manner, we assumed that the observed spatial position and cell-type identity
81 associated with each bead reflects the spatial position and cell-type identity of the soma of the
82 cell within the immediate vicinity of that bead. We applied CRAWDAD to evaluate 10,098 cells
83 representing 19 annotated cell-types across length scales ranging from 100µm to 1000µm (Fig
84 2b, Online Methods). Among the significantly colocalized cell-type pairs identified were
85 Purkinje neurons and Bergmann glia, (Fig 2b-d, Supp Fig 2a), which are known to interact at
86 close distances within the Purkinje cell layer of the cerebellum[8]. Likewise, the neighborhood
87 around Purkinje neurons was significantly depleted of oligodendrocytes and vice versa (Fig 2b-d,
88 Supp Fig 2a). As such, we may interpret these cell-types as being spatially separated, consistent
89 with the known spatially distinct layer structure of the cerebellum. In contrast, other spatial
90 analysis approaches do not distinguish between these expected cell-type spatial relationships
91 (Supp Fig 2d-f, Supp Note 1).

92

93 Further, without restricting to one tissue, we assessed a single-cell resolution spatial
94 transcriptomics dataset of a 5000µm-by-5000µm section of a whole developing embryo assayed
95 by seqFISH[4] (Fig 2e). We applied CRAWDAD to evaluate cell-type spatial relationships for
96 19,416 cells representing 22 cell types across 14 different length scales ranging from 100µm to
97 1000µm (Fig 2f, Online Methods). Consistent with the original publication's observations,

98 CRAWDAD identified significant spatial colocalization between the endothelium and
99 haematoendothelial progenitors as well as between the intermediate and lateral plate mesoderm
100 cells at the whole tissue scale. However, given CRAWDAD's multi-scale characterization, we
101 were also able to further identify the endothelium and haematoendothelial progenitors as being
102 significantly colocalized from a finer length scale (Fig 2g-h, Supp Fig 2b, from ~250µm) than
103 the intermediate and lateral plate mesoderm cells (Fig 2i-j, Supp Fig 2c, from ~900µm). This
104 potentially reflects the developmental and functional relationship between these cell-type pairs;
105 where endothelium and haematoendothelial progenitors are tightly intertwined in terms of both
106 their developmental origins and their functional roles in the blood circulatory system[9], the
107 intermediate and lateral plate mesoderm cells share a developmental origin but ultimately form a
108 variety of spatially and functionally distinct organ structures and appendages[10]. Again, such
109 differences between cell-type spatial relationships are difficult to discern using on other spatial
110 analysis methods (Supp Fig 2g-l). In this manner, CRAWDAD can quantitatively delineate cell-
111 type spatial relationships across multiple length scales.
112
113 Finally, having demonstrated that CRAWDAD is able to recapitulate expected cell-type spatial
114 relationships in both simulated and real spatial omics data of tissues, we sought to apply
115 CRAWDAD to characterize cell-type spatial relationships in spatial omics data of the human
116 spleen. The spleen is a highly structured organ where cell types interact to filter blood and
117 initiate immune responses. Delineating the spatial organization of cell types in the spleen can
118 provide insights into how these different cellular populations may achieve such diverse
119 immunologic functions. Therefore, we focused on six single-cell resolution spatial proteomics
120 datasets of the human spleen from three individuals assayed by CODEX as part of The Human

121 BioMolecular Atlas Program (HuBMAP)[9,11]. We first performed graph-based clustering for one
122 representative 9400µm by 9100µm section of a human spleen with 154,446 cells to identify 12
123 cell types based on 28 protein markers (Figure 3a-c, Online Methods). Applying CRAWDAD,
124 we observe cell-type spatial relationships to broadly recapitulate known splenic architecture with
125 cell-types primarily colocalizing at a coarse length scale into two major compartments defined
126 by follicle B cells and red pulp B cells corresponding to the white pulp (WP) and red pulp (RP),
127 respectively[12] (Fig 3d). Consistent with the functional roles of these compartments, cell types
128 identified to colocalize with red pulp B cells include macrophages, neutrophils, and monocytes
129 that may remove old and dead red blood cells within the RP[13]. Likewise, cell types identified to
130 colocalize with follicle B cells include CD4+ memory T cells and podoplanin-expressing cells
131 within the WP (Fig 3d). Podoplanin-expression has been previously associated with T-cell zone
132 reticular cells[14] and observed to surround large arteries[11] within the WP.
133
134 To determine whether such cell-type spatial colocalization relationships are consistent across
135 tissue sections and individuals, we further repeated these analyses with 837,952 cells from five
136 additional spleen samples both within and across individuals ranging from 9400µm-by-9100µm
137 to 9100µm-by-12100µm in size. To ensure all datasets were annotated in a uniform manner, we
138 applied batch-correction[15] and used a linear discriminant analysis model to transfer cell-type
139 annotations to these new datasets (Online Methods, Supp Fig 3). We then applied CRAWDAD
140 to identify similar cell-type spatial relationships corresponding to the WP and RP compartments
141 both within and across individuals (Fig 3d, Supp Fig 4a). In this manner, CRAWDAD can
142 reproducibly identify functionally relevant cell-type spatial relationships in tissues within and
143 across individuals.

144

145 Although our clustering analysis identified a group of cells enriched by CD4 and CD45RO
146 proteins, which we interpreted as CD4+ memory T cells, we were unable to distinguish follicular
147 helper T cells, a specialized subset of CD4+ memory T cells that play a critical role in the
148 adaptive immune response, due to limitations of our 28 protein-marker panel. However, we
149 know that follicular helper T cells interact with B cells during the B cell maturation and
150 differentiation process within the B cell follicles. As such, we hypothesized that follicular helper
151 T cells would represent a spatially-defined subset of these CD4+ memory T cells that are
152 colocalized with follicular B cells[16]. To identify putative follicular helper T cells, we therefore
153 applied the same binomial testing framework used in CRAWDAD's pairwise spatial relationship
154 testing to identify CD4+ memory T cells that are statistically enriched around follicle B cells
155 (Fig 3e, Supp Fig 4b, Online Methods). We observed that the distribution of these putative
156 follicular helper T cells as a percentage of all CD4+ memory T cells (34% to 50%, Supp Fig 4c)
157 is consistent with previous characterizations within the spleen from healthy donors achieved
158 through orthogonal fluorescence-activated cell sorting approaches[17,18]. In this manner,
159 CRAWDAD enables spatial enrichment testing to identify functionally distinct spatially defined
160 cell-type subsets.

161

162 As atlasing efforts such as HuBMAP[19] and others continue to profile the spatial organization of
163 cells within tissues, we anticipate that CRAWDAD will be a useful tool in facilitating the
164 identification and characterization of cell-type colocalization relationships in complex tissues to
165 advance our understanding of the relationship between cell-type organization and tissue function.

166

Online Methods

CRAWDAD (Cell-type Relationship Analysis Workflow Done Across Distances) Overview

CRAWDAD characterizes cell-type spatial relationship trends by comparing observed pairwise cell-type spatial relationships to a set of null distributions in which cell-type labels have been shuffled at different length scales.

