Cross-expression meta-analysis of 695 brain samples reveals coordinated gene expression across spatially adjacent cells

Ameer Sarwar[1] (ameer.sarwar@mail.utoronto.ca), Mara Rue[2] (mara.rue@alleninstitute.org) Leon French[3] (leon.french@utoronto.ca), Helen Cross[2] (crossh@carleton.edu), Sarah Choi[3] (sr.choi@mail.utoronto.ca), Xiaoyin Chen[2] (xiaoyin.chen@alleninstitute.org) and Jesse Gillis[3]* (jesse.gillis@utoronto.ca)

[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

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. Rather than focus on how these cells (samples) communicate with each other, we reframe the problem to investigate how genes (features) coordinate their expression between neighboring cells. To study these phenomena – called cross-expression – we compare all genes to find pairs that coordinate their expression between adjacent cells, thereby avoiding curating gene lists or annotating cell types. Our end-to-end method recovers ligand-receptor pairs as cross-expressing genes and finds gene combinations that mark anatomical regions, complementing marker gene-based region annotation. Leveraging the overlapping genes across different panels, we use multiple atlas-scale adult mouse brain datasets (~25 million cells, 695 samples, 8 technologies) to create an integrated, meta-analytic cross-expression network, whose communities are enriched in spatial processes such as synaptic signaling and G protein coupled receptor activity. Highlighting cross-expression's biological utility, our network shows that genes *Drd1* and *Gpr6*, which are individually implicated in Parkinson's disease (PD) and are being pursued as therapeutic targets, are cross-expressed within the striatum, hinting at their joint role in PD pathophysiology. We provide an efficient R package (https://github.com/gillislab/CrossExpression/) to computationally analyze and visually explore cross-expression patterns, which allow us to better understand how genes coordinate their expression in space to perform tissue-level functions.

Keywords

spatial transcriptomics; cross-expression; cell-neighbor relations; gene expression coordination; spatial signaling; meta-analytic network

Background

Spatial transcriptomic technologies record cells' gene expression alongside their physical locations, enabling us to understand how they influence one another within the tissue [1]. Focusing on select genes, imaging-based platforms profile expression at the single cell level, giving us a high-resolution snapshot of spatial gene expression [2–8]. They have facilitated numerous studies on defining local spatial patterns [9–12], finding gene covariation in spatial niches [13–19], elucidating cell-cell interactions using ligand-receptor expression [20–28], and determining spatial cell type heterogeneity and tissue structure [29–31]. These efforts have resulted in a greater understanding of tissue biology, culminating in the generation and exploration of reference atlases [32–35].

Imaging-based platforms can now profile over a thousand genes in millions of cells [2–8], generating large amounts of data ripe for biological discovery. These data can be analyzed at a high resolution, where we can investigate individual genes, single cells, and their spatial relations. The dominant framework in this space is the cell-cell communication methods [20–28], which infer co-localized ligands and receptors in neighboring cells. These approaches typically rely on curated ligand-receptor databases to infer interactions between cell types. Despite rapid progress in this area, these methods have important limitations. First, the reliance on extant databases limits novel discoveries to well-known ligands and receptors, thereby overlooking potential interaction partners present in the gene panel. Second, the use of cell types requires that the annotations are sufficiently accurate, but this is currently an open problem owing to the limited size of the gene panels, often requiring 'label transfer' from matched single-cell RNA-seq datasets. Indeed, the communication inference is made at the aggregate, cell type level when the underlying data allows a fully 'bottom-up' analysis at the single cell level. Third, the cell-centric perspective makes it difficult to integrate datasets across experiments, limiting the inferences to single studies. Because large-scale datasets are rapidly being collected, an approach is needed to develop a 'common map' between them to discover reproducible biological signals.

The gene-centric framework is an alternative to the cell-cell inference approaches. By prioritizing genes (features) over cells (samples), it allows comparisons across datasets using genes shared between their panels. Moreover, when the cell-cell interaction methods use ligand-receptor co-localization, they implicitly leverage the gene-centric perspective, where inferences are made if the relevant genes are expressed in neighboring cells more frequently than expected by chance. A number of studies have explicitly used the gene-centric approach. For example, Haviv et al. (2024) model how gene-gene co-expression within the same cells changes across spatial niches [14]. Studies by Miller et al. (2021) and Li et al. (2023) consider gene-gene coordination between spatially adjacent cells using the global bivariate Moran's *I* statistic [18,19]. Whereas Haviv et al. (2024) ignore gene-gene coordination between neighboring cells, Miller et al. (2021) and Li et al. (2023) require ligand-receptor databases, thus limiting their scope. Additionally, the spatial correlation methods, such as the global bivariate Moran's *I* statistic, are known to increase false positives because they do not account for "within location" association. For example, if genes A and B are both co-expressed in neighboring cells, then they would appear correlated across neighbors even when the two cells function fully autonomously.

Here, we introduce "cross-expression", which models the degree to which genes coordinate their expression across neighboring cells. Avoiding the use of curated databases and cell type labels, our fully end-to-end and bottom-up method uses the raw gene expression and cell location matrices to determine which gene pairs, among all possible pairs, coordinate their expression between neighboring cells. We explicitly account for spurious associations induced by "within cell" co-expression patterns, thereby revealing genuine coordination once cell intrinsic processes have been controlled. A consequence of our method is that it recovers ligand-receptor pairs as cross-expressed genes, e.g., *Sst* and *Sstr2*, thus extending the well-developed cell-cell communication framework. Its gene-centric approach facilitates integrative analysis, where we create a cross-expression meta-analytic network using 13 datasets spanning ~25 million cells, 695 brain samples, and 8 technological platforms. Our network shows highly modular structure,

with communities enriched for gene ontology (GO) terms like synaptic signaling, neurotransmitter regulation, cell adhesions, G protein receptor activity, and other spatial processes. Highlighting its biological utility, the network reveals that across numerous samples the genes *Drd1* and *Gpr6*, which are individually known to play roles in Parkinson's disease (PD) and are being pursued as therapeutic targets, are cross-expressed in the striatum, hinting at their joint function in the central anatomical locus in PD pathophysiology. We also show that the cross-expression patterns in one dataset reliably predict those in other datasets as well as that similar anatomical regions have similar cross-expression profiles, indicating that our method reliably detects subtle variations in gene expression across batches and anatomical space. To facilitate the analysis and exploration of cross-expression patterns, we provide a fully vectorized R package that processes hundreds of thousands of gene pairs across millions of cells within minutes on a standard personal laptop, allowing in-depth analyses of how genes coordinate their expression in space to perform tissue-level functions.

Results

Cross-expression overview

Cross-expression is defined as the degree to which two genes coordinate their expression across neighboring cells (Fig. 1a). Specifically, it asks whether the expression of gene A in a cell is associated with the expression of gene B in the cell's (nearest) neighbor. To properly capture these patterns, we control for co-expression or, more generally, the association induced between neighboring locations due to the association present within the same location. In particular, if genes A and B are co-expressed in both the neighboring cells, then they would appear correlated across these cells even when the cells function completely independently. We control for this by defining cross-expression mutually exclusively, namely gene A is expressed in the cell without gene B and gene B is expressed in the neighbor without gene A.

We quantify cross-expression in three statistically robust and highly interpretable ways. First, we use the hypergeometric test to compute p-values (Fig. 1b). We begin by finding each cell's nearest neighbor to create a cell-to-neighbor 'mapping'. While cells almost never have two nearest (equidistant) neighbors, two or more cells can have the same neighbor. The hypergeometric test is performed on these cell-neighbor pairs. For genes A and B, we count the number of pairs where the cell expresses gene A (sample size) and the number of pairs where the neighbor expresses gene B (possible successes). We then compute the intersection (observed successes), namely the number of pairs where the cell expresses A and the neighbor expresses B. Together with the total number of pairs (population size), we use the hypergeometric cumulative distribution function (CDF) to analytically compute the probability of observing as many or more successes, giving us a p-value for the cross-expression of genes A and B.

Second, we quantify the effect size by comparing the number of cross-expressing pairs to the number of co-expressing cells (Fig. 1c). The central idea is that if gene B is randomly assigned to cells while gene A is held constant, then we expect some co-expression and some cross-expression simply by chance, where their expected ratio is 1. This ratio is our effect size, with values greater than 1 indicating more cross- than co-expression, and we represent it till the *n*-th neighbor using concentric rings in the bullseye plot.

Third, we compute a Pearson correlation between gene A in each A-expressing cell and gene B in its nearest neighbor. This yields a directed set of cell-neighbor pairs, capturing how B varies across the local context of A. Some neighbors are shared by multiple cells, reflecting spatial centrality; this weighting mirrors the biological structure of the tissue, where certain cells may influence more of their surroundings. Although this induces non-uniform contributions, each pair is treated independently from the target cell's perspective, preserving statistical validity for correlation.

Although we focus on individual cells, groups of cells may form spatial niches and gene expression may be coordinated between niches. To assess 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 the creation of cross-expression networks (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 useful analytical framework for deeply exploring spatial transcriptomic data.

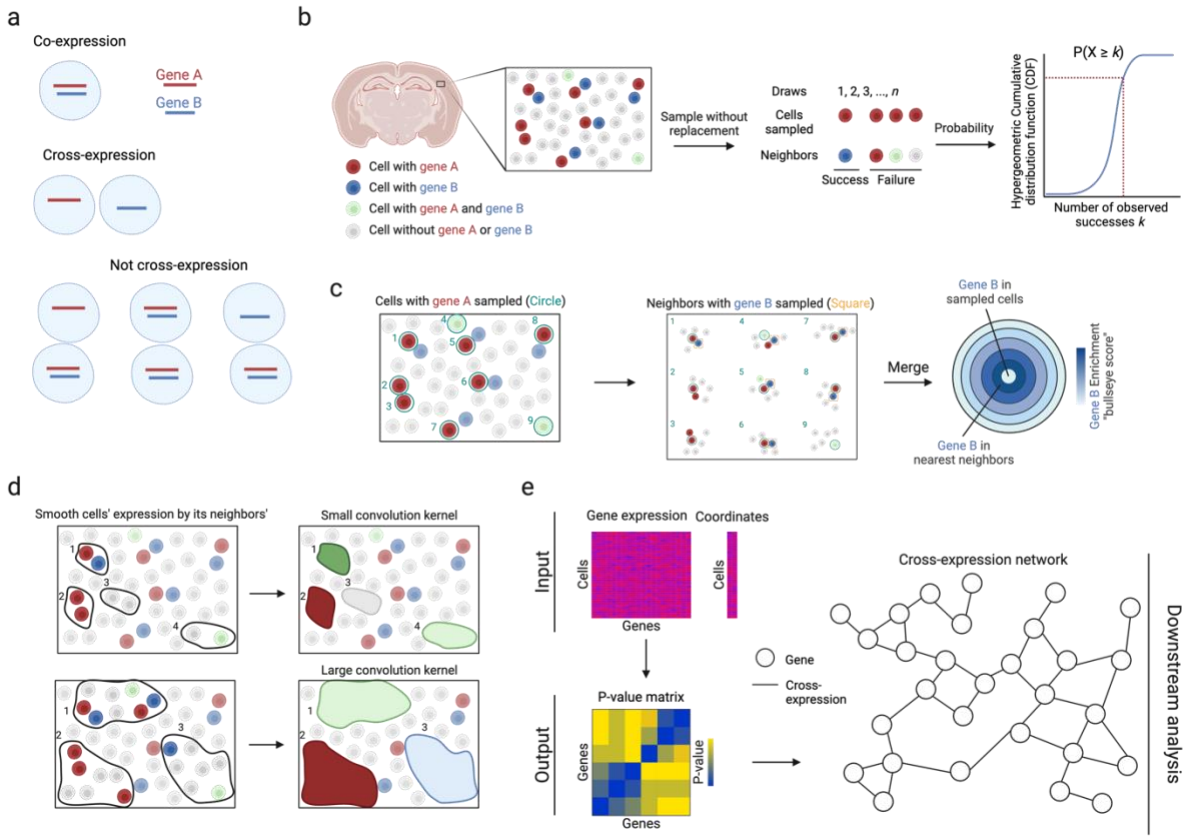


Fig. 1 | Cross-expression analysis. a, Cross-expression is the mutually exclusive expression of genes between neighboring cells. If either cell expresses both genes, the cell pair is not considered to cross-express. **b**, The probability that two genes cross-express is modeled by the hypergeometric distribution, where the 'successful trials' are the cell-neighbor pairs where the cell expresses gene A and the neighbor expresses gene B. **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 the creation of cross-expression networks. 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 16 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 the enzymes that synthesize the corresponding ligands are often expressed in nearby cells [20–28], we reasoned that these genes should show cross-expression. As an example, we find that across the cortical somatosensory nose region and visceral areas (Fig. 2a), the neuropeptide somatostatin *Sst* and its cognate receptor *Sstr2* are cross-expressed (Fig. 2b, left, p-values ≤ 0.01 and 0.05, respectively). Indeed, these genes are consistently expressed across neighboring cells (Fig. 2b, right), a pattern that is otherwise difficult to discover without prior knowledge. While cross-expression recovers ligand-receptor genes, it does not imply that they form synaptic or endocrine connections, since spatial transcriptomic measurements are restricted to the cell body.

Next, we explore the bullseye plots, which allow us to quantify the effect size by comparing cross-expression with co-expression. For *Sst* and *Sstr2* in the somatosensory nose (2,015 cells) and visceral regions (1,603 cells), we see a bullseye pattern with low co-expression and high

200 cross-expression that decreases for distant neighbors (Fig. 2c). Specifically, for these regions the
201 bullseye score ratio between the first neighbor and the central cell is 1.8 and 1.6, respectively,
202 whereas the ratio between the averaged second-to-tenth neighbor and the central cell is 1.3 and
203 1.2. These findings suggest that for central cells expressing one gene in a pair, a higher proportion
204 of adjacent neighbors, but not the more distant ones, express the cognate gene within the local
205 spatial niche, underscoring the specificity and resolution with which patterns of coordinated gene
206 expression can be recovered. We next compare the bullseye plots for gene pairs with and without
207 cross-expression (Fig. 2d), finding that the former match the patterns just described. To quantify
208 this, we compare the bullseye scores of the nearest neighbors with those of cells expressing gene
209 A, discovering that this ratio is much greater for genes that cross-express than for those that do
210 not (Fig. 2d, inset, Mann-Whitney U test, p-value ≤ 0.001, median ratios: 1.5 and 0.9,
211 respectively). Notably, this ratio is approximately 1 for genes that do not cross-express,
212 suggesting that here gene B is expressed in neighbors and cells alike. Hence, the bullseye
213 approach intuitively visualizes and quantifies the effect size, making it suitable for downstream
214 analysis, such as comparing cross-expression between different regions.

215 We next conducted brain-wide analysis and found that 20% of possible ligand-receptor
216 gene pairs and 4% of non-signaling gene pairs are cross-expressed, thus generating novel
217 candidates that potentially encode functionally relevant interactions. In fact, these patterns are
218 spatially enriched, where most gene pairs cross-express in a few slices and some cross-express
219 in multiple slices (Extended Data Fig. 1a). We now highlight some notable examples of cross-
220 expression for both signaling and non-signaling genes. The dopamine receptor D_1 (*Drd1*) and
221 proenkephalin (*Penk*) are strongly cross-expressed (Extended Data Fig. 1b), with discernible
222 spatial enrichment in the striatal regions. *Drd1* is involved in the reward system [36,37] while *Penk*
223 generates opioids that modulate fear response [38] and nociception [39,40], suggesting that these
224 genes may be involved in avoidance behavior. Indeed, Penk is strongly co-expressed with the
225 dopamine receptor D_2 (*Drd2*) (Pearson's R = 0.72 in scRNA-seq striatal data; *Drd2* is not in our

226 gene panel), indicating that the D1 and D2 neurons are spatially intermingled, allowing them to
227 play interrelated roles in motor control [41]. Additionally, we find that the somatostatin receptor
228 *Sstr2* cross-expresses with vasoactive intestinal polypeptide receptor 1 (*Vipr1*/*VPAC1*) in the
229 cortex (Extended Data Fig. 1c), suggesting a potential complementary interaction in modulating
230 local neuronal circuits and influencing neuroendocrine signaling pathways [42]. Beyond the
231 signaling genes, we note that the fibril-associated *Col19a1* (collagen type XIX alpha 1 chain), a
232 gene involved in maintaining the extracellular matrix (ECM) integrity [43,44], cross-expresses with
233 *C1ql3* (complement C1q like 3) (Extended Data Fig. 1d), whose secretion in the ECM facilitates
234 synapse homeostasis and the formation of cell-cell adhesion complexes [45,46]. Finally, our
235 analysis reveals that *Marcksl1* (myristoylated alanine-rich C-kinase substrate), which is involved
236 in adherens junctions and cytoskeletal processes [47,48], cross-expresses with actin beta (*Actb*)
237 (Extended Data Fig. 1e), hinting at their involvement in local tissue architecture [49]. Taken
238 together, the cross-expression analysis not only reveals expected relationships between signaling
239 molecules, but it also discovers genes implicated in the tissue microenvironment. Accordingly,
240 cross-expression is an unbiased, data-driven framework for finding genes with orchestrated
241 spatial expression profiles, with potential for novel discovery increasing with varying sizes and
242 compositions of the gene panels.

