Cross-expression analysis reveals patterns of coordinated gene expression in spatial transcriptomics

Ameer Sarwar[1], Mara Rue[2], Leon French[3], Helen Cross[2], Xiaoyin Chen[2], and Jesse Gillis[3]*

[1]Department of Cell and Systems Biology and Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

[2]Allen Institute for Brain Science, Seattle, WA, USA

[3]Department of Physiology and Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

*Corresponding Author, jesse.gillis@utoronto.ca

Abstract

Spatial transcriptomics promises to transform our understanding of tissue biology by molecularly profiling individual cells *in situ*. A fundamental question they allow us to ask is how nearby cells orchestrate their gene expression. To investigate this, we introduce cross-expression, a novel framework for discovering gene pairs that coordinate their expression across neighboring cells. Just as co-expression quantifies synchronized gene expression within the same cells, cross-expression measures coordinated gene expression between spatially adjacent cells, allowing us to understand tissue gene expression programs with single cell resolution. Using this framework, we recover ligand-receptor partners and discover gene combinations marking anatomical regions. More generally, we create cross-expression networks to find gene modules with orchestrated expression patterns. Finally, we provide an efficient R package to facilitate cross-expression analysis, quantify effect sizes, and generate novel visualizations to better understand spatial gene expression programs.

Spatial transcriptomics records cells' gene expression alongside their physical locations, enabling us to understand how they influence one another within the broader tissue context[1]. Focusing on select genes, imaging-based platforms profile expression at the single cell level, giving a high-resolution snapshot of spatial gene expression[2–7]. They have facilitated numerous studies on defining local spatial patterns[8–11], finding gene covariation in spatial niches[12–16], elucidating cell-cell interactions using ligand-receptor expression[17–27], and determining spatial cell type heterogeneity and tissue structure[10,28–30]. These efforts have resulted in a greater understanding of tissue biology, culminating in the generation of reference atlases[31–34].

Imaging-based platforms can now profile up to a few thousand genes in millions of cells[2–7], generating large amounts of data ripe for biological discovery. Historically, large-scale global gene networks have been instrumental in uncovering fundamental biological processes by leveraging the power of high-throughput data to compute gene-gene interactions[35,36]. A promising approach to exploiting large-scale single-cell RNA-seq data is gene co-expression analysis[37–39], which investigates how genes covary within cells and therefore discovers modules of functionally related genes. Extending this concept to spatial transcriptomics, a recent study[12] characterized gene covariation within well-defined spatial niches, finding continuous gradients during spinal cord development and localizing cortical somatostatin-positive interneuron subtypes. Another study[40] used co-expression to create hierarchical tissue structures, revealing the multi-scale organization of the hippocampus. While these studies fruitfully apply gene co-expression within cells to characterize tissue structure at multiple spatial scales, the co-expression framework is silent on patterns between cells as, for example, when a gene expression program in one cell gives rise to a complementary pattern in a neighboring cell.

Here we introduce cross-expression, a novel conceptual and statistical framework to understand coordinated gene expression as a network *between* neighboring cells. Whereas co-expression captures gene covariation within the same cells, cross-expression measures their coordination between neighboring cells, thereby highlighting how gene expression is orchestrated across the tissue. By developing methods to focus on the conjugate network to cell-cell interaction networks, we are able to investigate novel features that characterize individual genes, cells, and shared patterns across both. For example, we create a cross-expression network, finding that *Gpr20*, a G protein-coupled receptor that appears to line the blood vessels, is a central gene with high node degree and defines visible spatial tracts. Within the same network, we discover an interacting subset of genes enriched in astrocyte-mediated regulation of vascular processes, an essential biological function requiring spatially proximal gene expression. Investigating the relationship between cross-expression and cell type composition, we find that cross-expression

is frequently driven by compositional differences, where 64% of cross-expressing cell pairs have different cell subtype labels. Using cross-expression to discover anatomical marker combinations, we find gene pairs that cross-express specifically in cells that are located in the thalamus even though the individual genes are highly expressed in other regions. To investigate known gene patterns, we use BARseq[34,41] to collect mouse whole-brain data with a gene panel containing select ligands and receptors, finding that these genes are highly cross-expressed, thus confirming previous reports that typically target known interaction partners. Collectively, our unbiased framework fully leverages the spatial dimension at the cellular resolution to discover novel genes with coordinated expression, helping us better understand how cells influence one another in tissue.

Results

Cross-expression overview

Just as co-expression between two genes in single cell data can be conceptualized as the degree to which knowing one gene's expression in a given cell predicts the other gene's expression in the same cell, cross-expression is the degree to which knowing the expression of one gene in a given cell predicts the expression of another gene in a spatially related cell, typically the neighbor. One trivial case where this can occur is when two cells exhibit the same expression pattern; here, the prediction of the neighboring cell effectively captures co-expression and cell type composition. To exclude this, we define cross-expression as the predictions of neighbor expression where co-expression alone provides no performance. Specifically, cross-expression occurs where there is a consistent pattern in which gene A is expressed in a cell without gene B, and gene B is expressed in its neighbor without gene A (Fig. 1a).

To quantify cross-expression (Fig. 1b), we first consider cell-neighbor pairs where the cells express gene A. We next test if the neighbors express gene B. If many do, given gene B's incidence in the population at large, then these genes are said to cross-express. Additionally, we quantify the effect size by comparing the number of neighbors expressing gene B to the number of cells co-expressing genes A and B (Fig. 1c). Using this procedure on *n* nearest neighbors, we filter for a bullseye-like distribution with low co-expression (center) and high cross-expression (rings). In subsequent text, "gene A" and "gene B" refer to their expression in the central (or reference) and neighboring (or spatially adjacent) cells, respectively, unless indicated otherwise.

To explore cross-expression, we use imaging-based spatial transcriptomics[2–7] for several reasons. First, these platforms profile gene expression at the single cell level, allowing us to ask how individual cells influence each other. Second, they share common steps, such as transcript

identification and cell segmentation, that allow consistent downstream analysis and interpretation. Third, these platforms have been used to generate large amounts of data[2–5], making them suitable for developing and validating the computational framework underlying cross-expression.

Although we focus on individual cells, groups of cells may form spatial niches and gene expression may be coordinated between niches. To assay cross-expression at this coarser resolution, we average a gene's expression in a cell with its expression in the neighbors (Fig. 1d), thus smoothing it within a spatial niche, with the number of neighbors forming the niche size. Accordingly, cross-expression can be compared across niches by, for example, finding associations between smoothed niche-specific gene expression profiles.

To enable these analyses, we provide an efficient software package in R that requires the gene expression and cell location matrices as inputs, and outputs a gene-gene p-value matrix that facilitates downstream analyses, such as cross-expression network construction (Fig. 1e). The package also contains functions for computing effect sizes, making bullseye plots, smoothing gene expression, viewing cross-expressing cells *in situ*, and assessing if cross-expression is spatially enriched. Collectively, the cross-expression framework uses spatial information to discover how genes coordinate their expression across neighboring cells, thereby providing a novel analytical framework for deeply exploring spatial transcriptomic data.

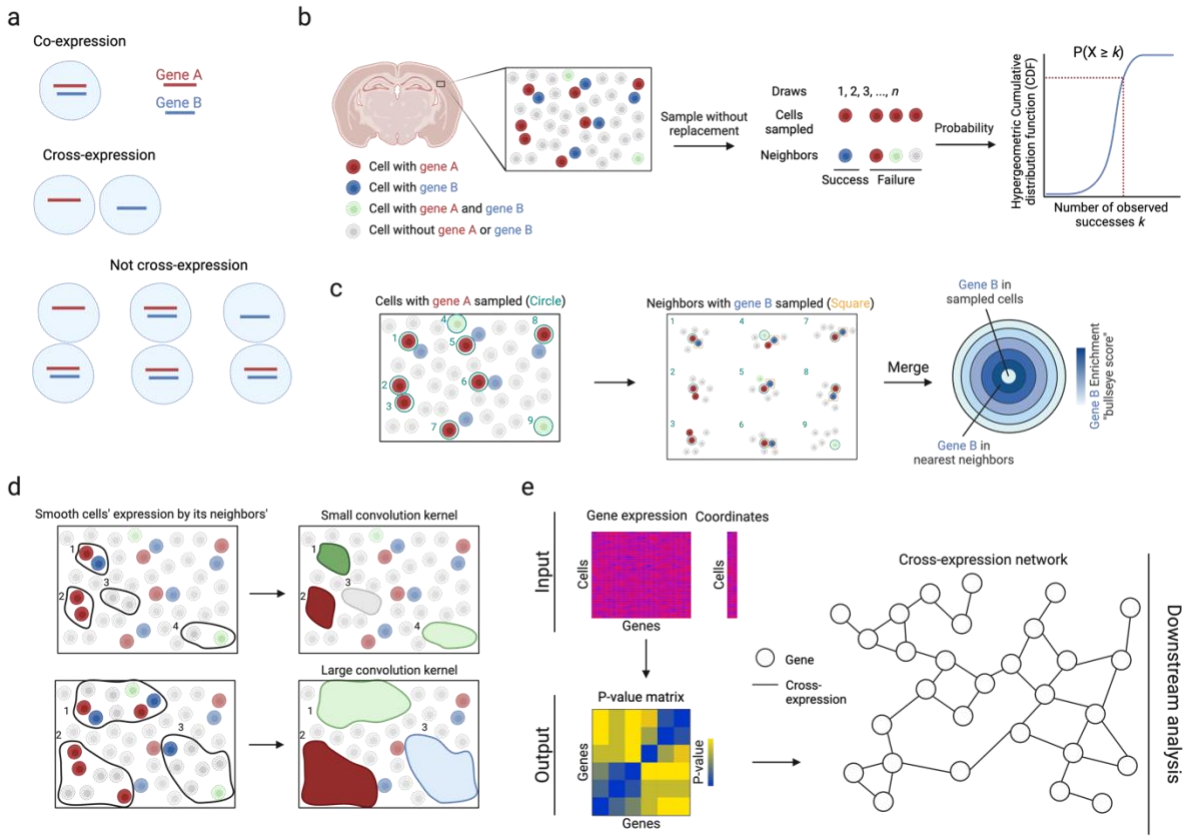


Fig. 1 | Cross-expression analysis. a, Cross-expression is the mutually exclusive expression of genes between neighboring cells. If either cell expresses both genes, the cell pair is not considered to cross-express. **b**, The probability that two genes cross-express is modeled by the hypergeometric distribution, where all the cells expressing gene A are sampled and their neighbors expressing gene B are deemed as 'successful trials'. **c**, Cross-expression is compared to co-expression to quantify the effect size, where the number of neighbors with gene B is compared to the number of cells co-expressing genes A and B. 'Sampled cells' (center) are those expressing gene A and neighbors are concentric rings, with the order indicating the *n*-th neighbor. **d**, Averaging gene expression between cells and their neighbors smooths it, extending cross-expression analysis from cell pairs to regions. Number of neighbors is the kernel size. **e**, Software inputs are the gene expression and cell location matrices, and the output is a p-value matrix, which enables downstream analyses, such as cross-expression network construction. Created with BioRender.com.

Cross-expression recovers ligand-receptor pairs and reveals coordinated gene expression profiles across the tissue

To study cross-expression, we used BARseq (barcoded anatomy resolved by sequencing) to collect data from a whole mouse brain. This dataset profiled expression in 1 million cells across 15 sagittal slices, using a gene panel of 104 cortical cell type markers and 25 ligands and receptors, including neuropeptides, their receptors, and monoamine neuromodulatory receptors. Because receptors and their corresponding ligands are often expressed in nearby cells[17–27], we

133 reasoned that these genes should show cross-expression. As an example, we find that across
134 the cortical somatosensory nose region and visceral areas (Fig. 2a), the neuropeptide
135 somatostatin *Sst* and its cognate receptor *Sstr2* are cross-expressed (Fig. 2b, left, p-values ≤ 0.01
136 and 0.05, respectively). Indeed, these genes are consistently expressed across neighboring cells
137 (Fig. 2b, right), a pattern that is otherwise difficult to discover without prior knowledge.

138 Next, we explore the bullseye plots, which allow us to quantify the effect size by comparing
139 cross-expression with co-expression. For *Sst* and *Sstr2* in the somatosensory nose (2,015 cells)
140 and visceral regions (1,603 cells), we see a bullseye pattern with low co-expression and high
141 cross-expression that decreases for distant neighbors (Fig. 2c). Specifically, for these regions the
142 bullseye score ratio between the first neighbor and the central cell is 1.8 and 1.6, respectively,
143 whereas the ratio between the averaged second-to-tenth neighbor and the central cell is 1.3 and
144 1.2. These findings suggest that for central cells expressing one gene in a pair, a higher proportion
145 of adjacent neighbors, but not the more distant ones, express the other gene within the local
146 spatial niche, underscoring the specificity and resolution with which patterns of coordinated gene
147 expression can be recovered. We next compare the bullseye plots for gene pairs with and without
148 cross-expression (Fig. 2d), finding that the former match the patterns just described. To quantify
149 this, we compare the bullseye scores of the nearest neighbors with those of cells expressing gene
150 A, discovering that this ratio is much greater for genes that cross-express than for those that do
151 not (Fig. 2d, inset, Mann-Whitney U test, p-value ≤ 0.001, median ratios: 1.5 and 0.9,
152 respectively). Notably, this ratio is approximately 1 for genes that do not cross-express,
153 suggesting that here gene B is expressed in neighbors and cells alike. Hence, the bullseye
154 approach intuitively visualizes and quantifies the effect size, making it suitable for downstream
155 analysis, such as comparing cross-expression between different regions.

156 We next conducted brain-wide analysis and found that 20% of possible ligand-receptor
157 gene pairs and 4% of non-signaling gene pairs are cross-expressed, thus generating novel
158 candidates that potentially encode functionally relevant interactions. In fact, these patterns are
159 spatially enriched, where most gene pairs cross-express in a few slices and some cross-express
160 in multiple slices (Extended Data Fig. 1a). We now highlight some notable examples of cross-
161 expression for both signaling and non-signaling genes. The dopamine receptor D_1 (*Drd1*) and
162 proenkephalin (*Penk*) are strongly cross-expressed (Extended Data Fig. 1b), with discernible
163 spatial enrichment in the striatal regions. *Drd1* is involved in the reward system[42,43] while *Penk*
164 generates opioids that modulate fear response[44] and nociception[45,46], suggesting that these genes
165 may be involved in avoidance behavior. Indeed, Penk is strongly co-expressed with the dopamine
166 receptor D_2 (*Drd2*) (Pearson's R = 0.72 in scRNA-seq striatal data; *Drd2* is not in our gene panel),

167 indicating that the D1 and D2 neurons are spatially intermingled, allowing them to play interrelated
168 roles in motor control[47]. We also find that the somatostatin receptor *Sstr2* cross-expresses with
169 vasoactive intestinal polypeptide receptor 1 (*Vipr1*/*VPAC1*) in the cortex (Extended Data Fig. 1c),
170 suggesting a potential complementary interaction in modulating local neuronal circuits and
171 influencing neuroendocrine signaling pathways[48,49]. Beyond the signaling genes, we note that the
172 fibril-associated *Col19a1* (collagen type XIX alpha 1 chain), a gene involved in maintaining the
173 extracellular matrix (ECM) integrity[50,51], cross-expresses with *C1ql3* (complement C1q like 3)
174 (Extended Data Fig. 1d), whose secretion in the ECM facilitates synapse homeostasis and the
175 formation of cell-cell adhesion complexes[52,53]. Finally, our analysis reveals that *Marcksl1*
176 (myristoylated alanine-rich C-kinase substrate), which is involved in adherens junctions and
177 cytoskeletal processes[54,55], cross-expresses with actin beta (*Actb*) (Extended Data Fig. 1e),
178 hinting at their possible involvement in local tissue architecture[56]. Taken together, the cross-
179 expression analysis not only reveals expected relationships between signaling molecules, but it
180 also discovers genes implicated in the tissue microenvironment. Accordingly, cross-expression is
181 an unbiased framework for finding genes with orchestrated spatial expression profiles, with
182 potential for novel discovery increasing as the gene panel gets larger.