Creating null distributions at different length scales

To generate empirical null distributions against which observed cell-type spatial relationships can be compared to evaluate for statistical significance, CRAWDAD employs a grid-based cell-type label shuffling strategy. Given a tissue containing cells represented as x-y spatial coordinates with cell-type annotations, we partition the tissue into square grids of area r^2. Then for all cells that reside within the same grid, cell-type labels are shuffled to create a shuffled null distribution dataset for the given spatial length scale r. These shuffled null distributions are created for multiple rs (ranging from 50 to 500 by default) in order to achieve a set of empirical null distribution at different length scales. We further create multiple permutations (10 by default) at each length scale by using a grid-offsetting approach in order to mitigate the influence of specific grid divisions. Specifically, when using permutations, the portion of the tissue that does not fit perfectly into the grids is divided by the number of permutations to be used as offset.

Computing cell-type spatial relationship trends

To evaluate the statistical significance of an observed pairwise cell-type spatial relationship, CRAWDAD uses a binomial testing framework. For a given reference cell-type, CRAWDAD defines neighboring cells as those within a Euclidean distance of a user-defined neighbor distance d (default: 100mm) of any reference cell-type cells. CRAWDAD then calculates the proportion of neighboring cells that are members of a given query cell-type:

$$p_1 = \frac{y_1}{n}$$

where n is the total number of total neighboring cells for a given reference cell-type, and y_1 is the total number of cells of a given query cell-type that are neighboring cells for a given reference cell-type.

This proportion is also computed for the previously created null distributions:

$$p_2 = \frac{y_2}{n}$$

where y_2 is the total number of cells of the query cell-type that are neighboring cells in the shuffled null distribution datasets.

A two-sided binomial proportion test is then performed to test the equality of p_1 and p_2 against the alternative that they are not equal:

$$H_0: p_1 = p_2 \text{ versus } H_A: p_1 \neq p_2$$

This is defined by a test statistic Z, which is:

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{2p(1-p)}{n}}}$$

where:

$$p = \frac{y_1 + y_2}{2n}$$

Defining the significance threshold

We used a significance threshold of 5% (p-value < 0.05 or |Z-score| > 1.96) corrected for multiple testing (Bonferroni method by default) based on the number of unique reference-neighbor cell-type pairs evaluated. In addition, when there were multiple permutations, we considered a Z-score significant if the mean value obtained across permutations is beyond the significance threshold.

Summarizing and visualizing spatial relationship trends

To visually summarize results for all cell-type pairs, CRAWDAD creates a dot plot to show the length scale at which the spatial relationship trend for each cell-type pair first becomes statistically significant, encoded as the size of dots, as well as the associated multiple-testing corrected Z-score, encoded as the color hue of dots. Notably, such summaries do not fully capture the dynamics of pairwise cell-type spatial relationships as a function of length scale. Specifically, if a pairwise cell-type spatial relationship becomes significant at a scale of n, but not significant again at a scale of m (m > n), only the n scale will be represented in the plot. If a relationship is not significant, there will be no dot associated with it.

Therefore, to visualize results for a specific pairwise cell-type spatial relationship, CRAWDAD further creates trend plots that show the computed multiple-testing corrected Z-score value as a function of the evaluated length scales. When there are various permutations, the mean multiple-testing corrected Z-score across all permutations and +/- an error bar of one standard deviation is visualized.

Subset assignment of cells

CRAWDAD can further assign cells of a given cell-type to subsets based on the enrichment or depletion of cell-types within their neighbors using the same binomial testing framework. Given

a tissue containing cells represented as x-y spatial coordinates with cell-type annotations, we again define a neighbor distance d. Given a reference and a query cell type, for every cell of the reference cell-type we perform a one-sided binomial exact test to assess if the proportion of query cell-type among the cell's neighbors are significantly greater than the proportion of the query cell-type in the population. Cells of the given reference cell-type with a p-value below a certain user-defined significance threshold can then be further subtyped as cells of the reference cell-type that are enriched with cells of the query cell type.

Analysis of simulated data

To simulate a spatially resolved omics dataset, we created 8000 cells randomly positioned from 0 to 2000 microns in the x and y coordinates. We chose four points P = [(250, 250), (250, 750), (750, 250), (750, 750)] to be the centers of the circular neighborhoods defining the cell populations, where d(P) represents the distance from a cell to its closest point in P. We labeled the cells with d(P) <= 100 microns as cell types B or C (50% chance each). Then, we labeled the cells with 100 < d(P) < 300 microns as cell type A. Finally, all the other cells were labeled cell type D. For CRAWDAD, a set of shuffled null distribution were created at length scales of 100 to 1000 separated by 50 microns for 10 permutations. To identify significant trends with multiple testing correction, we used a Z-score threshold of 2.96.

281

282 **Analysis of the developing mouse embryo**

283

284 A pre-processed subset of the seqFISH data of an 8-12 somite stage embryo[4] (Embryo 1) was
285 obtained from the original publication. This dataset contained 19416 cells with x-y coordinates
286 and 22 cell-type annotations. Coordinates were originally in micrometers and converted to
287 microns by multiplying by 1000.

288

289 For CRAWDAD, a set of shuffled null distributions were created at length scales of 100, 200,
290 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 3000, 6000 microns for 10 permutations. A
291 neighbor distance of 100 microns was used to evaluate every pairwise combination of cell-types
292 at each length scale. To identify significant trends with multiple testing correction, we used a Z-
293 score threshold of 3.88.

294

295

296

297 **Analysis of the mouse cerebellum**

298

299 A pre-processed subset of a Slide-seq dataset of the mouse cerebellum was obtained from the
300 original publication[7]. This dataset contained 10,098 beads with x-y coordinates and 19 cell-type
301 annotations previously predicted by RCTD[7]. Poorly represented cell types defined as those being
302 annotated in less than 20 beads (Choroid, Candelabrum, Ependymal, Globular, Macrophages)

303 were not considered in the CRAWDAD cell type colocalization, resulting in 14 remaining cell
304 types.
305
306 For CRAWDAD, a set of shuffled null distributions were created at length scales of 100, 200,
307 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 3000, 6000 microns for 10 permutations. A
308 neighbor distance of 100 microns was used to evaluate every pairwise combination of cell-types
309 at each length scale. To identify significant trends with multiple testing correction, we used a Z-
310 score threshold of 3.81.
311
312
313
314 **Analysis of the human spleen**
315
316 Pre-processed and compensated Akoya CODEX datasets of 6 human spleens tissue sections from
317 3 different donors were downloaded from the HuBMAP Data Portal
318 (https://portal.hubmapconsortium.org/) corresponding to dataset IDs: HBM389.PKHL.936,
319 HBM772.XXCD.697, HBM556.KSFB.592, HBM825.PBVN.284, HBM568.NGPL.345, and
320 HBM342.FSLD.938. These datasets contained protein expression for 28 markers, x-y
321 coordinates, and cell segmentation area measurements for 154,446, 150,311, 152,896, 130,584,
322 177,777, and 226,384 segmented cells, respectively. Protein expression was normalized by cell
323 area and $\log_{10}$ transformed with a pseudocount of 1. Using dataset HBM389.PKHL.936, cells
324 were clustered via Louvain graph-based clustering[20] based on a nearest neighbor-graph
325 with $k=50$. Protein expression values were summed together for cells assigned to each cluster, and

