



Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomics data with nonuniform cellular densities

Brendan F Miller, Dhananjay Bambah-Mukku, Catherine Dulac, et al.

Genome Res. published online May 25, 2021

Access the most recent version at doi:[10.1101/gr.271288.120](https://doi.org/10.1101/gr.271288.120)

P<P	Published online May 25, 2021 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Open Access	Freely available online through the <i>Genome Research</i> Open Access option.
Creative Commons License	This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution 4.0 International license), as described at http://creativecommons.org/licenses/by/4.0/ .
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

1 **Characterizing spatial gene expression heterogeneity in spatially resolved single-cell**
2 **transcriptomics data with non-uniform cellular densities**

3 Brendan F. Miller^{1,2}, Dhananjay Bambah-Mukku^{3,4}, Catherine Dulac^{3,4}, Xiaowei Zhuang^{3,5,6}, Jean Fan^{3,5,6},
4 †, *

5 ¹Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218

6 ²Center for Computational Biology, Johns Hopkins University, Baltimore, MD 21211

7 ³Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138

8 ⁴Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

9 ⁵Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

10 ⁶Department of Physics, Harvard University, Cambridge, MA 02138

11 †Present Address: Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218

12

13

14

15 *Corresponding author

16

17 Running title: MERINGUE

18

19 Correspondence should be addressed to:

20 Jean Fan (jeanfan@jhu.edu)

21

22 Key words: spatial transcriptomics, single-cell transcriptomics, bioinformatics, heterogeneity, cellular
23 communication, expression patterns

24 **Abstract**

25 Recent technological advances have enabled spatially resolved measurements of expression
26 profiles for hundreds to thousands of genes in fixed tissues at single-cell resolution. However, scalable
27 computational analysis methods able to take into consideration the inherent 3D spatial organization of cell
28 types and non-uniform cellular densities within tissues are still lacking. To address this, we developed
29 MERINGUE, a computational framework based on spatial auto-correlation and cross-correlation analysis
30 to identify genes with spatially heterogeneous expression patterns, infer putative cell-cell communication,
31 and perform spatially informed cell clustering in 2D and 3D in a density-agnostic manner using spatially
32 resolved transcriptomics data. We applied MERINGUE to a variety of spatially resolved transcriptomics
33 datasets including multiplexed error-robust fluorescence in situ hybridization (MERFISH), spatial
34 transcriptomics, Slide-Seq, and aligned in situ hybridization (ISH) data. We anticipate that such statistical
35 analysis of spatially resolved transcriptomics data will facilitate our understanding of the interplay
36 between cell state and spatial organization in tissue development and disease.

37

38

39 **Introduction**

40 Characterization of the spatial context of cells and their cellular states is essential to
41 understanding the connection between tissue organization and function, particularly in complex organs
42 such as the mammalian brain. Furthermore, spatial context plays an important role in development and
43 organ formation in multicellular organisms, as well as in aberrant processes such as cancer (Crosetto et al.
44 2015). While advances in single-cell sequencing technologies can be used to discover transcriptionally
45 distinct subpopulations of cells in an unbiased manner, current protocols require dissociating cells from
46 tissue, thereby losing valuable spatial context (Crosetto et al. 2015). Thus, how these subpopulations of
47 cells are organized in space and how they may interact with each other remains an open question in many
48 systems.

49 To preserve informative spatial context, recent advances in imaging-based approaches have
50 enabled *in situ*, spatially resolved transcriptomic profiling with single-cell resolution (Zhuang 2021). In
51 addition, approaches based on spatially resolved RNA capture followed by sequencing, such as spatial
52 transcriptomics and Slide-seq provide spatially resolved, untargeted transcriptomic profiling at the pixel
53 level, with pixel size of 10-100 μ m (Larsson et al. 2021). Such high throughput data generation, both in
54 terms of the number of genes and number of cells assayed, demands scalable computational methods that
55 take advantage of this new spatial dimension to efficiently identify statistically significant spatial patterns
56 and relationships. In addition, as these methods are applied to increasingly complex tissues, statistical
57 analyses must be able to accommodate the non-uniform cell density induced by biological factors, such as
58 the presence of multiple, often spatially organized, cell-types inherent to tissues, as well as technical
59 factors, such as distortions from tissue sectioning.

60 Three statistical methods, SpatialDE, Trendsceek, and SPARK have previously been developed to
61 identify spatial gene expression heterogeneity, defined as an uneven, aggregated or patterned, spatial
62 distribution of gene expression magnitudes (Svensson et al. 2018; Edsgård et al. 2018; Sun et al. 2020).
63 Briefly, SpatialDE identifies spatial gene expression heterogeneity by decomposing a gene's expression
64 variance into a spatial and a non-spatial component using a spatial variance term that incorporates the

65 pairwise distances between cells. Trendsceek characterizes spatial gene expression heterogeneity by
66 testing a gene's expression for dependence with the pairwise distances between cells. SPARK identifies
67 spatial gene expression heterogeneity that best fits the observed gene expression trends using multiple
68 linear spatial models based on different Gaussian and periodic kernel functions that incorporate distances
69 between cells. Thus, each method directly incorporates information regarding cell distances, which could
70 present a challenge for analyses within tissues where cells are distributed with non-uniform densities. For
71 example, where local cell density is higher and the distance between cells are smaller, randomly varying
72 gene expression may give rise to apparent spatial aggregation due to cellular aggregation (Supplemental
73 Fig. S1A). Likewise, spatial variation in cellular density could also potentially mask spatial variation in
74 gene expression (Supplemental Fig. S1B). It is, however, important to identify variations in gene
75 expression magnitudes across cells that do not arise from variations in cellular density. Alternative
76 approaches such as SpaOTsc can accommodate non-uniform cellular densities if provided with geodesic
77 distances (Cang and Nie 2020). Briefly, using such density agnostic geodesic distances relating cells in
78 space, SpaOTsc uses optimal transport to estimate how much information about each gene's expression
79 magnitude can be provided by another gene's expression magnitude in order to identify groups of genes
80 with similar spatial patterning. However, this approach does not provide a statistical framework to
81 distinguish between significantly spatially heterogeneous genes versus non-significant or non-spatially
82 heterogeneous genes. Furthermore, cells in tissues inherently exist in a 3-dimensional context, yet
83 computational approaches capable of taking into consideration z-axis information, often at differing
84 length-scales such as multiple non-contiguous tissue sections, have yet to be demonstrated. Here, we
85 developed MERINGUE, a density-agnostic method for identifying spatial gene expression heterogeneity
86 using spatial auto-correlation and cross-correlation analyses. Using a variety of spatially resolved
87 transcriptomics datasets, we demonstrate that MERINGUE is able to identify biologically relevant spatial
88 gene expression patterns in both 2D and 3D in a manner that is independent of cell density.

89

90 **Results**

91

92 **Overview of MERINGUE**

93 Given a set of spatial positions such as those corresponding to single cells, MERINGUE first
 94 represents these cells as neighborhoods using Voronoi tessellation (Fig. 1A). In Voronoi tessellation,
 95 planes are partitioned into neighborhoods where a neighborhood for a cell consists of all points closer to
 96 that cell than any other (Okabe et al. 1992). Cells are then considered adjacent if their neighborhoods
 97 share an edge. For biological interpretability, we further require adjacent cells to be within a certain
 98 spatial distance in space in order to accommodate realistic length scales of cellular interactions. This
 99 neighborhood representation of cells accommodates varying neighborhood sizes and distances between
 100 cells and thus can characterize cell-types and tissues with non-uniform densities. We also find that such
 101 neighborhood adjacency relationships to be more stable than k -nearest-neighbor or k -mutual-nearest-
 102 neighbor relationships since such relationships require k to be specified beforehand and a single k value
 103 may not be appropriate for all densities and regions within a spatially resolved dataset (Supplemental Fig.
 104 S2A). MERINGUE encodes these adjacency relationships using a binary adjacency weight matrix W ,
 105 with a weight of 1 if two datasets are adjacent and 0 otherwise (Fig. 1A). Such adjacency relationships are
 106 not restricted to 2D and thus can accommodate 3D information, such as from imaging of multiple slices
 107 of the same tissue or 3D volumetric imaging of a tissue block, if available (Wang et al. 2018; Lee et al.
 108 2015).

109 Next, to identify genes with spatially heterogeneous expression, given a matrix of normalized
 110 gene expression magnitudes for the same set of spatially resolved cells, MERINGUE uses this adjacency
 111 weight matrix W in calculating Moran's I, a global measure of spatial auto-correlation popular in geo-
 112 spatial analysis, for each gene's expression magnitude (x) across the population of N cells (Moran 1950):

$$Moran's\ I = \frac{N}{\sum_i \sum_j W_{ij}} \frac{\sum_i \sum_j W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

113 When a gene's expression magnitude (x) between spatially adjacent cells ($W_{ij} = 1$) are positively
 114 correlated, Moran's I will be positive (Fig. 1B), indicative of spatial gene expression heterogeneity.

115 Moran's I has a closed form, allowing p-values to be derived without reliance on computationally
 116 intensive permutations (Supplemental Fig. S2B, (Moran 1950)).

117 To further characterize the scale of significant spatial gene expression heterogeneity, using the
 118 same matrix of normalized gene expression magnitudes and adjacency weight matrix W , MERINGUE
 119 calculates a Local Indicators of Spatial Association (LISA) for each gene (Anselin 2010):

$$LISA_i = N \frac{(x_i - \bar{x}) \sum_j^N W_{ij} (x_j - \bar{x})}{\sum_i^N (x_i - \bar{x})^2}$$

120 When a gene's expression values (x) in a given cell (i) is positively correlated with that cell's spatially
 121 adjacent neighbors, the cell's LISA for the given gene will be highly positive. Again, LISA has a closed
 122 form, allowing p-values to be derived quickly. As such, MERINGUE defines the percent of cells with
 123 statistically significant LISAs as the percent of cells driving a spatially heterogeneous gene expression
 124 pattern. This use of LISA guards against the identification of spatially heterogeneous genes driven by
 125 small hotspots or outliers. Simulations suggest that false positives may be effectively eliminated by
 126 restricting to spatial heterogeneity driven by more than 5% of cells (Supplemental Fig. S2C).

127 Finally, to summarize genes into primary spatial patterns, MERINGUE calculates a spatial cross-
 128 correlation index between all pairs of genes identified with significant spatially heterogeneous expression
 129 driven by a sufficient percent of cells:

$$SCI = \frac{N}{2 \sum_i^N \sum_j^N W_{ij}} \frac{\sum_i^N \sum_j^N W_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2} \sqrt{\sum_j^N (y_j - \bar{y})^2}}$$

130 When one gene's expression magnitude (x) in a given cell (i) is positively correlated with another gene's
 131 expression magnitude (y) in the cell's spatially adjacent neighbors (j), the SCI for this gene pair will be
 132 positive. MERINGUE computes this spatial cross-correlation index for all gene pairs to derive a spatial
 133 cross-correlation matrix that is then used for hierarchical clustering and dynamic tree cutting to group
 134 these genes into primary spatial patterns (Fig. 1C, (Langfelder et al. 2008)).

135 In addition, MERINGUE further builds on this spatial cross-correlation index to identify spatially
 136 cross-correlated gene expression patterns that may be indicative of cell-cell communication. In particular,

137 communicating cell-types may express higher levels of particular ligand genes while being spatially
138 adjacent to cells that express higher levels of corresponding receptor genes or vice versa. Thus, to identify
139 such gene expression patterns that may be indicative of putative cell-cell communication, MERINGUE
140 constructs an adjacency weight matrix W to only include adjacency relationships between cell-types and
141 calculates the spatial cross-correlation statistics for known receptor and ligand genes (Ramilowski et al.
142 2015a). In this manner, when a receptor gene's expression magnitude (x) in a given cell (i) of cell-type
143 A is positively correlated with the corresponding ligand gene's expression magnitude (y) in cells of cell-
144 type B among the cell's spatially adjacent neighbors (j), the inter-cell-type SCI for this cell-type pair will
145 be highly positive. Statistical significance can then be assessed by permutation testing (Fig. 1D).