243 We have thus far investigated cross-expression between cells and their neighbors. Yet,
244 gene expression may be coordinated between more distant neighbors or between large spatial
245 niches. The former is facilitated by changing the rank of the nearest neighbor tested. The latter is
246 enabled by smoothing a gene's expression in a cell by averaging it with its expression in nearby
247 cells, as shown for cortical layer 4 marker *Rorb* (Fig. 2e) and layer 6 marker *Foxp2* (Extended
248 Data Fig. 1f) in the auditory cortex [32].

249 Although cross-expression may appear at varying length scales, we focus our analyses at
250 the cellular level to investigate its signature between individual cells.

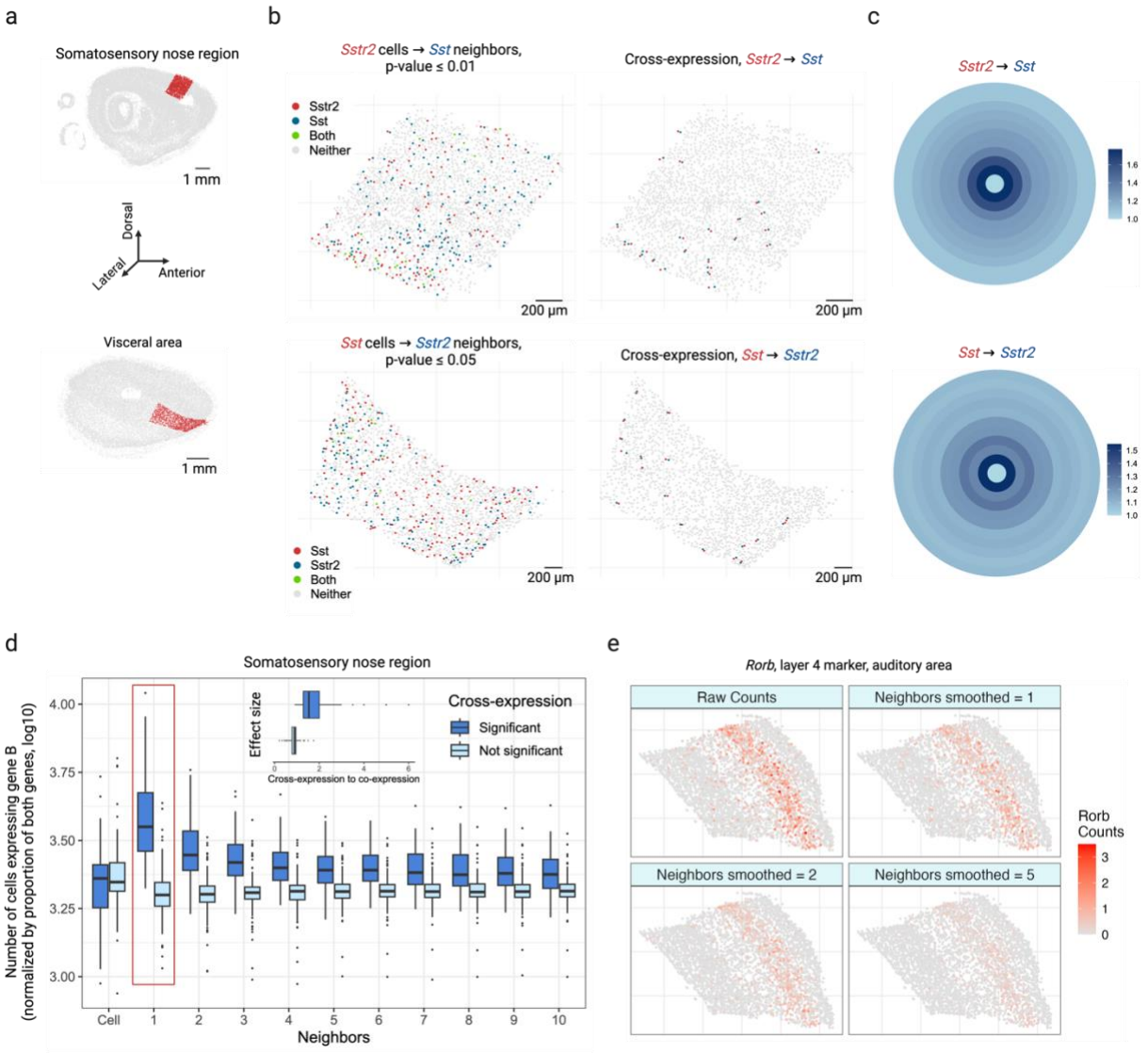


Fig. 2 | Cross-expression analysis reveals coordinated gene expression between neighboring cells. a, Sagittal brain slices showing cortical somatosensory nose region (top) and visceral area (bottom) as randomly selected regions of interest. **b**, Neuropeptide somatostatin *Sst* and its cognate receptor *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.

266 **Cross-expression is driven by subtle and consistent cell subtype compositional**
267 **differences**

268 Having seen that cross-expression recovers coordinated spatial gene expression, we now explore
269 its relationship with cell type heterogeneity. For this purpose, we use another BARseq dataset
270 [35] that was recently used to create a mouse cortical cell type atlas using the same 104 excitatory
271 marker genes as before. Here, we observe that genes cross-express between cells of the same
272 and of different types. For example, *Gfra1* and *Foxp2* are cross-expressed within the same cell
273 type L4/5 IT (intratelencephalic) and between different cell types Car3 or CT (corticothalamic) and
274 L4/5 IT (Fig. 3a). In general, genes vary greatly in terms of the cell type labels of cross-expressing
275 cell pairs (Fig. 3b). For instance, for some gene pairs 40% of the cell pairs have the same cell
276 type label while in others as many as 90% of the cell pairs belong to different cell types (Extended
277 Data Fig. 2a). Moreover, some genes involve many while others involve few cross-expressing
278 cells. For example, in the analyzed data the median number of cross-expressing cell pairs is
279 2,378, and 27% of genes involve over 4,000 while only 5% involve 400 or fewer pairs (Extended
280 Data Fig. 2b), indicating that the density of gene cross-expression is highly variable. Interestingly,
281 cell type purity – the proportion of cell pairs with the same type – decreases as more cell pairs
282 cross-express (Fig. 3c, Spearman's $\rho = -0.46$), highlighting a potential role for spatially
283 intermingled cell types in patterns of cross-expression.

284 To assess the influence of spatial cell type composition more broadly, we use our
285 hierarchical cell type taxonomy [35], where types at a higher-level divide into subtypes at a lower
286 level. Using cross-expressing glutamatergic cells, we find that 64% of the pairs consist of different
287 cell subtypes (Fig. 3d, right-tailed Wilcoxon signed-rank test, different labels ≥ same labels, p-
288 value ≤ 0.0001, Extended Data Fig. 2c), suggesting that subtle cell type differences drive cross-
289 expression. However, for cross-expressing GABAergic cells, we find that only 44% of the pairs
290 have different cell subtype labels (Extended Data Fig. 2d-e right-tailed Wilcoxon signed-rank test,
291 different labels ≥ same labels, p-value = 1), reflecting the fact that our gene panel is optimized to

detect cell subtype differences between excitatory, but not inhibitory, neurons. Crucially, we observe that cells of one type consistently cross-express with cells of another type (Fig. 3e, Extended Data Fig. 2f), indicating that cross-expression recapitulates patterns of cell type composition. Since cell type labels are assigned based on the expression of many genes, repeated spatial proximity of cell types results in the cross-expression of their marker genes.

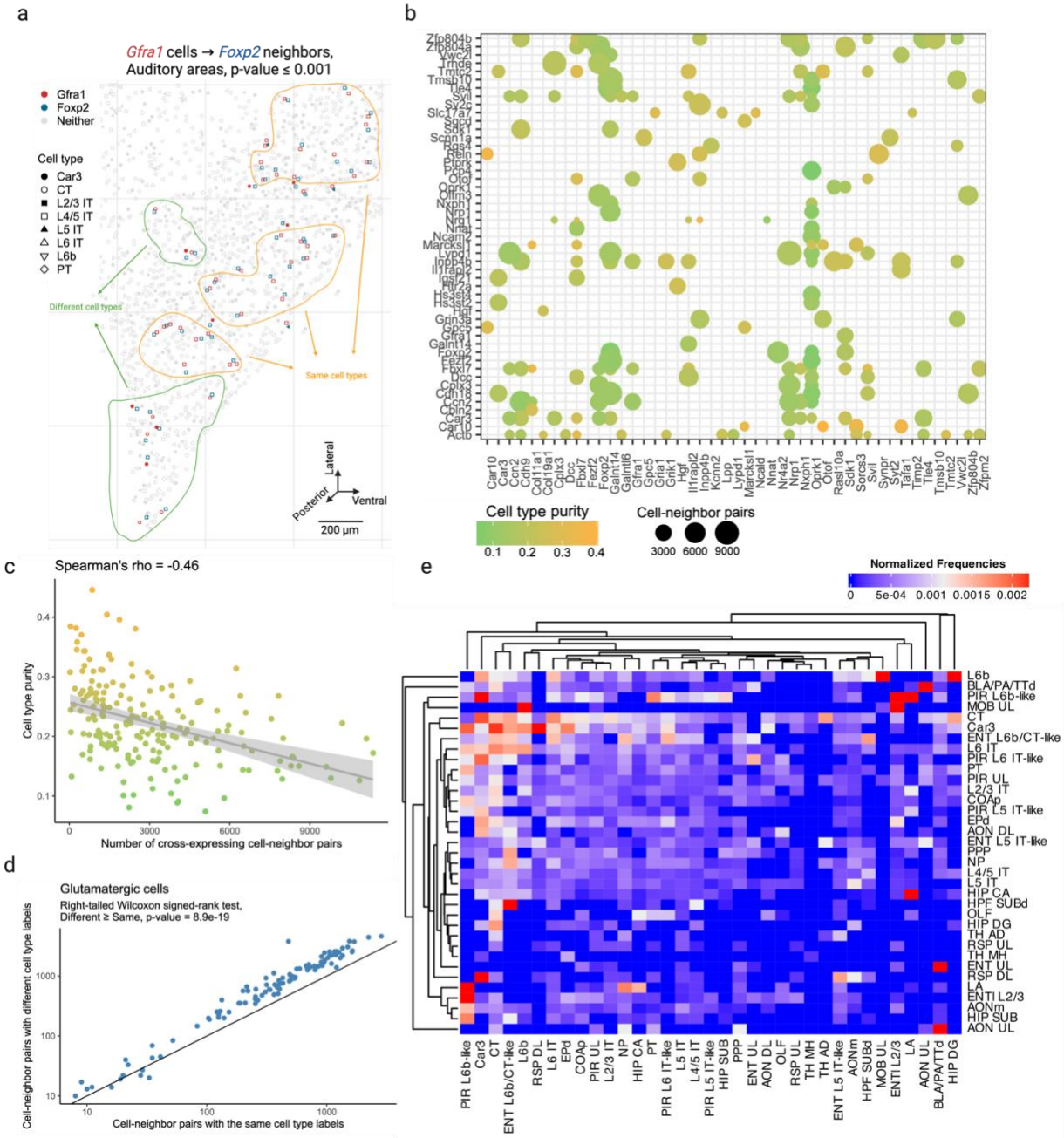


Fig. 3 | Cross-expression patterns are discovered independently of cell type labels but are driven by cell type heterogeneity. a, Cells of the same (yellow) and different (green) types cross-express genes *Gfra1* and *Foxp2* in the auditory cortex. Discovering cross-expression relations between this or any other gene pair does not require cell type labels. **b**, Genes are cross-expressed across numerous cells, 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 offers a common framework for analyzing multiple studies and the meta-analytic network reveals uniquely spatial biological processes

After analyzing ligand-receptor genes and cell type variability in terms of cross-expression, we asked whether its gene-centric approach can be used to simultaneously analyze multiple studies. To this end, we downloaded 13 adult mouse brain datasets (see methods), which were collected using 8 different platforms (4 commercial and 4 laboratory-based), spanning ~25 million cells (after dataset-specific quality control; median 724,790 cells) across 695 large samples/ slices (hemi-coronal, coronal, or sagittal) (Fig. 4a). Although the gene panels had varying sizes and compositions, they were sufficiently overlapping to allow integrative analyses (Fig. 4b; median pairwise overlap is 77 genes), reflecting that well-studied and informative genes are enriched across panels [50].

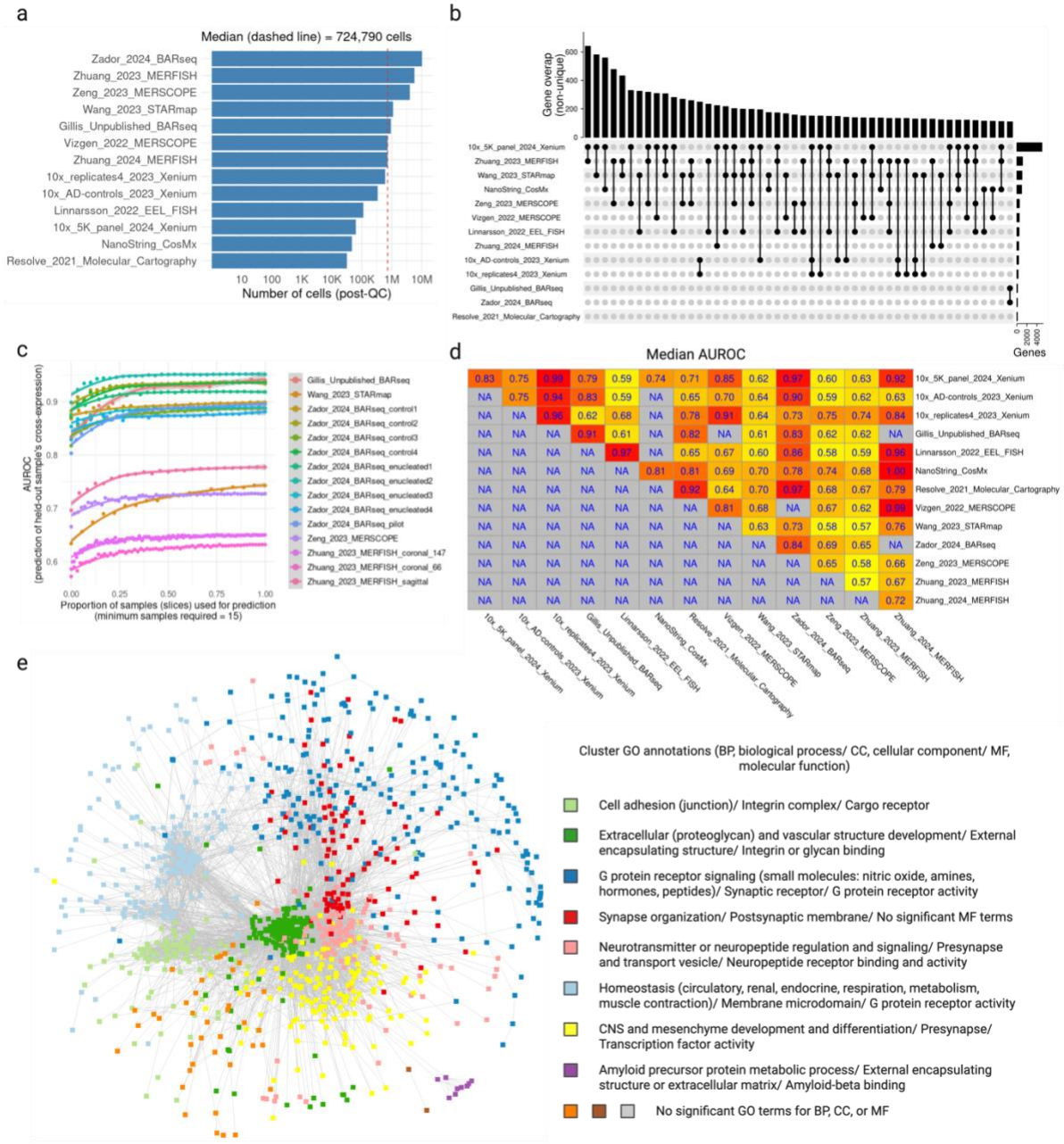
We first investigated cross-expression replicability within datasets. Here, we predicted the p-values of a held-out slice using the averaged correlations of the remaining slices. Our cross-validation-based classification revealed that the predictions (AUROCs) improved and then plateaued as more samples were included (Fig. 4c), suggesting that the sample ensemble progressively became more representative, e.g., due to the inclusion of anatomically adjacent slices. While the improvements were consistent across datasets, the AUROC scores varied between them, indicating the presence of potential batch effects, including differences in data quality between samples, enrichment of spatially variable or localized genes in the panels, and

329 large gaps between the slices, etc. In general, the cross-expression patterns within datasets are
330 broadly replicable, where including more samples improves performance, possibly due to better
331 representation of the underlying anatomical regions.