183 We have thus far investigated cross-expression between cells and their neighbors. Yet,
184 gene expression may be coordinated between more distant neighbors or between large spatial
185 niches. The former is facilitated by changing the rank of the nearest neighbor tested. The latter is
186 enabled by smoothing a gene's expression in a cell by averaging it with its expression in nearby
187 cells, as shown for cortical layer 4 marker *Rorb* (Fig. 2e) and layer 6 marker *Foxp2* (Extended
188 Data Fig. 1f) in the auditory cortex[34].

189 Although cross-expression may appear at varying length scales, we focus our analyses at
190 the single cell level to investigate its signature at the finest resolution.

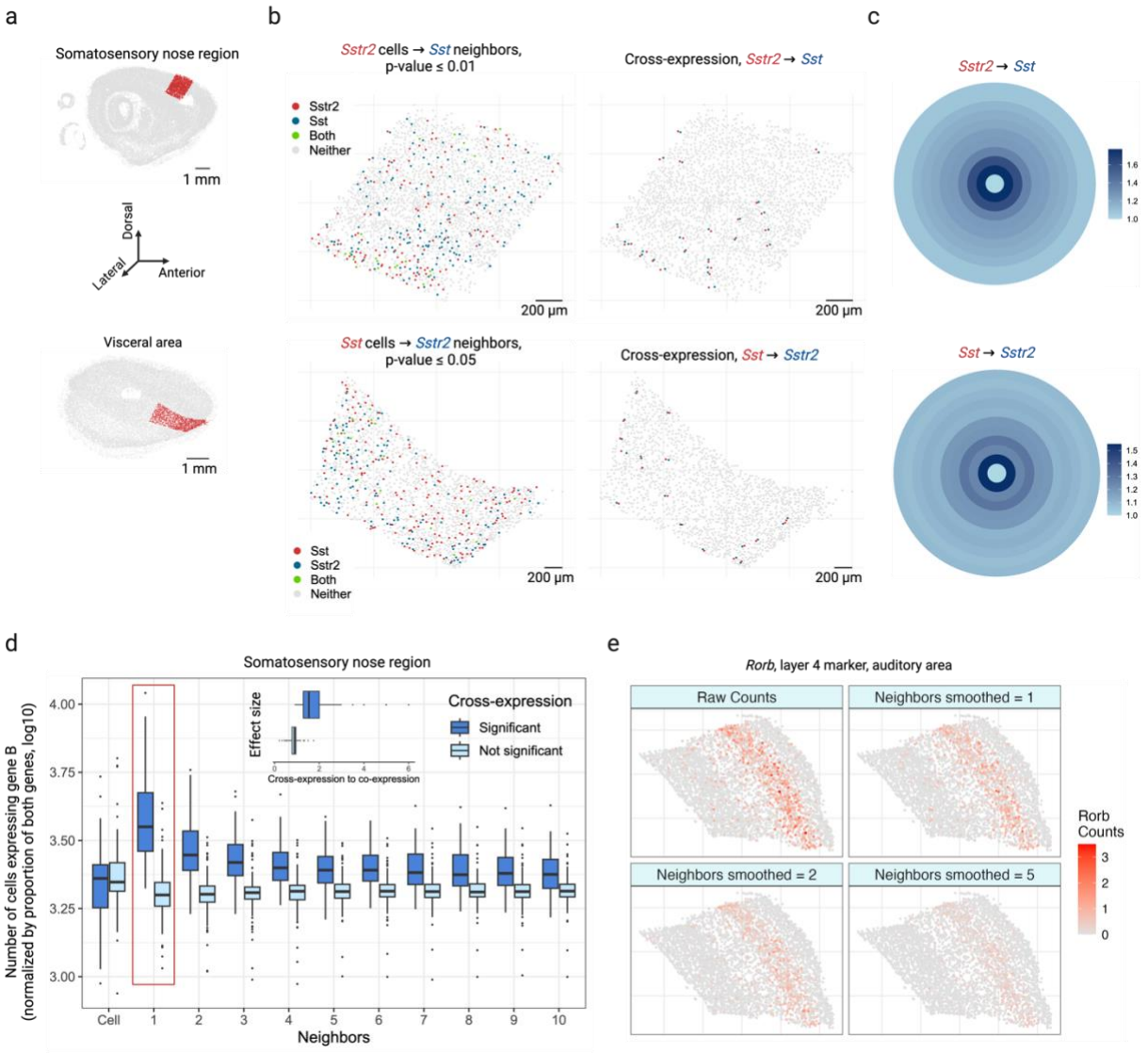


Fig. 2 | Cross-expression analysis reveals coordinated gene expression between neighboring cells. a, Sagittal brain slices showing cortical somatosensory nose region (top) and visceral area (bottom) as randomly selected regions of interest. **b**, Neuropeptide somatostatin *Sst* and its cognate receptor *Sstr2* cross-express in regions shown in (a). Points indicate cells and colors indicate gene expression (left), with cross-expressing cell pairs highlighted (right). **c**, Bullseye scores for *Sst* and *Sstr2* in the regions shown in (a, b). The scores are reported as ratio of cross- to co-expression. **d**, Bullseye scores for cross-expressing (significant) and non-cross-expressing (not significant) gene pairs in the somatosensory nose region. 'Cell' corresponds to the central ring in (c), and the red rectangle highlights the first neighbor/ ring. Inset, ratio of bullseye scores for the first neighbor to the central cells for cross-expressing and non-cross-expressing genes. Central line, median; box limits, first and third quartiles; whiskers, ±1.5x interquartile range; points, outliers. **e**, Smoothed gene expression for different numbers of neighbors for the auditory cortical layer 4 marker gene *Rorb*. Created with BioRender.com.

206 **Cross-expression is driven by subtle and consistent cell subtype compositional**
207 **differences**
208 Having seen that cross-expression recovers coordinated spatial gene expression, we now explore
209 its relationship with cell type heterogeneity. For this purpose, we use another BARseq dataset[34]
210 that was recently used to create a mouse cortical cell type atlas using the same 104 excitatory
211 marker genes as before. Here, we observe that genes cross-express between cells of the same
212 and of different types. For example, *Gfra1* and *Foxp2* are cross-expressed within the same cell
213 type L4/5 IT (intratelencephalic) and between different cell types Car3 or CT (corticothalamic) and
214 L4/5 IT (Fig. 3a). In general, genes vary greatly in terms of the cell type labels of cross-expressing
215 cell pairs (Fig. 3b). For instance, for some gene pairs 40% of the cell pairs have the same cell
216 type label while in others as many as 90% of the cell pairs belong to different cell types (Extended
217 Data Fig. 2a). Moreover, some genes involve many while others involve few cross-expressing
218 cells. For example, in the analyzed data the median number of cross-expressing cell pairs is
219 2,378, and 27% of genes involve over 4,000 while only 5% involve 400 or fewer pairs (Extended
220 Data Fig. 2b), indicating that the density of gene cross-expression is highly variable. Interestingly,
221 cell type purity – the proportion of cell pairs with the same type – decreases as more cell pairs
222 cross-express (Fig. 3c, Spearman's $\rho = -0.46$), highlighting a potential role for spatially
223 intermingled cell types in patterns of cross-expression.
224 To assess the influence of spatial cell type composition more broadly, we use our
225 hierarchical cell type atlas[30], where types at a higher-level divide into subtypes at a lower level.
226 Using cross-expressing glutamatergic cells, we find that 64% of the pairs consist of different cell
227 subtypes (Fig. 3d, right-tailed Wilcoxon signed-rank test, different labels ≥ same labels, p-value ≤
228 0.0001, Extended Data Fig. 2c), suggesting that subtle cell type differences drive cross-
229 expression. However, for cross-expressing GABAergic cells, we find that only 44% of the pairs
230 have different cell subtype labels (Extended Data Fig. 2d-e right-tailed Wilcoxon signed-rank test,
231 different labels ≥ same labels, p-value = 1), reflecting the fact that our gene panel is optimized to
232 detect cell subtype differences between excitatory, but not inhibitory, neurons. Crucially, we
233 observe that cells of one type consistently cross-express with cells of another type (Fig. 3e,
234 Extended Data Fig. 2f), indicating that cross-expression recapitulates patterns of cell type
235 composition. Since cell type labels are assigned based on the expression of many genes,
236 repeated spatial proximity of cell types is one mechanism that generates cross-expression.

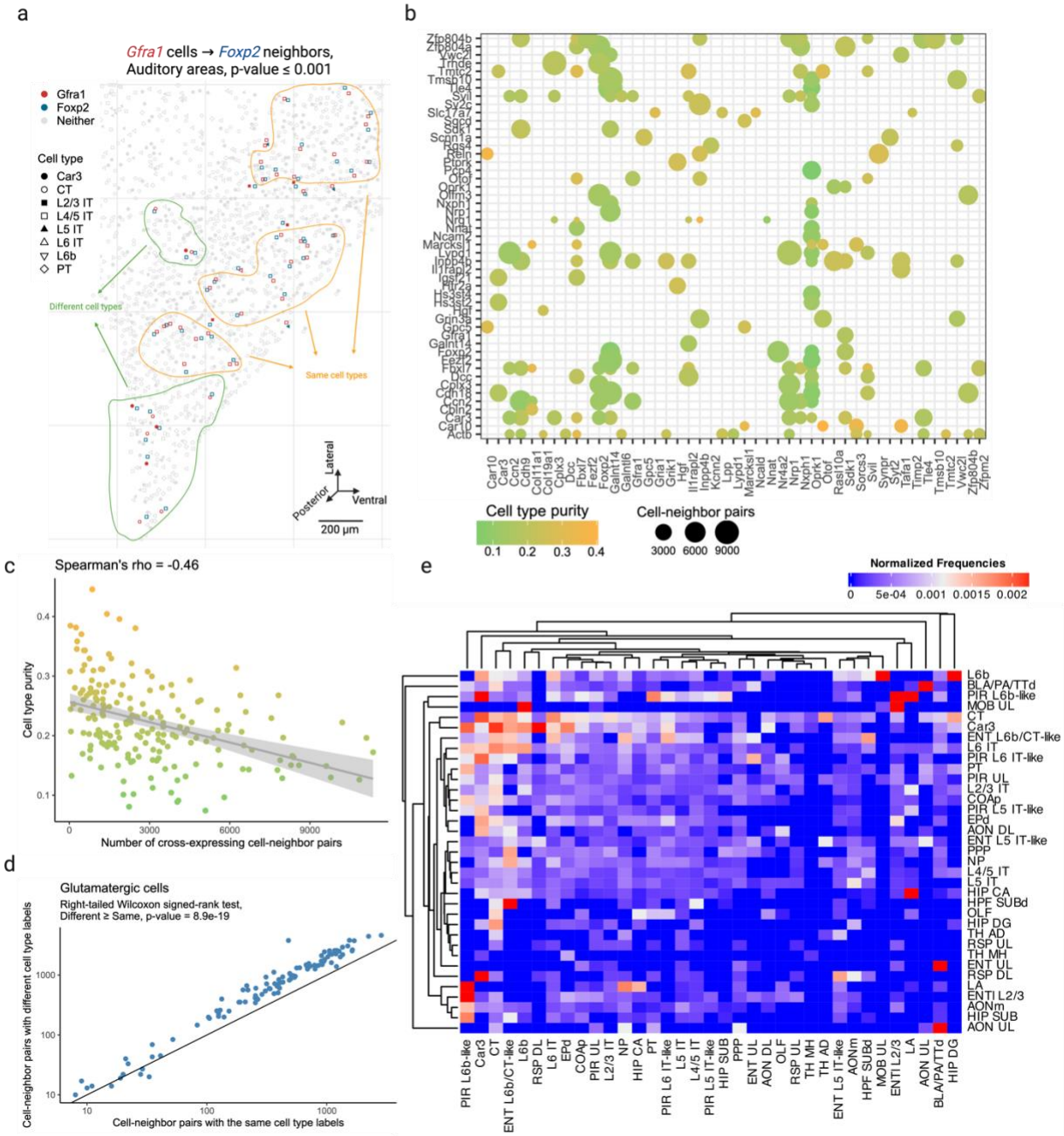


Fig. 3 | Cross-expression patterns are discovered independently of cell type labels but are driven by cell type heterogeneity. a, Cells of the same (yellow) and different (green) types cross-express genes *Gfra1* and *Foxp2* in the auditory cortex. Discovering cross-expression relations between this or any other gene pair does not require cell type labels. **b**, Numerous cells cross-express for each gene pair, with the dot size indicating the number of cell-neighbor pairs and the color showing the proportion of pairs with the same label (cell type purity). **c**, Cell type purity against the number of cross-expressing cell-neighbor pairs. Each point is a gene pair from (b), and shaded area is 95% confidence interval. **d**, Number of cell-neighbor pairs with the same or different cell subtype labels given that they were both labeled 'glutamatergic' at the higher level

of the cell type hierarchy. Each point is a cross-expressing gene pair. **e**, Heatmap showing the normalized frequencies of cell type label combinations between cross-expressing cells. Created with BioRender.com.

Cross-expression discovers anatomical marker combinations that delineate the thalamus and cortical layer VI

Having found patterns of cross-expression within regions, we next tested for spatial organization that reflects anatomical structure. While some anatomical structures, such as cortical layers, have well-defined markers, others are difficult to characterize due to lack of marker genes. We asked whether cross-expressing genes can delineate anatomical regions. An important difference between cross-expression and co-expression is that the former will generally increase independence/dimensionality within the dataset while the latter will decrease it, providing a much larger scope for useful combinatorial markers. Since a gene panel of size *N* contains $\binom{N}{2}$ pairs, we reasoned that the quadratic space likely contains suitable marker combinations. To assess this, we used Vizgen's MERFISH (multiplexed error-robust fluorescent in situ hybridization) data (data availability section) obtained from coronal mouse brain slices, with a panel of 483 genes, yielding 116,403 gene pairs. We registered the brain slices to Allen Common Coordinate Framework version 3 (CCFv3) atlas[57] to obtain region annotations for each cell, giving us a reference against which the marker-delineated regions could be compared.

Surprisingly, we found that cross-expression between *Lgr6* and *Adra2b* delineates the thalamus even though these genes are widely expressed in the brain (Fig. 4a). Specifically, while 48% of *Lgr6*- and 57% of *Adra2b*-expressing cells are thalamic, 91% of their cross-expressing cell pairs are in the thalamus (Extended Data Fig. 3a), underscoring the spatial enrichment of their cross-expression signature (Extended Data Fig. 3b). We find that *Lgr6* also cross-expresses with *Ret* in the thalamus despite brain-wide expression of both genes (Fig. 4b, Extended Data Fig. 3c). Next, we examined whether *Adra2b* and *Ret*, both of which cross-express with *Lgr6*, show enriched co-expression in the thalamus. We find that they are indeed co-expressed within the thalamus but not in rest of the brain (Fig. 4c), e.g., 89% of their co-expressing cells are in the thalamus, thus serving as robust combinatorial markers.

To evaluate whether the combinatorial marker-based approach is reliable, we asked whether single gene markers, when assessed for cross-expression, rediscover the anatomical locations. Using the BARseq cortical cell type atlas data[34], we assessed cross-expression between cortical layer 6 marker *Foxp2* and ubiquitously expressed gene *Cdh13*. We discover that cross-expression between these genes delineates layer 6 boundary (Fig. 4d), further supporting the view that combinatorial anatomical markers can be discovered using cross-expression.