146

147 **MERINGUE identifies genes with spatially heterogeneous expression patterns and is robust to**
148 **changes in cellular densities**

149 As a proof of principle, we first applied MERINGUE to Spatial Transcriptomics (ST) data of the
150 mouse main olfactory bulb (MOB) and Slide-Seq data of the mouse cerebellum (Ståhl et al. 2016;
151 Rodriques et al. 2019). Briefly, for ST and Slide-seq, RNAs from tissue sections are captured onto an
152 array of DNA barcoded spots or a monolayer of DNA barcoded beads, respectively. By resolving the
153 DNA barcodes, both approaches enable matching of detected RNA abundances with their original
154 spatially resolved spots or beads, resulting in RNA-sequencing measurements with uniformly gridded
155 two-dimensional positional information. To validate MERINGUE, we expected that identified spatially
156 heterogeneous genes in the MOB should mark transcriptionally distinct and spatially organized cell layers
157 or combinations of cell layers (Fig. 2A, Supplemental Fig. S3A). Indeed, when we applied MERINGUE
158 to analyze 7365 genes among 260 spots, of the 834 identified as significantly spatially heterogeneous
159 genes (adjusted p-value < 0.05) driven by more than 5% of spots (Fig. 2B, Supplemental Fig. S3B,
160 Supplemental Table 1), 90% (754/834) overlapped with genes that are significantly differentially
161 expressed genes across cell layers (adjusted p-value < 0.05) as identified from ANOVA testing.
162 Furthermore, these 834 spatially heterogeneous genes can be further partitioned into 5 primary spatial

163 patterns that correspond to cell layers and combinations of cell layers as expected (Fig. 2C, Supplemental
164 Fig. S3C). One well-characterized aspect of spatial organization in the MOB involves the convergence of
165 axonal projections from olfactory receptor neurons expressing a given olfactory receptor (*Olfir*) into
166 glomerular neuropils at fixed locations in the glomerular cell layer of the olfactory bulb (Ressler et al.
167 1994; Vassar et al. 1994; Mombaerts et al. 1996). Therefore, as an additional validation, we evaluated
168 whether *Olfir* genes were spatially heterogeneous in a pattern that corresponds to the glomerular and
169 surrounding cell layers. While individual *Olfir* genes are very lowly expressed such that detection was
170 generally limited to only a few copies in a few spots (Supplemental Fig. S3D), rendering assessment of
171 spatial heterogeneity for individual *Olfir* genes infeasible, by aggregating the expression of all detected
172 *Olfir* genes, we validate that MERINGUE was able to identify significant spatial heterogeneity (p-value =
173 0.0000283). The spatial expression pattern further corresponded approximately to the glomerular and
174 surrounding cell layer as expected (Supplemental Fig. S3E). For SlideSeq data of the mouse cerebellum,
175 we applied MERINGUE to analyze 9762 genes among 1589 beads previously annotated to correspond to
176 the Purkinje layer (Supplemental Fig. S4A). We validate that *Aldoc* (zebrin II) is identified as among the
177 most significantly spatially heterogeneous genes (adjusted p-value < 0.05, > 5% beads, Supplemental
178 Table 2), consistent with observations from the original publications (Rodrigues et al. 2019).

179 We next compared MERINGUE to previously published computational methods for analyzing
180 spatially resolved transcriptomics data, SpatialDE and SPARK (Svensson et al. 2018; Sun et al. 2020).
181 We applied each method to analyze 7365 genes among 260 spots in the MOB to identify spatially
182 heterogeneous genes (see Supplemental Methods). We found the resulting significance of spatial
183 heterogeneity in terms of $-\log_{10}(\text{adjusted p-value})$ to be highly correlated across genes between all tested
184 computational methods (R=0.914 between MERINGUE and SpatialDE, R=0.898 between MERINGUE
185 and SPARK, Supplemental Fig. S5A-B). The resulting set of significantly spatially heterogeneous genes
186 identified by each tested computational method using a common significance threshold (adjusted p-value
187 < 0.05) were also highly overlapping (Supplemental Fig. S5C). We further evaluated the computational
188 efficiency of each method in terms of runtime and memory usage as a function of the number of genes

189 and the number of cells in the dataset (see Supplemental Methods). We found that MERINGUE achieves
190 improved computational efficiency compared to previously published computational methods
191 (Supplemental Fig. S5D-E). Thus, MERINGUE is capable of identifying spatially heterogeneous genes
192 consistent with previously published approaches in scalable manner.

193 We developed MERINGUE to accommodate the non-uniform cellular densities inherent to
194 tissues. Thus, we reasoned that changes in cellular densities should not substantially impact
195 MERINGUE's ability to identify spatially heterogeneous genes. To assess MERINGUE's robustness to
196 spatial variations in cellular densities, we artificially induced non-uniformity in the spatial distribution of
197 ST spots by distorting their positional coordinates (see Supplemental Methods, Supplemental Fig. S6A).
198 Due to its use of a distance-agnostic binary weight matrix, MERINGUE's resulting significance of spatial
199 heterogeneity across genes was highly correlated between the uniform and non-uniform case as expected
200 (Spearman $\rho = 0.862$, Supplemental Fig. S6B). Likewise, while MERINGUE was able to identify 834
201 significantly spatially heterogeneous genes (adjusted p-value < 0.05 , $> 5\%$ of spots) in the uniform
202 density case, 544 (65%) of these genes were recovered in the non-uniform density case with the same
203 adjusted p-value and spot percentage thresholds. The discrepancies between the uniform and non-uniform
204 cases can be largely attributed to changes in the binary weight matrix (Supplemental Fig. S6C). Because
205 SpatialDE and SPARK incorporate Euclidean distances between cells in their evaluation of spatial
206 patterns, we reasoned that spatial variations in cellular density would impact their ability to identify
207 spatially heterogeneous genes. We thus applied the same uniform and artificially induced non-uniform
208 case comparison. As expected, the resulting significance of spatial heterogeneity across genes were less
209 well correlated between the uniform and non-uniform density case for both SpatialDE (Spearman $\rho =$
210 0.427) and SPARK (Spearman $\rho = 0.418$) (Supplemental Fig. S6D). Likewise, while SpatialDE was able
211 to identify 360 significantly (adjusted p-value < 0.05) spatially heterogeneous genes in the uniform
212 density case, only 56 (16%) of these genes were recovered in the non-uniform case with the same
213 adjusted p-value threshold. Similarly, while SPARK was able to identify 664 significantly (adjusted

214 combined p-value < 0.05) spatially heterogeneous genes in the uniform case, only 66 (10%) of these
215 genes were recovered in the non-uniform case with the same adjusted p-value threshold.

216

217 **MERINGUE integrates 3D and multi-layer tissue information**

218 While spatially resolved transcriptomics measurements generally provide positional information
219 in the imaging (x-y) plane, z-direction information can be obtained through optically scanning through
220 imaging planes or sequential tissue sections. To demonstrate integration of z-direction information, we
221 first applied MERINGUE to 3D *in situ* hybridization (ISH) data, aligned across multiple stage 6
222 *Drosophila melanogaster* embryos for 84 selected marker genes (Karaiskos et al. 2017; Fowlkes et al.
223 2008) (Fig. 2D, see Supplemental Methods). The role of spatial patterning in shaping cellular identities
224 has been well established in the *Drosophila melanogaster* embryo and as such, the 84 marker genes were
225 previously chosen for their known spatial patterning. Indeed, we validate that all 84 genes are identified
226 by MERINGUE as significantly spatially heterogeneous (adjusted p-value < 0.05, > 5% spots) as
227 expected. We further validated that these genes could be grouped by MERINGUE into 14 primary spatial
228 patterns that correspond to known regionally confined developmental fates and layers of the segmentation
229 gene network (Karaiskos et al. 2017; Ingham 1988) (Fig. 2E, Supplemental Fig. S7, Supplemental Table
230 3). For example, pattern 2 corresponds to the mesoderm and includes mesoderm determinant gene twist
231 (*twi*), while pattern 13 corresponds to the thoracic segments and includes known gap gene Kruppel (*Kr*)
232 (Leptin 1991; Preiss et al. 1985). Similarly, patterns 9 and 12 correspond to two spatially alternating
233 striped patterns that include known pair-rule genes even-skipped (*eve*) and odd-skipped (*odd*) respectively
234 (Macdonald et al. 1986; Coulter et al. 1990).

235 Alternatively, z information may be derived through serial sections. We thus next applied
236 MERINGUE to spatial transcriptomics data of four consecutive histological sections of a human breast
237 cancer biopsy ((Stahl et al. 2016), see Supplemental Methods). Analyzing each section independently, we
238 identified 414 genes that exhibit significant spatial variability (adjusted p-value < 0.05, > 5% spots) in at
239 least one section out of 6214 genes tested (Supplemental Table 4). As the distance between cells across

240 serial sections are greater than the distances between cells within sections, we sought to identify spatial
241 patterns consistent across layers by testing for spatial correlation between mutual nearest neighbors in
242 space across sections (Supplemental Fig. S8A). Such a multi-layer integrated approach confirmed 242
243 significantly spatially heterogeneous genes as being consistent across sections (Supplemental Table 4,
244 Supplemental Fig. S8B). Of the remaining 172 genes that were identified as significantly spatially
245 heterogeneous in individual sections but not across sections, visual inspection showed that although these
246 genes exhibit spatial variability within sections, there was minimal correspondence across sections
247 (Supplemental Fig. S8C). Such transcriptional patterns may be indicative of layer-specific subpopulations
248 or transcriptional features. For structurally stereotypic tissues, consistency across tissue sections may be
249 used as an additional criterion for identifying functionally relevant spatial patterns. Thus, MERINGUE is
250 capable of accommodating 3D information to identify spatially heterogeneous genes in 3D as well as
251 genes with spatial expression patterns consistent across serial sections.

252

253 **MERINGUE identifies spatial patterns in the mouse hypothalamic preoptic region using spatially** 254 **resolved single-cell gene expression data by MERFISH**

255 Particularly in complex organs such as the mammalian brain, the ability to identify and
256 interrogate the spatial organization of cell-types may provide additional insights into potential functional
257 roles underlying the spatial organization of neuronal populations (y Cajal 1911; Amaral and Witter 1989;
258 Arber 2012). We applied MERINGUE to analyze spatially resolved single-cell transcriptomics data of the
259 hypothalamic preoptic region obtained using multiplexed error-robust fluorescence in situ hybridization
260 (MERFISH) (Moffitt et al. 2018). Briefly, MERFISH allows individual RNA molecules in cells to be
261 imaged and identified by using a combinatorial labeling strategy that encodes RNA species with error-
262 robust barcodes that can be read out bit-by-bit using sequential rounds of single-molecule fluorescence in
263 situ hybridization (Chen et al. 2015). MERFISH has enabled simultaneous detection and identification of
264 thousands of targeted RNA species, which can then be segmented into cells to provide spatially resolved
265 single cell transcriptome measurements (Chen et al. 2015; Xia et al. 2019). Moffitt, Bambah-Mukku, et

266 al. previously used a 155 gene panel to characterize the hypothalamic preoptic region (1.8 mm × 1.8 mm
267 × 0.6 mm, Bregma +0.26 to -0.34) in adult mice to identify 31 excitatory and 39 inhibitory neuronal
268 subtypes in addition to non-neuronal cell-types using graph-based community-detection clustering
269 analysis that relies solely on the gene expression of profiles of cells without considering the spatial
270 information (Moffitt et al. 2018).