332 Next, we examined whether cross-expression correlations in one dataset can predict the
333 p-values in another dataset using the genes shared between their panels. We found high
334 replicability for some datasets and modest performance for others (Fig. 4d), indicating that
335 integrative meta-analysis is possible using at least a subset of these data. Although our results
336 here are reported as median AUROCs, the performance can suffer if (say) in one dataset most
337 slices are anterior while in the other most are posterior. Consequently, our subsequent analyses
338 focus on replicability of cross-expression between (anatomically homologous) slices across
339 datasets.

340 Finding that cross-expression is broadly replicable between datasets, we created a meta-
341 analytic network, where the nodes represent genes and the edges indicate cross-expression (Fig.
342 4e). Our network includes gene pairs only when they are cross-expressed in two or more datasets,
343 thus excluding those cross-expressed in just one dataset. The network shows high modularity,
344 with communities enriched for gene ontology (GO) terms representing spatially-mediated
345 biological processes. For example, the community 'synapse organization' (red) is connected with
346 the 'neurotransmitter regulation' community (pink), recapitulating ligand-receptor interactions
347 present in these data. Likewise, the 'cell adhesion' community (light green) is connected to the
348 'extracellular and vascular' community (dark green), implicating genes potentially involved in
349 structural and supportive processes. The 'G protein coupled receptor signaling' community (dark
350 blue) is diffusely connected to multiple other communities, reflecting these genes' diverse
351 physiological roles, such as in hormonal and metabolic processes. Similarly, the 'CNS
352 development' community shows distributed connectivity, likely revealing the cross-expression of
353 cell type markers with multiple other marker and non-marker genes, e.g., due to transcription
354 factors (indirectly) specifying cell type identity [51]. Lastly, the network contains a small 'amyloid

precursor protein' community (purple) composed of Alzheimer's disease (AD) genes cross-expressed with *ApoE*, which helps form amyloid plaques and neurofibrillary tangles [52]. Together, the cross-expression framework allows us to use the shared genes to perform integrative analyses across multiple studies, with our meta-analytic network highlighting genes involved in distinctively spatial biological processes.



362 **Fig. 4 | Cross-expression facilitates integrative meta-analysis of numerous samples across**
363 **13 studies. a**, Number of cells in each dataset after applying dataset-specific quality control
364 metrics. Red line is the median. **b**, Upset plot showing the overlaps between the gene panels of
365 different datasets. **c**, Predictions of held-out slices' p-values (significant or non-significant) using
366 the average correlation of the remaining slices. Performance (AUROC) is reported as the function
367 of the number of slices used to average the correlation (cross-validation). **d**, Median performance
368 (AUROC) when predicting the p-values of a slice in one dataset using the correlation of slice(s)
369 from another dataset using the genes shared between their panels. **e**, Cross-expression meta-
370 analytic network where the nodes are genes, with edges present if the genes are cross-expressed
371 in two or more datasets. The network is clustered into communities, with the colors and labels
372 indicating gene ontology (GO) annotations, where slash ('/') separates biological process (BP),
373 cellular components (CC), and molecular function (MF). Created with BioRender.com.
374

375 **Parkinson's disease (PD) genes *Drd1* and *Gpr6* show highly reproducible and localized**
376 **cross-expression patterns in the striatum**

377 Our meta-analytic network revealed that cross-expressing genes are involved in spatially-
378 mediated biological processes. Since genes were included in the network if they cross-expressed
379 in two or more datasets, we sought to identify an example gene pair that is both highly
380 reproducible and biologically informative. This process is inherently dependent on the
381 researchers' interests, so we focused on *Drd1* and *Gpr6*, which are implicated in Parkinson's
382 disease (PD) due to their hypo- and hyperactivity, respectively [53–60]. Our datasets are
383 registered to the Allen Institute's CCFv3 mouse brain region atlas [61], which allowed us to
384 arrange samples from multiple brains and different datasets in the same anatomical space.
385 Looking at coronal and sagittal sections, we found these genes' cross-expression highly
386 concentrated within a restricted range and effectively absent outside it, suggesting that it is
387 localized within specific anatomical regions (Fig. 5a). Next, we overlaid the Allen CCFv3 atlas on
388 the combined cross-expression signal, finding that it is located within the striatum (Fig. 5b,
389 protruded region), the central locus in PD pathophysiology [62]. Consistent with this observation,
390 we found that their cross-expression within slices is located in the striatal regions but not
391 elsewhere, with its patterns following gross neuroanatomy in the anterior-to-posterior and lateral-
392 to-medial directions (Fig. 5b, panels 1-9 and P-S). Accordingly, the genes *Drd1* and *Gpr6* show

393 coordinated expression between cells in the striatum, a property observed in multiple slices
394 across different datasets.

395 The cross-expression of these genes can help facilitate subsequent studies. Briefly, *Drd1*
396 is the dopamine receptor D1, whose stimulation by dopamine initiates movement, whereas *Gpr6*
397 is the G protein coupled receptor 6, whose constitutive activity prevents movement initiation. In
398 PD, the dopaminergic neurons in the substantia nigra insufficiently activate *Drd1*, and *Gpr6*
399 exhibits higher baseline activity, leading to severe difficulty in starting movement, the cardinal
400 symptom of PD [53–60]. While extant therapies increase dopamine (L-DOPA as Levodopa) to
401 stimulate *Drd1* [63,64], where the drug's effectiveness decreases over time and causes sides
402 effects, recent clinical trials have explored the inverse agonist CVN424 to inhibit *Gpr6* [55–60].
403 Although at present these approaches have not been pursued in tandem, their localized cross-
404 expression suggests that the drugs' staggered or co-delivery might offer complementary,
405 potentially amplified, effects. In general, once reproducible cross-expression signatures are
406 discovered, ideally across different samples and studies, one can further investigate gene pairs
407 of interest, making cross-expression a basic tool for biological research using spatial
408 transcriptomic data.

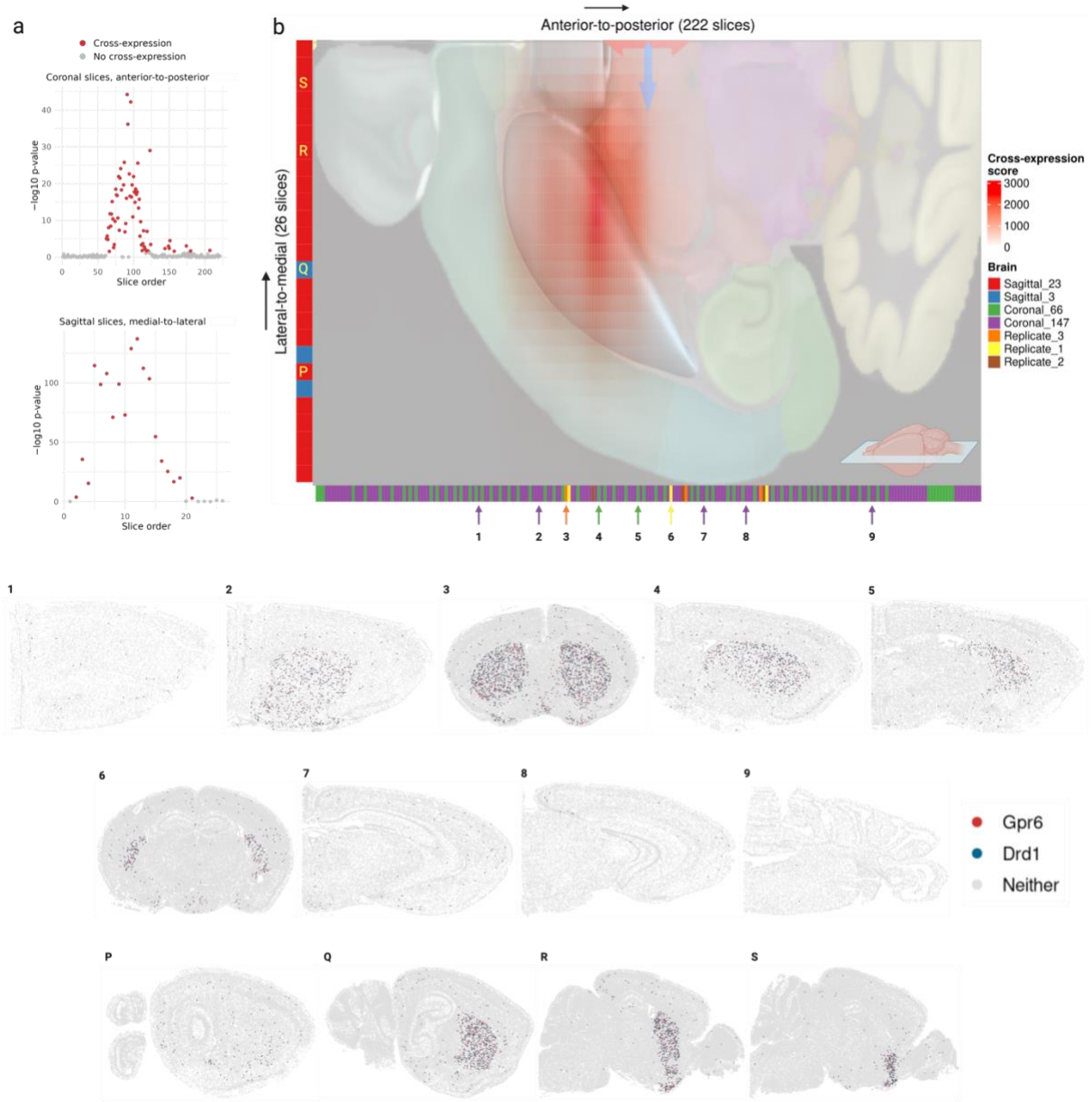


Fig. 5 | Cross-expression of genes *Drd1* and *Gpr6* in the striatum across multiple brains and different studies. a, Cross-expression signature (–log10 p-value) of *Drd1* and *Gpr6* genes in the coronal sections ordered in the anterior-to-posterior direction using the Allen Institute CCFv3 mouse brain atlas coordinates (top). Bottom, same as top but with sagittal slices ordered in the medial-to-lateral direction. **b**, Cross-expression score (outer product of coronal and sagittal slices' –log10 p-values) overlaid with the Allen Institute CCFv3 brain region annotations, with the striatum as the protruded (blue) area. Different colors represent the source brain of each slice. Numbers (1-9) and letters (P-S) indicate the position of example coronal and sagittal slices, respectively, within the overall gross neuroanatomy. Created with BioRender.com.

420 **Cross-expression network reveals *Gpr20* as a central gene and discovers possible**
421 **interaction partners between astrocytes and the brain microvasculature**
422 Having assessed multiple datasets, we now analyze a single study to show how cross-expression
423 patterns can be used alongside additional information, such as cell type labels. We first
424 supplemented our network formalism by including (second-order) edges between two genes if
425 they independently cross-expressed with a third same gene (Fig. 6a). Using the Vizgen MERFISH
426 data, we created a cross-expression network (Fig. 6b), which contains 200 genes with 382 first-
427 order, 217 second-order, and 107 dual-order edges. We observe that *Gpr20*, a G protein-coupled
428 receptor, is a central gene with a high node degree of 40 while the other genes form a median of
429 4.8 edges (Extended Data Fig. 4a). We performed gene ontology (GO) enrichment for genes
430 cross-expressed with *Gpr20*, finding functional groups like 'regulation of macromolecule
431 biosynthetic process', 'regulation of gene expression', and 'regulation of metabolic process'
432 (Extended Data Fig. 4b, all p-values ≤ 0.05). While some of these genes are co-expressed with
433 astrocytic and microglial cell type markers (Extended Data Fig. 4c), their global co-expression
434 with the endothelial marker is higher, where the co-expression profiles were computed using
435 neighbors cross-expressed with *Gpr20* rather than the entire dataset (Extended Data Fig. 4d,
436 Mann-Whitney U test, endothelial vs others, all p-values ≤ 0.01; remaining pairwise comparisons,
437 all p-values > 0.05).
438 Noting that the neighbors of *Gpr20*-positive cells are involved in the microvasculature, we
439 next viewed the spatial distribution of cells expressing *Gpr20*, finding that they form contiguous
440 linear streaks resembling blood vessels (Fig. 6c; anterior slice from mouse brain 1 shown). To test
441 this observation, we looked at whether the neighbors of *Gpr20*-positive cells also express this
442 gene and compared it to randomly selected cells, which constitute the expectation that *Gpr20* is
443 uniformly expressed across space. Consistent with the visualization, we find that cells with *Gpr20*
444 are surrounded by neighbors that also express this gene, a pattern that disappears for neighbor
445 order of 50 or more cells (Fig. 6d-e, area under curve (AUC), neighbors vs random cells, 0.69 vs

446 0.49; right-tailed Wilcoxon signed-rank test, neighbors vs random cells, p-value ≤ 0.0001). Having
447 seen that cells with *Gpr20* possibly reflect blood vessels, we asked whether these cells are
448 themselves vascular or whether they line the vasculature, especially since the cells that cross-
449 express with *Gpr20* are endothelial. We find *Gpr20* is poorly co-expressed with *Igfr1* (Pearson's
450 R = 0.0024), the vascular/endothelial marker [65–67] in our gene panel, suggesting that it lines
451 but does not mark the blood vessels. Moreover, it is lowly co-expressed with other cell type
452 markers (average Pearson's R, astrocytes = –0.0027, microglia = 0.0018, oligodendrocytes = –
453 0.022, neurons = –0.0025), eschewing cell type characterization. Taken together, *Gpr20*, a salient
454 topological feature of our cross-expression network, seems to be expressed in diverse cell types
455 that line the blood vessels, reflecting its possible role in the microvasculature.

456 Cross-expression driven by cell types might be particularly common when two genes
457 which cross-express with a third gene are co-expressed together, reflecting some common
458 transcriptional program jointly cross-expressing with neighboring cells. To investigate this, we
459 reduced co-expression further by specifying that cross-expressing genes must show lack of
460 significant co-expression, a procedure that yielded a subnetwork, which we further curated by
461 removing genes with fewer than two edges. Indeed, we find that two genes that independently
462 cross-express with another gene tend to be co-expressed (Fig. 5f, Extended Data Fig. 5a) and,
463 as expected, belong to the same cell types, as revealed by their co-expression with cell type
464 marker genes (Extended Data Fig. 5b). Confirming these results, the subnetwork genes are
465 enriched in GO groups like 'endothelial cell proliferation', 'positive regulation of vascular
466 endothelial growth factor production', and 'regulation of endothelial cell migration' (Fig. 6g, all p-
467 values ≤ 0.05). These results indicate that while cross-expressing genes are present in specific
468 cell types, the relations between them are functionally suggestive as opposed to simply reflecting
469 cell type compositional differences, especially since the cell type markers are not cross-
470 expressed. For example, the astrocytic *EGFR* (epidermal growth factor receptor) cross-expresses
471 with the vascular *Flt4*/*VEGFR-3* (FMS-like tyrosine kinase 4), *Tek*/*Tie2* (TEK tyrosine kinase/

472 angiopoietin-1), and *Tie1* (tyrosine kinase with immunoglobulin-like and EGF-like domain 1).
473 These three vascular receptors promote angiogenesis via the *VEGF* (vascular epidermal growth
474 factor) ligand [68,69], prevent endothelial cell apoptosis [70,71], and negatively regulate
475 angiogenesis [72], respectively, thus reflecting their potential role in the brain microvasculature in
476 coordination with the astrocytes, whose endfeet ensheathe the blood microvessels to constitute
477 the blood-brain barrier (BBB).