Indeed, the layer 6 boundary recovered by cross-expression captures additional L6 IT neurons whereas *Foxp2*-based boundary overlooks these cells, indicating that combinatorial markers can refine extant anatomical regions. More generally, this process leverages the spatial enrichment of cross-expression, where the distance between cross-expressing cells is smaller than the distance between cross-expressing and randomly selected cells. Once spatial enrichment is discovered, our framework can help refine anatomical regions and link them to patterns of coordinated expression across cells that are independent of co-expression.

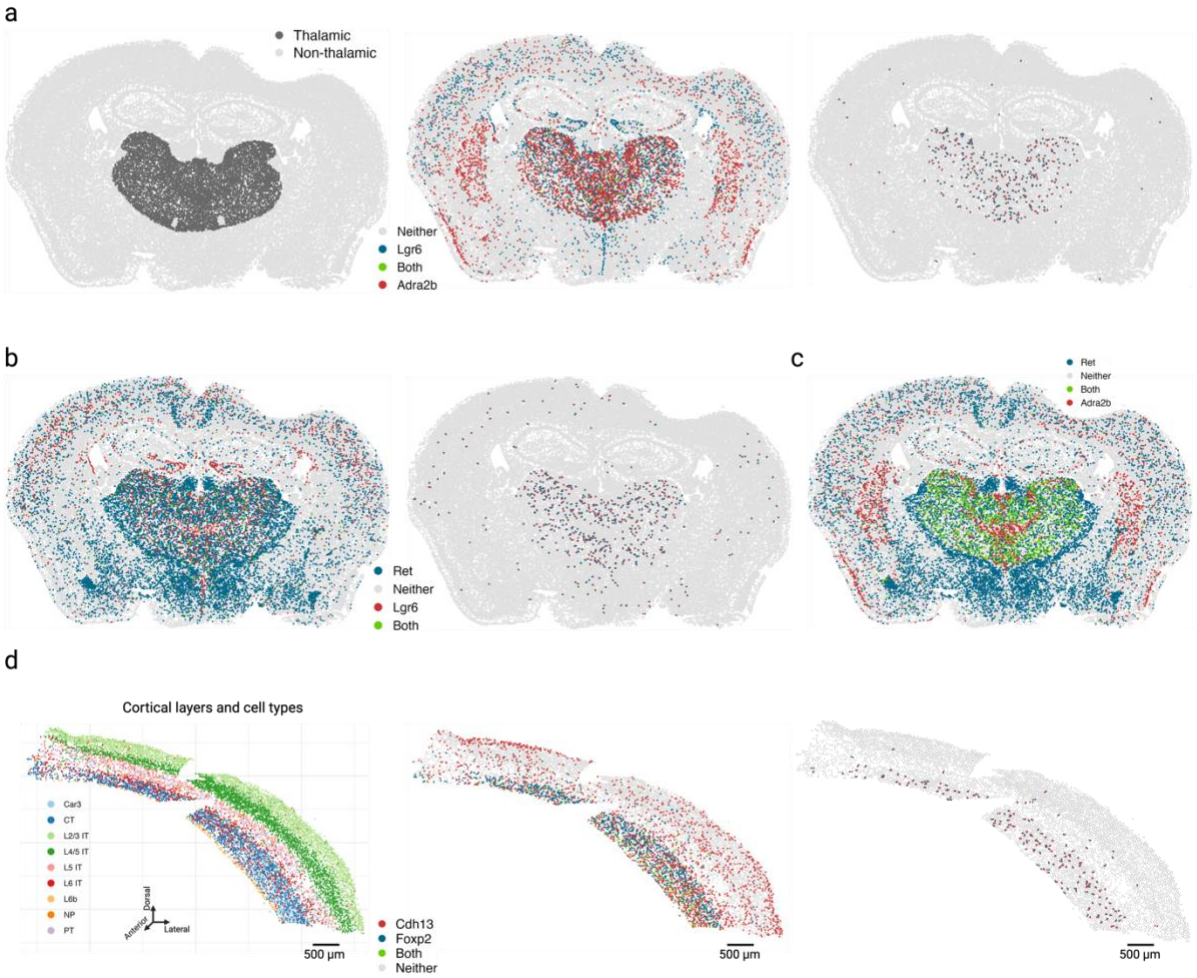


Fig. 4 | Cross-expression can discover combinatorial anatomical markers. a, Comparing the thalamus (left) to the rest of the brain, genes *Lgr6* and *Adra2b* are widely expressed across multiple brain regions (middle) but are preferentially cross-expressed in the thalamus (right). **b**, Same as in (a) but for genes *Lgr6* and *Ret*. **c**, Genes cross-expressing with *Lgr6* in (a) and (b) co-express in the thalamus. **d**, Cross-expression of *Cdh13* with cortical layer 6 marker *Foxp2* (middle) recapitulates layer 6 boundaries (right, cf. left). Created with BioRender.com.

Cross-expression network reveals *Gpr20* as a central gene and discovers possible interaction partners between astrocytes and the brain microvasculature

So far, we have assessed cross-expression between gene pairs to discover ligand-receptor interactions, cell type differences, and anatomical markers. However, each gene may cross-express with many others and thus form clusters of genes with coordinated expression. These relationships can be analyzed using networks, where nodes represent genes and edges indicate cross-expression (Fig. 5a, left). Within a network, if nodes A and B connect to node C but not to each other, they form a second-order edge (Fig. 5a, right). Both types of relationships are important, as in genetic interaction networks, because genes are joined not only by similarity but also by a form of complementarity. Representing cross-expression as a network is therefore a potentially powerful formalism, especially because it allows for the application of a substantial body of existing gene network methods.

Using the MERFISH data, we created a cross-expression network (Fig. 5b), which contains 200 genes with 382 first-order, 217 second-order, and 107 dual-order edges. We observe that *Gpr20*, a G protein-coupled receptor, is a central gene with a high node degree of 40 while the other genes form a median of 4.8 edges (Extended Data Fig. 4a). We performed gene ontology (GO) enrichment for genes cross-expressed with *Gpr20*, finding functional groups like 'regulation of macromolecule biosynthetic process', 'regulation of gene expression', and 'regulation of metabolic process' (Extended Data Fig. 4b, all p-values ≤ 0.05). While some of these genes are co-expressed with astrocytic and microglial cell type markers (Extended Data Fig. 4c), their global co-expression with the endothelial marker is higher, where the co-expression profiles were computed using neighbors cross-expressed with *Gpr20* rather than the entire dataset (Extended Data Fig. 4d, Mann-Whitney U test, endothelial vs others, all p-values ≤ 0.01; remaining pairwise comparisons, all p-values > 0.05).

Noting that the neighbors of *Gpr20*-positive cells are involved in the microvasculature, we next viewed the spatial distribution of cells expressing *Gpr20*, finding that they form contiguous linear streaks resembling blood vessels (Fig. 5c; anterior slice from mouse brain 1 shown). To test this observation, we looked at whether the neighbors of *Gpr20*-positive cells also express this gene and compared it to randomly selected cells, which constitute the expectation that *Gpr20* is uniformly expressed across space. Consistent with the visualization, we find that cells with *Gpr20* are surrounded by neighbors that also express this gene, a pattern that disappears for neighbor order of 50 or more cells (Fig. 5d-e, area under curve (AUC), neighbors vs random cells, 0.69 vs 0.49; right-tailed Wilcoxon signed-rank test, neighbors vs random cells, p-value ≤ 0.0001). Having seen that cells with *Gpr20* possibly reflect blood vessels, we asked whether these cells are

330 themselves vascular or whether they line the vasculature, especially since the cells that cross-
331 express with *Gpr20* are endothelial. We find *Gpr20* is poorly co-expressed with *Igfr1* (Pearson's
332 R = 0.0024), the vascular/endothelial marker[58–60] in our gene panel, suggesting that it lines but
333 does not mark the blood vessels. Moreover, it is lowly co-expressed with other cell type markers
334 (average Pearson's R, astrocytes = –0.0027, microglia = 0.0018, oligodendrocytes = –0.022,
335 neurons = –0.0025), eschewing cell type characterization. Taken together, *Gpr20*, a salient
336 topological feature of our cross-expression network, seems to be expressed in diverse cell types
337 that line the blood vessels, reflecting its possible role in the modulation of the microvasculature.
338 Cross-expression driven by cell types might be particularly common when two genes
339 which cross-express with a third gene are co-expressed together, reflecting some common
340 transcriptional program jointly cross-expressing with neighboring cells. To investigate this, we
341 reduced co-expression further by specifying that cross-expressing genes must show lack of
342 significant co-expression, a procedure that yielded a subnetwork, which we further curated by
343 removing genes with fewer than two edges. Indeed, we find that two genes that independently
344 cross-express with another gene tend to be co-expressed (Fig. 5f, Extended Data Fig. 5a) and,
345 as expected, belong to the same cell types, as revealed by their co-expression with cell type
346 marker genes (Extended Data Fig. 5b). Confirming these results, the subnetwork genes are
347 enriched in GO groups like 'endothelial cell proliferation', 'positive regulation of vascular
348 endothelial growth factor production', and 'regulation of endothelial cell migration' (Fig. 5g, all p-
349 values ≤ 0.05). These results indicate that while cross-expressing genes are present in specific
350 cell types, the relations between them are functionally suggestive as opposed to simply reflecting
351 cell type compositional differences, especially since the cell type markers are not cross-
352 expressed. For example, the astrocytic *EGFR* (epidermal growth factor receptor) cross-expresses
353 with the vascular *Flt4*/*VEGFR-3* (FMS-like tyrosine kinase 4), *Tek*/*Tie2* (TEK tyrosine kinase/
354 angiopoietin-1), and *Tie1* (tyrosine kinase with immunoglobulin-like and EGF-like domain 1).
355 These three vascular receptors promote angiogenesis via the *VEGF* (vascular epidermal growth
356 factor) ligand[61,62], prevent endothelial cell apoptosis[63,64], and negatively regulate angiogenesis[65],
357 respectively, thus reflecting their potential role in the brain microvasculature in coordination with
358 the astrocytes, whose endfeet ensheathe the blood microvessels to constitute the blood-brain
359 barrier (BBB).
360 Within the same subnetwork, the astrocytic gene *Ppp1r3g* (protein phosphatase 1
361 regulatory subunit 3G), which helps convert glucose to glycogen[66], cross-expresses with *Epha2*
362 (ephrin type-A receptor 2), whose activity makes the BBB more permeable[67], likely enabling
363 glucose's transport and eventual conversion into glycogen, thereby making this cross-expression

364 relation relevant for energy metabolism. Indeed, this observation can be used to generate
365 hypotheses about the (directional) relationship between energy needs within a local
366 microenvironment and the remodeling of the microvasculature, making cross-expression a
367 powerful approach with which to form testable hypotheses. More broadly, the cross-expression
368 framework can be combined with well-established approaches, such as network analysis, to
369 generate biological insights from high-throughput spatial transcriptomics data.

370 Next, we asked whether cross-expression networks change across the brain. Because
371 gene expression is regional, slices from various areas should show cross-expression between
372 distinct genes. We assessed this by forming networks for each slice in our sagittal BARseq data.
373 As expected, we find that adjacent slices have similar networks than distant slices (Fig. 5h,
374 Spearman's $\rho = -0.9$), a trend also seen in our BARseq coronal data (Extended Data Fig. 6a,
375 Spearman's $\rho = -0.87$) but not when the two datasets are mixed and the "distance" reflects the
376 difference in the order of slices (Extended Data Fig. 6b, Spearman's $\rho = 0.094$). Hence, cross-
377 expression is sensitive to broad spatial variation in gene expression.

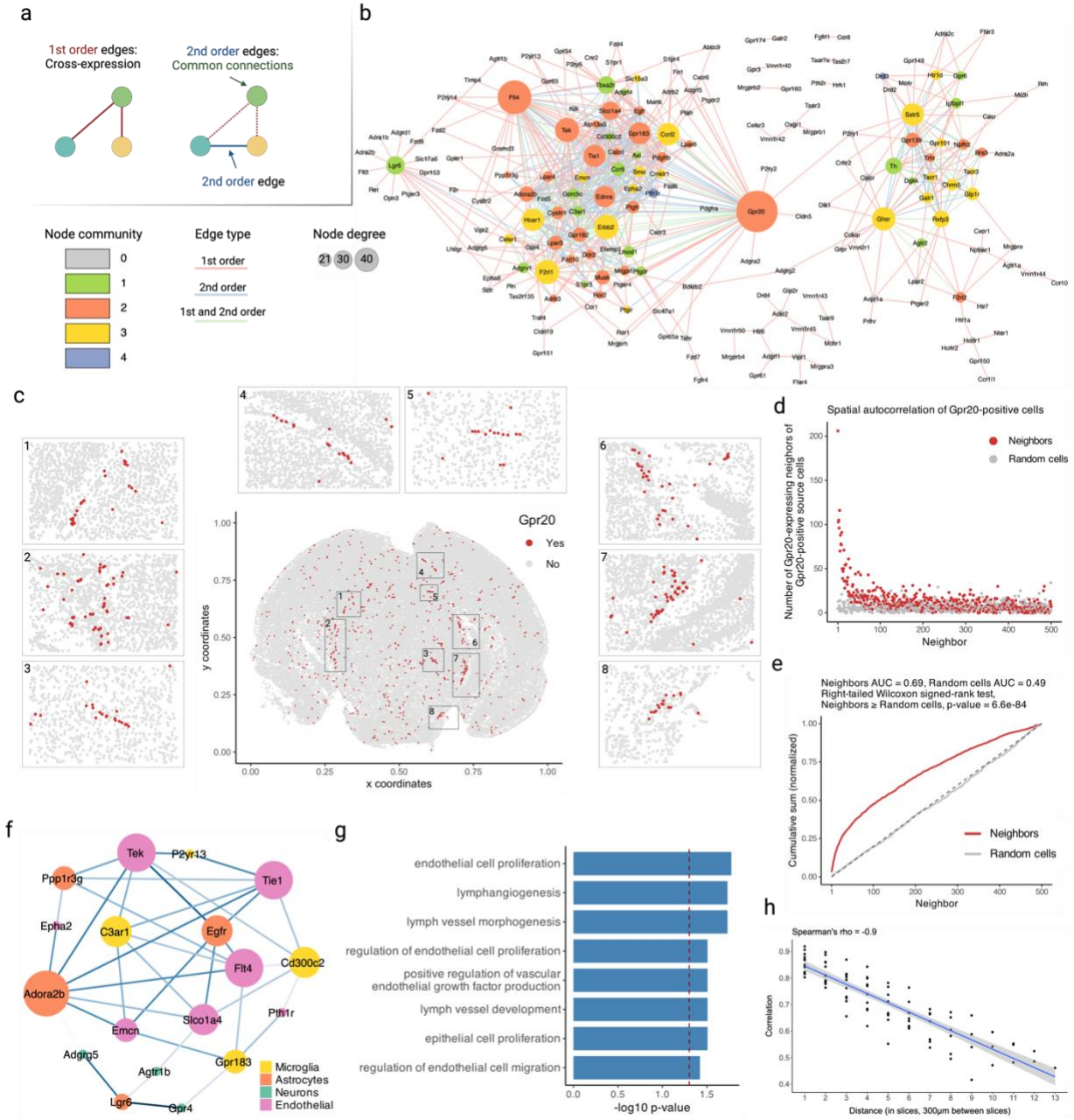


Fig. 5 | Networks of cross-expression. a, Cross-expression (edges) between genes (nodes) forms a network (left), where second-order edges (right) between genes share a node (first-order). **b**, Example cross-expression network, with first-order node degree represented by size and edge color showing first-, second-, or dual-order (first-order and second-order) connections. Threshold for the second-order edges is 4. Node color shows community membership assigned by Louvain clustering the second-order network. **c**, Cells are colored based on *Gpr20* expression. Numbered rectangles in the central figure correspond to zoomed-in versions. **d**, Number of neighbors with *Gpr20* given that the source cells also express this gene. **e**, Cumulative sums (after L1 normalization) from (d) for true and randomly selected neighbors. Identity line is dashed. **f**, Subnetwork created from (b) by pruning edges with significant co-expression and then removing

nodes with degree 1. Nodes are colored by cell types based on their co-expression with marker genes. **g**, GO functional groups for genes in the subnetwork in (f). **h**, Similarity in the network structures of nearby and distant brain slices. Shaded area is 95% confidence interval. Created with BioRender.com.