271 We applied MERINGUE to analyze the 155 genes along with 5 blank control barcodes, DAPI,
272 and poly-dT signals as negative controls within each cell-type and subtype to identify additional aspects
273 of spatial heterogeneity. Applying a rigorous approach to identify genes with spatially heterogeneous
274 expression patterns that are consistent across tissue layers and reproducible across animals (see
275 Supplemental Methods), we were able to identify at least one such spatially heterogeneous gene in 34 out
276 of 83 cell-types and subtypes analyzed (Fig. 3A, Supplemental Fig. S9, Supplemental Table 5). None of
277 the blank control barcodes, DAPI, or poly-dT signals were identified as consistently spatially variable.
278 MERINGUE further identified significant spatial gene expression heterogeneity within neuronal subtypes
279 in both the anterior and posterior of the preoptic region. Likewise, spatial gene expression heterogeneity
280 was identified in both inhibitory and excitatory neuronal subtypes. These aspects of spatial heterogeneity
281 were consistent with previous published spatially-unaware variance and principal components-based
282 analyses and visual inspection (Moffitt et al. 2018).

283 By providing a quantitative framework to systematically identify and evaluate the statistical
284 significance of spatial gene expression heterogeneity, MERINGUE identified that cells of inhibitory
285 subtype I-6 in the anterior of the preoptic region can be partitioned into a superior and inferior spatial lobe
286 marked by higher and lower expression of *Sema3c* and *Necab1* respectively (Fig. 3A). These patterns are
287 consistent across adjacent tissue sections. Likewise, cells of inhibitory subtype I-11 in the posterior
288 preoptic region can be partitioned into a medial and lateral spatial group marked by lower expression of
289 *Gabra1* higher expression of *Nos1* and higher expression of *Gabra1* lower expression of *Nos1*,
290 respectively, and this partition is consistent across adjacent tissue sections (Fig. 3B). Alternatively, *Gad1*,
291 which marks inhibitory cells, is highly expressed among all cells and does not exhibit significant spatial

292 heterogeneity as expected. Inhibitory subtype I-11 was previously identified to be specifically activated
293 by male mating based on the expression of immediate early gene *c-Fos* (Moffitt et al. 2018). While
294 inhibitory subtype I-11 exhibited significant spatial heterogeneity in both male and female animals (Fig.
295 3C), we found the fraction of *Nos1*+ I-11 neurons to be significantly higher in males than females (Fig.
296 3D, Student's *t*-test p-value = 0.03656). *Esr1* and *Irs4* were also identified to be significantly spatially
297 heterogeneous in I-11 neurons in a pattern similar to *Nos1* (Supplemental Fig. S9). *Esr1* and *Irs4* have
298 been previously shown to display sex-differences in their expression (Xu et al. 2012; Moffitt et al. 2018).
299 These observations suggest the potential presence of a finer *Nos1*+ I-11 neuronal subpopulation that is
300 sexually dimorphic. Furthermore, while MERINGUE generally identified concordant spatial gene
301 expression heterogeneity in both male and female animals (Supplemental Fig. S10), *Tacr1* (Tachykinin
302 receptor 1 also known as Neurokinin 1 receptor) was identified as significantly spatially heterogeneous in
303 excitatory subtype E-8 neurons only in male and not female mice (Fig. 3E). No other tested neuronal
304 subtype was identified to exhibit such consistently statistically significant sexually dimorphic spatial
305 heterogeneity. Previously, E-8 neurons were identified to be activated in male mice during mating based
306 on expression of *c-Fos* (Moffitt et al. 2018). However, E-8 neurons did not exhibit a significant difference
307 in terms of their proportion to all cells between female and male mice (Student's *t*-test p-value = 0.268).
308 Likewise, we confirmed that the fraction of cells expressing *Tacr1* in E-8 neurons is not significantly
309 different between male and female mice (Student's *t*-test p-value = 0.429). However, when we quantified
310 the fraction of cells driving the spatial heterogeneity of *Tacr1* expression based on LISA, we observe a
311 significant difference between male and female mice (Fig. 3F, Student's *t*-test p-value = 0.01316). *Tacr1*
312 knockout mice have been previously observed to exhibit deficits in sexual behavior (Berger et al. 2012).
313 The sexually dimorphic spatial organization of *Tacr1* expression in E-8 neurons may thus suggest a
314 sexually dimorphic difference in connectivity responsible for its sexually dimorphic activation in sexual
315 behavior. In this manner, MERINGUE enables quantitative and systematic evaluation of spatial gene
316 expression heterogeneity within transcriptionally distinct cell-subtypes from single cell spatially resolved
317 transcriptomics data.

318

319 **Spatially informed clustering identifies transcriptionally and spatially distinct subtypes of cells**

320 Spatial organization may play an important role in shaping cellular identities. Likewise, we may
321 expect unsupervised clustering based on transcriptional profiles alone to recover spatially organized cell
322 populations. However, for the aligned ISH data of the *Drosophila melanogaster* embryo, we find such
323 gene-expression clustering analysis to aggregate cells expressing different pair-rule genes into a single
324 transcriptional cluster due to these cells sharing many other commonly upregulated and downregulated
325 genes despite their spatially distinct organization (Fig. 4A-B), consistent with previously published
326 analyses (Karaïskos et al. 2017). However, as our spatial analysis was able to distinguish between the two
327 alternating striped spatial patterns marked by expression of pair-rule genes *eve* and *odd* respectively (Fig.
328 2E), we sought to incorporate spatial information to help distinguish these spatially distinct but
329 transcriptionally similar groups of cells.

330 Briefly, as in expression-based clustering, we constructed a neighbor graph where nodes are cells
331 and nodes are connected with an edge if the represented cells that are within the k-most transcriptionally
332 similar cells for some user-selected resolution parameter k. We incorporated spatial information by
333 weighing the edges of the network by the distance (d) between two neighborhoods (i, j) in the adjacency
334 representation W (see Methods): $\frac{1}{d_{ij}+1} + 1$. Again, use of such a neighborhood representation can
335 accommodate the non-homogenous density of cells in tissues compared to a Euclidean distance-based
336 measure of spatial distance. In this manner, if two cells are closer in space (d_{ij} is small), their
337 transcriptional similarity will give greater weight in the graph-based clustering. Incorporating these
338 spatial weights into our graph-based clustering with all other parameters kept constant, we were able to
339 split the cluster of cells expressing either *eve* or *odd* into two subpopulations, as desired (Fig. 4C).
340 Moreover, such spatially informed clustering generally preserved all other subpopulations and did not
341 result in additional splitting for other subpopulations (Fig. 4D). Furthermore, we demonstrated using
342 simulated data how such incorporation of spatial information can be used to distinguish transcriptionally

343 identical but spatially distinct clusters of cells (Supplemental Fig. S11). In a biological setting, however,
344 cells from the same cell-type may populate spatially distinct locations but such distinct spatial locations
345 alone would not necessarily indicate the presence of finer subtypes. We thus suggest that such spatially
346 informed clustering to be complementary to differential expression analysis, whereby identified spatially
347 distinct cell subpopulations should be analyzed for significantly differentially expressed genes to ensure
348 the presence of significant, likely subtle, transcriptional differences. Therefore, by incorporating spatial
349 information, in conjunction with differential expression analysis, we can identify finer, transcriptionally
350 and spatially distinct subpopulations.

351 Having demonstrated that incorporation of spatial information via graph-weighting can be applied
352 to identify finer transcriptionally and spatially distinct subpopulations of cells, we next sought to apply
353 this approach to identify finer neuronal subtypes in the preoptic region profiled by MERFISH (Moffitt et
354 al. 2018). Focusing on inhibitory neurons, we performed spatially informed clustering analysis on all
355 inhibitory cells in the same animal and tissue layer and compared resulting clusters to previous
356 annotations (Fig. 4E-F, see Supplemental Methods). We found that among the eight most populous
357 inhibitory neuronal subtypes (clusters with >100 cells each), our spatially informed clustering was able to
358 produce comparable clusters with the exception of I-2 and I-11, which were each split into two subtypes
359 (Fig. 4G). I-11 was split into two subtypes, cluster 10 (C10) and cluster 5 (C5), that significantly
360 differentially expressed genes including *Nos1* (Fig. 4H, Supplemental Fig. S12A), consistent with our
361 observations of significant spatial heterogeneity in *Nos1* expression among I-11 neurons. Indeed, the two
362 I-11 subtypes appear to be spatially distinct with C10 positioned more medially and C5 more laterally in
363 the posterior preoptic region (Fig. 4I). Likewise, I-2 was split into two subtypes, cluster 3 (C3) and cluster
364 8 (C8), that significantly differentially upregulated genes including *Cplx3* and *Dgkk*, respectively (Fig. 4J,
365 Supplemental Fig. S12B-C). Previously, I-2 neurons were observed to overlap with both the sexually
366 dimorphic nucleus of the preoptic area (SDN-POA) as well as other anatomical nuclei such as the bed
367 nucleus of the stria terminalis (BNST) (Moffitt et al. 2018). By refining I-2 into two finer subtypes, C8 is
368 observed to overlap more so with the BNST, while C3 comparably more so with the SDN-POA (Fig. 4K,

369 Supplemental Fig. S12C). I-2 neurons were previously observed to exhibit sexually dimorphic activation
370 during mating and aggression based on the expression of immediate early gene *c-Fos* (Moffitt et al.
371 2018). When we compare activation of the two I-2 subtypes based on significant *c-Fos* expression, we
372 observe comparatively greater activation during mating behavior in one subtype than the other (Fig. 4L).
373 This suggests that the activation in I-2 neurons observed previously may be driven by one of the two I-2
374 subtypes. While tuning parameters for regular graph-based clustering without spatial information can also
375 achieve splitting of I-2 and I-11, other inhibitory neuronal clusters can become over split (Supplemental
376 Fig. S12D). Therefore, by incorporating spatial information via graph-weighting, MERINGUE provides
377 an alternative approach to tease apart spatially distinct subpopulations without impacting other
378 transcriptionally distinct subtypes.

379

380 **MERINGUE identifies putative cell-cell communication between cell-types**

381 Spatially resolved transcriptomics data offers opportunity to identify gene expression patterns that
382 may be indicative of putative cell-cell communication between spatially co-localized cell-types. Previous
383 computational approaches for inferring cell-cell communication from single-cell RNA-sequencing data
384 have relied on correlations or co-expression of receptor genes in one cell type and corresponding
385 expression of ligand genes in another cell type (Ramilowski et al. 2015b; Smillie et al. 2019; Vento-
386 Tormo et al. 2018; Fan et al. 2020). Spatially resolved transcriptomics data provides the opportunity to
387 infer potential cell-cell communication by identifying spatially complementary expression patterns
388 between genes corresponding to interacting surface proteins such as receptors and ligands on spatially
389 neighboring cells. To enable such analyses, we further build on MERINGUE's spatial cross-correlation
390 functionalities by developing an inter-cell-type spatial-cross correlation function to identify potential
391 complementary spatial patterns of gene expression across spatially co-localized cell-types (Fig. 1D,
392 Supplemental Fig. S13A-D). However, unlike the spatial auto-correlation function, this inter-cell-type
393 spatial-cross correlation function is not solvable and thus significance must be assessed using permutation
394 to derive a null model. We enhance computational efficiency by implementing a parallelized, adaptive

395 permutation testing approach and assess significance using a permutation-based random label null model.
396 We confirm using simulations that such a permutation-based assessment produces the expected type-I
397 error rate (Supplemental Fig. S13E).