478 Within the same subnetwork, the astrocytic gene *Ppp1r3g* (protein phosphatase 1
479 regulatory subunit 3G), which helps convert glucose to glycogen [73], cross-expresses with *Epha2*
480 (ephrin type-A receptor 2), whose activity makes the BBB more permeable [74], likely enabling
481 glucose's transport and eventual conversion into glycogen, thereby making this cross-expression
482 relation relevant for energy metabolism. Indeed, this observation can be used to generate
483 hypotheses about the (directional) relationship between energy needs within a local
484 microenvironment and remodeling of the microvasculature, making cross-expression a powerful
485 approach with which to form testable hypotheses.

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

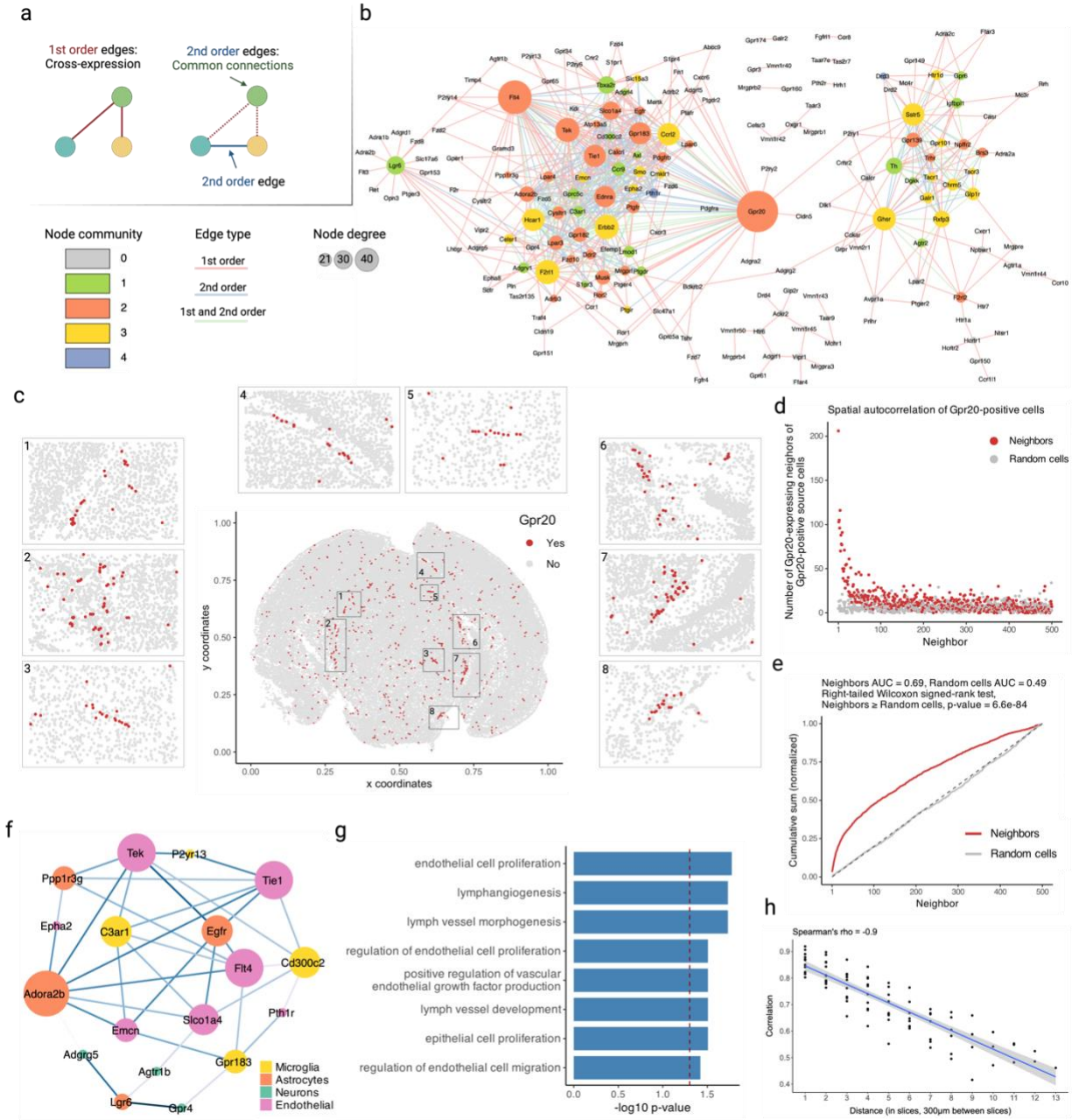


Fig. 6 | Networks of cross-expression. a, Cross-expression (edges) between genes (nodes) forms a network (left), where second-order edges (right) between genes share a node (first-order). **b**, Example cross-expression network, with first-order node degree represented by size and edge color showing 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 discovers anatomical marker gene combinations that delineate the thalamus and refine cortical layer VI boundaries

A key goal in biological research is finding marker genes, such as those that identify cell types, e.g., *Olig1* for oligodendrocytes [32], anatomical regions, e.g., *Rorb* for cortical layer IV [35], and functional modules, e.g., *Trpc4* for lateral septum in social behaviors [75]. In addition to using individual markers, one can use co-expression to discover marker gene combinations, an approach that uses single-cell or single-nucleus RNA-seq databases [76]. Although co-expression provides more combinations than individual markers, it relies on measurements from the same cells, thereby underutilizing the spatial relations between cells in spatial transcriptomics datasets.

Leveraging the spatial dimension to discover marker gene combinations, we asked whether cross-expressing genes can delineate anatomical regions, including putative functional modules. As an example, we found that cross-expression between *Lgr6* and *Adra2b* delineates the thalamus even though these genes are expressed throughout the brain (Fig. 7a). Specifically, while 48% of *Lgr6*- and 57% of *Adra2b*-expressing cells are thalamic, 91% of their cross-expressing cell-neighbor 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. 7b, 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 the rest of the brain (Fig. 7c), e.g., 89% of their co-expressing cells are in the thalamus, thus serving as robust combinatorial markers.

531 To evaluate whether the combinatorial marker-based approach is reliable, we asked
532 whether single gene markers, when assessed for cross-expression, rediscover the anatomical
533 locations. Using the BARseq cortical cell type atlas data [35], we assessed cross-expression
534 between cortical layer 6 marker *Foxp2* and ubiquitously expressed gene *Cdh13*. We discover that
535 cross-expression between these genes delineates layer 6 boundary (Fig. 7d), further supporting
536 the view that combinatorial anatomical markers can be discovered using cross-expression.
537 Indeed, the layer 6 boundary recovered by cross-expression captures additional L6 IT neurons
538 whereas *Foxp2*-based boundary overlooks these cells, indicating that combinatorial markers can
539 refine extant anatomical regions. More generally, this process leverages the spatial enrichment
540 of cross-expression, where we assess whether cross-expressing pairs are closer to other cross-
541 expressing pairs than to randomly selected cells. Our framework therefore discovers gene pairs
542 that annotate anatomical regions and refine extant boundaries, thus extending the single gene
543 and co-expression-based approaches.

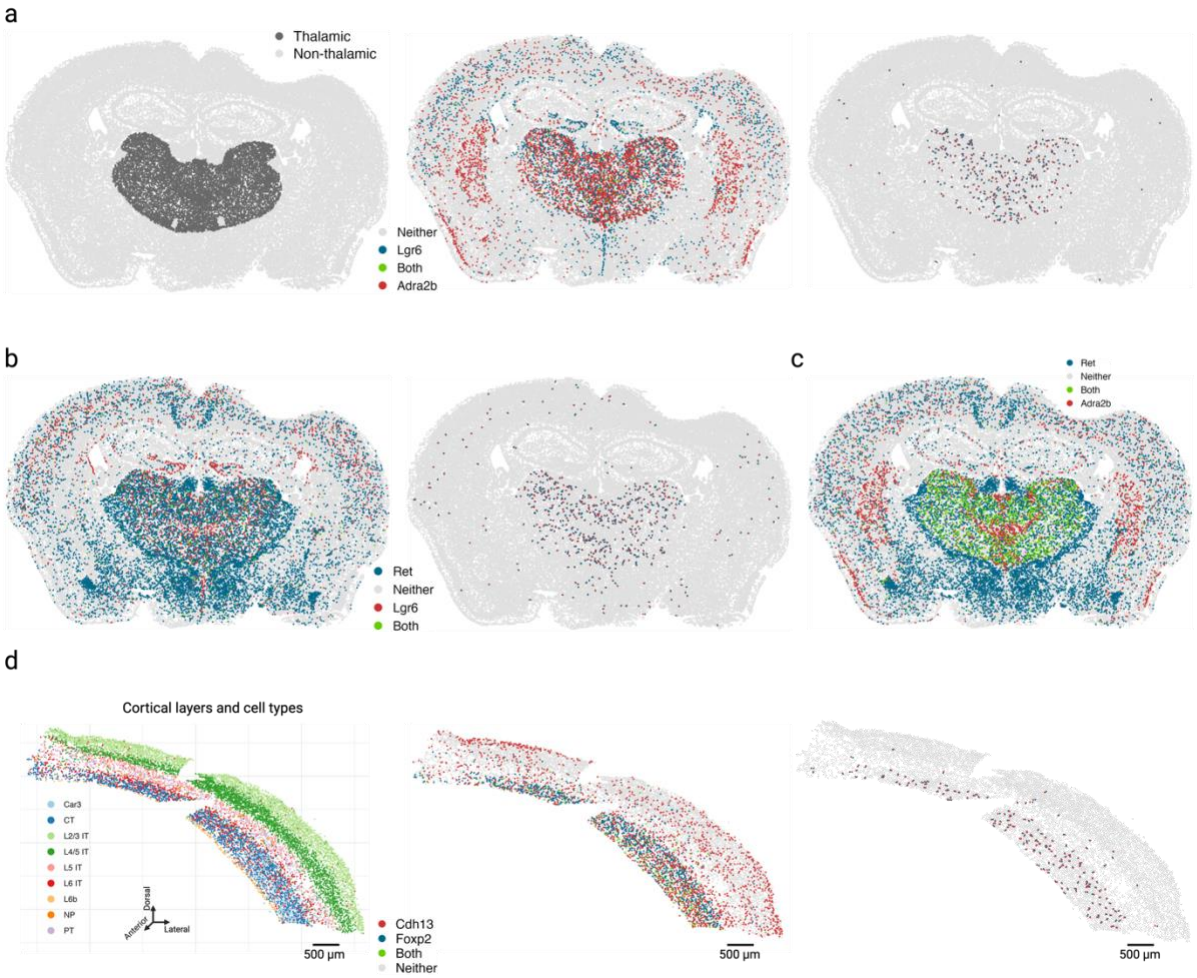


Fig. 7 | Cross-expression can discover combinatorial anatomical markers. a, Comparing the thalamus (left) to the rest of the brain, genes *Lgr6* and *Adra2b* are widely expressed across multiple brain regions (middle) but are preferentially cross-expressed in the thalamus (right). **b**, Same as in (a) 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 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 [2–8] 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.

558 We assess batch effects by comparing cross-expression between corresponding slices
559 across biological replicates. The MERFISH data contains three replicates with three slices each,
560 where the slices are sampled from roughly the same position. We find that the cross-expression
561 signature is highly similar across replicates. For example, the average correlation for the anterior
562 slices between the three replicates is 0.83 (Fig. 8a), with similar findings for the middle and
563 posterior slices (Extended Data Fig. 7a-b, Spearman's ρ = 0.81 and 0.8, respectively).

564 We next assessed the degree to which cross-expression within the BARseq sagittal or
565 coronal experiments [35] is similar to that between experiments. To this end, we compared cross-
566 expression patterns between brain slices. As expected, the cross-expression profiles are more
567 similar within brains than between brains (Fig. 8b, Mann-Whitney U tests, FDR-corrected, coronal
568 vs sagittal, p-value = 0.2, coronal vs mixed, p-value ≤ 0.001, and sagittal vs mixed, p-value ≤
569 0.001), suggesting that the sectioning procedure samples different brain regions and therefore
570 reveals distinct underlying gene expression profiles. Supporting this result, we find that the same
571 anatomical regions (per Allen CCFv3 brain atlas [61]) across brains have more similar cross-
572 expression profiles than do different regions within or between brains (Fig. 8c, Mann-Whitney U
573 test, different regions vs same regions, p-value ≤ 0.001). Noting that the sagittal and coronal
574 brains contain the same regions in the dorsal to ventral directions, we asked whether the cross-
575 expression is similar in this shared dimension. Here, we computed the density of cross-expressing
576 cells in the dorsal to ventral direction and compared these distributions across the two brains,
577 finding that 99% (without FDR correction) of the genes did not have significantly different density
578 profiles (Fig. 8d), suggesting that the cross-expression patterns are highly similar across batches
579 at the whole-brain level.

580 Having found that the cross-expression profiles are generally robust, we assessed cell
581 segmentation at a global level by comparing gene co-expression between the single cell RNA-
582 sequencing (scRNA-seq) [32] and spatial transcriptomic data. We reasoned that scRNA-seq does
583 not require segmentation and therefore captures genes co-expressed within the cell's boundaries

584 (Fig. 8e). Because cell segmentation alters transcript assignment, it could change co-expression
585 in spatial transcriptomic data. Reassuringly, we find a strong association between gene co-
586 expression in the scRNA-seq and spatial transcriptomic data (Fig. 8f, Pearson's R = 0.83). We
587 further examine whether this correlation is sufficiently strong by comparing co-expression
588 between scRNA-seq and single-nucleus RNA-sequencing (snRNA-seq) [77] (Fig. 8g, Pearson's
589 R = 0.86), finding agreement between the two comparisons (R = 0.83 vs. R = 0.86). These results
590 imply similar levels of technical variability between platforms while suggesting that gene co-
591 expression is congruent between scRNA-seq and spatial transcriptomic data.

592 The data in our work was processed using CellPose [78], a deep learning-based cell
593 segmentation algorithm. A recent benchmarking study showed that it outperforms other methods
594 on a variety of metrics [79]. In fact, it uses the nuclear stain DAPI as a cell landmark and forms
595 boundaries using cytoplasmic signal, such as the transcript distributions, making it the state-of-
596 the-art segmentation algorithm on a variety of assessments. Further, the cell segmentation
597 algorithms are continuously being improved [80], allowing users to re-segment and reanalyze
598 their data. Most importantly, the analysis conducted using the cross-expression framework may
599 suffer if segmentation is performed poorly, but the validity of the concept and the soundness of its
600 statistics do not rely on this potential artefact and, with rapid improvements in data quality, the
601 inferences drawn from it will become increasingly more reliable.

602 Moreover, we assessed the relationship between cross-expression and noise in gene
603 expression measurement. Since the algorithm requires binarizing the expression matrix, an
604 appropriate threshold needs to be applied prior to analysis. To count a gene as expressed in a
605 cell, we applied thresholds of 1 to 10 molecules, finding that the cross-expression patterns are
606 generally concordant across these noise levels (Extended Data Fig. 8a-b, median Pearson's R =
607 0.88). Importantly, our framework is agnostic to and compatible with multiple models of gene
608 expression noise [81], and once an appropriate threshold has been applied, the resultant
609 expression matrix can be used for cross-expression analysis.

610 Finally, we explored the patterns of cell-neighbor relations and found that over 60% of cells
611 are the nearest neighbors of exactly one cell but the remaining cells are the nearest neighbors of
612 two or more cells (Extended Data Fig. 9a). Patterns such as these may be biologically important
613 if the 'neighbor' cell plays a central role in the local microenvironment, so deviations from one-to-
614 one mappings should be captured by statistical analyses. To investigate that our results are
615 consistent across these patterns, we compared cross-expression in one-to-one against many-to-
616 one mappings and with the full dataset, finding an average Pearson's correlation of 0.96
617 (Extended Data Fig. 9b). Importantly, our procedure is consistent with the assumption of
618 independent sampling because while a cell may be the nearest neighbor of multiple cells, each
619 cell-neighbor pair is statistically independent.