326 then the values were scaled across protein expression measurements within each cluster. The
327 scaled expression values were used to assign cell-type annotations, which initially resulted in 13
328 cell types. Outer and inner Follicular B cells annotations were combined into a single Follicular
329 B cells annotation. Linear discriminant analysis (LDA) was performed to transfer cell-type labels
330 between the datasets. Prior to label transfer, paired donor datasets HBM556.KSFB.592 and
331 HBM568.NGPL.345, and HBM825.PBVN.284 and HBM342.FSLD.938 were harmonized[21] due
332 to observed batch effects. LDA cell-type label transfers were done between the following
333 reference and query datasets: HBM389.PKHL.936 -> HBM772.XXCD.697,
334 HBM389.PKHL.936 -> HBM556.KSFB.592, HBM389.PKHL.936 -> HBM825.PBVN.284,
335 HBM556.KSFB.592 -> HBM568.NGPL.345, HBM825.PBVN.284 -> HBM342.FSLD.938.
336
337 For CRAWDAD, a set of shuffled null distributions were created at resolutions of 100, 200, 300,
338 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 3000, 6000 microns for 10 permutations. A
339 neighbor distance of 100 microns was used to evaluate every pairwise combination of cell-types
340 at each scale. The corrected Z-score threshold used was 3.58.
341
342 CD4+ memory T cells were further subsetted based on their proximity to Follicle B cells using a
343 one-sided exact binomial test, a neighbor distance defined as 100 microns and a p-value
344 threshold of 0.05. With the subsets, we calculated the proportion of CD4+ memory T cells near
345 Follicle B cells with respect to all CD4+ memory T cells and compared across datasets.
346
347
348

Comparison to other methods

Ripley's Cross-K Analysis

We used the spatstat[22] (version 3.0-6) package to compute different Ripley's Cross-K values for each pairwise combination of cell types. We subtracted the theoretical K for a Poisson homogeneous processes from the isotropic edge corrected Ripley's Cross-K in visualizing results.

Squidpy's Neighborhood Enrichment

Using Squidpy[23] (version 1.2.3), we defined the spatial neighbors using a radius of 100. We calculated Squidpy's neighborhood enrichment using default parameters. The results were plotted as a heatmap of Z-scores.

Squidpy's Cooccurrence Probability

Using Squidpy[23] (version 1.2.3) and its co_occurrence function with default parameters' values, we calculated the co-occurrence probability of clusters for each cell type.

372 **Supplementary Note 1: Conceptual differences between CRAWDAD, Squidpy's Co-**
373 **occurrence ratio, and Ripley's K**
374
375 Squidpy's Co-occurrence ratio
376 The co-occurrence ratio implemented in Squidpy[23] is a ratio of the frequency of observing a
377 given cell type within a specific distance interval conditioned on a reference cell type, over the
378 frequency of observing the given cell type within a specific distance interval regardless of the
379 reference cell type. It differs from CRAWDAD in that cell type neighbors are no longer defined
380 by a radius distance of spatial units away from a reference cell. Instead, it assesses the frequency
381 of each cell type within a series of specific distance intervals, which can be thought of as "rings"
382 of a given thickness away from a reference cell. This theoretically provides information about
383 other cell type structures or patterns that are organized at certain distances away from a given
384 reference cell type. For expected cell type relationships, such as cell types that are known to
385 specifically cluster together, the co-occurrence ratio is expectedly high at small intervals away
386 from the reference cell type. However, it is difficult to define a significant co-occurrence ratio to
387 compare spatial relationships between different reference cell types and identify novel significant
388 cell type spatial relationships.
389
390 Ripley's K
391 Ripley's K is a descriptive statistic of how clustered a given point process is, or how clustered
392 multiple point processes are with respect to each other[24]. Ripley's K calculates an observed K
393 across multiple scales and compares this to a theoretical K, which assumes complete spatial
394 randomness and is based on a homogenous Poisson process. However, the underlying

assumption of complete spatial randomness for this theoretical distribution is likely too relaxed and not valid for tissues in which the cell types are inherently organized at some level. This likely leads to the observation of all cell types appearing to be highly clustered together when measured by Ripley's K, leading to false positives. This contrasts with CRAWDAD, which uses a permutation-based multi-scale shuffling approach to account for the inherent underlying structure and compartmentalization of cell types in tissues.

Acknowledgements

This material is based upon work supported by the HuBMAP Integration, Visualization & Engagement (HIVE) Initiative under Award Number 1OT2OD026673 (JF, BFM, RdSP) and the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35-GM142889 (JF, MA, LA) and U54AI-142766 (MA, TB, AP, CW).

Parts of this work was carried out at the Advanced Research Computing at Hopkins (ARCH) core facility (rockfish.jhu.edu), which is supported by the National Science Foundation (NSF) grant number OAC1920103.

Conflicts of interest

None declared.

References

1. Mitra, P. P. The Circuit Architecture of Whole Brains at the Mesoscopic Scale. *Neuron* **83**, 1273–1283 (2014).
2. Elhanani, O., Ben-Uri, R. & Keren, L. Spatial profiling technologies illuminate the tumor microenvironment. *Cancer Cell* **41**, 404–420 (2023).

426 3. Manz, B. N. & Groves, J. T. Spatial organization and signal transduction at intercellular
427 junctions. *Nat Rev Mol Cell Biol* **11**, 342–352 (2010).
428 4. Lohoff, T. *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse
429 organogenesis. *Nat Biotechnol* **40**, 74–85 (2022).
430 5. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide
431 expression at high spatial resolution. *Science (1979)* **363**, 1463–1467 (2019).
432 6. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX
433 Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).
434 7. Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics.
435 *Nat Biotechnol* **40**, 517–526 (2022).
436 8. Bellamy, T. C. Interactions between Purkinje neurones and Bergmann glia. *The*
437 *Cerebellum* **5**, 116–126 (2006).
438 9. Wu, Y. & Hirschi, K. K. Regulation of Hemogenic Endothelial Cell Development and
439 Function. *Annu Rev Physiol* **83**, 17–37 (2021).
440 10. Gilbert SF. *Developmental Biology*. (Sinauer Associates, 2000).
441 11. Currlin, S. *et al.* Immune, endothelial and neuronal network map in human lymph node
442 and spleen. *bioRxiv* 2021.10.20.465151 (2022) doi:10.1101/2021.10.20.465151.
443 12. Lewis, S. M., Williams, A. & Eisenbarth, S. C. Structure-function of the immune system
444 in the spleen. *Sci Immunol* **4**, (2019).
445 13. Nolte, M. A., ’t Hoen, E. N. M., Van Stijn, A., Kraal, G. & Mebius, R. E. Isolation of the
446 intact white pulp. Quantitative and qualitative analysis of the cellular composition of the
447 splenic compartments. doi:10.1002/1521-4141.