Cross-expression signal is replicable across datasets, and global co-expression between spatial and single cell datasets indicates reliable cell segmentation

Two sources of non-biological variation in spatial transcriptomics[2–7] are batch effects, which result from technical differences between experimental runs, and cell segmentation, which draws boundaries around and assigns transcripts to cells, a process that can alter gene expression profiles and affect downstream analysis, including cross-expression.

We assess batch effects by comparing cross-expression between corresponding slices across biological replicates. The MERFISH data contains three replicates with three slices each, where the slices are sampled from roughly the same position. We find that the cross-expression signature is highly similar across replicates. For example, the average correlation for the anterior slices between the three replicates is 0.83 (Fig. 6a), with similar findings for the middle and posterior slices (Extended Data Fig. 7a-b, Spearman's ρ = 0.81 and 0.8, respectively).

We next assessed the degree to which cross-expression within the BARseq sagittal or coronal experiments[34] is similar to that between experiments. To this end, we compared cross-expression patterns between brain slices. As expected, the cross-expression profiles are more similar within brains than between brains (Fig. 6b, Mann-Whitney U tests, FDR-corrected, coronal vs sagittal, p-value = 0.2, coronal vs mixed, p-value ≤ 0.001, and sagittal vs mixed, p-value ≤ 0.001), suggesting that the sectioning procedure samples different brain regions and therefore reveals distinct underlying gene expression profiles. Supporting this result, we find that the same anatomical regions (per Allen CCFv3 brain atlas[57]) across brains have more similar cross-expression profiles than do different regions within or between brains (Fig. 6c, Mann-Whitney U test, different regions vs same regions, p-value ≤ 0.001). Noting that the sagittal and coronal brains contain the same regions in the dorsal to ventral directions, we asked whether the cross-expression is similar in this shared dimension. Here, we computed the density of cross-expressing cells in the dorsal to ventral direction and compared these distributions across the two brains, finding that 99% (without FDR correction) of the genes did not have significantly different density profiles (Fig. 6d), suggesting that the cross-expression patterns are highly similar across batches at the whole-brain level.

Having found that the cross-expression profiles are generally robust, we assessed cell segmentation at a global level by comparing gene co-expression between the single cell RNA-

sequencing (scRNA-seq)[31] and spatial transcriptomic data. We reasoned that scRNA-seq does not require segmentation and therefore captures genes co-expressed within the cell's boundaries (Fig. 6e). Because cell segmentation alters transcript assignment, it could change co-expression in spatial transcriptomic data. Reassuringly, we find a strong association between gene co-expression in the scRNA-seq and spatial transcriptomic data (Fig. 6f, Pearson's R = 0.83). We further examine whether this correlation is sufficiently strong by comparing co-expression between scRNA-seq and single-nucleus RNA-sequencing (snRNA-seq)[68] (Fig. 6g, Pearson's R = 0.86), finding agreement between the two comparisons (R = 0.83 vs. R = 0.86). These results imply similar levels of technical variability between platforms while suggesting that gene co-expression is congruent between scRNA-seq and spatial transcriptomic data.

The data in our work was processed using CellPose[69], a deep learning-based cell segmentation algorithm. A recent benchmarking study[70] showed that it outperforms other methods on a variety of metrics. In fact, it uses the nuclear stain DAPI as a cell landmark and forms boundaries using cytoplasmic signal, such as the transcript distributions, making it the state-of-the-art segmentation algorithm on a variety of assessments. Further, the cell segmentation algorithms are continuously being improved[71], allowing users to re-segment and reanalyze their data. Most importantly, the analysis conducted using the cross-expression framework may suffer if segmentation is performed poorly, but the validity of the concept and the soundness of its statistics do not rely on this potential artefact and, with rapid improvements in data quality, the inferences drawn from it will become increasingly more reliable.

Moreover, we assessed the relationship between cross-expression and noise in gene expression measurement. Since the algorithm requires binarizing the expression matrix, an appropriate threshold needs to be applied prior to analysis. To count a gene as expressed in a cell, we applied thresholds of 1 to 10 molecules, finding that the cross-expression patterns are generally concordant across these noise levels (Extended Data Fig. 8a-b, median Pearson's R = 0.88). Importantly, our framework is agnostic to and compatible with multiple models of gene expression noise[72], and once an appropriate threshold has been applied, the resultant expression matrix can be used for cross-expression analysis.

Finally, we explored the patterns of cell-neighbor relations and found that over 60% of cells are the nearest neighbors of exactly one cell but the remaining cells are the nearest neighbors of two or more cells (Extended Data Fig. 9a). Patterns such as these may be biologically important if the 'neighbor' cell plays a central role in the local microenvironment, so deviations from one-to-one mappings should be captured by statistical analyses. To investigate that our results are consistent across these patterns, we compared cross-expression in one-to-one against many-to-

one mappings and with the full dataset, finding an average Pearson's correlation of 0.96 (Extended Data Fig. 9b). Importantly, our procedure is consistent with the assumption of independent sampling because while a cell may be the nearest neighbor of multiple cells, each cell-neighbor pair is statistically independently.

We enable these analyses by providing a highly efficient R package. A laptop with 16 GB RAM can test for cross-expression in large datasets containing hundreds of thousands of cells and thousands of genes within minutes (Fig. 6h). At present, most (commercial) imaging-based platforms cannot profile gene panels of this magnitude[2–7], though such capabilities are anticipated and are being developed[73]. Our software's performance makes it well-suited for analyzing current and future spatial transcriptomic datasets.

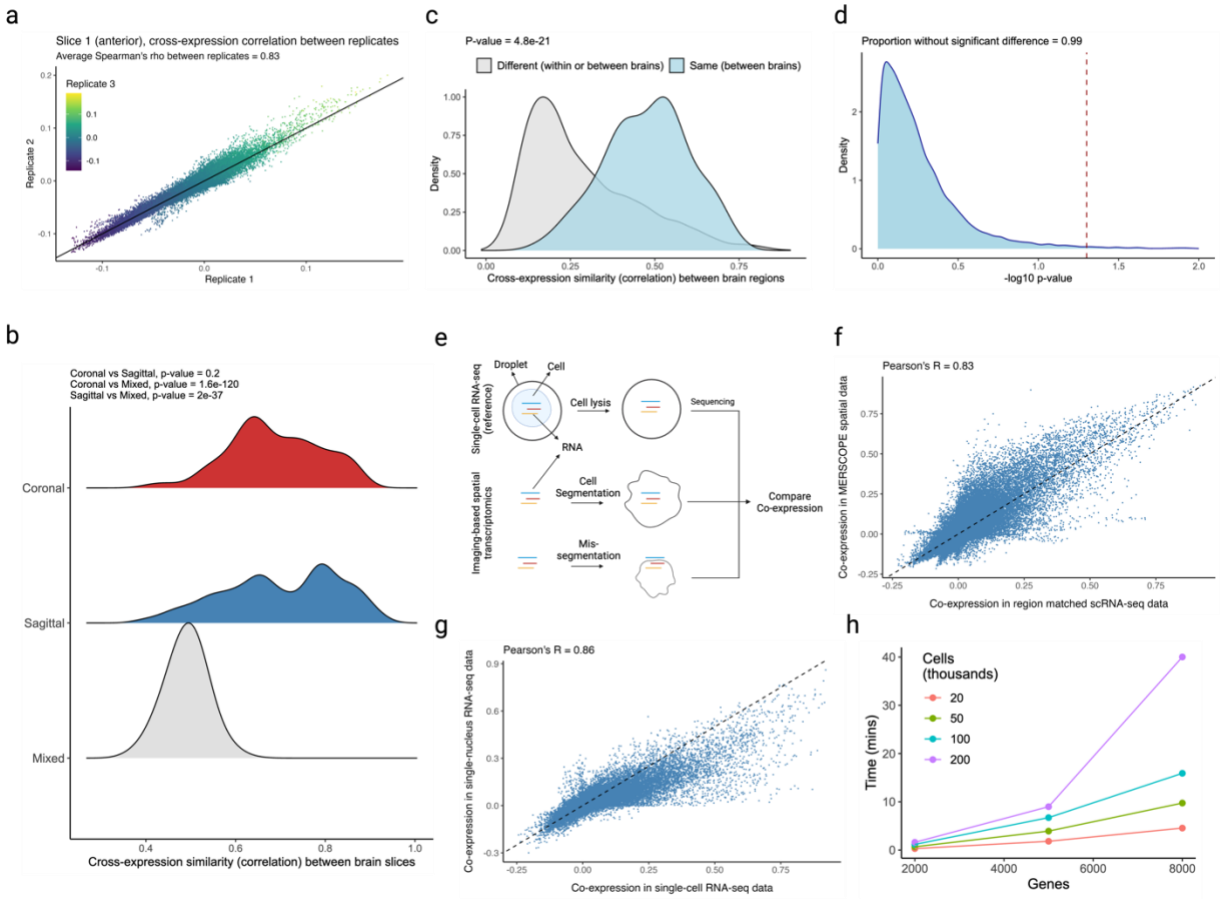


Fig. 6 | Assessing batch effects, cell segmentation, and software runtime. a, Correlation between cross-expression signatures across three biological replicates. **b**, Correlation between cross-expression signatures within (sagittal or coronal) and between (mixed) brains. Positive signal between brains likely reflects the fact that the sagittal and coronal brains both contain regions in the dorsal to ventral direction. **c**, Correlation between cross-expression signatures between the same anatomical regions across brains or between different anatomical regions

across or within brains. **d**, Density of cross-expressing cells in the dorsal to ventral directions is compared across the sagittal and coronal brains. Significant p-values (without FDR correction) indicate that a cross-expressing gene pair has different densities across the two brains. Red dotted line is the significance threshold at alpha = 0.05. **e**, Single cell RNA-sequencing (scRNA-seq) profiles cells' gene expression without cell segmentation. Co-expression between scRNA-seq and spatial transcriptomic data helps diagnose segmentation artifacts. **f**, Gene co-expression in spatial transcriptomic and in scRNA-seq data. Each point is a gene pair. **g**, Gene co-expression in single-nucleus RNA-sequencing (snRNA-seq) and in scRNA-seq data. Same gene panel is used in (f) and (g). **h**, Software runtime for varying numbers of genes and cells using a personal laptop with 16 GB RAM. Created with BioRender.com.

Discussion

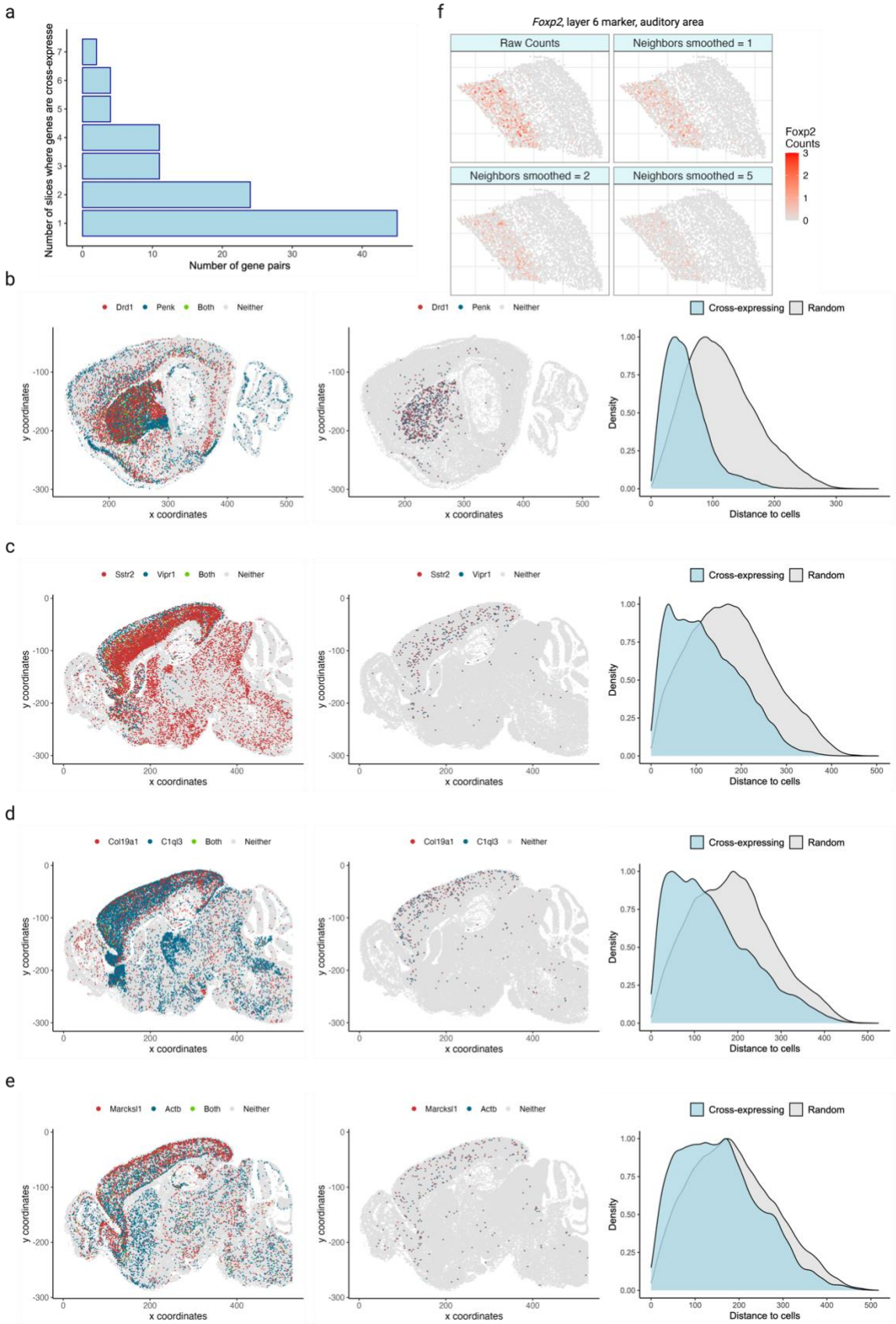
Cross-expression allows us to study gene-gene networks that reflect how cells influence each other through coordinated gene expression between neighboring cells. Using this framework, we recapitulated known ligand-receptor interactions at the single cell level, revealing biologically meaningful tissue phenotypes. We further showed that cross-expression can be discovered without cell type labels but often reflects cell subtype compositional differences. Moreover, it helps us discover paired markers for anatomical regions, such as the thalamus, and is amenable to network formulations, finding genes like the *Gpr20* as central nodes and revealing the relationships between astrocytes and brain microvasculature. Together, cross-expression is a powerful way of analyzing spatial transcriptomic data and allows us to study gene-gene relations between adjacent cells, thereby fully harnessing the high-throughput of these technologies.