620 We facilitate these analyses by providing a highly efficient R package. A laptop with 16 GB
621 RAM can test for cross-expression in large datasets containing hundreds of thousands of cells
622 and thousands of genes within minutes (Fig. 8h). At present, most (commercial) imaging-based
623 platforms cannot profile gene panels of this magnitude [2–8], though such capabilities are
624 anticipated. Our software's performance makes it well-suited for analyzing current and future
625 spatial transcriptomic datasets.

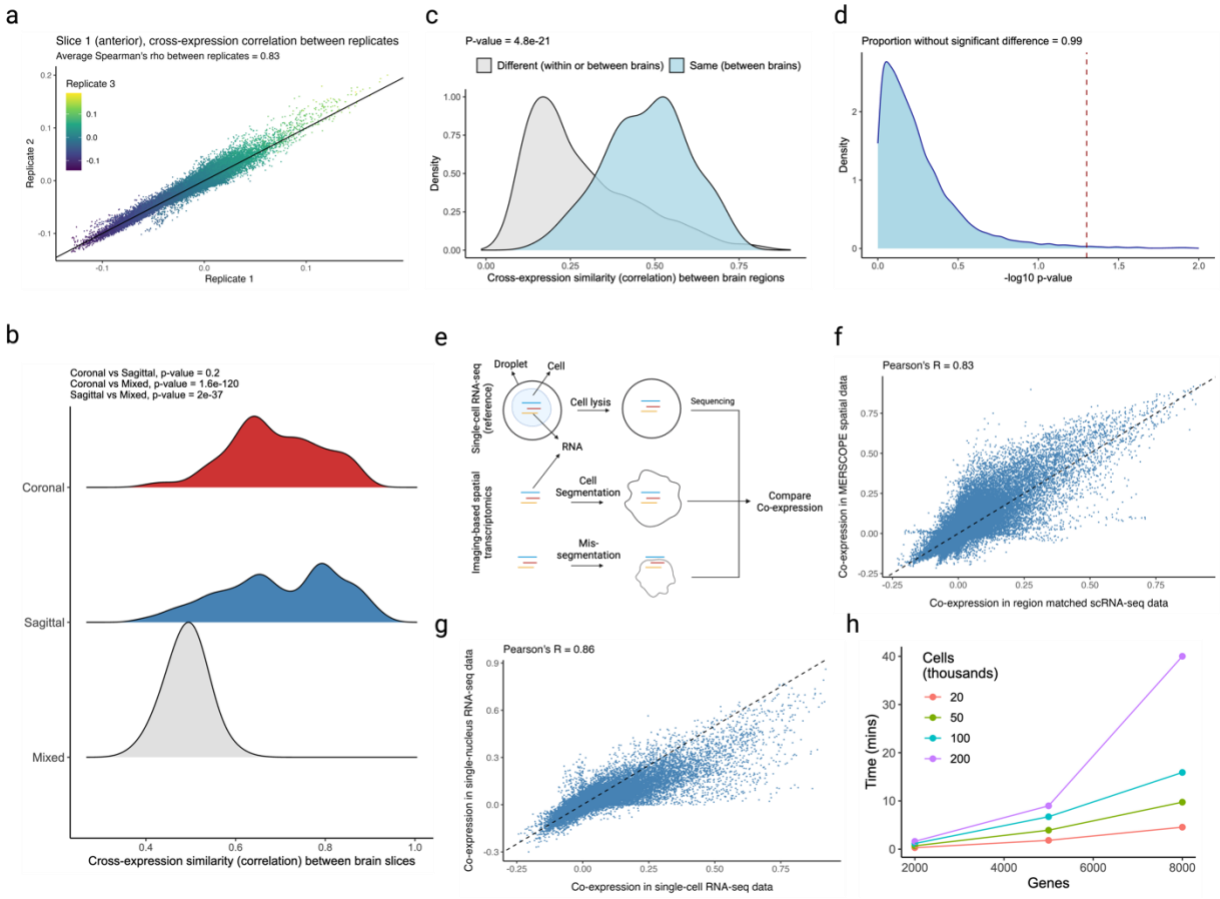


Fig. 8 | Assessing batch effects, cell segmentation, and software runtime. a, Correlation between cross-expression signatures across three biological replicates. **b**, Correlation between cross-expression signatures within (sagittal or coronal) and between (mixed) brains. Positive signal between brains 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.

645 **Discussion**

646 Cross-expression allows us to study gene-gene networks that reflect how nearby cells influence
647 each other by coordinating their gene expression. Using this framework, we recapitulated known
648 ligand-receptor interactions at the single cell level, revealing biologically meaningful tissue
649 phenotypes. We further showed that cross-expression is discovered without cell type labels but
650 often reflects cell subtype compositional differences in the form of marker gene cross-expression.
651 We also revealed that its gene-centric perspective enables integrative meta-analysis, where many
652 studies can be combined to find robust biological signals, such as the cross-expression of *Drd1*
653 and *Gpr6* in the striatum. Moreover, it helps us perform deeper analyses of individual studies,
654 revealing the relationships between astrocytes and the brain microvasculature, and discovers
655 paired markers for anatomical region annotation. Together, cross-expression is a powerful way of
656 analyzing spatial transcriptomic data and allows us to study gene-gene relations between
657 adjacent cells, thereby fully harnessing the high-throughput of these technologies.

658 The cross-expression framework complements current approaches analyzing spatial
659 transcriptomic data, such as those exploring niche-specific co-expression patterns [13–19].
660 Specifically, niche-specific cross-expression networks may be compared with co-expression
661 networks to examine if inter-cellular relations are associated with intra-cellular gene programs
662 and vice versa. This may be approached at different, potentially hierarchical spatial scales to
663 reveal spatial gene expression programs within the tissue. Moreover, the cross-expression
664 patterns can be quantified in multiple ways, such as using mutual information or graphlets,
665 allowing investigations into the best approaches that capture the signal of interest. For example,
666 just as co-expression relations can be measured using the Pearson's correlation coefficient,
667 cross-expression patterns may be investigated from numerous perspectives to discover the most
668 robust formalism. In this sense, the cross-expression framework introduced here is primarily a
669 way of conceptualizing gene-gene relations within spatial transcriptomics data, thereby serving
670 as a powerful framework for research in tissue biology. For instance, it can be used to study

671 cancer [82], where tumor progresses via signaling with the stromal tissue, as well as
672 neurodegenerative diseases like Alzheimer's [83] or senescence [84], where the progression of
673 pathology is spatially structured, making it a broadly useful approach for numerous problems.

674 Cross-expression is not restricted to imaging-based spatial transcriptomics. Instead, it can
675 be applied to any biological assay that provides cell-by-features and cell-by-coordinate matrices.
676 For example, it can be extended to spatial proteomics [85], with potential to discover ligand-
677 receptor interactions. Likewise, it may be applied to spatial translatomics [86] to focus on
678 translating mRNAs that are more likely to form functional proteins, making conclusions about cell-
679 cell relations more robust. In fact, with the increasing resolution of spatially barcoded RNA capture
680 based methods [87,88], the framework may be extended transcriptome-wide to understand
681 relations between spots at near single-cell resolution.

682 A key challenge in imaging-based spatial transcriptomics [2–8], including the datasets
683 used in this work, is the size and constitution of the gene panels, which sets an upper limit on
684 biological discovery. Although our framework will become more powerful as the quality of spatial
685 transcriptomic data, especially the gene panel, increases, care must be taken to not interpret the
686 results in mechanistic terms. Instead, the coordinated gene expression between neighboring cells
687 should be viewed as a target for further investigation. In this sense, the cross-expression
688 framework substantially narrows the space of gene-gene relations by identifying pairs that are
689 potentially biologically meaningful, making the problem experimentally tractable. Overall, cross-
690 expression offers a unique and powerful perspective on using spatial transcriptomic data for
691 driving biological discovery.

692

693 **Conclusions**

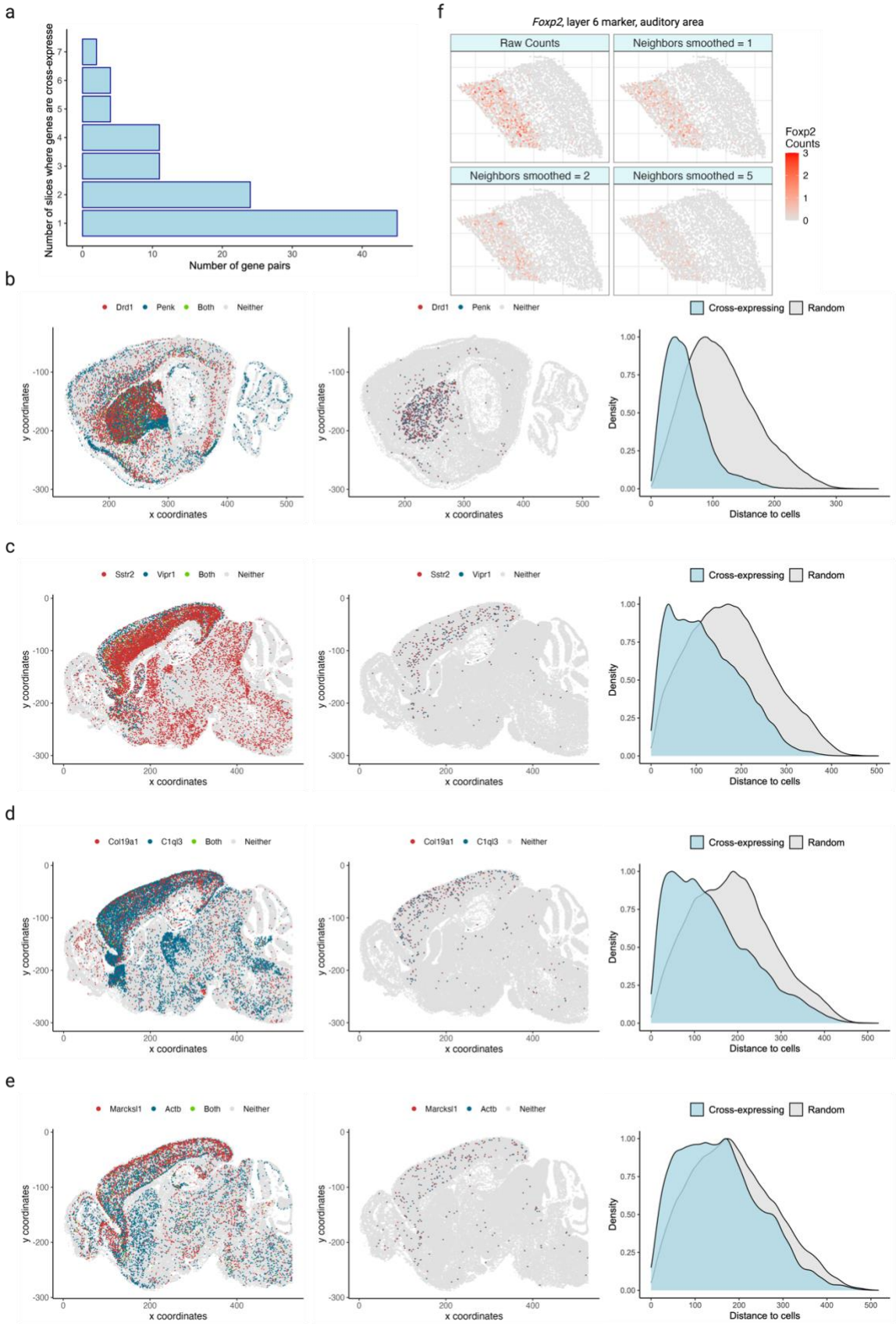
694 Cross-expression is a useful conceptual and analytical framework, which compares all genes and
695 identifies pairs that coordinate their expression between neighboring cells. The accompanying R
696 software efficiently facilitates these and other analyses as well as provides new set of

visualizations to deeply explore spatial transcriptomics data by leveraging coordinated gene expression at the single-cell resolution.

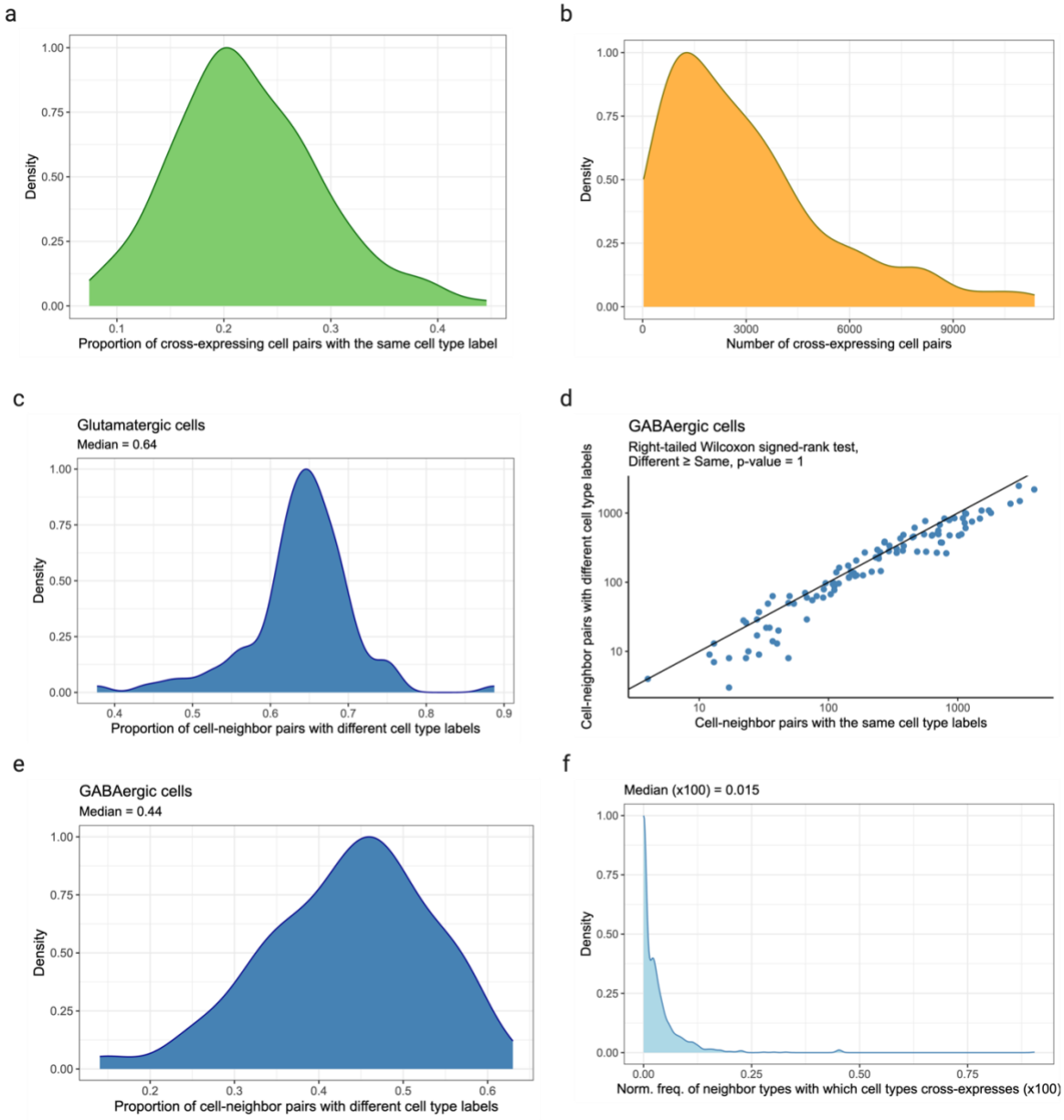
The package is available at https://github.com/gillislab/CrossExpression.