448 14. Cheng, H.-W. *et al.* Origin and differentiation trajectories of fibroblastic reticular cells in
449 the splenic white pulp. *Nat Commun* **10**, 1739 (2019).
450 15. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
451 Harmony. *Nat Methods* **16**, 1289–1296 (2019).
452 16. Garside, P. *et al.* Visualization of Specific B and T Lymphocyte Interactions in the Lymph
453 Node. *Science (1979)* **281**, 96–99 (1998).
454 17. Vecchione, A. *et al.* Reduced Follicular Regulatory T Cells in Spleen and Pancreatic
455 Lymph Nodes of Patients With Type 1 Diabetes. *Diabetes* **70**, 2892–2902 (2021).
456 18. Zhao, J. *et al.* Hyperactive Follicular Helper T Cells Contribute to Dysregulated Humoral
457 Immunity in Patients With Liver Cirrhosis. *Front Immunol* **10**, (2019).
458 19. Jain, S. *et al.* Advances and prospects for the Human BioMolecular Atlas Program
459 (HuBMAP). *Nat Cell Biol* (2023) doi:10.1038/s41556-023-01194-w.
460 20. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of
461 communities in large networks. (2008) doi:10.1088/1742-5468/2008/10/P10008.
462 21. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
463 Harmony. *Nat Methods* **16**, 1289–1296 (2019).
464 22. Baddeley, A., Rubak, E. & Turner, R. *Spatial Point Patterns: Methodology and*
465 *Applications with R*. (Chapman and Hall/CRC Press, 2015).
466 23. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat Methods* **19**,
467 171–178 (2022).
468
469

Figures

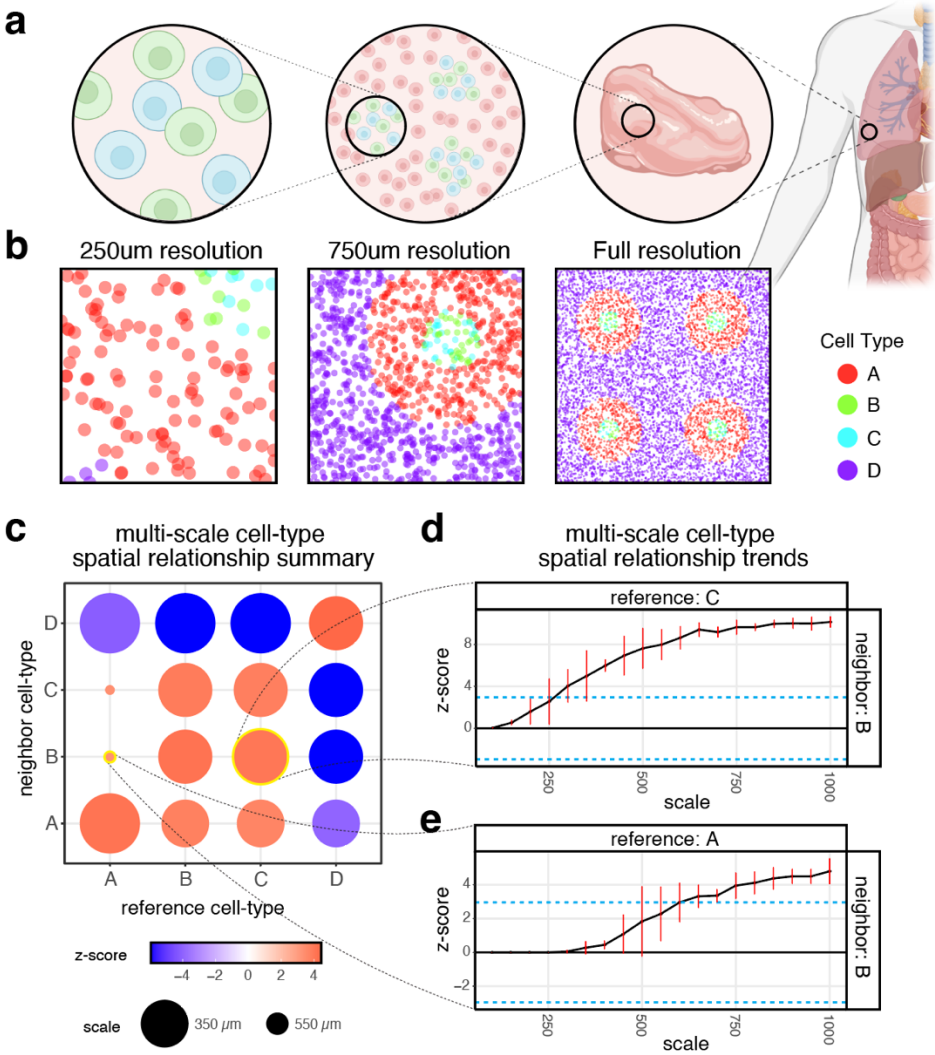
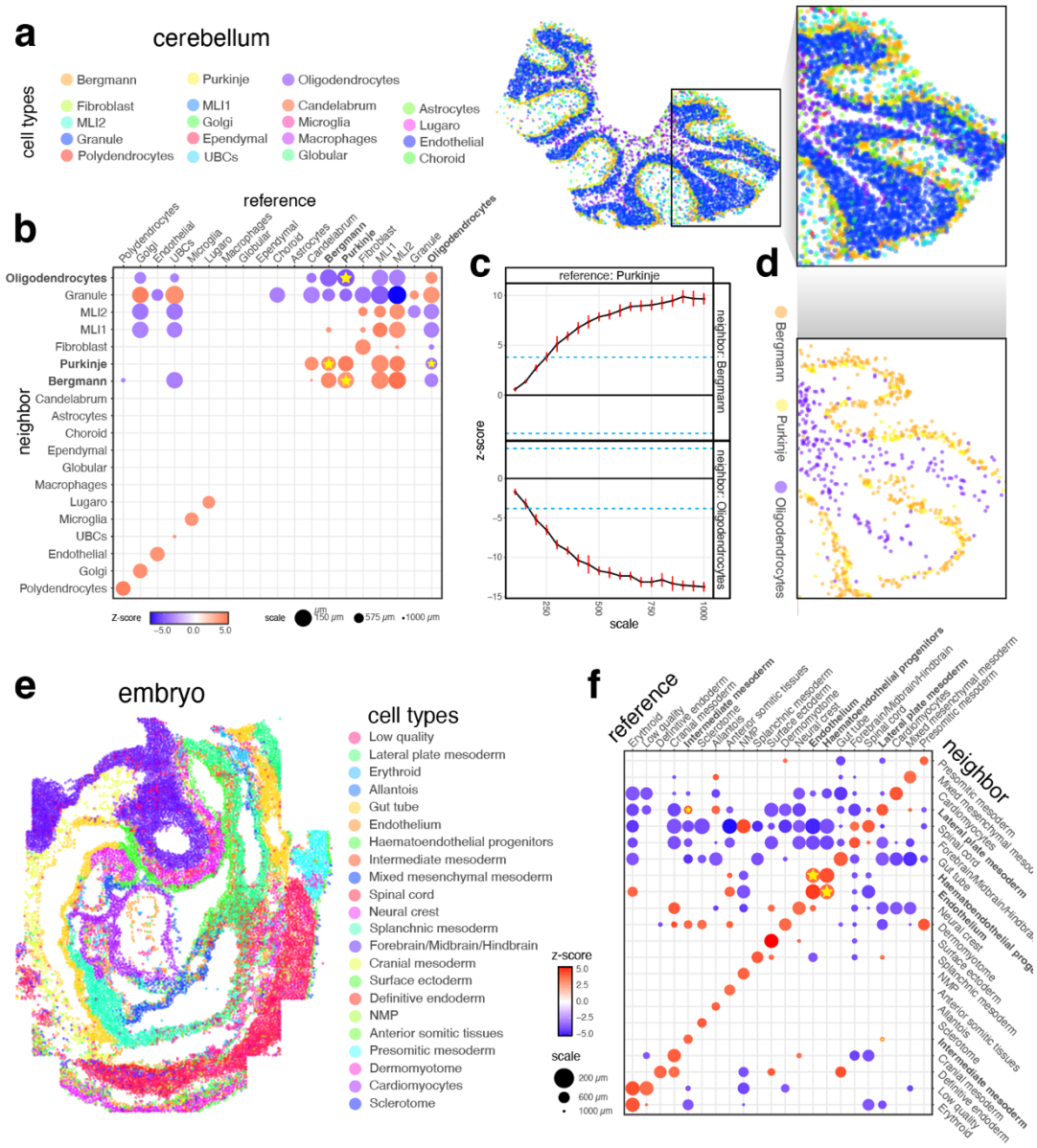
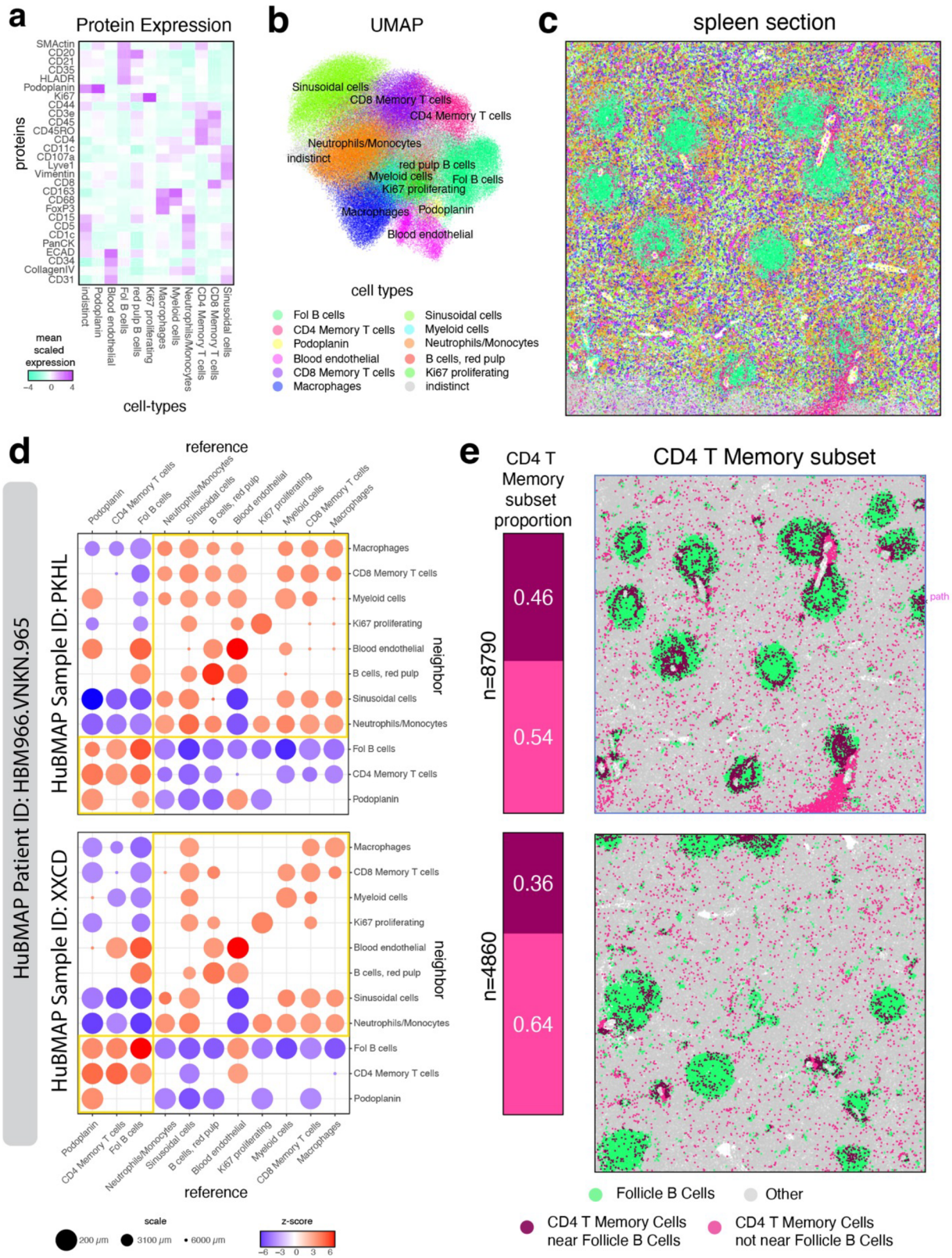