398 We first apply our approach identify gene expression patterns that may be indicative of putative
399 cell-cell communication between cells on beads corresponding to the Purkinje layer with cells on spatially
400 adjacent beads in Slide-Seq data of the mouse cerebellum (Fig. 5A). We use a set of over 2,500 known
401 receptor-ligand pairs previously supported by orthogonal biological validations (Ramilowski et al.
402 2015a). Restricting to well detected (CPM > 0 in more than 30 cells) receptor genes in the Purkinje layer
403 beads and well detected ligand genes in the spatially adjacent beads, we applied MERINGUE to test for
404 significant spatial cross correlation between all receptor and ligand gene pairs. We identified statistically
405 significant inter-cell-type spatial cross-correlation between expression of *Ptprz1* (Protein Tyrosine
406 Phosphatase Receptor Type Z1 i.e. PTP ζ) in beads corresponding to Purkinje layer and expression of its
407 ligand *Ptn* (secreted growth factor pleiotrophin) in spatially adjacent beads (Fig. 5B-C). *Ptprz1* has been
408 previously identified to be expressed by Purkinje neurons while *Ptn* has been previously identified to
409 distribute along Bergmann glial fibers in postnatally developing cerebellum (Matsumoto et al. 1994).
410 Although this Slide-Seq dataset does not provide single-cell resolution, we confirm significant co-
411 expression of *Ptprz1* with Purkinje cell specific promoter *Pcp2* (Fisher's exact p-value = 2.3×10^{-18}),
412 suggestive that the *Ptprz1* expression may be attributed to Purkinje cells within the Purkinje layer beads.
413 Likewise, we confirm significant co-expression of *Ptn* with *Slc1a3* (Glutamate Aspartate Transporter i.e.
414 GLAST), a glutamate transporter expressed by Bergmann glia (Fisher's exact p-value = 4.5×10^{-36}). In
415 contrast, restricting to well detected ligand genes in the Purkinje layer beads and well detected receptor
416 genes in the spatially adjacent beads (Fig. 5D-E), we do not identify significant spatial cross correlation
417 between any receptor and ligand gene pairs, including between *Ptn* expression in beads corresponding to
418 Purkinje layer and *Ptprz1* expression in spatially adjacent beads, indicative of the cell-type specificity of
419 inferred receptor-ligand interactions. Previous studies with cerebellar slice culture systems have shown
420 that Ptn-Ptprz1 signaling is involved in the morphogenesis of Purkinje dendrites (Tanaka et al. 2003). The

421 identification of such putative cell-cell communication between Bergmann glia and Purkinje cells may be
422 suggestive of the potential for glial signals to actively regulate neuronal function and contribute to
423 sustained plasticity in adult brains (Barres 2008).

424 We next sought to identify gene expression patterns that may be indicative of putative cell-cell
425 communication between cell-types using single-cell resolution MERFISH data of the preoptic region.
426 Previously, Moffitt, Bambah-Mukku, et al. visually noted that aromatase (*Cyp19a1*) enriched inhibitory I-
427 2 neurons displayed substantial spatial overlap with estrogen receptor (*Esr1*) enriched neuronal subtypes
428 (Moffitt et al. 2018). Aromatase is an enzyme that converts testosterone to estrogen, thereby modulating
429 steroid hormone signaling in the preoptic region. The spatial organization of these aromatase-enriched
430 neuronal subtypes with *Esr1*-enriched cells suggest that estrogen synthesized by these aromatase-
431 expressing neurons may be interacting with estrogen receptors on spatially adjacent *Esr1*-expressing cells
432 in a paracrine manner. To quantitatively assess to putative cell-cell communication between neuronal
433 subtypes and spatially neighboring cells via such paracrine signaling, we applied MERINGUE to assess
434 for significant spatial cross-correlation between *Cyp19a1* expression in all neuronal subtypes and *Esr1*
435 expression in spatially adjacent neurons. Indeed, we are able to identify statistically significant spatial
436 cross-correlation between *Cyp19a1* expression in I-2 neurons and *Esr1* expression in spatially adjacent
437 neurons in a manner that is consistent across tissue layers and reproducible across animals (Fig. 6A). In
438 contrast, neuronal subpopulations I-13 also express *Cyp19a1* but do not exhibit significant spatial cross-
439 correlation with *Esr1* in surrounding cells (Fig. 6B-C, Supplemental Fig. S14A). Furthermore, we also
440 apply MERINGUE to test for spatial cross-correlation between *Cyp19a1* expression in all neuronal
441 subtypes and androgen receptor (*Ar*) expression in spatially adjacent neurons and do not identify
442 consistently significant associations (Supplemental Fig. S14B), thereby highlighting the non-randomness
443 of the *Esr1* juxtaposition. This thus highlights MERINGUE's potential to quantitatively and
444 systematically identify complementary gene expression patterns that may be indicative of cell-cell
445 communication.

446

447 **Discussion**

448 Spatially resolved transcriptomic measurements demand computational approaches to identify
449 and characterize significant spatial gene expression heterogeneity. Here, we presented MERINGUE as a
450 density-agnostic approach to characterize spatially heterogeneous gene expression by identifying genes
451 with spatially auto-correlated expression and gene-pairs with spatially cross-correlated expression. We
452 validate our approach by analyzing spatially resolved transcriptomics data from both sequencing and
453 imaging-based methods in 2D and 3D to recover known biologically relevant spatial patterns. Our
454 analysis of the mouse preoptic region by MERFISH revealed sexually dimorphic spatial organization of
455 *Tacr1* expression in excitatory E-8 neurons and identified additional neuronal subpopulations within
456 inhibitory I-2 and I-11 neurons with spatially distinct organization that may play roles in murine sexual
457 behavior. MERINGUE is highly scalable and computationally efficient compared to previous spatial
458 analysis methods (Supplemental Fig. S15). Furthermore, MERINGUE is robust to spatial variations in
459 cellular density and can thus better accommodate non-uniform cellular densities common in tissues.

460 In comparison with previously published spatial gene expression analysis methods, while
461 MERINGUE identifies and groups spatially heterogeneous genes into primary spatial patterns, it does not
462 interpret identified spatial patterns based on pre-defined aggregated or alternating spatial patterns. In this
463 manner, we find MERINGUE to be complementary to previously published spatial gene expression
464 analysis methods in characterizing the spatial patterns of spatially heterogeneous genes. Likewise, we find
465 MERINGUE to be complementary to expression-based clustering analysis in order to identify additional
466 aspects of spatial heterogeneity within cell clusters or shared spatial gradients across cell clusters. In
467 addition, in analyzing spatially resolved single cell gene expression datasets obtained from different
468 technologies, MERINGUE may also be applied in combination with different normalization and error
469 model schemes such as cell volume-based normalization for imaging data (Moffitt et al. 2018), cell
470 density normalization for ST data (Saiselet et al. 2020). Furthermore, for zero-inflated transcriptomics
471 measurements, additional drop-out error modeling or imputation of drop-outs may be applied prior to
472 MERINGUE analysis (Kharchenko et al. 2014; Hou et al. 2020).

473 Finally, while MERINGUE uses spatial cross-correlation analysis to identify gene expression
474 patterns that may be indicative of putative cell-cell communication, such inference is based on spatial
475 proximity, which restricts inferred interactions to short-range interactions or chemical cues. This is
476 limiting for tissues such as the mammalian brain where neuronal communication and interactions often
477 span long distances due to long axons and dendritic processes. We anticipate that additionally combining
478 single-cell transcriptomics profiling with neuronal tracing could derive new binary weight matrices that
479 would fit into MERINGUE's analysis framework, enabling study of a more comprehensive cell-cell
480 interactions in a spatially resolved manner. Likewise, in the future, computational approaches such as
481 MERINGUE, in combination with systematic biological perturbations, can help elucidate the mechanisms
482 responsible for these spatial patterns and enhance our understanding of the spatial organization of and
483 communications between cell-types and cell-states within tissues.

484

485

486 **Methods**

487

488 **MERINGUE approach**

489

490 **Data preprocessing and quality control**

491 Data must be corrected for sequencing depth differences and other technically driven variation of
 492 expression magnitude prior to MERINGUE. CPM normalization without log transforming was applied to
 493 all spatial transcriptomics datasets. For MERFISH data, RNA counts were normalized per cell by the
 494 imaged volume of each cell per the originally published analysis (Moffitt et al. 2018).

495

496 **Adjacency weight matrix**

497 Given a set of spatial positional coordinates for spatially resolved datasets, such as cells, MERINGUE
 498 represents these cells as connected neighborhoods in space using an adjacency weight matrix W , where:

$$W_{ij} = \begin{cases} 1 & \text{if } cell_i \text{ and } cell_j \text{ are adjacent} \\ 0 & \text{if } cell_i \text{ and } cell_j \text{ are not adjacent} \end{cases}$$

499 Cells are defined as adjacent using Delaunay triangulation. Note, the Delaunay triangulation of a discrete
 500 set of points, in this case cells in space, is equivalent to the Voronoi diagram for the same set of points
 501 (Okabe et al. 1992). This approach is thus equivalent to defining cells as adjacent if they have Voronoi
 502 polygons, as inferred from Voronoi tessellation, that share an edge. For biological interpretability,
 503 adjacency relationships beyond a certain spatial distance can also be ignored. Delaunay triangulation can
 504 also accommodate 3-dimensional data.

505

506 **Identifying significantly spatially heterogeneous genes**

507 We define spatially heterogeneous genes as genes with uneven, often aggregated or patterned, spatial
 508 distribution of expression magnitudes. MERINGUE identifies such spatially heterogeneous genes by
 509 computing Moran's I (Moran 1950):

$$I = \frac{N}{\sum_i \sum_j W_{ij}} \frac{\sum_i \sum_j W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

510 for each gene given its normalized gene expression vector x across a population of N cells using the
 511 adjacency weight matrix W described previously to detect for positive spatial autocorrelation.
 512

513 The expected value of I under the null hypothesis of no spatial autocorrelation can be solved by
 514 computing the first moment (Getis 1995) and simplified to:

$$E(I) = \frac{-1}{N-1}$$

515 Likewise, variance can be derived using the second moment and simplified to:

$$Var(I) = \frac{N * S_4 + S_3 * S_5}{(N-1)(N-2)(N-3)W^2} - \left(\frac{-1}{N-1}\right)^2$$

516 Where:

$$S_1 = \frac{1}{2} \sum_i \sum_j (W_{ij} + W_{ji})^2$$

$$S_2 = \sum_i \left(\sum_j W_{ij} + \sum_j W_{ji} \right)^2$$

$$S_3 = \frac{\frac{\sum_i (x_i - \bar{x})^4}{N}}{\left(\frac{\sum_i (x_i - \bar{x})^2}{N}\right)^2}$$

$$S_4 = (N^2 - 3 * N + 3) * S_1 - N * S_2 + 3 * W^2$$

$$S_5 = (N^2 - N) * S_1 - 2 * N * S_2 + 6W^2$$

$$W = \sum_i \sum_j w_{ij}$$

517 We implement these calculations in C++ using Rcpp (Eddelbuettel and François 2011).

518

519 In a given dataset, we evaluate all genes for spatial heterogeneity and apply the Benjamini-Hochberg
 520 procedure to correct for multiple testing and control for false discovery (Benjamini and Hochberg 1995).

521

522 We assume here that the expression magnitudes represented by each neighborhood is comparable such
 523 that observed differences in gene expression levels across neighborhoods are not the result of different
 524 sequencing depths or other technical confounders. In this manner, x must already be normalized to
 525 control for variability in sequencing depth or other technical confounders, where appropriate, prior to
 526 analysis with Moran's I. Likewise, as Morans's I is not defined for constant signals, and thus genes
 527 without any expression variability are omitted from analysis.

528

529 If the data is produced by a mechanism that inherently induces some autocorrelation, such as high
 530 variability between spatially segregated batches or presence of noisy hotspots, then such a null hypothesis
 531 would not be appropriate, and evaluation of significance must be done using permutation. We show at
 532 least for a random subset of genes in our tested datasets that the null hypothesis is appropriate and thus
 533 results in essentially identical p-values regardless of approach (Supplemental. Fig. 2B).