Extended Data Figures

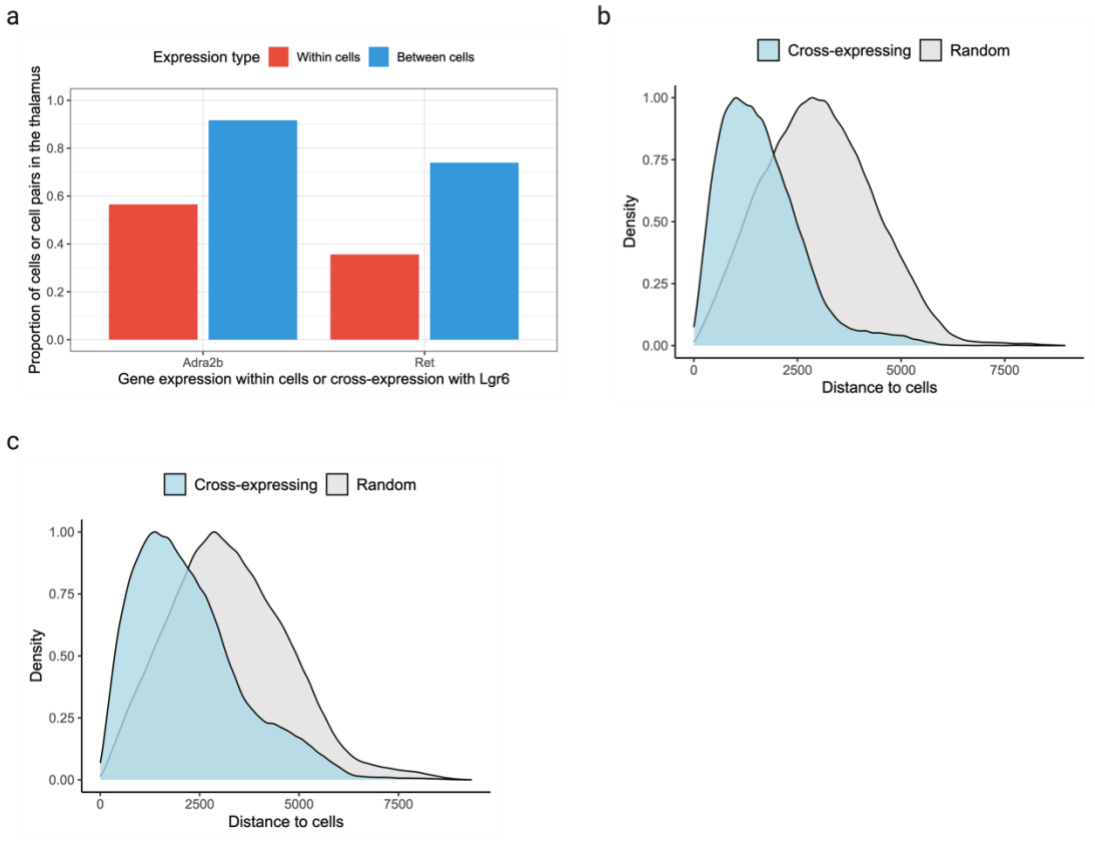
Extended Data Fig. 1 | Cross-expression across tissue slices and regions for ligand-receptor and non-signaling genes. a, Distribution of the number of gene pairs cross-expressed in different slices. Dataset has 16 slices sampled sagittally from the left hemisphere of a mouse brain. **b-e**, Cells are colored by gene expression (left) and cross-expressing cells are highlighted (center). Right, distances between cross-expressing cells are compared with those between cross-expressing and randomly selected cells. 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). Genes include ligands and receptors (b, c) and non-ligands and non-receptors (d, e). **f**, Smoothed gene expression for different numbers of neighbors for the auditory 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

712

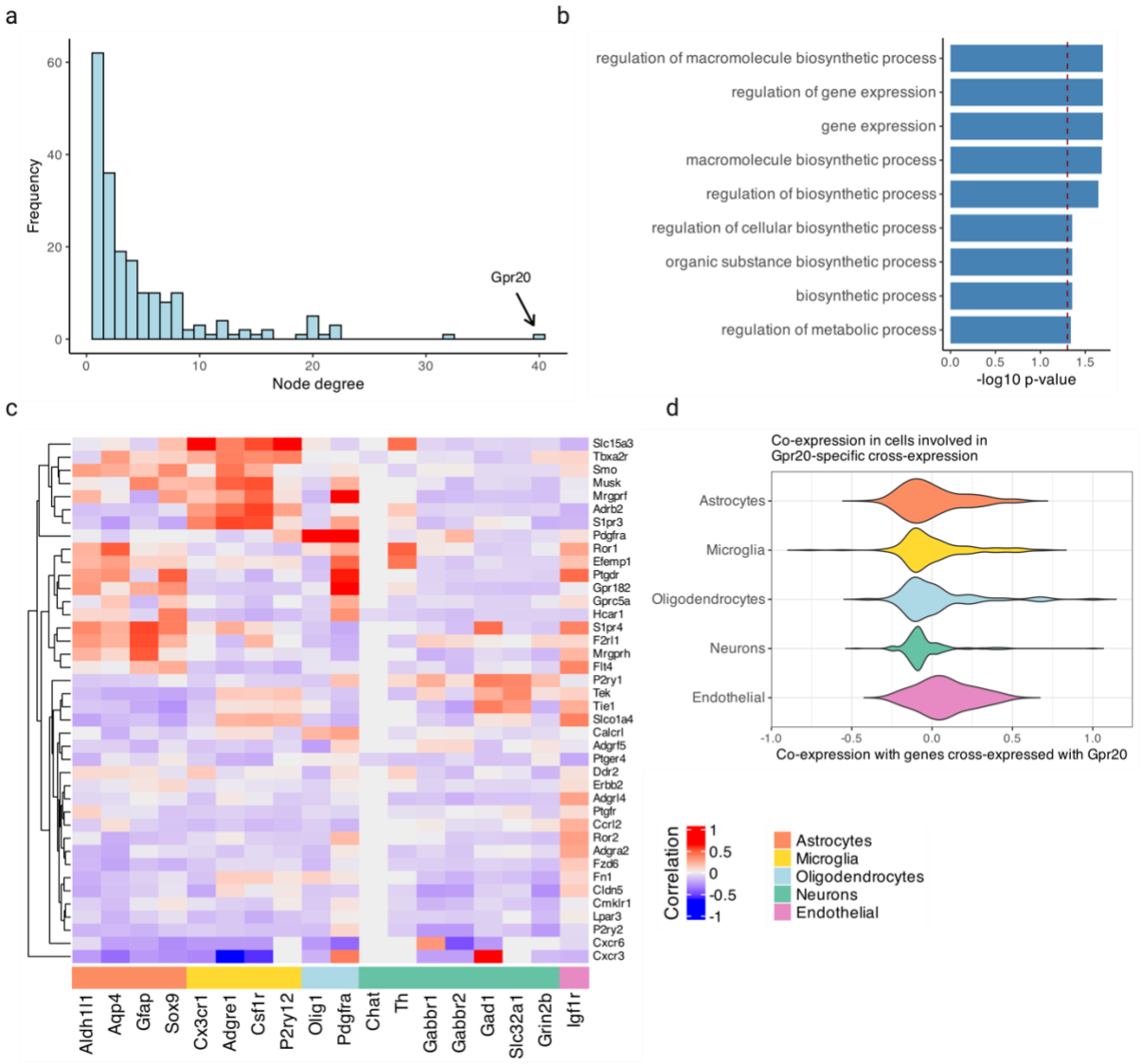


Extended Data Fig. 2 | Relationship between cross-expression and cell type heterogeneity. **a**, Proportion of cross-expressing cell pairs belonging to the same cell type label. **b**, Number of cell-neighbor pairs involved in cross-expression. **c**, Proportion of cell-neighbor pairs with different 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.

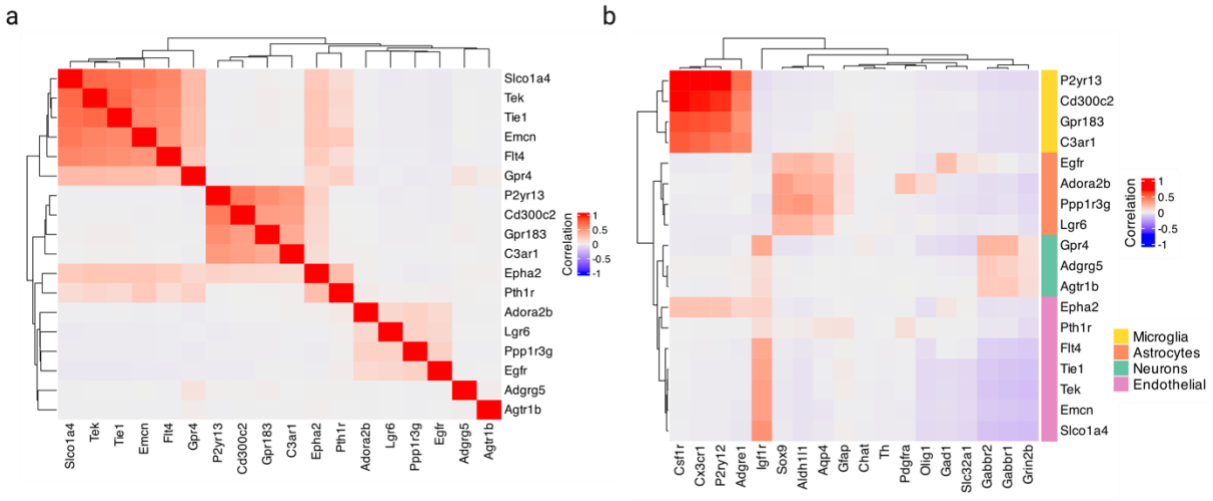


722

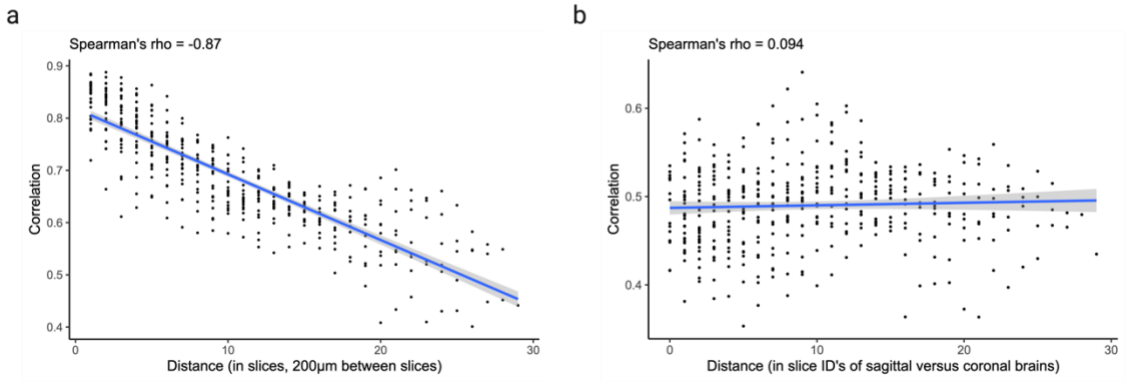
723 **Extended Data Fig. 3 | Combinatorial anatomical markers discovered using spatially**
724 **enriched cross-expression. a**, Proportion of *Adra2b*- and *Ret*-expressing cells in the thalamus
725 (red) and the proportion of cell pairs in the thalamus (blue) when cross-expressing with *Lgr6*. **b-**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



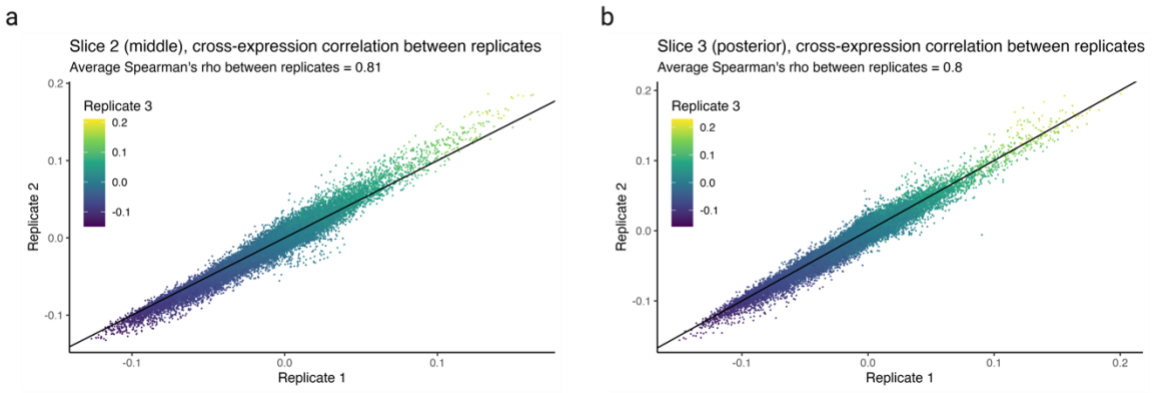
Extended Data Fig. 4 | Exploration of *Gpr20* and its cross-expressing genes. a, Distribution of node degree, with *Gpr20* highlighted. **b**, Gene ontology (GO) functional groups for genes cross-expressed with *Gpr20*. **c**, Co-expression of genes cross-expressed with *Gpr20* (right) against cell type 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.



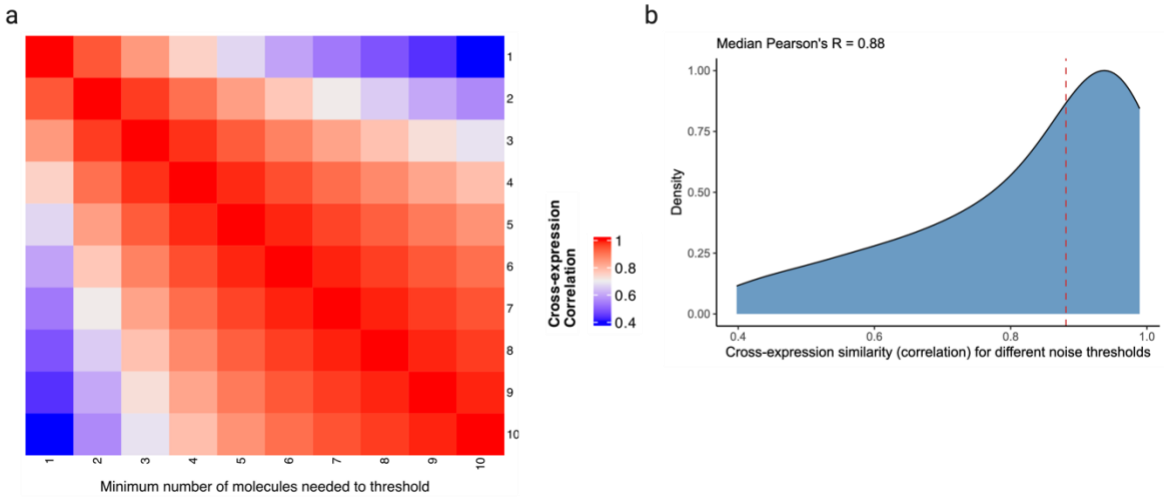
Extended Data Fig. 5 | Exploration of the MERFISH cross-expression (sub)network. a, Co-expression of genes in the subnetwork. **b**, Co-expression between genes in the subnetwork (right) and cell type marker genes (bottom). Created with BioRender.com.



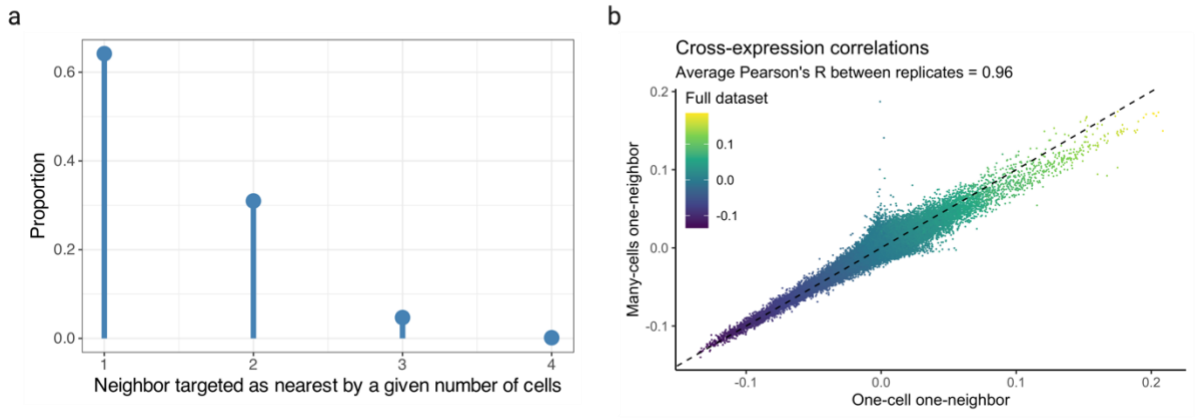
Extended Data Fig. 6 | Cross-expression network similarity between slices. a, Slice-specific cross-expression networks compared and shown as a function of distance between slices. **b**, Same as in (a) but slice-specific networks compared between sagittal and coronal datasets, where the “distance” is 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 calculate the probability using the hypergeometric approach

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

where N is the population size, K is the number of possible successes, n is the number of samples or draws, k is the number of observed successes, and the form $\binom{a}{b}$ is the binomial coefficient.