Figure 1. Motivating Cell-type Relationship Analysis Workflow Done Across Distances (CRAWDAD) using simulated data. **a.** Illustration of the cell-type spatial relationships found at different length scales. **b.** Simulated spatial omics tissue data, visualized at different scales. Each point is a cell, colored by cell type. **c.** Summary visualization of CRAWDAD's multi-scale cell-type spatial relationship analysis of the simulated data. The size of the dot represents the scale in which a neighbor cell-type first reaches a significant spatial relationship with respect to a reference cell-type. The color of the dot is the z-score at such scale (enriched or depleted). The

yellow stars represent the mutual enrichment between this cell-type pair, and therefore, their colocalization. **d.** The multi-scale spatial relationship trend plot for reference cell-type C and neighbor cell-type B. The horizontal dotted lines represent the z-score significance threshold corrected for multiple testing (Z-score = ±2.96). If a neighbor cell type has a trend that crosses the positive z-score threshold at a certain scale, then this neighbor type is deemed to be significantly enriched at that scale given the reference cell type. If the trend crosses the negative z-score threshold, then the neighbor cell type is deemed to be significantly depleted at that scale given the reference cell type. The red bars denote the standard error for the z-score estimated using permutations. **e.** The multi-scale spatial relationship trend plot for reference cell-type A and neighbor cell-type B.


a
cerebellum
cell types
Bergmann
Purkinje
Oligodendrocytes
Fibroblast
MLI1
Candelabrum
Astrocytes
MLI2
Golgi
Microglia
Lugaro
Granule
Ependymal
Macrophages
Endothelial
Polydendrocytes
UBCs
Globular
Choroid
b
reference
neighbor
Z-score
scale
c
reference: Purkinje
neighbor: Bergmann
neighbor: Oligodendrocytes
z-score
scale
d
Bergmann
Purkinje
Oligodendrocytes
e
embryo
cell types
Low quality
Lateral plate mesoderm
Erythroid
Allantois
Gut tube
Endothelium
Haematoendothelial progenitors
Intermediate mesoderm
Mixed mesenchymal mesoderm
Spinal cord
Neural crest
Splanchnic mesoderm
Forebrain/Midbrain/Hindbrain
Cranial mesoderm
Surface ectoderm
Definitive endoderm
NMP
Anterior somitic tissues
Presomitic mesoderm
Dermomyotome
Cardiomyocytes
Sclerotome
z-score
scale
f
reference
neighbor
g
colocalization at finer scale
Endothelium
Haematoendothelial progenitors
h
reference: Endothelium
neighbor: Haemato-endothelial progenitors
z-score
scale
i
colocalization at coarser scale
Lateral plate mesoderm
Intermediate mesoderm
j
reference: Intermediate mesoderm
neighbor: Lateral plate mesoderm
z-score
scale