The cross-expression framework complements current approaches analyzing spatial transcriptomic data, such as those exploring niche-specific co-expression patterns[12–16]. Specifically, niche-specific cross-expression networks may be compared with co-expression networks to examine if inter-cellular relations are associated with intra-cellular gene programs and vice versa. This may be approached at different, potentially hierarchical spatial scales to reveal spatial gene expression programs within the tissue. Moreover, the cross-expression patterns can be quantified in multiple ways, such as using mutual information or graphlets, allowing investigations into the best approaches that capture the signal of interest. For example, just as co-expression relations can be measured using the Pearson's correlation coefficient, cross-expression patterns may be investigated from numerous perspectives to discover the most robust formalism. In this sense, the cross-expression framework introduced here is primarily a novel way of conceptualizing gene-gene relations within spatial transcriptomics data, thereby serving as a powerful framework for research in tissue biology. For instance, it can be used to study cancer[74], where tumor progresses via signaling with the stromal tissue, as well as

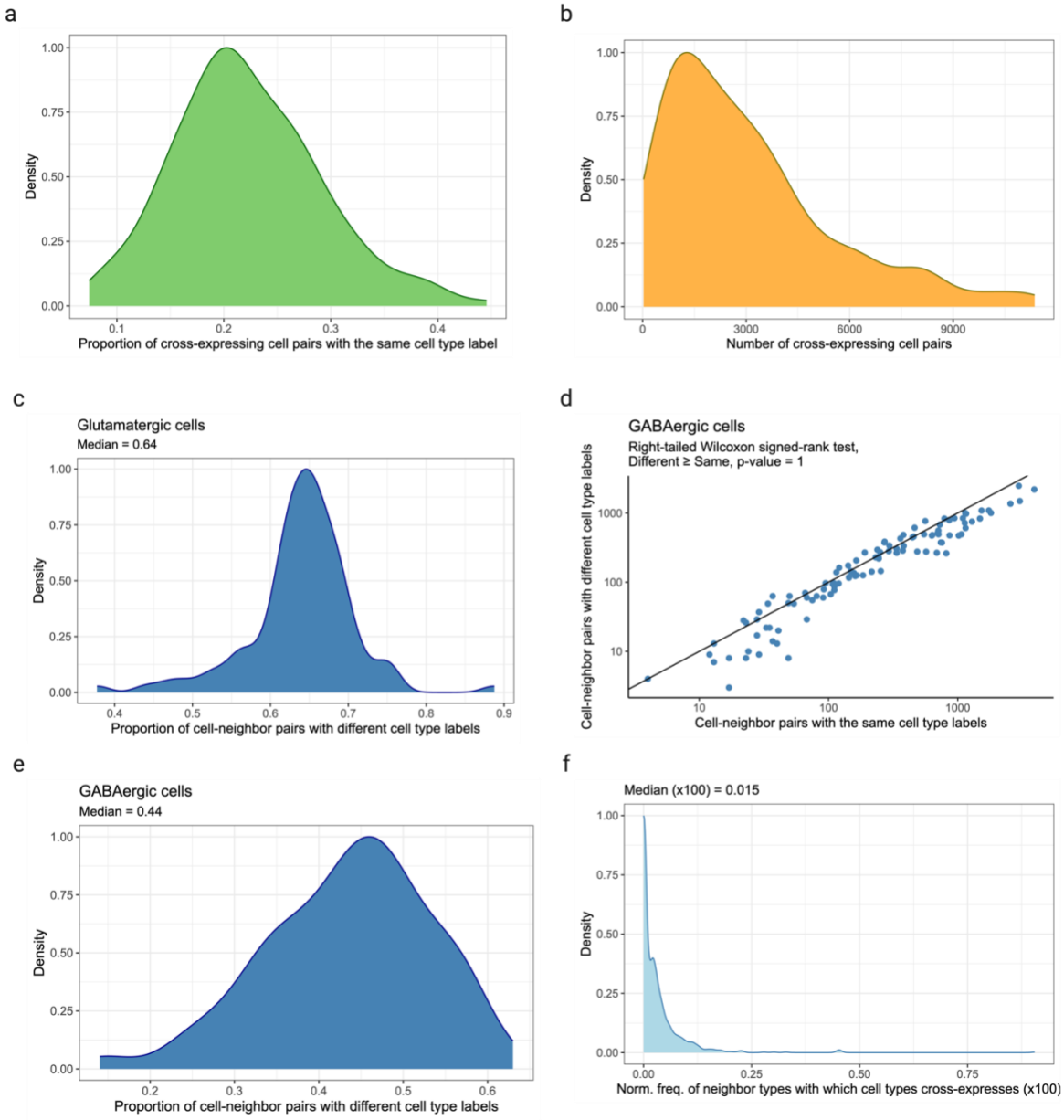
511 neurodegenerative diseases like Alzheimer's[75] or senescence[76], where the progression of
512 pathology is spatially mediated, making it a broadly useful approach for numerous problems.
513 Cross-expression is not restricted to imaging-based spatial transcriptomics. Instead, it can
514 be applied to any biological assay that provides cell-by-features and cell-by-coordinate matrices.
515 For example, it can be extended to spatial proteomics[77], with potential to discover ligand-receptor
516 interactions. Likewise, it may be applied to spatial translatomics[78] to focus on translating mRNAs
517 that are more likely to form functional proteins, making conclusions about cell-cell relations more
518 robust. In fact, with the increasing resolution of spatially barcoded RNA capture based methods[3–
519 7], the framework may be extended transcriptome-wide to understand relations between spots at
520 near single cell resolution.
521 A key challenge in imaging-based spatial transcriptomics[2–7], including the datasets used
522 in this work, is the size and constitution of the gene panel, which sets an upper limit on biological
523 discovery. Although our framework will become more powerful as the quality of spatial
524 transcriptomic data, especially the gene panel, increases, care must be taken to not interpret the
525 results in mechanistic terms. Instead, the coordinated gene expression between neighboring cells
526 should be viewed as a target for experimental validation. In this sense, the cross-expression
527 framework radically narrows the space of gene-gene relations by identifying pairs that are
528 potentially biologically meaningful, making the problem experimentally tractable. Overall, cross-
529 expression is a powerful addition to the growing list of analytical techniques and, most importantly,
530 it offers a unique perspective on using spatial transcriptomic data for driving biological discovery.
531

532 **Extended Data Figures**

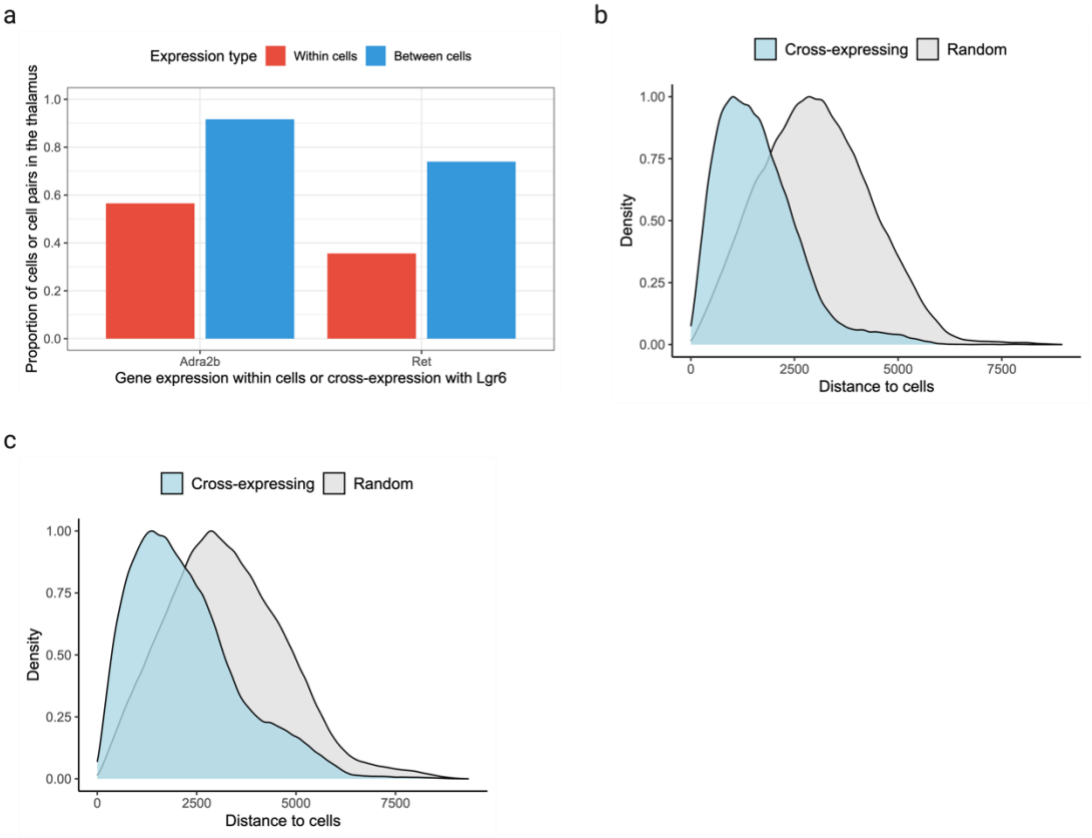
533 **Extended Data Fig. 1 | Cross-expression across tissue slices and regions for ligand-**
534 **receptor and non-signaling genes. a**, Distribution of the number of gene pairs cross-expressed
535 in different slices. Dataset has 15 slices sampled sagittally from the left hemisphere of a mouse
536 brain. **b-e**, Cells are colored by gene expression (left) and cross-expressing cells are highlighted
537 (center). Right, distances between cross-expressing cells are compared with those between
538 cross-expressing and randomly selected cells. Smaller distances mean that cross-expressing
539 cells are nearer each other (spatial enrichment) than expected by chance (p-values ≤ 0.01, left-
540 tailed Mann-Whitney U test). Genes include ligands and receptors (b, c) and non-ligands and non-
541 receptors (d, e). **f**, Smoothed gene expression for different numbers of neighbors for the auditory
542 cortical layer 6 marker gene *Foxp2*. Created with BioRender.com.


a
Number of slices where genes are cross-expresse
Number of gene pairs
f
Foxp2, layer 6 marker, auditory area
Raw Counts
Neighbors smoothed = 1
Neighbors smoothed = 2
Neighbors smoothed = 5
Foxp2 Counts
b
Drd1 Penk Both Neither
Drd1 Penk Neither
Cross-expressing Random
y coordinates
x coordinates
Density
Distance to cells
c
Sstr2 Vipr1 Both Neither
Sstr2 Vipr1 Neither
d
Col19a1 C1ql3 Both Neither
Col19a1 C1ql3 Neither
e
Marcksl1 Actb Both Neither
Marcksl1 Actb Neither

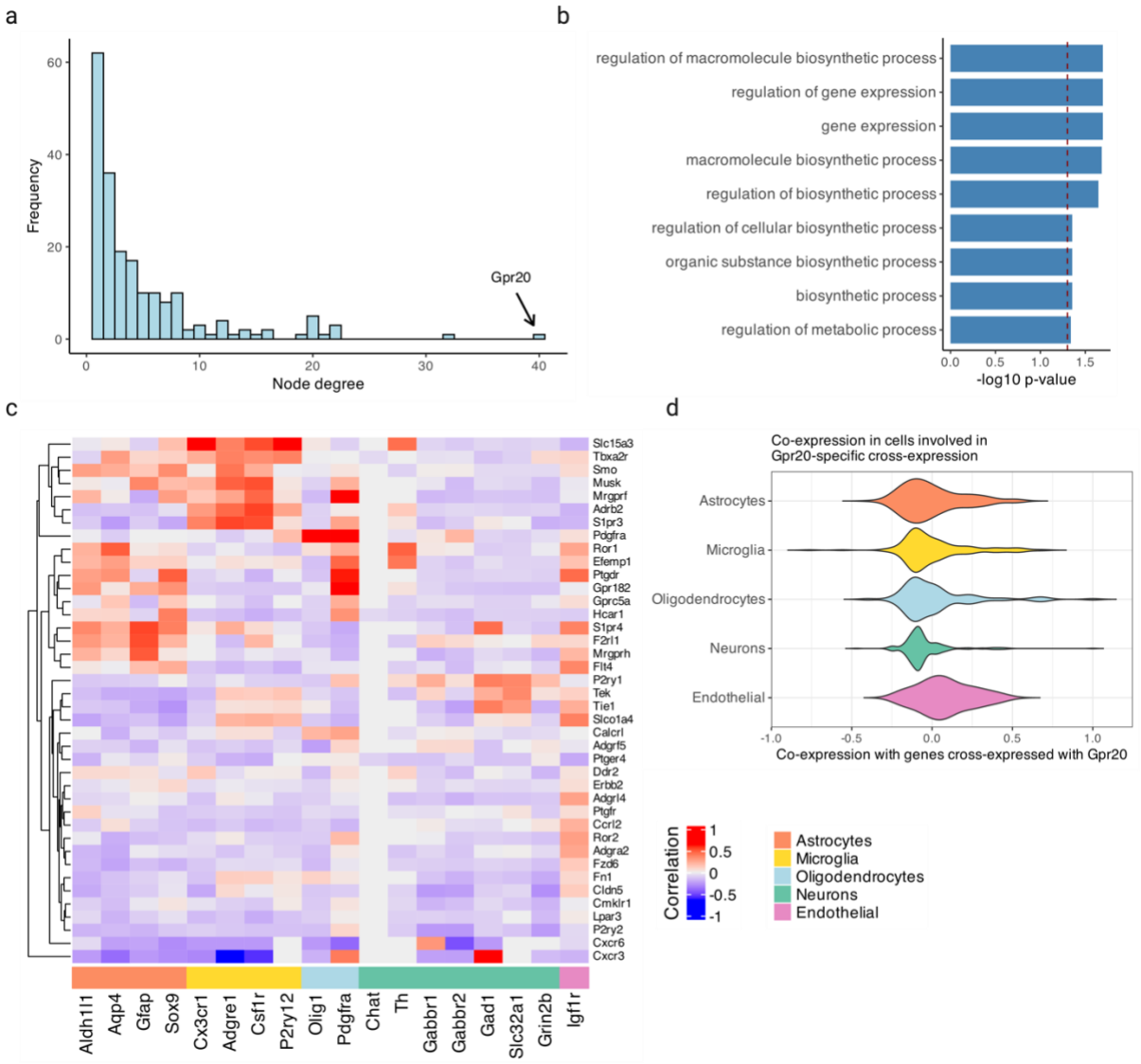
543



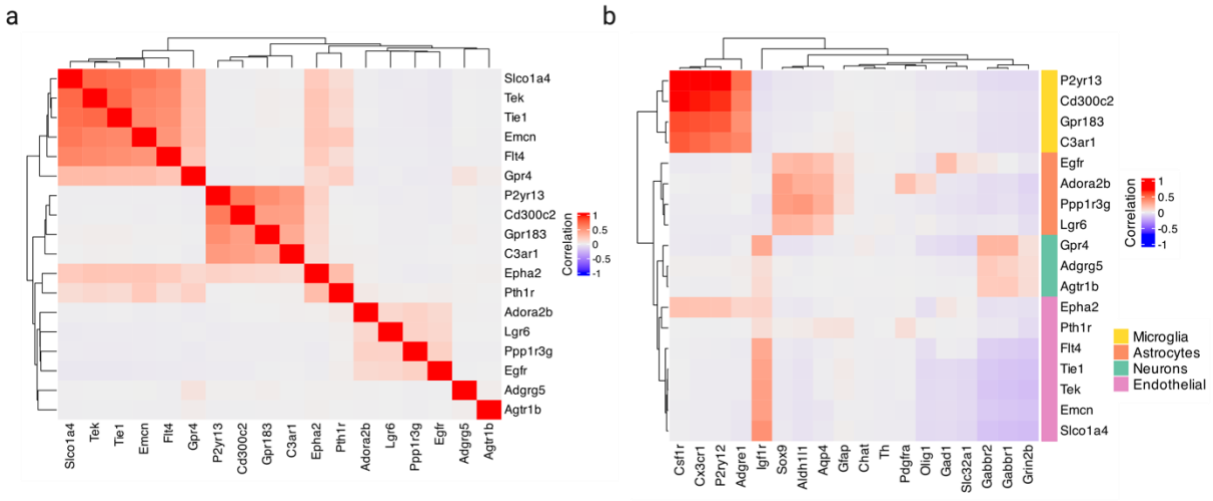
Extended Data Fig. 2 | Relationship between cross-expression and cell type heterogeneity. **a**, Proportion of cross-expressing cell pairs belonging to the same cell type label. **b**, Number of cell-neighbor pairs involved in cross-expression. **c**, Proportion of cell-neighbor pairs with different cell subtype labels given that both were labeled 'glutamatergic' at the higher level in the cell type hierarchy. **d**, Number of cell-neighbor pairs with the same or different cell subtype label given that both were labeled 'GABAergic' at the higher level in the cell type hierarchy. Each point is a cross-expressing gene pair. **e**, Same as in (c) but for 'GABAergic' cells. **f**, Proportion of neighbor cell types against which cell types cross-express. Created with BioRender.com.