534

535 **Characterizing the scale of significantly spatially heterogeneous genes**

536 For a given gene x identified as significantly spatially heterogeneous, MERINGUE next quantifies the
 537 scale of the spatial pattern by calculating the Local Indicators of Spatial Association (Anselin 2010)
 538 (LISA) for each neighborhood i :

$$I_i = N \frac{(x_i - \bar{x}) \sum_j^N W_{ij} (x_j - \bar{x})}{\sum_i^N (x_i - \bar{x})^2}$$

539 LISA relates to Moran's I via:

$$I = \sum_i^N \frac{I_i}{N}$$

540 And as such, LISA also contains a closed form that can be solved for its expected value and standard
 541 deviation under the null hypothesis of no spatial autocorrelation. We define the scale of a gene's spatial
 542 pattern as the percentage of cells with a LISA that is statistically significant i.e. has a p-value below an

543 alpha threshold (default: 0.05). Downstream analyses can be restricted to spatially heterogeneous genes of
 544 a sufficient scale, defined by default as 5% of cells.

545

546 Again, these calculations are implemented in C++ using Rcpp (Eddelbuettel and François 2011).

547

548 For visualization purposes, we further implement a signed LISA score:

$$sI_i = \text{sign}(x_i - \bar{x}) * N \frac{(x_i - \bar{x}) \sum_j^N w_{ij} (x_j - \bar{x})}{\sum_i^N (x_i - \bar{x})^2}$$

549

550 **Primary pattern determination using spatial cross-correlation analysis**

551 After identifying significantly spatially heterogeneous genes of a sufficient scale, MERINGUE groups
 552 these genes into primary spatial patterns. We calculate a spatial cross-correlation index (*SCI*) between all
 553 pairs of these genes. For N cells, gene x , and gene y , the *SCI* can be calculated as:

$$SCI = \frac{N}{2 \sum_i^N \sum_j^N W_{ij}} \frac{\sum_i^N \sum_j^N W_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2} \sqrt{\sum_j^N (y_j - \bar{y})^2}}$$

554 The *SCI* for all pairs of genes forms a spatial cross-correlation matrix, which we use as the basis for
 555 hierarchical clustering. Clusters of genes are then identified using dynamic tree cutting (Langfelder et al.
 556 2008) such that highly spatially cross-correlated genes fall into the same clusters, thus comprising the
 557 primary spatial patterns. By default, the hybrid dynamic tree cutting approach is used.

558

559 We visualize these primary patterns by interpolating across spatial regions not covered by cells using
 560 Akima interpolation (Akima 1996a, 1996b).

561

562 **Spatially informed clustering**

563 To identify spatially distinct but transcriptionally similar subpopulations, we begin with graph-based
 564 expression clustering. Specifically, we construct a k -nearest neighbor graph on the reduced principal

565 components space derived from normalized gene expression. In such a graph, each node is a cell and they
 566 are connected with an edge if they are among the k -nearest neighbors based on transcriptional similarity.
 567 To introduce spatial information, we weigh the edges of the graph based on the geodesic spatial distance
 568 between the two nodes' cells. The geodesic spatial distance is computed based on the adjacency matrix W
 569 where two cells would have a spatial distance of 1 if they are neighbors or 2 if they are neighbors of
 570 neighbors and so forth. We then transform the spatial distance into a weight that is inversely proportional
 571 to the distance such that cells closer together (i.e. with a small distance) will be given higher weight and
 572 cells farther apart (i.e. large distance) will be given a smaller weight: $weight = \frac{1}{distance + \alpha} + \beta$ where α
 573 and β are pseudocounts to guard against excessively large and small weights respectively. By default, we
 574 used $\alpha = \beta = 1$, though the unit and magnitude of both α and β will depend on the unit and magnitude of
 575 *distance*. We then apply Louvain graph-based clustering to the resulting weighted graph (Phyu and Myat
 576 Min 2019).

577

578 **Inference of cell-cell communication using inter-cell-type spatial cross-correlation analysis**

579 To infer cell-cell communication between spatially co-localized cell-types, MERINGUE focuses on
 580 identifying complementary gene expression patterns between known receptor-ligand pairs (Ramilowski et
 581 al. 2015b).

582

583 For each receptor-ligand pair, we compute an inter-cell-type spatial cross-correlation (*iSCI*) between
 584 expression of receptor x for the N cells of cell-type A and the expression of ligand y for the M cells of
 585 cell-type B:

586

$$iSCI = \frac{N + M}{2 \sum_i^A \sum_j^B W_{ictij}} \frac{\sum_i^N \sum_j^M W_{ictij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i^A (x_i - \bar{x})^2} \sqrt{\sum_j^B (y_j - \bar{y})^2}}$$

587

588 Here, the inter-cell-type adjacency weight matrix $W_{ict\ ij} = 1$ if a cell of cell-type A and a cell of cell-type
589 B are inferred to be adjacent or vice versa and $W_{ict\ ij} = 0$ otherwise to capture only spatial cross-
590 correlation patterns between the two cell-types.

591
592 We assess statistical significance by comparing the observed *iSCI* with the likelihood of observing such
593 an extremely positive value under a permutation-based random labeling model randomly permuting cell
594 labels. To enhance computational efficiency, we allow for parallelization across multiple cores and use an
595 adaptive permutation testing approach whereby receptor-ligand pairs are first assessed for significance
596 with 100 permutations by default, and putatively significant hits with permutation p-values $< 1/100$ are
597 then reassessed with 1000 permutations and so forth. Additional gene pairs with known interacting
598 products such as hormone-receptors can also be evaluated by this approach.

599

600 **Interactive application**

601 An interactive application built on Shiny (Chang et al, 2020) can be launched directly from R sessions to
602 enable interactive visual exploration of MERINGUE results and statistics.

603

604 **Software availability**

605 MERINGUE is programmed in C++ and available as an open-source R software package (R Core Team,
606 2020) with the source code available in the Supplemental Material and on GitHub at

607 <https://github.com/JEFworks-Lab/MERINGUE>. Additional documentation and tutorials are available at

608 <https://JEF.works/MERINGUE>.

609

610

611 **Acknowledgements**

612 We thank Stephen W. Eichhorn and Jeffrey R. Moffitt for their help in understanding the preoptic region
613 MERFISH data and their early input into computational methodological development, and Guiping Wang
614 for her feedback and help in proof-reading this manuscript. The computations in this paper were run on
615 the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard
616 University. This work is supported in part by the CRUK IMAXT Grand Challenge grant (to X.Z.). J.F.
617 acknowledges support from the National Institutes of Health Pre-Doc to Post-Doc Transition Award
618 (K00CA222750). D. B.-M. acknowledges support from the National Institutes of Health Pathway to
619 independence support (K99HD092542). C.D and X.Z. are Howard Hughes Medical Institute
620 investigators.

621

622

623 **Author contributions**

624 J.F. and X.Z. conceived the research. J.F. led computational work under the guidance of X.Z. B.F.M.
625 contributed to computational work under the guidance of J.F. All authors participated in interpretation and
626 writing the manuscript.

627

628

629 **Competing Interests**

630 X.Z. is a co-founder and consultant of Vizgen, Inc.

631

632

633

634

635 **References**

- 636 Akima H. 1996a. Algorithm 760: Rectangular-grid-data surface fitting that has the accuracy of a bicubic
637 polynomial. *ACM Transactions on Mathematical Software (TOMS)* **22**: 357–361.
- 638 Akima H. 1996b. Algorithm 761: Scattered-data surface fitting that has the accuracy of a cubic
639 polynomial. *ACM Transactions on Mathematical Software (TOMS)* **22**: 362–371.
- 640 Amaral DG, Witter MP. 1989. The three-dimensional organization of the hippocampal formation: A
641 review of anatomical data. *Neuroscience* **31**: 571–591.
- 642 Anselin L. 2010. Local Indicators of Spatial Association-LISA. *Geographical Analysis* **27**: 93–115.
- 643 Arber S. 2012. Motor Circuits in Action: Specification, Connectivity, and Function. *Neuron* **74**: 975–989.
- 644 Barres BA. 2008. The Mystery and Magic of Glia: A Perspective on Their Roles in Health and Disease.
645 *Neuron* **60**: 430–440.
- 646 Benjamini Y, Hochberg Y. 1995. *Controlling the False Discovery Rate: A Practical and Powerful*
647 *Approach to Multiple.*
- 648 Berger A, Tran AH, Dida J, Minkin S, Gerard NP, Yeomans J, Paige CJ. 2012. Diminished pheromone-
649 induced sexual behavior in neurokinin-1 receptor deficient (TACR1(-/-)) mice. *Genes, brain, and*
650 *behavior* **11**: 568–76.
- 651 Cang Z, Nie Q. 2020. Inferring spatial and signaling relationships between cells from single cell
652 transcriptomic data. *Nature Communications* **11**.
- 653 Chang W, Cheng J, Allaire JJ, Xie Y, and McPherson J (2020). shiny:
654 Web Application Framework for R. R package version 1.5.0.
655 <https://CRAN.R-project.org/package=shiny>
- 656 Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. 2015. Spatially resolved, highly multiplexed
657 RNA profiling in single cells. *Science* **348**: aaa6090–aaa6090.
658 <http://www.ncbi.nlm.nih.gov/pubmed/25858977> (Accessed November 1, 2019).

- 659 Coulter DE, Swaykus EA, Beran-Koehn MA, Goldberg D, Wieschaus E, Schedl P. 1990. Molecular
660 analysis of odd-skipped, a zinc finger encoding segmentation gene with a novel pair-rule expression
661 pattern. *The EMBO journal* **9**: 3795–804.
- 662 Crosetto N, Bienko M, van Oudenaarden A. 2015. Spatially resolved transcriptomics and beyond. *Nature*
663 *Reviews Genetics* **16**: 57–66.
- 664 Eddelbuettel D, François R. 2011. Rcpp: Seamless R and C++ Integration. *Journal of Statistical*
665 *Software* **40**: 1–18.
- 666 Edsgård D, Johnsson P, Sandberg R. 2018. Identification of spatial expression trends in single-cell gene
667 expression data. *Nature Methods* **15**: 339–342.
- 668 Fan J, Slowikowski K, Zhang F. 2020. Single-cell transcriptomics in cancer: computational challenges
669 and opportunities. *Experimental & Molecular Medicine* 1–14.
- 670 Fowlkes CC, Hendriks CLL, Keränen SVE, Weber GH, Rübél O, Huang M-Y, Chatoor S, DePace AH,
671 Simirenko L, Henriquez C, et al. 2008. A quantitative spatiotemporal atlas of gene expression in the
672 *Drosophila* blastoderm. *Cell* **133**: 364–74.
- 673 Getis A. 1995. Cliff, A.D. and Ord, J.K. 1973: Spatial autocorrelation. London: Pion. *Progress in Human*
674 *Geography* **19**: 245–249.
- 675 Hou W, Ji Z, Ji H, Hicks SC. 2020. A systematic evaluation of single-cell RNA-sequencing imputation
676 methods. *Genome Biology* **21**: 218.
677 <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02132-x>
- 678 Ingham PW. 1988. *The molecular genetics of embryonic pattern formation in Drosophila*.
- 679 Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP.
680 2017. The *Drosophila* embryo at single-cell transcriptome resolution. *Science (New York, NY)* **358**:
681 194–199.
- 682 Kharchenko P V, Silberstein L, Scadden DT. 2014. Bayesian approach to single-cell differential
683 expression analysis. *Nature methods* **11**: 740–2.