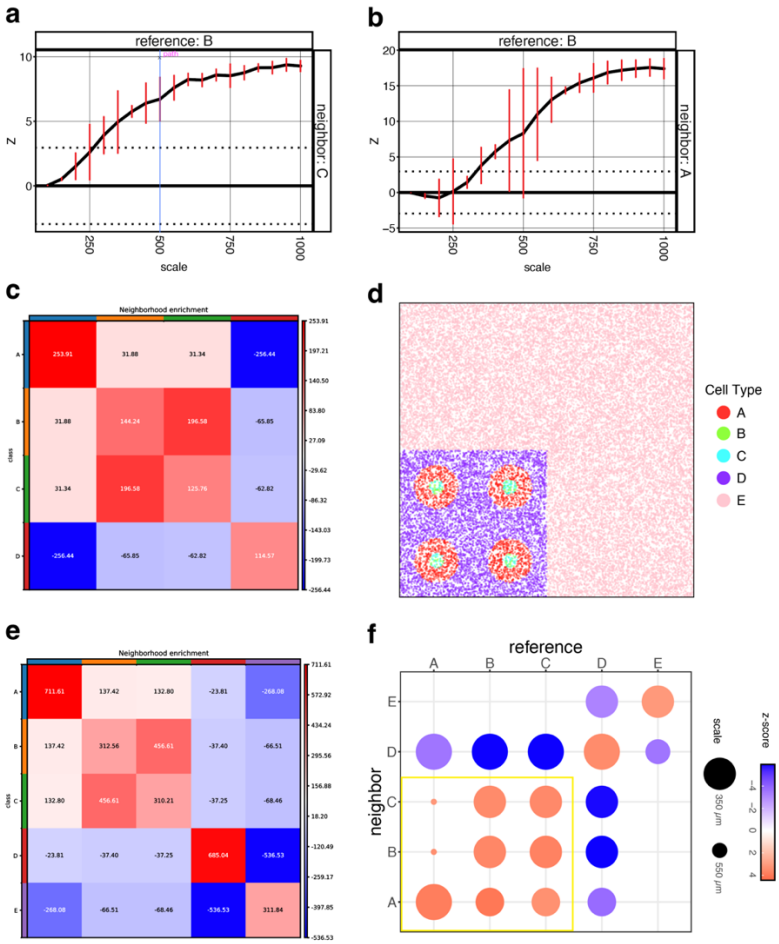
490 **Figure 2.** CRAWDAD characterizes cell-type spatial relationships in the mouse cerebellum and
491 embryo assayed by spatial transcriptomics. **a.** Cell types present in the cerebellum assayed by
492 Slide-seq. **b.** Summary visualization of the multi-scale cell-type spatial relationship analysis of
493 the cerebellum data. **c.** The multi-scale cell-type spatial relationship trend plot for Purkinje
494 neurons as the reference cell-type and Bergman glia (top) and Oligodendrocytes (bottom) as
495 neighbor cell-types. **d.** Zoom-in of Purkinje neurons, Bergman glia, and Oligodendrocytes in the
496 cerebellum. **e.** Cell types present in the embryo assayed by seqFISH. **f.** Summary of the multi-
497 scale cell-type spatial relationship analysis of the embryo data. **g.** Zoom-in of Endothelium and
498 the Hematoendothelial Progenitor cells. **h.** The multi-scale cell-type spatial relationship trend
499 plot for Endothelium cells as the reference cell type and Hematoendothelial Progenitor cells as
500 the neighbor cell type. **i.** Zoom-in of Intermediate Mesoderm and the Lateral Plate Mesoderm
501 cells. **j.** The multi-scale cell-type spatial trend plot for Intermediate Mesoderm cells as the
502 reference cell type and Lateral Plate Mesoderm cells as the neighbor cell type.


a
Protein Expression
proteins
SMActin
CD20
CD21
CD35
HLADR
Podoplanin
Ki67
CD44
CD3e
CD45
CD45RO
CD4
CD11c
CD107a
Lyve1
Vimentin
CD8
CD163
CD68
FoxP3
CD15
CD5
CD1c
PanCK
ECAD
CD34
CollagenIV
CD31
indistinct
Podoplanin
Blood endothelial
Fol B cells
red pulp B cells
Ki67 proliferating
Macrophages
Myeloid cells
Neutrophils/Monocytes
CD4 Memory T cells
CD8 Memory T cells
Sinusoidal cells
mean scaled expression
−4 0 4
cell-types
b
UMAP
Sinusoidal cells
CD8 Memory T cells
CD4 Memory T cells
Neutrophils/Monocytes
indistinct
red pulp B cells
Myeloid cells
Fol B cells
Ki67 proliferating
Macrophages
Podoplanin
Blood endothelial
cell types
Fol B cells
CD4 Memory T cells
Podoplanin
Blood endothelial
CD8 Memory T cells
Macrophages
Sinusoidal cells
Myeloid cells
Neutrophils/Monocytes
B cells, red pulp
Ki67 proliferating
indistinct
c
spleen section
d
HuBMAP Patient ID: HBM966.VNKN.965
HuBMAP Sample ID: PKHL
HuBMAP Sample ID: XXCD
reference
neighbor
Podoplanin
CD4 Memory T cells
Fol B cells
Neutrophils/Monocytes
Sinusoidal cells
B cells, red pulp
Blood endothelial
Ki67 proliferating
Myeloid cells
CD8 Memory T cells
Macrophages
scale
200 μm
3100 μm
6000 μm
z-score
−6 −3 0 3 6
e
CD4 T Memory subset proportion
CD4 T Memory subset
n=8790
0.46
0.54
n=4860
0.36
0.64
path
Follicle B Cells
Other
CD4 T Memory Cells near Follicle B Cells
CD4 T Memory Cells not near Follicle B Cells