774 Equation 1 outlines all the ways in which success can be observed—$\binom{K}{k}$—*and* (product
775 rule) all the ways in which failure can be obtained—$\binom{N-K}{n-k}$—normalized by all possible ways of
776 generating our sample $\binom{N}{n}$, making the outcome probabilistic by bounding it between [0,1].
777 Traditionally, the *n* samples are assessed for the presence of some property *k*. Here, we sample
778 cell-neighbor *pairs* and ask whether the cell expresses gene A while the neighbor expresses gene
779 B. Thus, the sample size *n* is the number of pairs where the cells express gene A, the number of
780 observed successes *k* is the number of pairs where the cells express gene A and the neighbors
781 express gene B, and the number of success states *K* is the number of pairs where the neighbors
782 express gene B. The population size *N* is the total number of cell-neighbor pairs, including those
783 that co-express A and B and those that express neither gene. To calculate the probability of *k* or
784 more successes, we use the hypergeometric cumulative distribution function (CDF)

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

785 A value lower than alpha α indicates an unusually large number of pairs where the neighbors
786 express gene B and the cells express gene A, making their cross-expression significant.
787

788 **Statistics of cross-expression between all gene pairs**

789 We need to assess cross-expression across all gene pairs, which rise quadratically by $\binom{N}{2}$ or
790 $\frac{N(N-1)}{2}$ for *N* genes. To efficiently explore this space, we implement the procedure above using
791 matrix operations and specialized packages in R.

792 We begin with a cells-by-genes expression matrix **E** and a cells-by-coordinates location
793 matrix **L**, where the coordinates in our data are cell centroids on two-dimensional slices. We input
794 **L** into `RANN` package's function `nn2` with search type as `standard`, which implements a kd-tree
795 algorithm to explore data subspaces and efficiently finds the *n*-th neighbors [89,90]. Using the
796 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. (To use a distance-based approach, one can use the *n*-th neighbor insofar as the average distance to this neighbor is smaller than the threshold.)

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 to calculate probabilities. 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 if the genes are expressed, 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 $\mathbf{C}_{ij}$ 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 false discovery rate (FDR) correction on the entire p-value distribution [91]. For each gene pair, we then use the lower FDR-corrected p-value as the final output **P** to indicate whether or not these genes cross-express.

Cross-expression networks

We can threshold and binarize **P** at a pre-selected alpha α to form an adjacency matrix **N**, where '1' indicates cross-expression (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 correlations

To compute correlations between genes, we first use the "masks" $\mathbf{X}$ and $\mathbf{Y}$ (eq. 4.1 and 4.2) and perform elementwise multiplication with non-binarized $\mathbf{E}$ and $\mathbf{E}'$, respectively. This makes the gene counts "0" if cell-neighbor pairs co-express the gene pair of interest. We then compute the correlation between all pairs of columns (genes) of $\mathbf{E}$ and $\mathbf{E}'$, yielding a gene-by-gene correlation matrix. We report the results as the average of this matrix with its transpose, giving a symmetric gene-gene correlation matrix as the output.

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 [89,90], 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 tissue sections 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*-degrees

$$\theta = n_{degrees} \times \frac{\pi}{180} \tag{10.1}$$

$$x' = \cos(\theta) \cdot x - \sin(\theta) \cdot y \tag{10.2}$$

$$y' = \sin(\theta) \cdot x + \cos(\theta) \cdot y \tag{10.3}$$

880 where x' and y' are the cell coordinates after counterclockwise rotation. Rotation does not change
881 the distances between cells, so x' and y' can be used for downstream analysis.
882
883 **Spatial enrichment of cross-expression**
884 Cross-expressing cells may be distributed across the tissue or show spatial localization. To
885 quantify their enrichment, we first average the distance between cell-neighbor pairs. We next
886 compare the distances between all cross-expressing cells to the distances between cross-
887 expressing and randomly selected cells. If the former distance is significantly smaller than the
888 latter distance, then cross-expression is spatially enriched.
889
890 **Data acquisition and preprocessing**
891 *BARseq datasets*
892 We collected a sagittal mouse brain data (P56 male) from the left hemisphere (20μm thick
893 sections, 300μm distance between slices) and measured 133 genes, including ligand-receptor
894 pairs (neuropeptides, neuropeptide receptors, monoamine receptors such as cholinergic,
895 adrenergic, serotonergic, and dopaminergic). This dataset includes 16 slices, which were assayed
896 as previously described [35]. We removed cells expressing fewer than 5 genes or with less than
897 20 counts.
898 We downloaded the BARseq coronal mouse brain dataset [35], which contains 9 brains,
899 with 4 control and binocular enucleation pairs as well as a pilot brain. These data contain 289
900 coronal slices and 109 genes. They were quality controlled just like the sagittal dataset.
901
902 *10x Genomics Xenium datasets*
903 We obtained three datasets from the Xenium platform https://www.10xgenomics.com/datasets.
904 The first dataset contains 1 coronal slice with 5006 genes obtained from

905 https://www.10xgenomics.com/datasets/xenium-prime-fresh-frozen-mouse-brain. We removed
906 cells with fewer than 50 counts.
907 The second dataset contains 3 coronal age-matched AD-control FFPE brain pairs, where
908 the ages are 2.5 months, 5.7 months, and 13.4 (control) versus 17.9 months (AD). This data
909 includes 346 genes and was obtained from https://www.10xgenomics.com/datasets/xenium-in-
910 situ-analysis-of-alzheimers-disease-mouse-model-brain-coronal-sections-from-one-hemisphere-
911 over-a-time-course-1-standard. We removed cells with fewer than 50 counts.
912 The third dataset includes 4 coronal replicates with 248 genes. Three of these were
913 obtained from https://www.10xgenomics.com/datasets/fresh-frozen-mouse-brain-replicates-1-
914 standard and the remaining from https://www.10xgenomics.com/datasets/fresh-frozen-mouse-
915 brain-for-xenium-explorer-demo-1-standard. We removed cells with fewer than 30 counts.
916
917 *EEL FISH dataset*
918 We obtained a sagittal brain section data assaying 440 genes [92]. We removed cells with fewer
919 than 15 counts.
920
921 *NanoString CosMx dataset*
922 We obtained a coronal brain section data from https://nanostring.com/products/cosmx-spatial-
923 molecular-imager/ffpe-dataset/cosmx-smi-mouse-brain-ffpe-dataset/. This dataset assays 950
924 genes. We removed cells with fewer than 100 counts.
925
926 *Resolve Biosciences Molecular Cartography dataset*
927 We downloaded one hemi-coronal slice data from https://resolvebiosciences.com/open-
928 dataset/?dataset=mouse-brain-2021, which assayed 99 genes. We removed cells with fewer than
929 30 counts.
930

931 *Vizgen MERFISH brain receptor map datasets*

932 We obtained Vizgen MERSCOPE's mouse brain receptor map from
933 https://info.vizgen.com/mouse-brain-data. This data contains three coronal slices from three
934 replicates, with the middle slice covering the center of the brain. It assays 483 genes. We filtered
935 cells with fewer than 50 counts and those lacking brain region annotations (see below). We
936 registered the slices to the Allen CCFv3 (Common Coordinate Framework version 3) brain region
937 atlas [61]. To facilitate this, we annotated the cells using Seurat [93]. Here, we created a Seurat
938 object and used `SCTransform` with the `clip.range` between -10 and 10. We then ran Principal
939 Component Analysis (PCA), setting the number of components to 30 and specifying the features
940 as genes. Next, we used `FindNeighbors` and `FindClusters` with the resolution set to 0.3.
941 The clusters are cell type labels, which help us identify brain structures during registration. For
942 registration, we used QuickNii (v3 2017) [94] to linearly align the slice to the Allen CCFv3 atlas
943 using discernible regions like the hippocampus and the ventricles as anchors. We then used
944 VisuAlign (v0.9) [94] to non-linearly transform the slice to improve alignment with the atlas. This
945 procedure assigns a brain region annotation to every cell.

946

947 *STARmap dataset*

948 The STARmap dataset contains 20 coronal, sagittal, or spinal sections, where 1022 genes were
949 assayed [34]. We removed cells with fewer than 30 counts.

950

951 *Allen Institute MERSCOPE atlas dataset*

952 We obtained MERSCOPE dataset from the Allen Institute brain atlas, which contains 500 genes
953 measured across 59 coronal slices [32]. We removed cells with fewer than 15 unique genes or
954 lower than 40 counts.

955

956 *MERFISH brain atlas datasets*

957 The first dataset assays 1122 genes across 239 coronal and sagittal slices [33]. We removed cells
958 in the top and bottom 1% by total counts.
959 The second dataset assays 374 genes across age-matched (4 weeks, 24 weeks, 90
960 weeks) controls and LPS-infected coronal brains, with 49 total slices [95]. (LPS is
961 lipopolysaccharide bacterial infection.) We removed cells expressing fewer than 5 genes or with
962 less than 20 counts.
963
964 *Single-cell RNA-seq (scRNA-seq) and single-nucleus RNA-seq (snRNA-seq) datasets*
965 We obtained the scRNA-seq [32] and snRNA-seq [77] data from the Brain Initiative Cell Census
966 Network (BICAN) cell type atlases. These data were collected from dissected tissue regions,
967 giving us the cells' coarse anatomical origin. We removed cells with a doublet score of 30 or above
968 and randomly selected 10,000 cells from each region for subsequent analysis.
969
970 **Ligand-receptor cross-expression**
971 We aimed to find cross-expression between known ligand-receptor pairs. In our sagittal data, we
972 selected two slices and within each slice we chose a cortical region. These choices were made
973 randomly. In practice, we chose the visceral area (VISC) in slice 3 and the somatosensory nose
974 region (SSp-n) in slice 5. Next, we selected the well-known neuropeptide somatostatin *Sst* and
975 its cognate receptor *Sstr2* as the candidate pair. Finding their cross-expression significant, we
976 show their expression on tissue and highlight cross-expressing cells. We also compute their
977 bullseye scores and report them as a ratio of cross- to co-expression across 10 neighbors.
978
979 **Cross-expression and cell type heterogeneity**
980 We explore the relationship between cross-expression and cell type heterogeneity using the
981 BARseq coronal data [35]. First, we use *Gfra1* and *Foxp2* to highlight cross-expressing cells and
982 map different cell types to distinct shapes. Second, we count the number of cross-expressing cell-

983 neighbor pairs for numerous genes. Since each cell has a cell type label, we compute cell type
984 purity as the proportion of pairs with the same label. Third, we use the cell type hierarchy to assess
985 if cell-neighbor pairs with the same H1 label have the same H3 label. We first find cross-
986 expressing gene pairs using the entire dataset. Next, using cell pairs with the 'glutamatergic' H1
987 label, we compute the number of pairs with the same or different H3 labels. We perform a similar
988 analysis using cells labelled as 'GABAergic' at the H1 level. Finally, we compute the frequencies
989 with which cell type label combinations are associated between neighboring cells and normalize
990 this by the expected frequencies of those cell type pairs in the population.
991
992 **Replicability of cross-expression**
993 We evaluate the replicability of our data using the hypergeometric and correlation-based
994 approaches, which provide independent ways of quantifying cross-expression. We predict the
995 binary p-values (significant or not, alpha = 0.05) using correlations and report the performance as
996 AUROC. For performance within datasets, we predict the labels of a held-out slice using the
997 average correlation of the remaining slices using a leave-one-out cross-validation procedure, with
998 the results reported as AUROC against the number of slices/ samples used. For performance
999 between datasets, we predict the labels of slices in one study using correlations of slices from
1000 another study and report the performance as median AUROC between study pairs. This process
1001 requires using gene pairs shared between studies' panels.
1002
1003 **Creation and annotation of meta-analytic cross-expression network**
1004 Our cross-expression meta-analytic network included connections (edges) between genes
1005 (nodes) if they were cross-expressed in two or more studies. We used the `igraph` package [96]
1006 to create a network and Leiden clustered it [97], with "modularity" as the objective function ran for
1007 100 iterations. Since Leiden clustering is somewhat stochastic, we performed this process 100
1008 times, giving us 100 distinct partitions/ community assignments per gene. To find consensus

1009 communities, we calculated the adjusted Rand index (ARI) [98], which tells us the degree to which
1010 two clustering results agree with each other, between all pairs of partitions. Next, we found the
1011 average ARI for each partition and choose the one with the highest ARI, which gives us the
1012 'consensus' communities. Importantly, we observed that multiple partitions had the same highest
1013 average ARI, and upon further inspection, we discovered that these partitions were identical,
1014 giving credence to our final community assignments.

1015 After assigning genes to communities, we performed gene ontology (GO) [99,100]
1016 enrichment per community using the union of the studies' gene panels as the background. We
1017 merged significant GO terms together using Revigo [101], which gave us the biological processes
1018 (BP), cellular components (CC), and molecular functions (MF) of each community. We performed
1019 minor manual curation of these terms to generate our final GO annotations. The meta-analytic
1020 network was visualized using Cytoscape (v3.10.1) [102,103].

1021

1022 **Networks of cross-expression when studying *Gpr20***

1023 Using the MERFISH data, we computed cross-expression p-values between all genes and
1024 binarized the matrix at $\alpha \leq 0.05$ to create an adjacency matrix. We calculate the node degree as
1025 the number of edges formed by each gene and create a network with second-order edges (shared
1026 connections) as outlined in equation 6.2. We set the threshold for second-order edges to 4,
1027 meaning that two genes are connected if they share at least 4 first-order edges, ensuring that the
1028 higher-order network is robust. Next, we use the `igraph` package [96] to perform Louvain
1029 clustering [104] (with default parameters) on the second-order network and thus assign genes to
1030 communities.

1031 We visualize the network using Cytoscape (v3.10.1) [102,103], mapping node size to
1032 degree, color to node community, and edge color to edge type (first-order, second-order, or both).
1033 We use the "organic" layout and apply "remove overlaps" from the `yFiles` app [103] and tweak

1034 the network to further reduce overlaps. Finally, we use the `Legend Creator` app [103] to render
1035 a legend with node degree size and community assignment.

1036 Because our network revealed *Gpr20* as topologically salient, we performed gene
1037 ontology (GO) [99,100] enrichment analysis on genes that cross-expressed with it ('test set').
1038 Here, we used the entire gene panel (except *Gpr20*) as the background set and used the
1039 hypergeometric test to determine if it significantly overlapped with the test set, giving us p-values
1040 for each GO functional group. We report FDR-corrected p-values. Additionally, for each gene
1041 cross-expressed with *Gpr20*, we used the cells involved in cross-expression, rather than the entire
1042 dataset, to compute co-expression with cell type marker genes and compared these global
1043 profiles between marker types.

1044 Since the cells expressing *Gpr20* visually showed spatial autocorrelation, we assessed
1045 their neighbors as well as randomly chosen cells for the expression of *Gpr20*. We L1-normalized
1046 the counts for both groups, rendering them into probability distributions, and computed cumulative
1047 sums. To calculate the area under curve (AUC), we scaled the neighbor order between 0 and 1
1048 and used the `trapz` function from R's `pracma` package to calculate the AUC [105].

1049 Within the main network, we introduce a further constraint that cross-expressing genes
1050 must lack significant co-expression. We curate the subnetwork by removing genes with node
1051 degree of 1 and assign cell type labels based on genes' co-expression with marker genes. Like
1052 before, we perform GO enrichment using the subnetwork genes as the test set and the gene
1053 panel as the background set, and report FDR-corrected p-values.

1054 To assess whether cross-expression networks are more similar between adjacent slices
1055 than between distant slices, we compute slice-specific cross-expression networks and calculate
1056 Spearman's correlation between these networks. The correlations are plotted against distances
1057 between slices, where the "distance" is the difference in the slice order. As a control, we compute
1058 the Spearman's correlations between slice-specific networks obtained from different brains and
1059 plot this against the "distance" between the slice ID's.

1060 **Discovering combinatorial anatomical marker genes**

1061 We observed that cross-expression discovers anatomical marker genes that delineate the
1062 thalamus. To quantitatively assess this, we made a mask by combining the following regions:
1063 anterior group of the dorsal thalamus (ATN), intralaminar nuclei of the dorsal thalamus (ILM),
1064 lateral group of the dorsal thalamus (LAT), medial group of the dorsal thalamus (MED), midline
1065 group of the dorsal thalamus (MTN), ventral group of the dorsal thalamus (VENT), and ventral
1066 posterior complex of the thalamus (VP). Importantly, we compared every brain region annotation
1067 in our data with Allen CCFv3 atlas [61] and judged the ones presented here to best mark the
1068 thalamic regions. This allowed us to calculate the number of cells expressing each gene within or
1069 outside the thalamus. For cross-expressing cells, we considered a pair as thalamic if both cells
1070 were part of the regional mask. More generally, potential combinatorial markers can be discovered
1071 by assessing if their cross-expression is spatially enriched.