- 684 Langfelder P, Zhang B, Horvath S. 2008. Defining clusters from a hierarchical cluster tree: the Dynamic
685 Tree Cut package for R. *Bioinformatics* **24**: 719–720.
- 686 Larsson L, Frisén J, Lundeberg J. 2021. Spatially resolved transcriptomics adds a new dimension to
687 genomics. *Nature Methods* **18**: 15–18. <https://doi.org/10.1038/s41592-020-01038-7>
- 688 Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS,
689 Aach J, et al. 2015. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling
690 in intact cells and tissues. *Nature Protocols* **10**: 442–458.
- 691 Leptin M. 1991. twist and snail as positive and negative regulators during *Drosophila* mesoderm
692 development. *Genes & development* **5**: 1568–76.
- 693 Macdonald PM, Ingham P, Struhl G. 1986. Isolation, structure, and expression of even-skipped: A second
694 pair-rule gene of *Drosophila* containing a homeo box. *Cell* **47**: 721–734.
- 695 Matsumoto K, Wanaka A, Mori T, Taguchi A, Ishii N, Muramatsu H, Muramatsu T, Tohyama M. 1994.
696 Localization of pleiotrophin and midkine in the postnatal developing cerebellum. *Neuroscience*
697 *letters* **178**: 216–20.
- 698 Moffitt JR, Bambach-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, Perez JD, Rubinstein ND, Hao J,
699 Regev A, Dulac C, et al. 2018. Molecular, spatial and functional single-cell profiling of the
700 hypothalamic preoptic region. *Science (New York, NY)* eaau5324.
- 701 Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. 2016. High-throughput single-cell gene-
702 expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proceedings of*
703 *the National Academy of Sciences* **113**: 11046–11051.
- 704 Mombaerts P, Wang F, Dulac C, Chao SK, Nemes A, Mendelsohn M, Edmondson J, Axel R. 1996.
705 Visualizing an olfactory sensory map. *Cell* **87**: 675–86.
- 706 Moran PAP. 1950. Notes on Continuous Stochastic Phenomena. *Biometrika* **37**: 17.
- 707 Okabe A, Boots BN, Sugihara K. 1992. *Spatial tessellations* □: *concepts and applications of Voronoi*
708 *diagrams*. Wiley.

- 709 Phyu KS, Myat Min M. 2019. Graph-based Community Detection in Social Network. In *2019 IEEE/ACIS*
710 *18th International Conference on Computer and Information Science (ICIS)*, pp. 12–17, IEEE.
- 711 Preiss A, Rosenberg UB, Kienlin A, Seifert E, Jäckle H. 1985. Molecular genetics of Krüppel, a gene
712 required for segmentation of the *Drosophila* embryo. *Nature* **313**: 27–32.
- 713 R Core Team (2020). R: A language and environment for statistical computing. R
714 Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 715 Ramilowski JA, Goldberg T, Harshbarger J, Kloppman E, Lizio M, Satagopam VP, Itoh M, Kawaji H,
716 Carninci P, Rost B, et al. 2015b. A draft network of ligand-receptor-mediated multicellular
717 signalling in human. *Nature Communications* **6**.
- 718 Ressler KJ, Sullivan SL, Buck LB. 1994. Information coding in the olfactory system: Evidence for a
719 stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**: 1245–1255.
- 720 Rodrigues SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen
721 F, Macosko EZ. 2019. Slide-seq: A scalable technology for measuring genome-wide expression at
722 high spatial resolution. *Science (New York, NY)* **363**: 1463–1467.
723 <http://www.ncbi.nlm.nih.gov/pubmed/30923225> (Accessed September 19, 2019).
- 724 Saiselet M, Rodrigues-Vitória J, Tourneur A, Craciun L, Spinette A, Larsimont D, Andry G, Lundeberg J,
725 Maenhaut C, Detours V. 2020. Transcriptional output, cell types densities and normalization in
726 spatial transcriptomics. *Journal of Molecular Cell Biology*.
- 727 Smillie CS, Biton M, Ordovas-Montanes J, Sullivan KM, Burgin G, Graham DB, Herbst RH, Rogel N,
728 Slyper M, Waldman J, et al. 2019. Intra- and Inter-cellular Rewiring of the Human Colon during
729 Ulcerative Colitis. *Cell* **178**: 714-730.e22.
- 730 Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M,
731 Westholm JO, Huss M, et al. 2016. Visualization and analysis of gene expression in tissue sections
732 by spatial transcriptomics. *Science (New York, NY)* **353**: 78–82.
733 <http://www.ncbi.nlm.nih.gov/pubmed/27365449>

- 734 Sun S, Zhu J, Zhou X. 2020. Statistical analysis of spatial expression patterns for spatially resolved
735 transcriptomic studies. *Nature Methods* **17**: 193–200.
- 736 Svensson V, Teichmann SA, Stegle O. 2018. SpatialDE: identification of spatially variable genes. *Nature*
737 *Methods* **15**: 343–346.
- 738 Tanaka M, Maeda N, Noda M, Marunouchi T. 2003. A chondroitin sulfate proteoglycan PTPzeta
739 /RPTPbeta regulates the morphogenesis of Purkinje cell dendrites in the developing cerebellum. *The*
740 *Journal of neuroscience* □: *the official journal of the Society for Neuroscience* **23**: 2804–14.
- 741 Vassar R, Chao SK, Sitcheran R, Nuñez JM, Vosshall LB, Axel R. 1994. Topographic organization of
742 sensory projections to the olfactory bulb. *Cell* **79**: 981–91.
- 743 Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, Park J-E,
744 Stephenson E, Polański K, Goncalves A, et al. 2018. Single-cell reconstruction of the early
745 maternal–fetal interface in humans. *Nature* **563**: 347–353.
- 746 Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan
747 C, Liu J, et al. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states.
748 *Science (New York, NY)* **361**: eaat5691.
- 749 Xia C, Fan J, Emanuel G, Hao J, Zhuang X. 2019. Spatial transcriptome profiling by MERFISH reveals
750 subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proceedings of*
751 *the National Academy of Sciences of the United States of America* 201912459.
752 <http://www.ncbi.nlm.nih.gov/pubmed/31501331>.
- 753 Xu X, Coats JK, Yang CF, Wang A, Ahmed OM, Alvarado M, Izumi T, Shah NM. 2012. Modular
754 Genetic Control of Sexually Dimorphic Behaviors. *Cell* **148**.
- 755 y Cajal S. 1911. Histologie du système nerveux de l’homme et des vertébrés. *Maloine, Paris* **2**: 153–173.
- 756 Zhuang X. 2021. Spatially resolved single-cell genomics and transcriptomics by imaging. *Nature*
757 *Methods* **18**: 18–22..
- 758
- 759

760

761 **Figure Legends**

762

763 **Figure 1. Overview of MERINGUE. A.** MERINGUE encodes spatial relationships among spatially
764 resolved datasets, such as cells, using a binary adjacency weight matrix W . Two cells are considered
765 adjacent if their neighborhoods inferred from Voronoi tessellation share an edge (left). The binary
766 adjacency weight matrix W is visualized by plotting all cells in space with a red line connecting cells if
767 cells are adjacent and no line otherwise (right). **B.** MERINGUE identifies genes with spatially
768 heterogeneous expression using W to compute and evaluate the significance of a spatial auto-correlation
769 index I for each gene. When a gene's expression magnitude between spatially adjacent cells are highly
770 correlated, I will be positive, indicative of spatial heterogeneity. Three simulated gene expression profiles
771 are provided to illustrate examples of high and low spatial heterogeneity with red indicating high
772 expression and blue indicating low expression. **C.** MERINGUE groups identified spatially heterogeneous
773 genes into primary spatial patterns by computing a spatial cross-correlation index for every gene pair. The
774 resulting spatial cross-correlation matrix is used to construct a hierarchical dendrogram (top). Dynamic
775 tree cutting is applied to partition genes into patterns (bottom). Groups of genes are z-scored and
776 averaged, with empty regions filled in using Akima interpolation to visualize final patterns. **D.**
777 MERINGUE identifies gene expression patterns that may be indicative of putative cell-cell
778 communication using an inter-cell-type weight matrix W_{ict} between two cell-types, which can then be
779 used to compute an inter-cell-type spatial cross-correlation index $iSCI$ between two genes. Two cell-types
780 A and B are shown as green triangles and orange squares respectively. W_{ict} is visualized with a red line
781 for cells of cell-type A spatially adjacent to cells of cell-type B (top left). Cell-type A cells express gene A
782 at variable levels while cell-type B cells express gene B at variable levels, with red indicating high
783 expression and blue indicating low expression (top right). Cell-type A cells do not express gene B and
784 cell-type B cells do not express gene A, resulting in a generally negative Pearson's correlation (R)
785 between the two genes (bottom left). However, expression of gene A in cells of cell-type A is highly

786 correlated with the expression of gene B in spatially adjacent cells of cell-type B, resulting in a positive
 787 *iSCI*. The significance of this *iSCI* is assessed by permutation (bottom right).

788

789 **Figure 2. Application of MERINGUE to 2D spatial transcriptomics data of the main olfactory bulb**

790 **(MOB) and 3D aligned ISH data of the *Drosophila melanogaster* embryo. A.** Spatially un-aware

791 single-cell clustering analysis identifies 5 transcriptionally distinct clusters corresponding to various

792 known cell layers in the MOB. Spatial spots are colored based on their inferred cell layer annotation. **B.**

793 MERINGUE identifies genes with significantly spatially heterogeneous expression in the MOB. Select

794 genes are shown. **C.** MERINGUE groups genes with significantly spatially heterogeneous expression in

795 the MOB into 5 primary spatial patterns. Select patterns are shown. **D.** MERINGUE's adjacency weight

796 matrix visualized for aligned 3D in situ hybridization data of the *Drosophila melanogaster* embryo. Each

797 point is an aligned cell. Cells are connected with a red line if they are inferred to be adjacent. A top view

798 and rotated side view are shown. **E.** MERINGUE groups genes into spatial patterns in the *Drosophila*

799 *melanogaster* embryo. Representative genes from select identified patterns are shown.

800

801 **Figure 3. MERINGUE identifies spatial heterogeneity within cell-types in the preoptic region of the**

802 **mouse hypothalamus using MERFISH. A.** (left) Expression of three sample genes in Inhibitory I-6

803 neurons in female naïve animal 7 (FN7). Each point is a cell. Cells are colored by expression with red

804 denoting high expression and blue denoting low expression. Cells that are not I-6 cells are colored in

805 grey. *Gad1* is highly expressed in all I-6 neurons while *Necab1* and *Sema3c* exhibit significant spatial

806 variation. (right) Expression of *Sema3c* in I-6 in adjacent tissue sections in FN7 exhibit similar spatial

807 patterning. **B.** (left) Expression of three sample genes in Inhibitory I-11 neurons in FN7. Again, *Gad1* is

808 highly expressed in all I-11 cells while *Gabra1* and *Nos1* exhibit significant spatial variation. (right)

809 Expression of *Nos1* in I-11 neurons in adjacent tissue sections in FN7 exhibit similar spatial patterning.

810 **C.** Expression of *Nos1* in I-11 neurons in representative male naïve animals MN5, MN8, and MN9 show

811 similar spatial patterns to the female animal in (B). **D.** Fraction of *Nos1*+ I-11 cells for male and female

812 animals across tissue layers from the anterior to posterior preoptic region. Each dot represents one tissue
813 layer in one animal. Lines represent fitted curves for males and female animals. **E.** Expression of *Tacr1*
814 in E-8 neurons in female (top) and male (bottom) animals. **F.** Scale spatial heterogeneity of *Tacr1* in E-8
815 neurons for male and female animals across tissue layers from the anterior to posterior preoptic region.
816 Each dot represents one tissue layer in one animal. Lines represent fitted curves for males and female
817 animals.