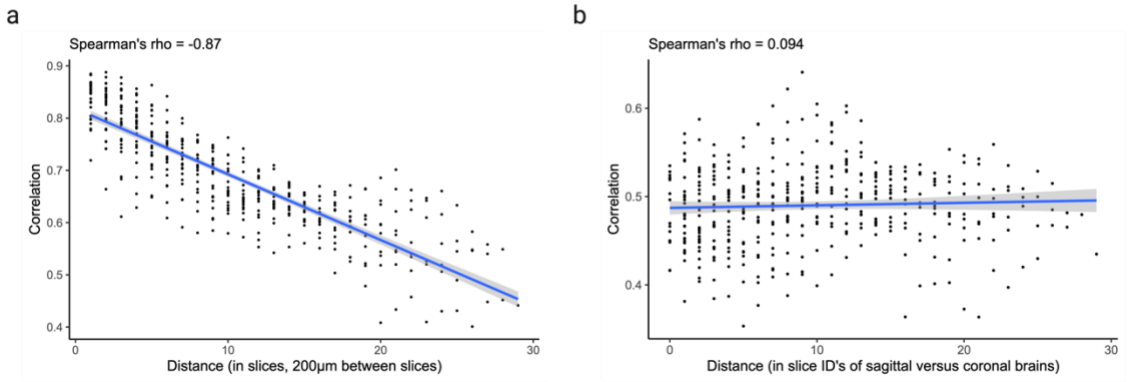
Extended Data Fig. 3 | Combinatorial anatomical markers discovered using spatially enriched cross-expression. a, Proportion of *Adra2b*- and *Ret*-expressing cells in the thalamus (red) and the proportion of cell pairs in the thalamus (blue) when cross-expressing with *Lgr6*. **b-c**, Distances between cross-expressing cells versus those between cross-expressing and randomly chosen cells for genes *Lgr6* and *Adra2b* (c) and for *Lgr6* and *Ret* (d). Smaller distances mean that cross-expressing cells are nearer each other (spatial enrichment) than expected by chance (p-values ≤ 0.01, left-tailed Mann-Whitney U test). Created with BioRender.com.



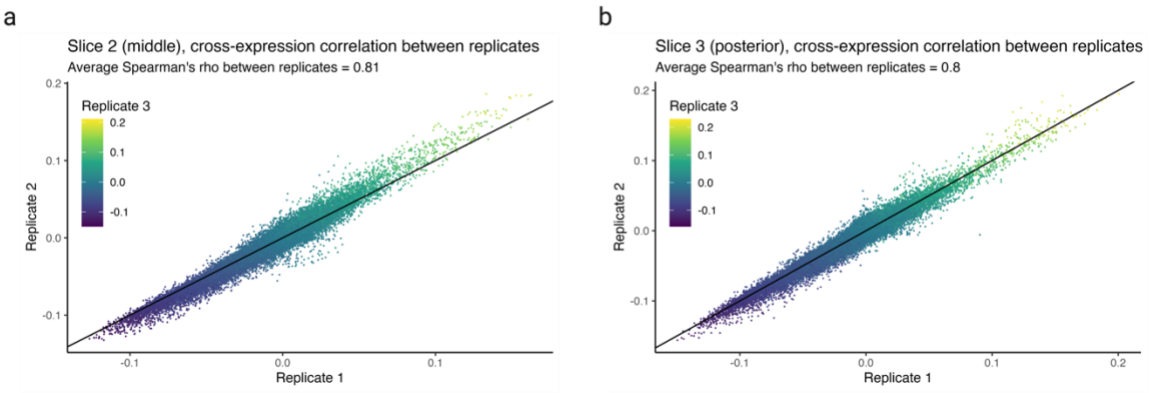
Extended Data Fig. 4 | Exploration of *Gpr20* and its cross-expressing genes. a, Distribution of node degree, with *Gpr20* highlighted. **b**, Gene ontology (GO) functional groups for genes cross-expressed with *Gpr20*. **c**, Co-expression of genes cross-expressed with *Gpr20* (right) against cell type marker genes (bottom). For each gene, co-expression was computed using cells involved in cross-expression and not the entire dataset. **d**, Distribution of cell type marker genes' co-expression across the genes in (c). Created with BioRender.com.



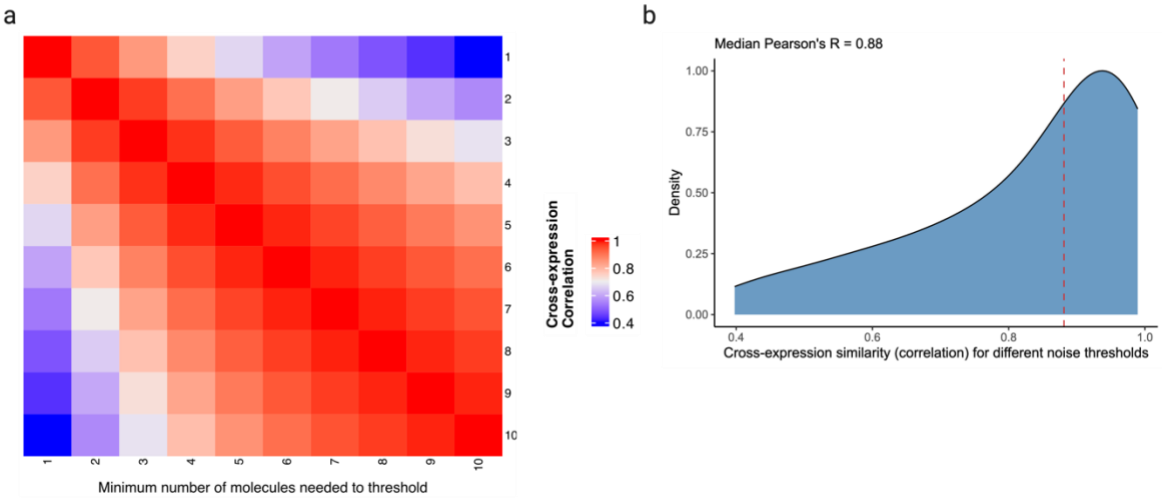
Extended Data Fig. 5 | Exploration of the MERFISH cross-expression (sub)network. a, Co-expression of genes in the subnetwork. **b**, Co-expression between genes in the subnetwork (right) and cell type marker genes (bottom). Created with BioRender.com.



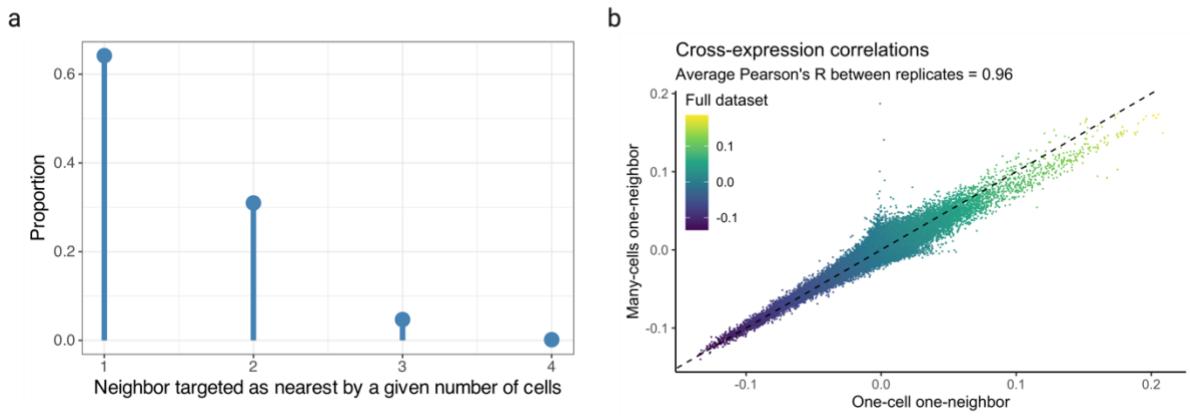
Extended Data Fig. 6 | Cross-expression network similarity between slices. a, Slice-specific cross-expression networks compared and shown as a function of distance between slices. **b**, Same as in (a) but slice-specific networks compared between sagittal and coronal datasets, where the “distance” is the difference in slice ID's. Shaded areas are 95% confidence intervals. Created with BioRender.com.



Extended Data Fig. 7 | Cross-expression network similarity between replicates. a-b, Cross-expression networks compared between three replicates for the middle (a) and posterior (b) slices. Created with BioRender.com.



Extended Data Fig. 8 | Cross-expression network similarity at different levels of gene expression noise thresholds. a, Cross-expression networks compared after applying different noise thresholds, which are the minimum number of molecules a gene must express within a cell to be considered as detected. **b**, Distribution of the network similarities across noise levels, with the median indicated using the dotted line. Created with BioRender.com.



Extended Data Fig. 9 | Patterns of cell-neighbor mappings and their relationship with cross-expression. a, Cells considered as 'nearest neighbor' by other cells reported as a proportion of total cell-neighbor relations. '1' is one-to-one mapping and '2-4' is many-to-one mapping. **b**, Cross-expression networks computed using one-to-one mappings, many-to-one mappings, and the full dataset (both mappings). Created with BioRender.com.

Online Methods

We first explain the theoretical underpinnings of our approach and outline the features of the associated R package. We then specify how these are used in various analyses.

Statistics of cross-expression between a gene pair

Cross-expression is the mutually exclusive expression of a gene pair across neighboring cells. To assess whether gene A's expression in cells and gene B's expression in their spatial neighbors is significant, we use a simple sampling procedure and model the probabilities using the hypergeometric distribution

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{1}$$

where *N* is the population size, *K* is the number of successes (or success states), *n* is the number of samples or draws, and *k* is the number of observed successes. The form $\binom{a}{b}$ is the binomial coefficient giving us the number of distinct *b*-sized groups from *a*-entries.

Equation 1 outlines all the ways in which success can be observed—$\binom{K}{k}$—*and* (product rule) all the ways in which failure can be obtained—$\binom{N-K}{n-k}$—normalized by all possible ways of generating our sample $\binom{N}{n}$, making the outcome probabilistic by bounding it between [0,1].

Traditionally, the *n* samples are assessed for the presence of some property *k*. Here, we sample cell-neighbor *pairs* conditioned on the cell expressing gene A and ask whether the neighbor expresses gene B. Thus, the sample size *n* is the number of cells with gene A, the number of observed successes *k* is the number of neighbors with gene B whose corresponding cells express gene A, and the number of success states *K* is the total number of neighbors with gene B. The population size *N* is the total number of cells, including those that co-express A and B and those that express neither gene. We are interested in the probability of observing *k* or more neighbors with B when *n* cells with A are sampled. To this end, we modify the hypergeometric cumulative distribution function (CDF)

$$P(X \geq k) = 1 - P(X < k) = 1 - \sum_{i=0}^{k-1} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}} \tag{2}$$

to calculate the probability of *k* or more successes. A value lower than alpha α indicates an unusually large number of neighbors expressing gene B when cells expressing gene A are sampled, implying statistically significant cross-expression between this pair.

Statistics of cross-expression between all gene pairs

We need to assess cross-expression across all gene pairs, which rise quadratically by $\binom{N}{2}$ or $\frac{N(N-1)}{2}$ for *N* genes. For example, a panel with 500 genes contains around 125,000 pairs whereas a panel with 1,000 genes has approximately 500,000 pairs. To efficiently explore this space, we implement the procedure above using matrix operations and specialized packages in R.

We begin with a cells-by-genes expression matrix **E** and a cells-by-coordinates location matrix **L**, where the coordinates in our data are cell centroids on two-dimensional slices. We input **L** into `RANN` package's function[79,80] `nn2` with search type as `standard`, which implements a kd-tree (or optionally a bd-tree) search algorithm to explores data subspaces and efficiently find the *n*-th neighbors. Using the neighbor indices, we re-order the expression matrix **E** to generate the neighbors-by-genes matrix **E′**. The value of *n* can be changed to generate paired gene expression matrices, where the corresponding rows represent cells and their *n*-th nearest neighbors.

Our aim is to use **E** and **E′** to compute *N* (population), *K* (neighbors with B), *n* (cells with A), and *k* (neighbors with B when their corresponding cells express gene A) for each gene pair. These four values can be inputted into R's `phyper` function for all gene pairs, facilitating efficient computation. The population size *N* is the total number of cells and is the same across all pairs. To compute *n*, we binarize **E** based on expression or lack thereof, and compute co-occurrences using the dot product

$$\mathbf{C} = \mathbf{E}^T \cdot \mathbf{E} \tag{3.1}$$

where $\mathbf{C}_{ii}$ is the number of cells expressing gene i and $\mathbf{C}_{ij}$ (for $i \neq j$) is the number of cells co-expressing genes i and j. We perform

$$\mathbf{U}_{ij} = \mathbf{C}_{ii} - \mathbf{C}_{ij} \tag{3.2}$$

where $\mathbf{U}_{ij}$ is the number of cells uniquely expressing gene i. We implement this by extracting the diagonal of $\mathbf{C}$, and "broadcast" it against its off-diagonal entries, thus aligning the corresponding values before subtraction. For each pair, this gives us the number of cells *n* uniquely expressing each gene. We perform an analogous calculation for *K* using **E′** instead of **E**, giving us the number of neighbors uniquely expressing each gene within a gene pair.

We now turn to *k*, the number of neighbors observed with gene B given that their corresponding cells express gene A. Using binarized matrices **E** and **E′**, we compute the number of cell-neighbor pairs such that the cells express gene A without gene B and the neighbors express gene B without gene A

$$\mathbf{X} = (\mathbf{E} \odot (1 - \mathbf{E}')) \tag{4.1}$$

$$\mathbf{Y} = ((1 - \mathbf{E}) \odot \mathbf{E}') \tag{4.2}$$

$$\mathbf{Q} = \mathbf{X}^T \cdot \mathbf{Y} \tag{5}$$

where $\odot$ is the Hadamard (elementwise) product and $\mathbf{Q}_{ij}$ is the number of cell-neighbor pairs with mutually exclusive expression. In $\mathbf{X}$, $\mathbf{E}$ contains '1' in cells where a gene is expressed and $1 - \mathbf{E}'$ contains '1' in neighbors where a gene is *not* expressed. Their elementwise product $\mathbf{X}$ has '1' to indicate genes' presence in cells and their absence in neighbors. $\mathbf{Y}$ shows the analogous procedure for genes' presence in the neighbor and their absence in cells. Hence, the dot product of $\mathbf{X}$ and $\mathbf{Y}$ gives $\mathbf{Q}$, a genes-by-genes asymmetric matrix, whose entries show the number of cell-neighbor pairs with mutually exclusive expression. ($\mathbf{Q}$ is asymmetric because the number of cell-neighbor pairs in the A-to-B and B-to-A directions are not always identical.) This is *k* or observed successes. These steps generate four number – *N*, *K*, *n*, and *k* – per gene pair. We input these into R's `phyper` function in accordance with equation (2), giving us corresponding p-values.

Since $\mathbf{Q}$ is asymmetric, we obtain two p-values per gene pair, one in the A-to-B and the other in the B-to-A direction. We perform Benjamini-Hochberg[81] false discovery rate (FDR) multiple test correction on the entire p-value distribution. For each gene pair, we then assess whether or not cross-expression is observed in either direction and use the lower FDR-corrected p-value as the final output, which is provided both as an edge list and as a gene-by-gene p-value matrix **P**.