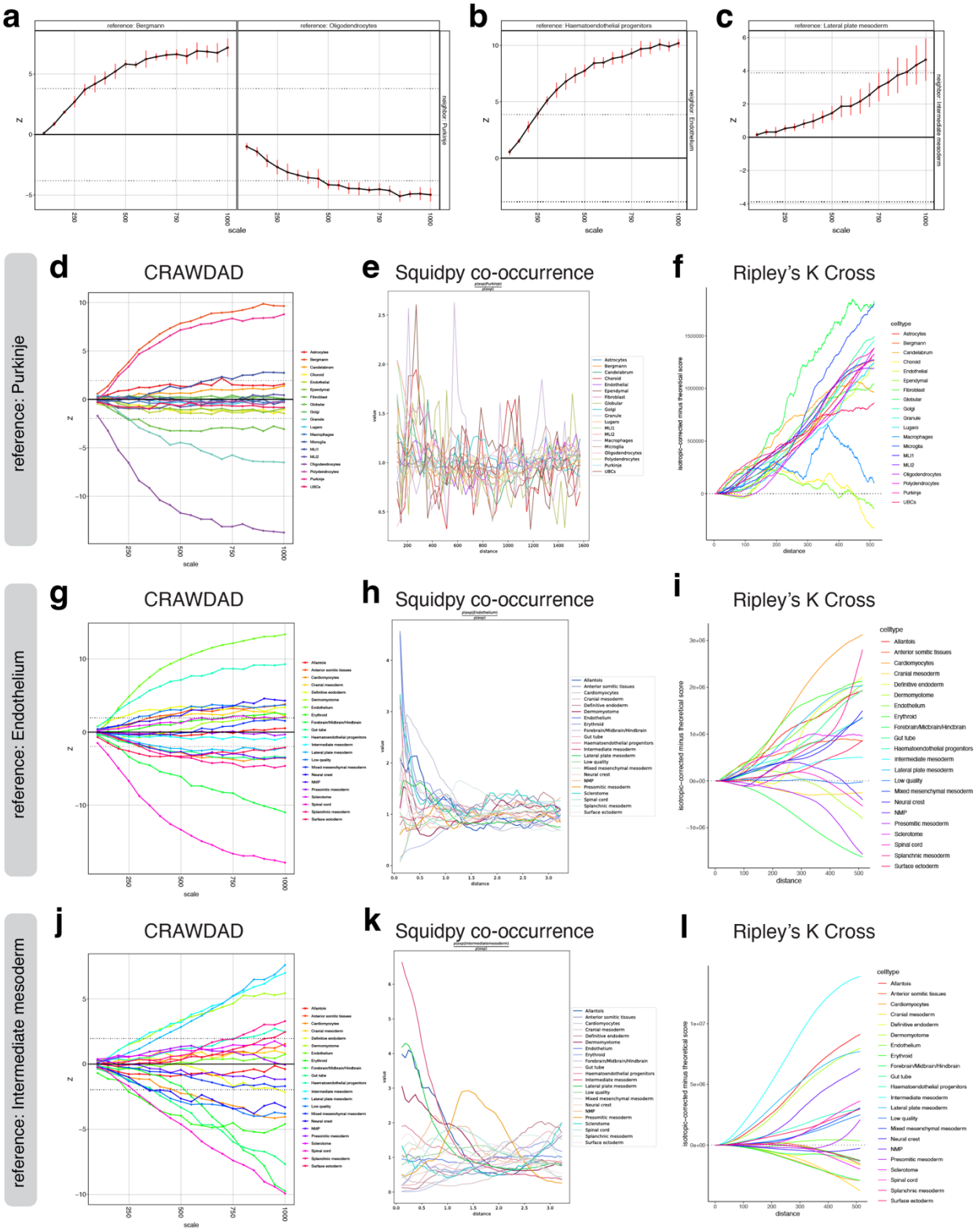
503

Figure 3. CRAWDAD characterize cell-type spatial relationships in the human spleen assayed by CODEX. **a.** Heatmap of marker protein expression for annotated cell types. **b.** UMAP reduced-dimension visualization of annotated cell-types. **c.** Spatial visualization of annotated cell-types. **d.** Summary visualization of the multi-scale cell-type spatial relationship analysis for tissue sections PKHL and XXCD from patient HBM966.VNKN.965. The small yellow square represents the white pulp, while the larger yellow square represents the red pulp. **e.** Subset of CD4+ Memory T cells near Follicle B cells. The number of CD4+ Memory T cells (n) and the proportion of subsets (left) and spatial visualization of subsets (right).

Supplementary Figures



Supplementary Figure 1. CRAWDAD and Squidpy's Neighborhood Enrichment on Simulated Data. **a.** CRAWDAD's multi-scale spatial relationship trend plot for reference cell-type B and neighbor cell-type C. **b.** CRAWDAD's multi-scale spatial relationship trend plot for reference cell-type B and neighbor cell-type A. **c.** Squidpy's neighborhood enrichment analysis on the simulated data. **d**. New simulated data that included cell-type E contouring the previous original dataset. **e.** Squidpy's neighborhood enrichment analysis on the new simulated data shows changing cell-type relationships. **f.** Summary visualization of CRAWDAD's multi-scale cell-type spatial relationship analysis of the new simulated data. The yellow box highlights that the cell-type relationships found in the analysis of the original simulated data were maintained.