818

819 **Figure 4. Spatially informed clustering distinguishes spatially distinct subpopulations of cells. A.**
820 Expression-based clustering of 3035 stage 6 *Drosophila melanogaster* embryo cells with 84 marker genes
821 by aligned ISH identifies approximately 5 transcriptionally distinct clusters. (top) UMAP embedding
822 colored by identified cluster annotations. (bottom) Spatial coordinates colored by identified cluster
823 annotations. **B.** Expression of select marker genes on the UMAP embedding with red denoting high
824 expression and blue denoting low expression. **C.** Spatially informed clustering splits expression-based
825 clusters in spatially coherent manner. Again (top) UMAP embedding colored by identified spatially
826 informed cluster annotations. (bottom) Spatial coordinates colored by identified spatially informed cluster
827 annotations. **D.** Correspondence between expression-based clusters in (A) and spatially informed clusters
828 in (C) highlights high correspondence between most clusters with the exception of one cluster being split
829 into two. **E.** UMAP embedding of populous inhibitory neuronal subtypes in one posterior preoptic tissue
830 section from one animal measured using MERFISH, where each point is a cell colored by the original
831 subtype annotations. **F.** Same UMAP embedding as (E) where each point is a cell colored by the spatially
832 informed clustering annotation. Black dashed lines highlight clusters that have now split. **G.**
833 Correspondence between expression-based clusters in (E) and spatially informed clusters in (F) highlights
834 high correspondence between most clusters with the exception of cells originally annotated as I-2 and I-
835 11 now being split into two. **H.** Same UMAP embedding as (E) where each point is a cell colored by
836 *Nos1* expression for cells originally annotated as I-11. **I.** Spatial location of cells within the tissue colored
837 by their spatially informed cluster assignment for cells originally annotated as I-11. **J.** Same UMAP

838 embedding as (E) where each point is a cell colored by *Clpx3* expression for cells originally annotated as
 839 I-2. **K.** Spatial location of cells within the tissue colored by their spatially informed cluster assignment for
 840 cells originally annotated as I-2. Regions corresponding to the BNST and SDN are highlighted with blue
 841 and red dashed lines respectively. Representative slice in representative animal shown. **L.** Percentage of
 842 activated cells based on *c-Fos* expression during female (FS) and male (MS) sexual behavior for spatially
 843 informed clusters C3 and C8 originally annotated as I-2. Boxes in the boxplot denote the median values
 844 and inner quartile ranges (IQR) and whiskers denote 1.5xIQR with additional outliers represented as
 845 points.

846

847 **Figure 5. MERINGUE identifies putative cell-cell communication in the cerebellum using SlideSeq**

848 **data.** **A.** Adjacency relationship between SlideSeq beads. Orange dots correspond to beads previously
 849 annotated as corresponding to the Purkinje layer. Green dots correspond to beads that are spatially
 850 adjacent. Grey lines connect each bead with its spatial neighbors and is agnostic to bead density. **B.** (left)
 851 Expression of receptor *Ptprz1* in beads annotated to correspond to the Purkinje layer. (right) Expression
 852 of corresponding ligand *Ptn* in spatially adjacent beads. Same select region highlighted. **C.** Barplot of -
 853 \log_{10} (adjusted p-value) for the inter-cell-type spatial cross correlation statistic of all receptors in Purkinje
 854 layer beads versus ligands in the spatially adjacent beads. Red line indicates alpha = 0.2 multiple testing
 855 corrected significance threshold. **D.** (left) Expression of ligand *Psap* in beads annotated to correspond to
 856 the Purkinje layer. (right) Expression of corresponding receptor *Gpr3711* in spatially adjacent beads.
 857 Same select region highlighted as (D). **E.** Barplot of $-\log_{10}$ (adjusted p-value) for the inter-cell-type spatial
 858 cross correlation statistic of all ligands in Purkinje layer beads versus receptors in the spatially adjacent
 859 beads.

860

861 **Figure 6. MERINGUE systematically and quantitatively evaluates for putative cell-cell**

862 **communication for neuronal subtypes in the preoptic region using MERFISH data.** **A.** Distribution
 863 of $-\log_{10}$ (p-values) for the spatial cross-correlation between aromatase (*Cyp19a1*) expression in neuronal

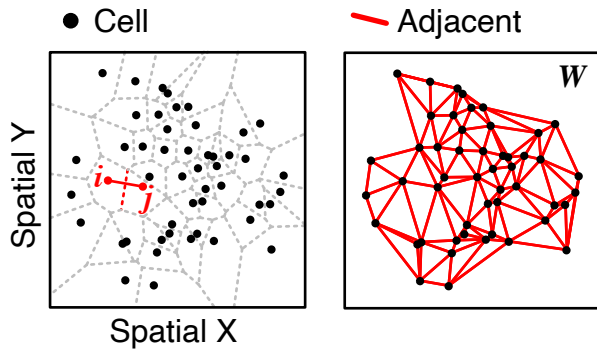
864 subtypes and *Esr1* expression in adjacent cells across animals. Boxes in the boxplot denote the median
865 values and inner quartile ranges (IQR) and whiskers denote 1.5xIQR with additional outliers represented
866 as points. Red dotted line is the $\alpha=0.05$ significance threshold. Generally, inhibitory neuron subtype I-
867 2 exhibits significant spatial cross-correlation between aromatase expression and *Esr1* expression in
868 adjacent cells in a manner that is consistent across animals. **B.** Aromatase (*Cyp19a1*) expression in I-2
869 neurons in one tissue slice in one animal with red indicating high expression and blue indicating low
870 expression. Representative slice and animal shown. Select areas are highlighted in the zoom-in. **C.** *Esr1*
871 expression in cells neighboring I-2 neurons in one tissue layer in one animal with red indicating high
872 expression and blue indicating low expression. Representative slice and animal shown. The same select
873 areas as (B) are highlighted in the zoom-in.

874

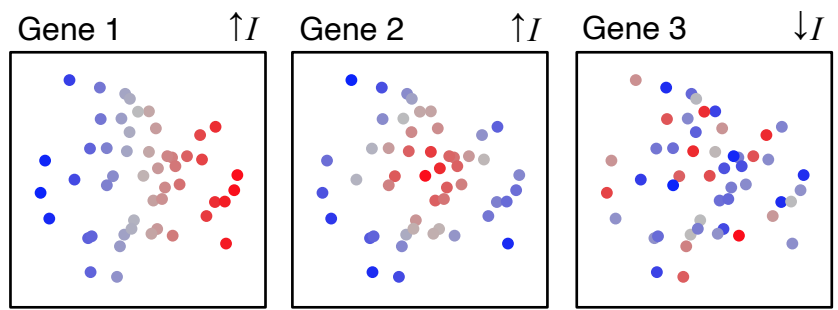
875

876

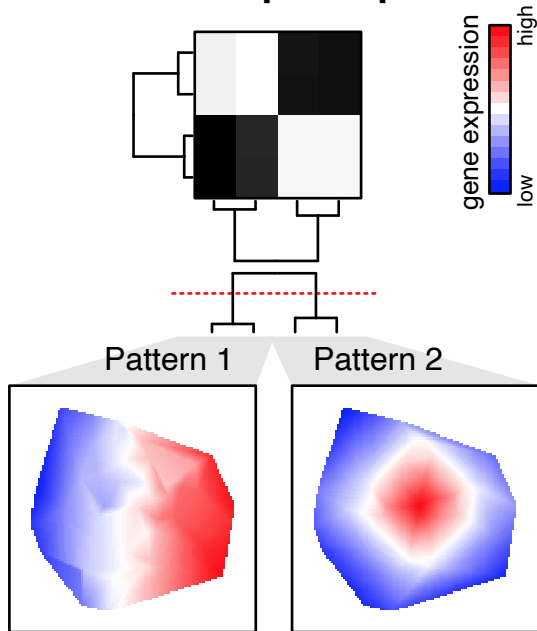
A encode spatial relationships



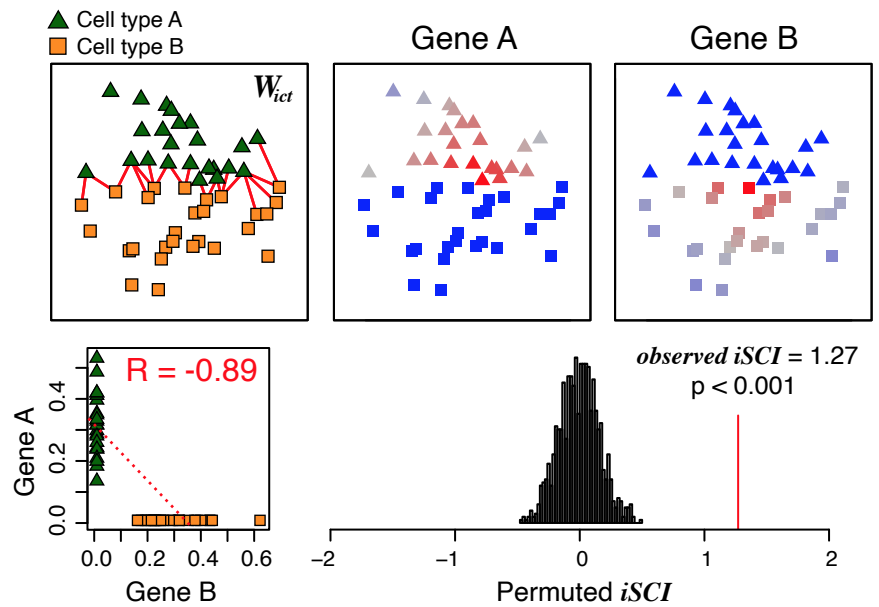
B identify spatial expression heterogeneity

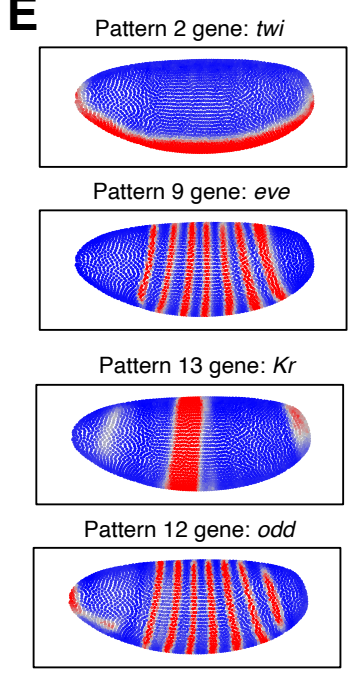
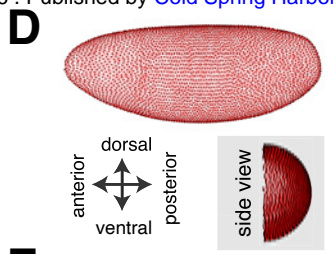
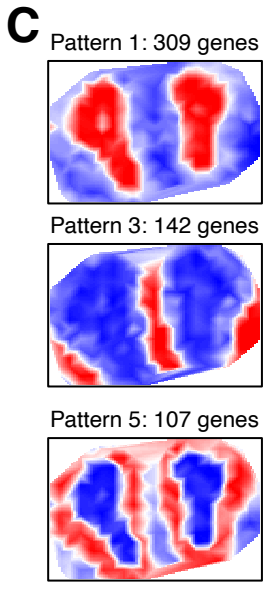
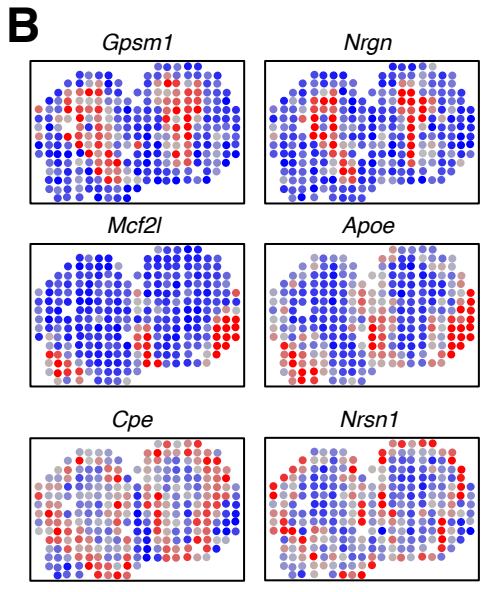
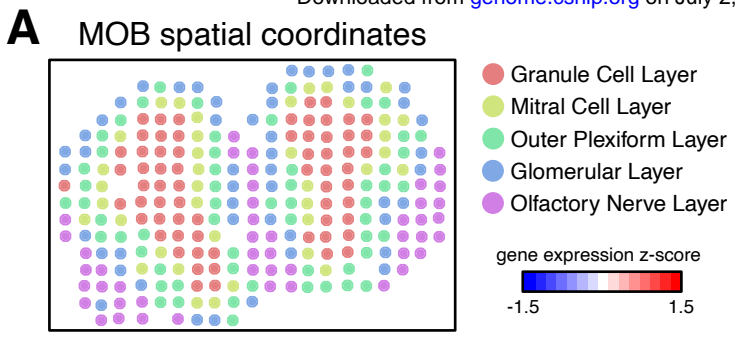