Cross-expression networks

We can threshold and binarize **P** at a pre-selected alpha α to form an adjacency matrix **N**, where '1' indicates connections (edges) between genes (nodes)

$$\mathbf{N}_{ij} = \begin{cases} 1 & \mathbf{P}_{ij} \leq \alpha \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

This allows us to perform cross-expression network analysis, where higher-order community structure is discovered using shared connections between genes

$$\mathbf{S}^R = \mathbf{N}^1 \cdot \mathbf{N}^2 \cdot \ldots \cdot \mathbf{N}^{R-1} \cdot \mathbf{N}^R \tag{6.2}$$

where we restrict R = 2 to discover second-order connections between genes.

Cross-expression at multiple length scales

Cross-expression is coordinated gene expression between neighboring cells. Yet, these patterns may be present at larger length scales, requiring us to understand associations between regions. To facilitate this, we smooth the expression of each gene in a cell by averaging it with its expression in *n* nearby cells. Using the `RANN` package, we find the indices of each cell's *n* nearest neighbors, and make the corresponding values '1' in the cells-by-cells matrix **C**

$$\mathbf{C}_{ij} = \sum_{k=1}^{s} \mathbb{I}\{(i,j) = (i_k, j_k)\} \tag{7.1}$$

where the indicator function $\mathbb{I}$ specifies

$$\begin{cases} 1 & \text{if } (i,j) = (i_k, j_k) \\ 0 & \text{otherwise} \end{cases}$$

and

$$s = c + (c \times n) \tag{7.2}$$

where *c* is the number of cells and *n* is the number of neighbors. Here, *s* is the total number of row-column indices *i-j* that *k* iterates over. We perform averaging using the expression matrix **E**

$$\mathbf{S} = \frac{1}{n}(\mathbf{C} \cdot \mathbf{E}) \tag{8}$$

where $\mathbf{S}_{ij}$ is the *j*-th gene's average value in *i*-th cell across *n* neighbors. The smoothed gene expression matrix **S** can be used for downstream analysis.

Bullseye scores as effect size

The bullseye scores quantify the effect size by comparing cross-expression with co-expression. Here, the number of neighbors with gene B is compared to the number of cells co-expressing genes A and B. We use binarized cell and neighbor expression matrices **E** and **E′**, respectively

$$\mathbf{B}_{\boldsymbol{n}} = \mathbf{E}^T \cdot \mathbf{E}'_n \tag{9.1}$$

where *n* is the *n*-th neighbor, giving us *n* gene-by-gene asymmetric matrices $\mathbf{B}_{\boldsymbol{n}}$. The *i*-th and *j*-th entries of $\mathbf{B}_{\boldsymbol{n}}$ indicate the number of *n*-th nearest neighbors expressing gene B when cells in **E** express gene A. $\mathbf{B}_{\boldsymbol{n}}$ is a co-occurrence matrix when $n = 0$. Viewing $\mathbf{B}_{\boldsymbol{n}}$ as a tensor with dimensions *i*, *j*, and *n*, for each gene pair we take the cumulative sum and normalize across the neighbors

$$\mathbf{B}_{ijn} = \frac{1}{n}\sum_{n'=1}^{n} \mathbf{B}_{ijn'} \quad \text{for } n \geq 1 \tag{9.2}$$

These matrices can be compared with $\mathbf{B}_{n=0}$ to find the ratio of cross-expression to co-expression and/or $\log_2$-transformed for further analysis. The output is provided as an array of matrices (tensor) or as an edge list, where columns represent different *n* neighbors.

Expression of gene pairs on tissue

A powerful way of viewing cross-expression is to plot the cells and color them by their gene expression. For a gene pair, a cell can express genes A, B, both, or neither. We make these plots for user-selected gene pairs using the expression matrix **E** and the cell coordinates matrix **L**. We can also exclusively highlight cross-expressing cell-neighbor pairs. Finally, the tissues are often not upright, partly due to their misorientation with respect to the glass slide, making it difficult to interpret the results. Accordingly, we rotate them using user-defined *n*-degree

$$\theta = n_{degrees} \times \frac{\pi}{180} \tag{11.1}$$

$$x' = \cos(\theta) \cdot x - \sin(\theta) \cdot y \tag{11.2}$$

$$y' = \sin(\theta) \cdot x + \cos(\theta) \cdot y \tag{11.3}$$

where *x′* and *y′* are the cell coordinates after counterclockwise rotation. Rotation does not change the distances between cells, so *x′* and *y′* can be used for downstream analysis.

Spatial enrichment of cross-expression

Cross-expressing cells may be distributed across the tissue or show spatial localization. To quantify their enrichment, we first average the distance between cell-neighbor pairs. We next compare the distances between all cross-expressing cells to the distances between cross-expressing and randomly selected cells. If the former distance is significantly smaller than the latter distance, then cross-expression is spatially enriched.

Data acquisition and preprocessing

MERFISH brain receptor map data

721 We obtained Vizgen MERSCOPE's mouse brain receptor map from
722 https://info.vizgen.com/mouse-brain-data. This data contains three coronal slices from three
723 replicates, with the middle slice covering the center of the brain. We analyzed slice 2 from replicate
724 2, which contains 483 genes and 84,172 cells. We filtered cells with fewer than 50 counts and
725 those lacking brain region annotations (see below), leaving around 82,000 cells. The gene panel
726 consists of cell type markers, G protein coupled receptors (GPCRs), and receptor tyrosine kinases
727 (RTKs). We registered the slice to the Allen CCFv3 (Common Coordinate Framework version 3)
728 brain region atlas[57]. To facilitate this, we annotated the cells using Seurat[82]. Here, we created a
729 Seurat object and used `SCTransform` with the `clip.range` between -10 and 10. We then ran
730 Principal Component Analysis (PCA), setting the number of components to 30 and specifying the
731 features as genes. Next, we used `FindNeighbors` and `FindClusters` with the resolution set
732 to 0.3. The clusters are cell type labels, which help us identify brain structures during registration.
733 (These labels were not used for any analyses.) For registration, we used QuickNii[83] (v3 2017) to
734 linearly align the slice to the Allen CCFv3 atlas using discernible regions like the hippocampus
735 and the ventricles as anchors. We then used VisuAlign[83] (v0.9) to non-linearly transform the slice
736 to improve alignment with the atlas. This procedure assigns a brain region annotation to every
737 cell. Finally, we rotated the image 40 degrees counterclockwise to make it upright.
738 The entire dataset contains 3 replicates with 3 slices each (anterior, middle, posterior),
739 yielding a total of 734,647 cells that we used for additional analyses.
740
741 *BARseq data*
742 The BARseq data was collected in an effort to create a mouse brain cortical cell type atlas[34]. Its
743 104 genes consist largely of excitatory cell type markers (109 total genes), and its 1,161,387 cells
744 were sampled across 40 slices. The cells were iteratively clustered into H1, H2, and H3 types,
745 providing a hierarchical cell type atlas. The H2 types were used during brain registration, which
746 was performed as described above. We filtered cells expressing fewer than 5 genes or with less
747 than 20 counts.
748 We also collected a sagittal mouse hemi-brain data (P56 male) from the left hemisphere
749 (20μm thick sections, 300μm distance between slices) with the same gene panel as the coronal
750 data but with 24 additional ligand-receptor pairs (neuropeptides, neuropeptide receptors,
751 monoamine receptors such as cholinergic, adrenergic, serotonergic, and dopaminergic). This
752 data yielded 133 genes assayed across 1,311,001 cells spanning 16 slices. It was collected for
753 this project and was processed similarly to the coronal data[34]. All experimental procedures were

754 carried out in accordance with the Institutional Animal Care and Use Committee at the Allen
755 Institute for Brain Science.
756
757 *Single-cell RNA-seq (scRNA-seq) and single-nucleus RNA-seq (snRNA-seq) data*
758 We obtained the scRNA-seq[31] and snRNA-seq[68] data from the Brain Initiative Cell Census
759 Network (BICAN) cell type atlases. These data were collected from dissected tissue regions,
760 giving us the cells' coarse anatomical origin. We removed cells with a doublet score of 30 or above
761 and randomly selected 10,000 cells from each region for subsequent analysis.
762
763 **Ligand-receptor cross-expression**
764 We aimed to find cross-expression between known ligand-receptor pairs. In our sagittal data, we
765 selected two slices and within each slice we chose a cortical region. These choices were made
766 randomly. In practice, we chose the visceral area (VISC) in slice 3 and the somatosensory nose
767 region (SSp-n) in slice 5. Next, we selected the well-known neuropeptide somatostatin *Sst* and
768 its cognate receptor *Sstr2* as the candidate pair. Finding their cross-expression significant, we
769 show their expression on tissue and highlight cross-expressing cells. We also compute their
770 bullseye scores and report them as a ratio of cross- to co-expression across 10 neighbors.
771
772 **Cross-expression and cell type heterogeneity**
773 We explore the relationship between cross-expression and cell type heterogeneity using the
774 BARseq coronal data[34]. First, we use *Gfra1* and *Foxp2* to highlight cross-expressing cells and
775 map different cell types to distinct shapes. Second, we count the number of cross-expressing cell-
776 neighbor pairs for numerous genes. Since each cell has a cell type label, we compute cell type
777 purity as the proportion of pairs with the same label. Third, we use the cell type hierarchy to assess
778 if cell-neighbor pairs with the same H1 label have the same H3 label. We first find cross-
779 expressing gene pairs using the entire dataset. Next, using cell pairs with the 'glutamatergic' H1
780 label, we compute the number of pairs with the same or different H3 labels. We perform a similar
781 analysis using cells labelled as 'GABAergic' at the H1 level. Finally, we compute the frequencies
782 with which cell type label combinations are associated between neighboring cells and normalize
783 this by the expected frequencies of those cell type pairs in the population.
784
785 **Discovering combinatorial anatomical marker genes**
786 We observed that cross-expression discovers anatomical marker genes that delineate the
787 thalamus. To quantitatively assess this, we made a mask by combining the following regions:

anterior group of the dorsal thalamus (ATN), intralaminar nuclei of the dorsal thalamus (ILM), lateral group of the dorsal thalamus (LAT), medial group of the dorsal thalamus (MED), midline group of the dorsal thalamus (MTN), ventral group of the dorsal thalamus (VENT), and ventral posterior complex of the thalamus (VP). Importantly, we compared every brain region annotation in our data with Allen CCFv3 atlas[57] and judged the ones presented here to best mark the thalamic regions. This allowed us to calculate the number of cells expressing each gene within or outside the thalamus. For cross-expressing cells, we considered a pair as thalamic if both cells were part of the regional mask. More generally, potential combinatorial markers can be discovered by assessing if their cross-expression is spatially enriched.

Our second exploration involved well-known genes *Foxp2* and *Cdh13*, which mark cortical layer 6 and show pan-layer expression in the cortex, respectively. These genes exhibited significant cross-expression, which was spatially enriched in layer 6, whose boundaries we identified using H2 cell type annotation. The spatial enrichment was viewed by comparing tissue plots with and without highlighting cross-expressing cells.

Networks of cross-expression

Using the MERFISH data, we computed cross-expression p-values between all genes and binarized the matrix at $\alpha \leq 0.05$ to create an adjacency matrix. We calculate the node degree as the number of edges formed by each gene and create a network with second-order edges (shared connections) as outlined in equation 6.2. We set the threshold for second-order edges to 4, meaning that two genes are connected if they share at least 4 first-order edges, ensuring that the higher-order network is robust. Next, we use the `igraph` package[84] to perform Louvain clustering (with default parameters) on the second-order network and thus assign genes to communities.

We visualize the network using Cytoscape[85] (v3.10.1), mapping node size to degree, color to node community, and edge color to edge type (first-order, second-order, or both). We use the "organic" layout and apply "remove overlaps" from the `yFiles` app[86] and tweak the network to further reduce overlaps. Finally, we use the `Legend Creator` app[86] to render a legend with node degree size and community assignment.

Because our network revealed *Gpr20* as topologically salient, we performed gene ontology[87,88] (GO) enrichment analysis on genes that cross-expressed with it ('test set'). Here, we used the entire gene panel (except *Gpr20*) as the background set and used the hypergeometric test to determine if it significantly overlapped with the test set, giving us p-values for each GO functional group. We report FDR-corrected p-values. Additionally, for each gene cross-expressed with *Gpr20*, we used the cells involved in cross-expression, rather than the entire dataset, to

822 compute co-expression with cell type marker genes and compared these global profiles between
823 marker types.
824 Since the cells expressing *Gpr20* visually showed spatial autocorrelation, we assessed
825 their neighbors as well as randomly chosen cells for the expression of *Gpr20*. We L1-normalized
826 the counts for both groups, rendering them into probability distributions, and computed cumulative
827 sums. To calculate the area under curve (AUC), we scaled the neighbor order between 0 and 1
828 and used the `trapz` function from R's `pracma` package to calculate the AUC.
829 Within the main network, we introduce a further constraint that cross-expressing genes
830 must lack significant co-expression. We curate the subnetwork by removing genes with node
831 degree of 1 and assign cell type labels based on genes' co-expression with marker genes. Like
832 before, we perform GO enrichment using the subnetwork genes as the test set and the gene
833 panel as the background set, and report FDR-corrected p-values.
834 To assess whether cross-expression networks are more similar between adjacent slices
835 than between distant slices, we compute slice-specific cross-expression networks and calculate
836 Spearman's correlation between these networks. The correlations are plotted against distances
837 between slices, where the "distance" is the difference in the slice order. As a control, we compute
838 the Spearman's correlations between slice-specific networks obtained from different brains and
839 plot this against the "distance" between the slice ID's.
840
841 **Cross-expression replicability across batches**
842 To assess the replicability of the cross-expression signature, we used the MERFISH dataset
843 containing 3 biological replicates (mouse brains) with 3 slices each, where the slices are sampled
844 from approximately the same location across the brains. We compared the slice-specific networks
845 between corresponding slices. Moreover, for slice-specific and brain region-specific networks, we
846 performed comparisons within the sagittal data and within the coronal data as well as between
847 these two datasets. Finally, observing that the dorsal to ventral direction is sampled in both the
848 coronal and the sagittal brains, we compared the densities of cross-expressing cells in the dorsal
849 to ventral directions across these datasets.
850
851 **Cell segmentation quality control assessment**
852 We assessed the quality of cell segmentation at a global level by comparing co-expression
853 between the scRNA-seq[31] and MERFISH spatial transcriptomic data. Since the scRNA-seq was
854 obtained from dissected brain regions, we established correspondence between these and the
855 brain region annotations in the MERFISH data. The regions used in both data are reported in

Supplementary Table 1. We included only those genes – and in the same order – as present in the MERFISH data. We calculated gene co-expression using Pearson's correlation and compared these across the two datasets.

To quantify variability between platforms, we compared gene co-expression between scRNA-seq and snRNA-seq[68] for the same genes – and in the same order – as above. Because the snRNA-seq was obtained from dissected brain regions, we established correspondence between these and the scRNA-seq data. The regions used in these data are reported in Supplementary Table 2. Like before, we quantified co-expression using Pearson's correlation and compared it across the two datasets.

Gene expression noise thresholds and cell-neighbor relations

Because gene expression measurement is noisy, we applied thresholds of 1 to 10 molecules, thus specifying the minimum number of counts per cell a gene must have to be considered expressed. We then compared cross-expression networks across these thresholds.

Additionally, a cell might be the nearest neighbor of one or more cells. To ensure that our framework captures this variability, we compare cross-expression networks for the one-to-one and many-to-one mappings with each other and with that of the full dataset.

Benchmarking the algorithm's speed

We assessed the speed of the cross-expression algorithm by duplicating our BARseq coronal data, where the gene panel ranged from 2,000 to 8,000 and the number of cells ranged from 20,000 to 200,000. We ran the cross-expression algorithm and calculated the time on a 16 GB Apple M1 Pro macOS Sonoma 14.5 laptop.