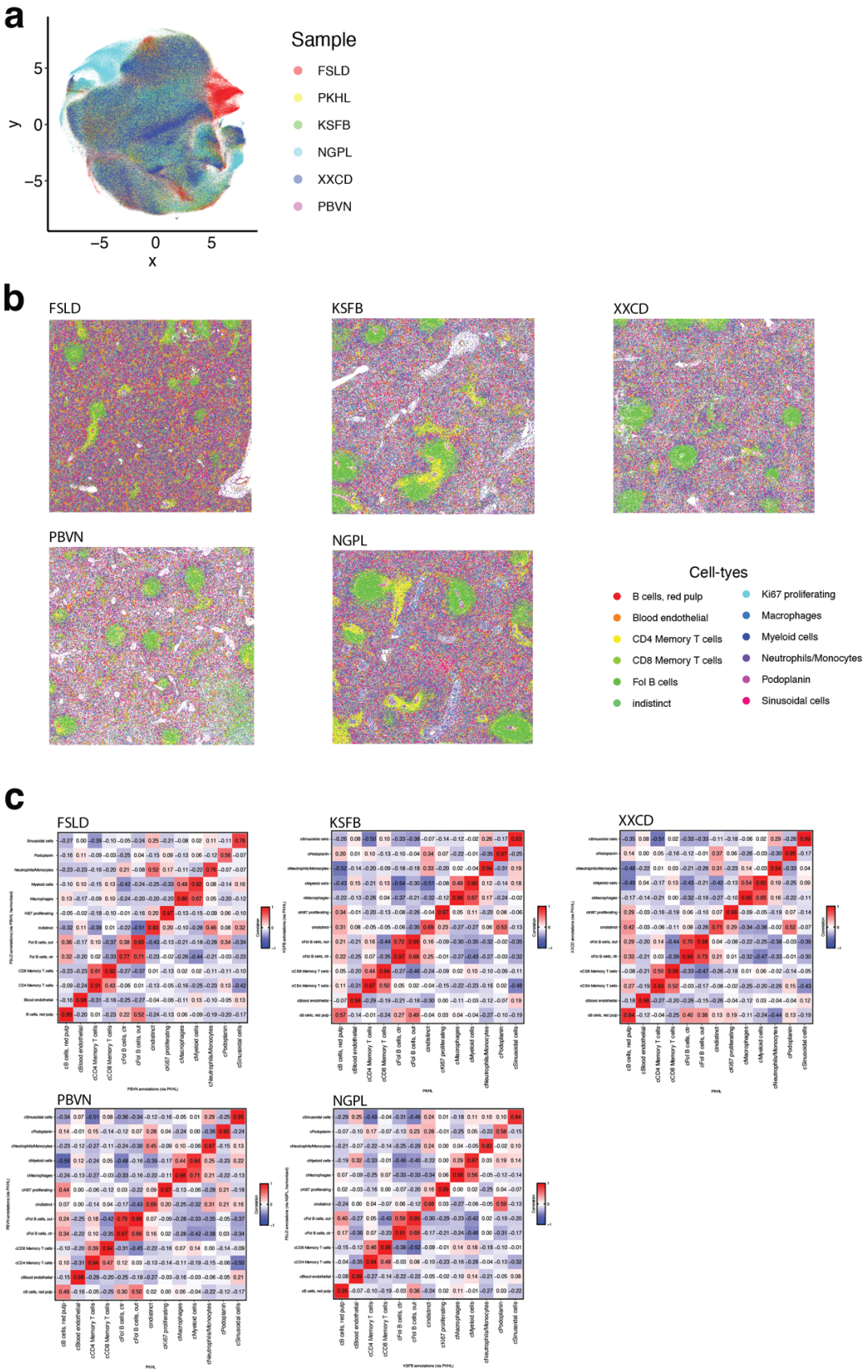


523

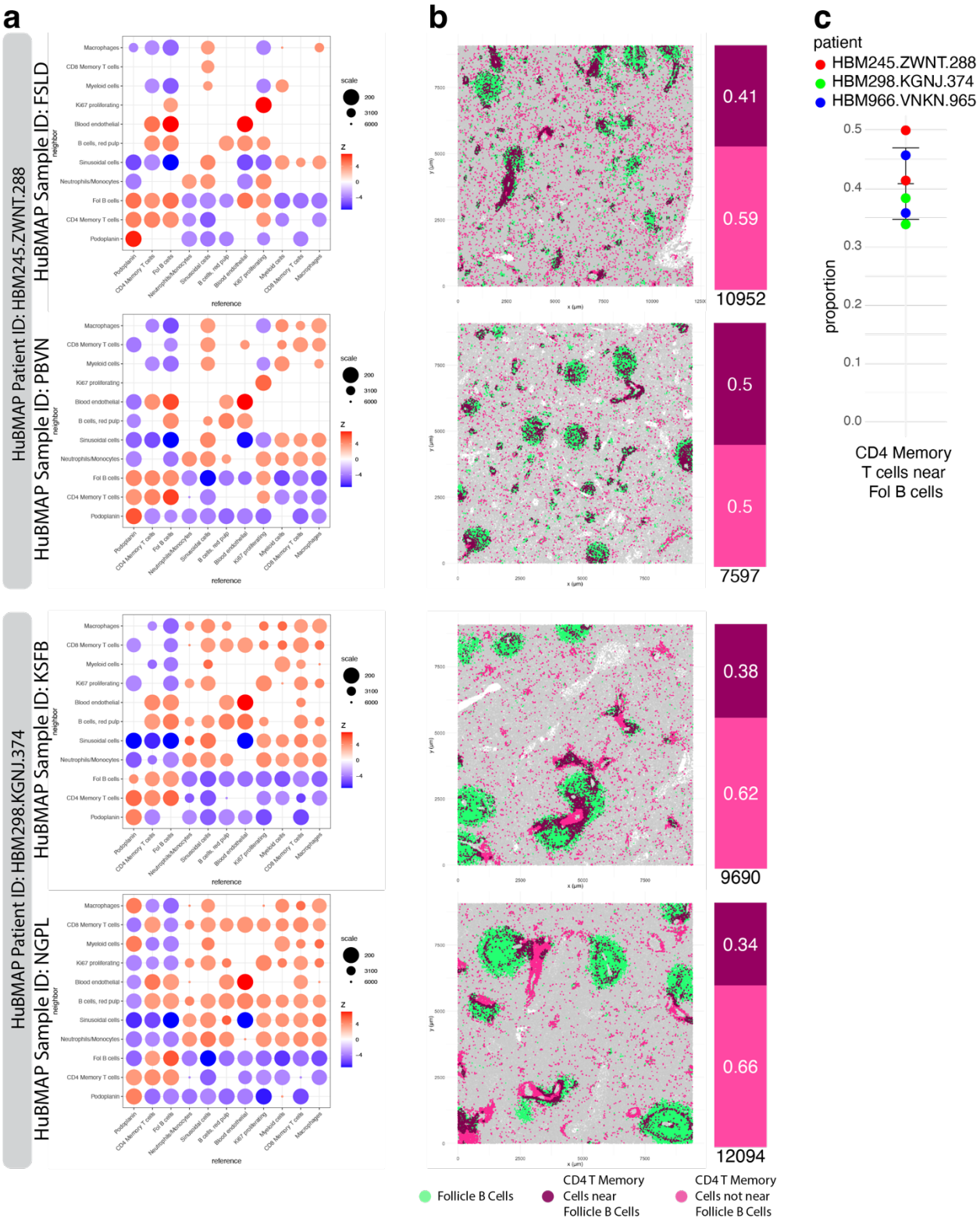
524 **Supplementary Figure 2.** Comparison of CRAWDAD, Squidpy's co-occurrence, and Ripley's

525 K Cross analyses. **d-f.** Cell-type neighborhood enrichment of Purkinje neurons versus all cell

526 types in the cerebellum Slide-seq dataset for CRAWDAD's multi-scale spatial relationship trend
527 plots (**d**), Squidpy's co-occurrence (**e**) and Ripley's K Cross analysis (**f**). For CRAWDAD, the
528 horizontal doted lines represent the significance threshold for enrichment (positive Z-score) and
529 separation (negative Z-score). For Ripley`s K cross, the trends above the horizontal dotted line
530 indicate colocalization while the ones bellow represent separation. **g-i.** Cell-type neighborhood
531 enrichment of Endothelium cells as reference on the Embryo dataset for CRAWDAD's multi-
532 scale spatial relationship trend plots (**g**), Squidpy's co-occurrence (**h**) and Ripley's K Cross
533 analysis (**i**). **j-l.** Cell-type neighborhood enrichment of Intermediate Mesoderm cells as reference
534 on the Embryo dataset for CRAWDAD's multi-scale spatial relationship trend plots (**j**),
535 Squidpy's co-occurrence (**k**) and Ripley's K Cross analysis (**l**).


a
Sample
FSLD
PKHL
KSFB
NGPL
XXCD
PBVN
x
y
b
FSLD
KSFB
XXCD
PBVN
NGPL
Cell-tyes
B cells, red pulp
Blood endothelial
CD4 Memory T cells
CD8 Memory T cells
Fol B cells
indistinct
Ki67 proliferating
Macrophages
Myeloid cells
Neutrophils/Monocytes
Podoplanin
Sinusoidal cells
c
FSLD
KSFB
XXCD
PBVN
NGPL

537 **Supplementary Figure 3.** CODEX Human Spleen Samples. a. Reduced-dimension visualization
538 of the samples. b. Spatial visualization of cell types in the different samples. c. Heatmaps
539 visualizing transcriptional correlations between label-transferred cell-types in spleen samples and
540 the originally annotated cell-types in reference spleen sample PKHL.
541



542

543 **Supplementary Figure 4.** Analysis of other spleen samples. **a.** Summary visualization of the

544 multi-scale cell-type spatial relationship analysis for datasets FSLD and PBVN from patient

545 HBM245.ZWNT.288 and for datasets KSFB and NGPL from patient HBM298.KGNJ.374. **b.**
546 On the left, the spatial visualization of the subsets of CD4+ Memory T cells near and not near
547 Follicle B cells and Follicle B cells. On the right, the number of CD4+ Memory T cells (n) and
548 the proportion of those that are near Follicle B cells. **c.** Proportion of CD4+ Memory T cells near
549 Follicle B cells over all CD4+ Memory T cells in each sample.