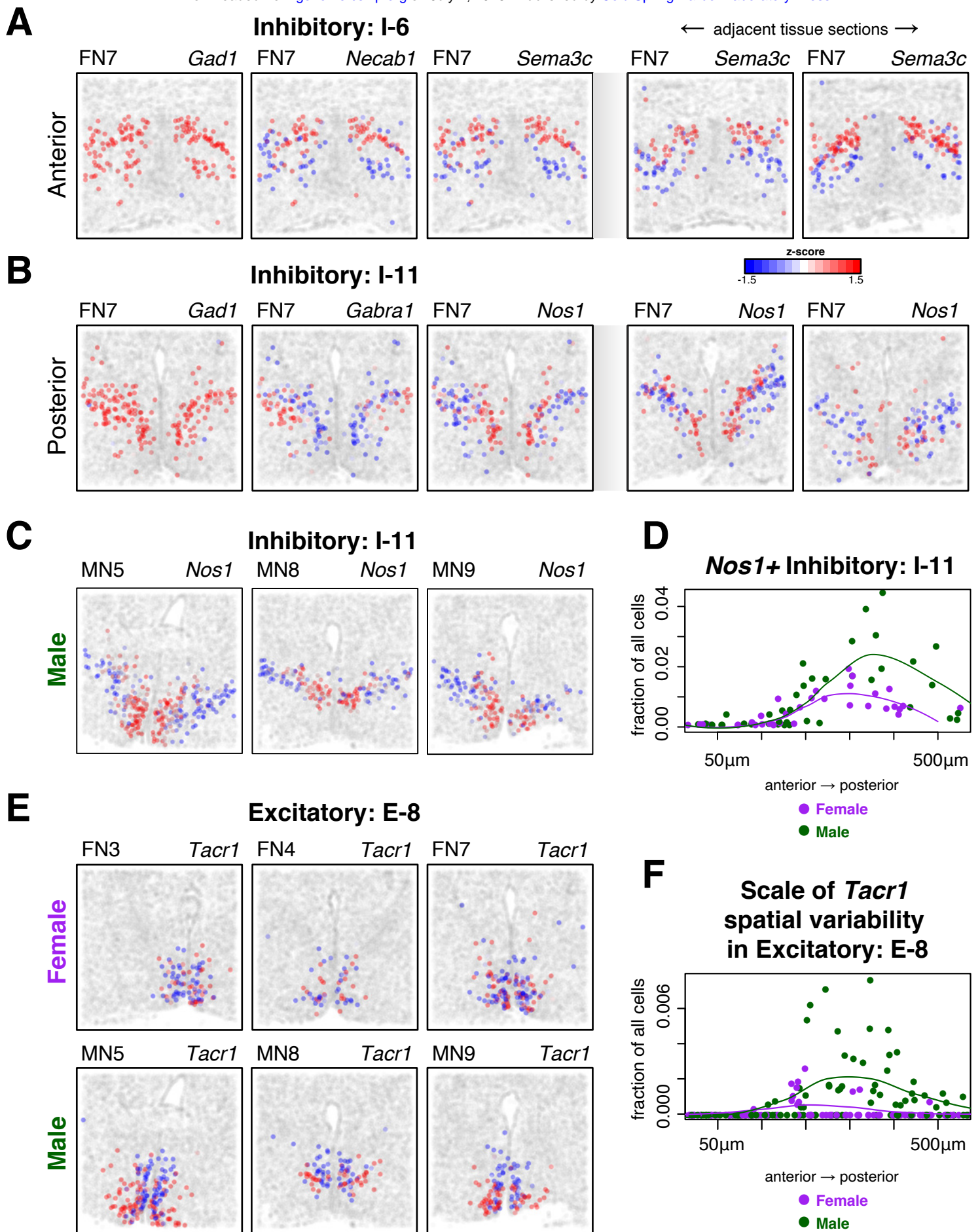
C summarize spatial patterns

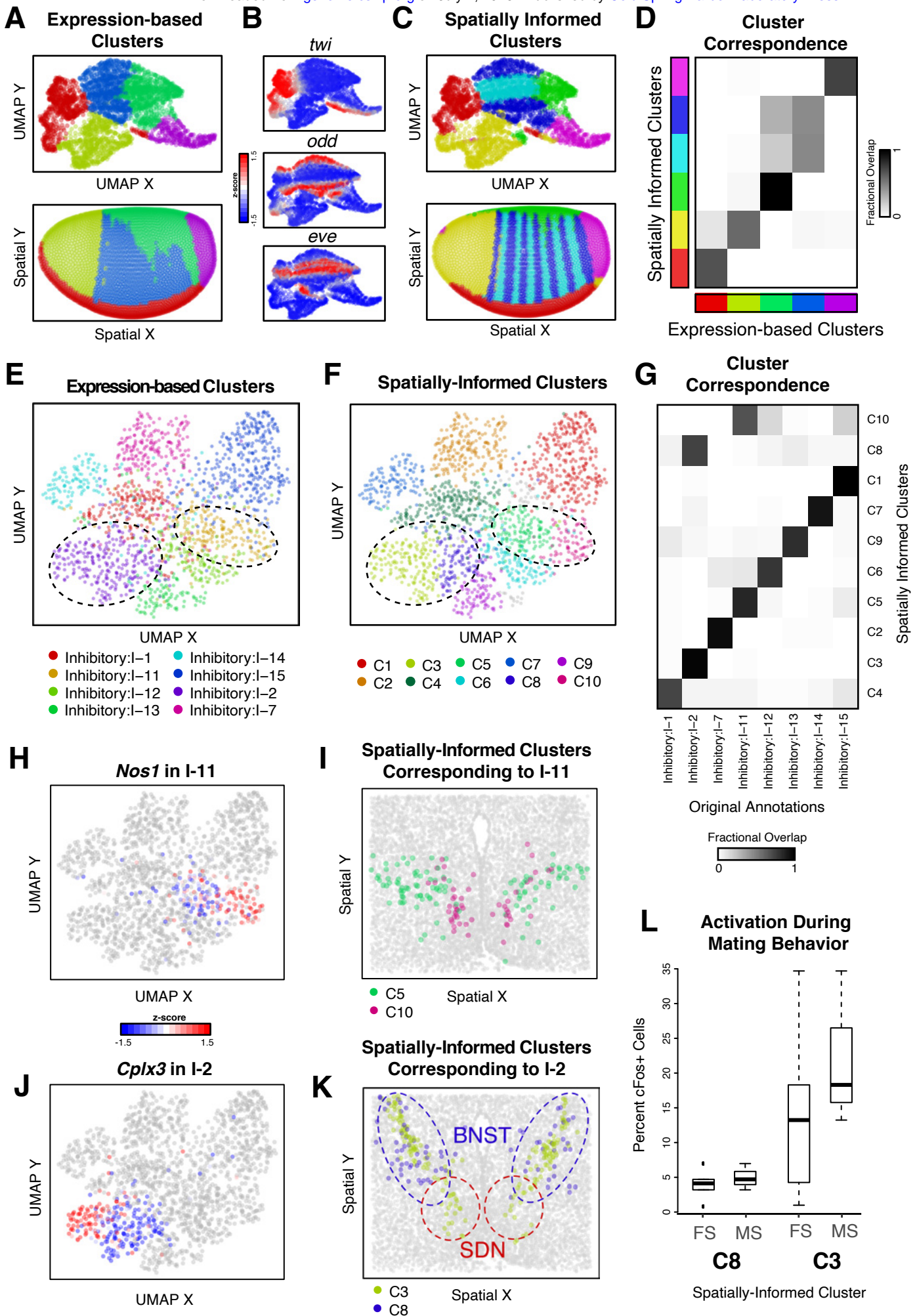


D identify gene expression patterns indicative of cell-cell communication



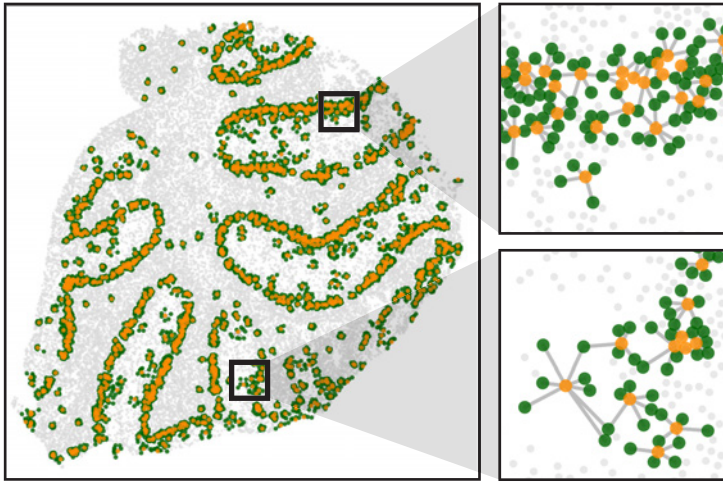






A

● Purkinje
● Adjacent

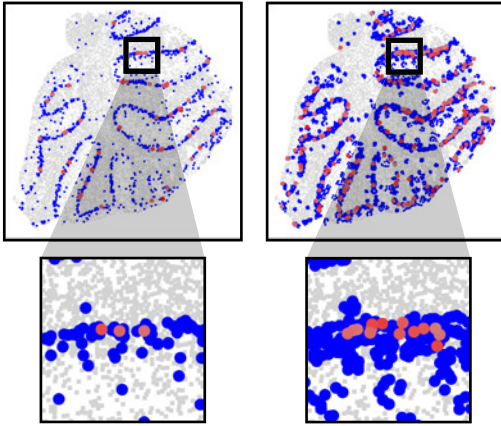


B

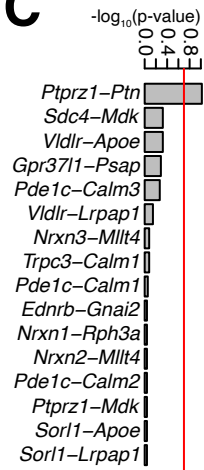
Ptprz1

Ptn

gene expression
low high



C

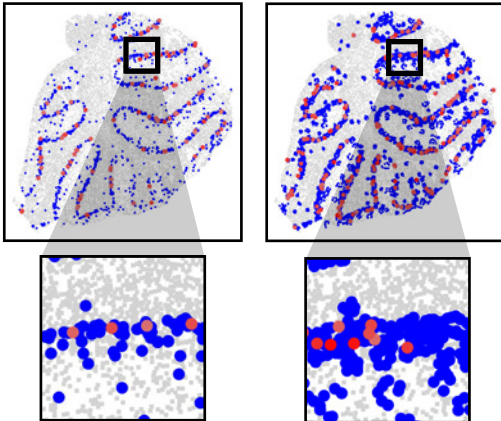


D

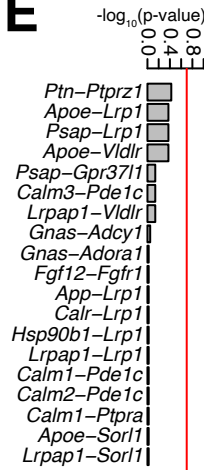
Psap

Gpr3711

gene expression
low high



E



A

significance of spatial cross-correlation between *Cyp19a1* in neuronal subtypes and *Esr1* in neighbors across animals