Data availability

The MERFISH/ MERSCOPE data was downloaded from Vizgen's mouse brain receptor map at https://info.vizgen.com/mouse-brain-data. The BARseq coronal data is deposited at the Brain Image Library (BIL) at https://api.brainimagelibrary.org/web/view?bildid=ace-cry-zip, with the cell and rolony level data at https://data.mendeley.com/datasets/8bhhk7c5n9/1. The BARseq sagittal data's sequencing images are being deposited to BIL. While it is being approved, we stored the cell-level gene expression and cell metadata at https://drive.google.com/drive/folders/1fk5JDeVJcE71iH1AalCT0il9PN9DTJJm?usp=drive_link. scRNA-seq is at https://alleninstitute.github.io/abc_atlas_access/descriptions/WMB_dataset.html and snRNA-seq at https://docs.braincelldata.org/downloads/index.html.

890 **Code availability**

891 The R package is available at https://github.com/ameersarwar/cross_expression

892

893 **Acknowledgements**

894 A.S. received funding from University of Toronto's FAST Fellowship and from NSERC's Canada
895 Graduate Scholarship. X.C. was supported by grants R01MH133181 and DP2MH132940. L.F.
896 received support from R01MH133181 and J.G. from both R01MH133181 and R01MH113005.
897 The content is solely the responsibility of the authors and does not necessarily represent the
898 official views of the National Institutes of Health.

899

900 **Author Contributions**

901 J.G. conceived the project. A.S. conducted the analyses, developed the software, and wrote the
902 first draft under supervision from J.G. M.R and H.C. collected the sagittal brain data under
903 supervision from X.C. L.F. managed, curated, and parsed the datasets. All authors interpreted the
904 results and edited the manuscript.

905

906 **Competing interests**

907 L.F. owns shares in Quince Therapeutics and has received consulting fees from PeopleBio Co.,
908 GC Therapeutics Inc., Cortexyme Inc., and Keystone Bio. The remaining authors declare no
909 competing interests.

910

911 **References**

912 1. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat. Methods* **18**, 9–
913 14 (2021).
914 2. Zhuang, X. Spatially resolved single-cell genomics and transcriptomics by imaging.
915 *Nat. Methods* **18**, 18–22 (2021).
916 3. Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling
917 technologies. *Nat. Rev. Genet.* **23**, 741–759 (2022).
918 4. Tian, L., Chen, F. & Macosko, E. Z. The expanding vistas of spatial transcriptomics.
919 *Nat. Biotechnol.* **41**, 773–782 (2023).
920 5. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science* **381**,
921 eabq4964 (2023).
922 6. Liu, L. *et al.* Spatiotemporal omics for biology and medicine. *Cell* **187**, 4488–4519
923 (2024).
924 7. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–
925 546 (2022).
926 8. Ma, Y. & Zhou, X. Accurate and efficient integrative reference-informed spatial domain
927 detection for spatial transcriptomics. *Nat. Methods* **21**, 1231–1244 (2024).

9. Kim, J. *et al.* Unsupervised discovery of tissue architecture in multiplexed imaging. *Nat. Methods* **19**, 1653–1661 (2022).
10. Chidester, B., Zhou, T., Alam, S. & Ma, J. SpiceMix enables integrative single-cell spatial modeling of cell identity. *Nat. Genet.* **55**, 78–88 (2023).
11. Jerby-Arnon, L. & Regev, A. DIALOGUE maps multicellular programs in tissue from single-cell or spatial transcriptomics data. *Nat. Biotechnol.* **40**, 1467–1477 (2022).
12. Haviv, D. *et al.* The covariance environment defines cellular niches for spatial inference. *Nat. Biotechnol.* 1–12 (2024) doi:10.1038/s41587-024-02193-4.
13. Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* **17**, 193–200 (2020).
14. Ghazanfar, S. *et al.* Investigating higher-order interactions in single-cell data with scHOT. *Nat. Methods* **17**, 799–806 (2020).
15. Bernstein, M. N. *et al.* SpatialCorr identifies gene sets with spatially varying correlation structure. *Cell Rep. Methods* **2**, (2022).
16. Miller, B. F., Bambah-Mukku, D., Dulac, C., Zhuang, X. & Fan, J. Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome Res.* **31**, 1843–1855 (2021).
17. Armingol, E., Baghdassarian, H. M. & Lewis, N. E. The diversification of methods for studying cell–cell interactions and communication. *Nat. Rev. Genet.* **25**, 381–400 (2024).
18. Almet, A. A., Tsai, Y.-C., Watanabe, M. & Nie, Q. Inferring pattern-driving intercellular flows from single-cell and spatial transcriptomics. *Nat. Methods* 1–12 (2024) doi:10.1038/s41592-024-02380-w.
19. Yang, W. *et al.* Deciphering cell–cell communication at single-cell resolution for spatial transcriptomics with subgraph-based graph attention network. *Nat. Commun.* **15**, 7101 (2024).
20. Cang, Z. *et al.* Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat. Methods* **20**, 218–228 (2023).
21. Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).
22. Troulé, K. *et al.* CellPhoneDB v5: inferring cell-cell communication from single-cell multiomics data. Preprint at https://doi.org/10.48550/arXiv.2311.04567 (2023).
23. Liu, J. *et al.* CytoSignal Detects Locations and Dynamics of Ligand-Receptor Signaling at Cellular Resolution from Spatial Transcriptomic Data. 2024.03.08.584153 Preprint at https://doi.org/10.1101/2024.03.08.584153 (2024).
24. Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics. 2023.11.05.565674 Preprint at https://doi.org/10.1101/2023.11.05.565674 (2023).
25. Li, Z., Wang, T., Liu, P. & Huang, Y. SpatialDM for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns. *Nat. Commun.* **14**, 3995 (2023).
26. Yuan, Y. & Bar-Joseph, Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol.* **21**, 300 (2020).

27. Tang, Z., Zhang, T., Yang, B., Su, J. & Song, Q. spaCI: deciphering spatial cellular communications through adaptive graph model. *Brief. Bioinform.* **24**, bbac563 (2023).
28. Singhal, V. *et al.* BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nat. Genet.* **56**, 431–441 (2024).
29. Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342–1351 (2021).
30. Zubair, A. *et al.* Cell type identification in spatial transcriptomics data can be improved by leveraging cell-type-informative paired tissue images using a Bayesian probabilistic model. *Nucleic Acids Res.* **50**, e80 (2022).
31. Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317–332 (2023).
32. Zhang, M. *et al.* Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).
33. Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution. *Nature* **622**, 552–561 (2023).
34. Chen, X. *et al.* Whole-cortex in situ sequencing reveals input-dependent area identity. *Nature* 1–10 (2024) doi:10.1038/s41586-024-07221-6.
35. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95**, 14863–14868 (1998).
36. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
37. Harris, B. D., Crow, M., Fischer, S. & Gillis, J. Single-cell co-expression analysis reveals that transcriptional modules are shared across cell types in the brain. *Cell Syst.* **12**, 748-756.e3 (2021).
38. Crow, M., Suresh, H., Lee, J. & Gillis, J. Coexpression reveals conserved gene programs that co-vary with cell type across kingdoms. *Nucleic Acids Res.* **50**, 4302–4314 (2022).
39. Crow, M. & Gillis, J. Co-expression in Single-Cell Analysis: Saving Grace or Original Sin? *Trends Genet.* **34**, 823–831 (2018).
40. Walker, B. L. & Nie, Q. NeST: nested hierarchical structure identification in spatial transcriptomic data. *Nat. Commun.* **14**, 6554 (2023).
41. Sun, Y.-C. *et al.* Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nat. Neurosci.* **24**, 873–885 (2021).
42. Wu, X.-B. *et al.* Excitatory Projections from the Prefrontal Cortex to Nucleus Accumbens Core D1-MSNs and κ Opioid Receptor Modulate Itch-Related Scratching Behaviors. *J. Neurosci.* **43**, 1334–1347 (2023).
43. Xiao, P. *et al.* Ligand recognition and allosteric regulation of DRD1-Gs signaling complexes. *Cell* **184**, 943-956.e18 (2021).
44. Szklarczyk, K., Korostynski, M., Cieslak, P. E., Wawrzczak-Bargiela, A. & Przewlocki, R. Opioid-dependent regulation of high and low fear responses in two inbred mouse strains. *Behav. Brain Res.* **292**, 95–101 (2015).

1017 45. Minett, M. S. *et al.* Endogenous opioids contribute to insensitivity to pain in
1018 humans and mice lacking sodium channel Nav1.7. *Nat. Commun.* **6**, 8967 (2015).
1019 46. Wawrzczak-Bargieła, A. *et al.* Neuropathic Pain Dysregulates Gene Expression
1020 of the Forebrain Opioid and Dopamine Systems. *Neurotox. Res.* **37**, 800–814 (2020).
1021 47. Beaulieu, J.-M. & Gainetdinov, R. R. The Physiology, Signaling, and
1022 Pharmacology of Dopamine Receptors. *Pharmacol. Rev.* **63**, 182–217 (2011).
1023 48. Singh, V. *et al.* Somatostatin Receptor Subtype-2-Deficient Mice with Diet-
1024 Induced Obesity Have Hyperglycemia, Nonfasting Hyperglucagonemia, and
1025 Decreased Hepatic Glycogen Deposition. *Endocrinology* **148**, 3887–3899 (2007).
1026 49. Fabricius, D. *et al.* Characterization of Intestinal and Pancreatic Dysfunction in
1027 VPAC1-Null Mutant Mouse. *Pancreas* **40**, 861 (2011).
1028 50. Khaleduzzaman, M. *et al.* Structure of the Human Type XIX Collagen (*COL19A1*)
1029 Gene, Which Suggests It Has Arisen from an Ancestor Gene of the FACIT Family.
1030 *Genomics* **45**, 304–312 (1997).
1031 51. Calvo, A. C. *et al.* Type XIX collagen: a promising biomarker from the basement
1032 membranes. *Neural Regen. Res.* **15**, 988 (2020).
1033 52. Sticco, M. J. *et al.* C1QL3 promotes cell-cell adhesion by mediating complex
1034 formation between ADGRB3/BAI3 and neuronal pentraxins. *FASEB J.* **35**, e21194
1035 (2021).
1036 53. Martinelli, D. C. *et al.* Expression of C1ql3 in Discrete Neuronal Populations
1037 Controls Efferent Synapse Numbers and Diverse Behaviors. *Neuron* **91**, 1034–1051
1038 (2016).
1039 54. Van Itallie, C. M. *et al.* MARCKS-related protein regulates cytoskeletal
1040 organization at cell–cell and cell–substrate contacts in epithelial cells. *J. Cell Sci.* **131**,
1041 jcs210237 (2018).
1042 55. Kondrychyn, I. *et al.* Marcksl1 modulates endothelial cell mechanoresponse to
1043 haemodynamic forces to control blood vessel shape and size. *Nat. Commun.* **11**,
1044 5476 (2020).
1045 56. Shestakova, E. A., Singer, R. H. & Condeelis, J. The physiological significance of
1046 β-actin mRNA localization in determining cell polarity and directional motility. *Proc.*
1047 *Natl. Acad. Sci.* **98**, 7045–7050 (2001).
1048 57. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas.
1049 *Cell* **181**, 936-953.e20 (2020).
1050 58. Delafontaine, P., Song, Y.-H. & Li, Y. Expression, Regulation, and Function of
1051 IGF-1, IGF-1R, and IGF-1 Binding Proteins in Blood Vessels. *Arterioscler. Thromb.*
1052 *Vasc. Biol.* **24**, 435–444 (2004).
1053 59. Dallinga, M. G. *et al.* IGF2 and IGF1R identified as novel tip cell genes in primary
1054 microvascular endothelial cell monolayers. *Angiogenesis* **21**, 823–836 (2018).
1055 60. Imrie, H. *et al.* Novel Role of the IGF-1 Receptor in Endothelial Function and
1056 Repair: Studies in Endothelium-Targeted IGF-1 Receptor Transgenic Mice. *Diabetes*
1057 **61**, 2359–2368 (2012).
1058 61. Liang, W., Zheng, Y., Zhang, J. & Sun, X. Multiscale modeling reveals
1059 angiogenesis-induced drug resistance in brain tumors and predicts a synergistic drug
1060 combination targeting EGFR and VEGFR pathways. *BMC Bioinformatics* **20**, 203
1061 (2019).

62. Le, X. *et al.* Dual EGFR-VEGF Pathway Inhibition: A Promising Strategy for Patients With *EGFR*-Mutant NSCLC. *J. Thorac. Oncol.* **16**, 205–215 (2021).
63. Milner, C. S., Hansen, T. M., Singh, H. & Brindle, N. P. J. Roles of the receptor tyrosine kinases Tie1 and Tie2 in mediating the effects of angiopoietin-1 on endothelial permeability and apoptosis. *Microvasc. Res.* **77**, 187–191 (2009).
64. Niu, Q. *et al.* Inhibition of Tie-2 Signaling Induces Endothelial Cell Apoptosis, Decreases Akt Signaling and Induces Endothelial Cell Expression of the Endogenous Anti-Angiogenic Molecule, Thrombospondin-1. *Cancer Biol. Ther.* **3**, 402–405 (2004).
65. Savant, S. *et al.* The Orphan Receptor Tie1 Controls Angiogenesis and Vascular Remodeling by Differentially Regulating Tie2 in Tip and Stalk Cells. *Cell Rep.* **12**, 1761–1773 (2015).
66. Zhang, Y. *et al.* Regulation of Glucose Homeostasis and Lipid Metabolism by PPP1R3G-mediated Hepatic Glycogenesis. *Mol. Endocrinol.* **28**, 116–126 (2014).
67. Aaron, P. A., Jamklang, M., Uhrig, J. P. & Gelli, A. The blood–brain barrier internalises via the EphA2-tyrosine kinase receptor. *Cell. Microbiol.* **20**, e12811 (2018).
68. Langlieb, J. *et al.* The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**, 333–342 (2023).
69. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
70. Fu, X. *et al.* BIDCell: Biologically-informed self-supervised learning for segmentation of subcellular spatial transcriptomics data. *Nat. Commun.* **15**, 509 (2024).
71. Ma, J. *et al.* The multimodality cell segmentation challenge: toward universal solutions. *Nat. Methods* **21**, 1103–1113 (2024).
72. Choudhary, S. & Satija, R. Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol.* **23**, 27 (2022).
73. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
74. Yu, Q., Jiang, M. & Wu, L. Spatial transcriptomics technology in cancer research. *Front. Oncol.* **12**, (2022).
75. Chen, W.-T. *et al.* Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* **182**, 976-991.e19 (2020).
76. Gurkar, A. U. *et al.* Spatial mapping of cellular senescence: emerging challenges and opportunities. *Nat. Aging* **3**, 776–790 (2023).
77. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).
78. Zeng, H. *et al.* Spatially resolved single-cell translatomics at molecular resolution. *Science* **380**, eadd3067 (2023).
79. Bentley, J. L. Multidimensional binary search trees used for associative searching. *Commun ACM* **18**, 509–517 (1975).
80. Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R. & Wu, A. Y. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J ACM* **45**, 891–923 (1998).

81. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
82. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
83. Puchades, M. A., Csucs, G., Ledergerber, D., Leergaard, T. B. & Bjaalie, J. G. Spatial registration of serial microscopic brain images to three-dimensional reference atlases with the QuickNII tool. *PLOS ONE* **14**, e0216796 (2019).
84. Csárdi, G. *et al.* igraph for R: R interface of the igraph library for graph theory and network analysis. Zenodo https://doi.org/10.5281/zenodo.10681749 (2024).
85. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
86. Morris, J. H., Kuchinsky, A., Ferrin, T. E. & Pico, A. R. enhancedGraphics: a Cytoscape app for enhanced node graphics. Preprint at https://doi.org/10.12688/f1000research.4460.1 (2014).
87. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
88. The Gene Ontology Consortium *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).