1072 Our second exploration involved well-known genes *Foxp2* and *Cdh13*, which mark cortical
1073 layer 6 and show pan-layer expression in the cortex, respectively. These genes exhibited
1074 significant cross-expression, which was spatially enriched in layer 6, whose boundaries we
1075 identified using H2 cell type annotation. The spatial enrichment was viewed by comparing tissue
1076 plots with and without highlighting cross-expressing cells.

1077

1078 **Robustness of cross-expression to batch effects**

1079 To assess the replicability of the cross-expression signature, we used the MERFISH dataset
1080 containing 3 biological replicates (mouse brains) with 3 slices each, where the slices are sampled
1081 from approximately the same location across the brains. We compared the slice-specific networks
1082 between corresponding slices. Moreover, for slice-specific and brain region-specific networks, we
1083 performed comparisons within the BARseq sagittal data and within the coronal data [35] as well
1084 as between these two datasets. Finally, observing that the dorsal to ventral direction is sampled

1085 in both the coronal and the sagittal brains, we compared the densities of cross-expressing cells
1086 in the dorsal to ventral directions across these datasets.
1087
1088 **Cell segmentation quality control assessment**
1089 We assessed the quality of cell segmentation at a global level by comparing co-expression
1090 between the scRNA-seq [32] and MERFISH spatial transcriptomic data. Since the scRNA-seq
1091 was obtained from dissected brain regions, we established correspondence between these and
1092 the brain region annotations in the MERFISH data. The regions used in both data are reported in
1093 Supplementary Table 1. We included only those genes – and in the same order – as present in
1094 the MERFISH data. We calculated gene co-expression using Pearson's correlation and compared
1095 these across the two datasets.
1096 To quantify variability between platforms, we compared gene co-expression between
1097 scRNA-seq and snRNA-seq [77] for the same genes as above. Because the snRNA-seq was
1098 obtained from dissected brain regions, we established correspondence between these and the
1099 scRNA-seq data. The regions used in these data are reported in Supplementary Table 2. Like
1100 before, we quantified co-expression using Pearson's correlation and compared it across the two
1101 datasets.
1102
1103 **Gene expression noise thresholds and cell-neighbor relations**
1104 Because gene expression measurement is noisy, we applied thresholds of 1 to 10 molecules,
1105 thus specifying the minimum number of counts per cell a gene must have to be considered
1106 expressed. We then compared cross-expression networks across these thresholds.
1107 Additionally, a cell might be the nearest neighbor of one or more cells. To ensure that our
1108 framework captures this variability, we compare cross-expression networks for the one-to-one
1109 and many-to-one mappings with each other and with that of the full dataset.
1110

1111 **Benchmarking the algorithm's speed**

1112 We assessed the speed of the cross-expression algorithm by duplicating our BARseq coronal
1113 data, where the gene panel ranged from 2,000 to 8,000 and the number of cells ranged from
1114 20,000 to 200,000. We ran the cross-expression algorithm and calculated the time on a 16 GB
1115 Apple M1 Pro macOS Sonoma 14.5 laptop.

1116

1117 **Declarations**

1118 *Ethics approval and consent to participate*

1119 The experimental work in this study involved the collection of a mouse sagittal brain dataset using
1120 BARseq. All experimental procedures were carried out in accordance with the Institutional Animal
1121 Care and Use Committee at the Allen Institute for Brain Science.

1122

1123 *Consent for publication*

1124 Not applicable.

1125

1126 *Availability of data and materials*

1127 The datasets used in this study are available in the repositories or articles outlined in the "data
1128 acquisition and preprocessing" section. The BARseq sagittal data's sequencing images are being
1129 deposited to Brain Image Library (BIL). While it is being approved, we stored the cell-level gene
1130 expression and cell metadata at
1131 https://drive.google.com/drive/folders/1fk5JDeVJcE71iH1AalCT0il9PN9DTJJm?usp=drive_link.

1132

1133 *Competing interests*

1134 L.F. owns shares in Quince Therapeutics and has received consulting fees from PeopleBio Co.,
1135 GC Therapeutics Inc., Cortexyme Inc., and Keystone Bio. The remaining authors declare no
1136 competing interests.

Funding

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

Authors' contributions

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

Acknowledgements

Not applicable.

Authors' information (optional)

Not applicable.

Code availability

The R package is available at https://github.com/gillislab/CrossExpression/, including installation guide via Bioconductor and a tutorial with an example expression matrix and metadata.

References

1. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods. 2021;18:9–14.

2. Zhuang X. Spatially resolved single-cell genomics and transcriptomics by imaging. Nat Methods. 2021;18:18–22.

3. Moffitt JR, Lundberg E, Heyn H. The emerging landscape of spatial profiling technologies. Nat Rev Genet. 2022;23:741–59.

4. Tian L, Chen F, Macosko EZ. The expanding vistas of spatial transcriptomics. Nat Biotechnol. 2023;41:773–82.

5. Bressan D, Battistoni G, Hannon GJ. The dawn of spatial omics. Science. 2023;381:eabq4964.

6. Liu L, Chen A, Li Y, Mulder J, Heyn H, Xu X. Spatiotemporal omics for biology and medicine. Cell. 2024;187:4488–519.

7. Moses L, Pachter L. Museum of spatial transcriptomics. Nat Methods. 2022;19:534–46.

8. Ren J, Luo S, Shi H, Wang X. Spatial omics advances for in situ RNA biology. Mol Cell. 2024;84:3737–57.

9. Ma Y, Zhou X. Accurate and efficient integrative reference-informed spatial domain detection for spatial transcriptomics. Nat Methods. 2024;21:1231–44.

10. Kim J, Rustam S, Mosquera JM, Randell SH, Shaykhiev R, Rendeiro AF, et al. Unsupervised discovery of tissue architecture in multiplexed imaging. Nat Methods. 2022;19:1653–61.

11. Chidester B, Zhou T, Alam S, Ma J. SpiceMix enables integrative single-cell spatial modeling of cell identity. Nat Genet. 2023;55:78–88.

12. Birk S, Bonafonte-Pardàs I, Feriz AM, Boxall A, Agirre E, Memi F, et al. Quantitative characterization of cell niches in spatially resolved omics data. Nat Genet. 2025;57:897–909.

13. Jerby-Arnon L, Regev A. DIALOGUE maps multicellular programs in tissue from single-cell or spatial transcriptomics data. Nat Biotechnol. 2022;40:1467–77.

14. Haviv D, Remšík J, Gatie M, Snopkowski C, Takizawa M, Pereira N, et al. The covariance environment defines cellular niches for spatial inference. Nat Biotechnol. 2024;1–12.

15. Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. Nat Methods. 2020;17:193–200.

16. Ghazanfar S, Lin Y, Su X, Lin DM, Patrick E, Han Z-G, et al. Investigating higher-order interactions in single-cell data with scHOT. Nat Methods. 2020;17:799–806.

17. Bernstein MN, Ni Z, Prasad A, Brown J, Mohanty C, Stewart R, et al. SpatialCorr identifies gene sets with spatially varying correlation structure. Cell Rep Methods [Internet]. 2022 [cited 2024 Aug 19];2. Available from: https://www.cell.com/cell-reports-methods/abstract/S2667-2375(22)00271-5

1196 18. Miller BF, Bambah-Mukku D, Dulac C, Zhuang X, Fan J. Characterizing spatial gene
1197 expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform
1198 cellular densities. Genome Res. 2021;31:1843–55.

1199 19. Li Z, Wang T, Liu P, Huang Y. SpatialDM for rapid identification of spatially co-expressed
1200 ligand–receptor and revealing cell–cell communication patterns. Nat Commun. 2023;14:3995.

1201 20. Armingol E, Baghdassarian HM, Lewis NE. The diversification of methods for studying cell–
1202 cell interactions and communication. Nat Rev Genet. 2024;25:381–400.

1203 21. Almet AA, Tsai Y-C, Watanabe M, Nie Q. Inferring pattern-driving intercellular flows from
1204 single-cell and spatial transcriptomics. Nat Methods. 2024;1–12.

1205 22. Yang W, Wang P, Xu S, Wang T, Luo M, Cai Y, et al. Deciphering cell–cell communication at
1206 single-cell resolution for spatial transcriptomics with subgraph-based graph attention network.
1207 Nat Commun. 2024;15:7101.

1208 23. Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, et al. Screening cell–cell
1209 communication in spatial transcriptomics via collective optimal transport. Nat Methods.
1210 2023;20:218–28.

1211 24. Dries R, Zhu Q, Dong R, Eng C-HL, Li H, Liu K, et al. Giotto: a toolbox for integrative
1212 analysis and visualization of spatial expression data. Genome Biol. 2021;22:78.

1213 25. Troulé K, Petryszak R, Cakir B, Cranley J, Harasty A, Prete M, et al. CellPhoneDB v5:
1214 inferring cell–cell communication from single-cell multiomics data. Nat Protoc. 2025;1–29.

1215 26. Liu J, Manabe H, Qian W, Wang Y, Gu Y, Chu AKY, et al. CytoSignal Detects Locations and
1216 Dynamics of Ligand-Receptor Signaling at Cellular Resolution from Spatial Transcriptomic Data
1217 [Internet]. bioRxiv; 2024 [cited 2024 Aug 19]. p. 2024.03.08.584153. Available from:
1218 https://www.biorxiv.org/content/10.1101/2024.03.08.584153v1

1219 27. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell–cell communication from
1220 single-cell transcriptomics. Nat Protoc. 2025;20:180–219.

1221 28. Yuan Y, Bar-Joseph Z. GCNG: graph convolutional networks for inferring gene interaction
1222 from spatial transcriptomics data. Genome Biol. 2020;21:300.

1223 29. Singhal V, Chou N, Lee J, Yue Y, Liu J, Chock WK, et al. BANKSY unifies cell typing and
1224 tissue domain segmentation for scalable spatial omics data analysis. Nat Genet. 2024;56:431–
1225 41.

1226 30. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene
1227 expression, spatial location and histology to identify spatial domains and spatially variable
1228 genes by graph convolutional network. Nat Methods. 2021;18:1342–51.

1229 31. Zubair A, Chapple RH, Natarajan S, Wright WC, Pan M, Lee H-M, et al. Cell type
1230 identification in spatial transcriptomics data can be improved by leveraging cell-type-informative
1231 paired tissue images using a Bayesian probabilistic model. Nucleic Acids Res. 2022;50:e80.

1232 32. Yao Z, van Velthoven CTJ, Kunst M, Zhang M, McMillen D, Lee C, et al. A high-resolution
1233 transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature. 2023;624:317–
1234 32.

1235 33. Zhang M, Pan X, Jung W, Halpern AR, Eichhorn SW, Lei Z, et al. Molecularly defined and
1236 spatially resolved cell atlas of the whole mouse brain. Nature. 2023;624:343–54.

1237 34. Shi H, He Y, Zhou Y, Huang J, Maher K, Wang B, et al. Spatial atlas of the mouse central
1238 nervous system at molecular resolution. Nature. 2023;622:552–61.

1239 35. Chen X, Fischer S, Rue MCP, Zhang A, Mukherjee D, Kanold PO, et al. Whole-cortex in situ
1240 sequencing reveals input-dependent area identity. Nature. 2024;1–10.

1241 36. Wu X-B, Zhu Q, Gao M-H, Yan S-X, Gu P-Y, Zhang P-F, et al. Excitatory Projections from
1242 the Prefrontal Cortex to Nucleus Accumbens Core D1-MSNs and κ Opioid Receptor Modulate
1243 Itch-Related Scratching Behaviors. J Neurosci. 2023;43:1334–47.

1244 37. Xiao P, Yan W, Gou L, Zhong Y-N, Kong L, Wu C, et al. Ligand recognition and allosteric
1245 regulation of DRD1-Gs signaling complexes. Cell. 2021;184:943-956.e18.

1246 38. Szklarczyk K, Korostynski M, Cieslak PE, Wawrzczak-Bargiela A, Przewlocki R. Opioid-
1247 dependent regulation of high and low fear responses in two inbred mouse strains. Behav Brain
1248 Res. 2015;292:95–101.

1249 39. Minett MS, Pereira V, Sikandar S, Matsuyama A, Lolignier S, Kanellopoulos AH, et al.
1250 Endogenous opioids contribute to insensitivity to pain in humans and mice lacking sodium
1251 channel Nav1.7. Nat Commun. 2015;6:8967.

1252 40. Wawrzczak-Bargieła A, Ziółkowska B, Piotrowska A, Starnowska-Sokół J, Rojewska E, Mika
1253 J, et al. Neuropathic Pain Dysregulates Gene Expression of the Forebrain Opioid and Dopamine
1254 Systems. Neurotox Res. 2020;37:800–14.

1255 41. Beaulieu J-M, Gainetdinov RR. The Physiology, Signaling, and Pharmacology of Dopamine
1256 Receptors. Sibley DR, editor. Pharmacol Rev. 2011;63:182–217.

1257 42. Fabricius D, Karacay B, Shutt D, Leverich W, Schafer B, Takle E, et al. Characterization of
1258 Intestinal and Pancreatic Dysfunction in VPAC1-Null Mutant Mouse. Pancreas. 2011;40:861.

1259 43. Khaleduzzaman M, Sumiyoshi H, Ueki Y, Inoguchi K, Ninomiya Y, Yoshioka H. Structure of
1260 the Human Type XIX Collagen (*COL19A1*) Gene, Which Suggests It Has Arisen from an
1261 Ancestor Gene of the FACIT Family. Genomics. 1997;45:304–12.

1262 44. Calvo AC, Moreno L, Moreno L, Toivonen JM, Manzano R, Molina N, et al. Type XIX
1263 collagen: a promising biomarker from the basement membranes. Neural Regen Res.
1264 2020;15:988.

1265 45. Sticco MJ, Peña Palomino PA, Lukacsovich D, Thompson BL, Földy C, Ressl S, et al.
1266 C1QL3 promotes cell-cell adhesion by mediating complex formation between ADGRB3/BAI3
1267 and neuronal pentraxins. FASEB J. 2021;35:e21194.

1268 46. Martinelli DC, Chew KS, Rohlmann A, Lum MY, Ressl S, Hattar S, et al. Expression of C1ql3
1269 in Discrete Neuronal Populations Controls Efferent Synapse Numbers and Diverse Behaviors.
1270 Neuron. 2016;91:1034–51.

1271 47. Van Itallie CM, Tietgens AJ, Aponte A, Gucek M, Cartagena-Rivera AX, Chadwick RS, et al.
1272 MARCKS-related protein regulates cytoskeletal organization at cell–cell and cell–substrate
1273 contacts in epithelial cells. J Cell Sci. 2018;131:jcs210237.

1274 48. Kondrychyn I, Kelly DJ, Carretero NT, Nomori A, Kato K, Chong J, et al. Marcksl1 modulates
1275 endothelial cell mechanoresponse to haemodynamic forces to control blood vessel shape and
1276 size. Nat Commun. 2020;11:5476.

1277 49. Shestakova EA, Singer RH, Condeelis J. The physiological significance of β-actin mRNA
1278 localization in determining cell polarity and directional motility. Proc Natl Acad Sci.
1279 2001;98:7045–50.

1280 50. Crow M, Gillis J. Co-expression in Single-Cell Analysis: Saving Grace or Original Sin?
1281 Trends Genet. 2018;34:823–31.

1282 51. Xu M, Bai X, Ai B, Zhang G, Song C, Zhao J, et al. TF-Marker: a comprehensive manually
1283 curated database for transcription factors and related markers in specific cell and tissue types in
1284 human. Nucleic Acids Res. 2022;50:D402–12.

1285 52. Jackson RJ, Hyman BT, Serrano-Pozo A. Multifaceted roles of APOE in Alzheimer disease.
1286 Nat Rev Neurol. 2024;20:457–74.

1287 53. Jones-Tabah J, Mohammad H, Paulus EG, Clarke PBS, Hébert TE. The Signaling and
1288 Pharmacology of the Dopamine D1 Receptor. Front Cell Neurosci [Internet]. 2022 [cited 2025
1289 Apr 17];15. Available from:
1290 https://www.frontiersin.orghttps://www.frontiersin.org/journals/cellular-
1291 neuroscience/articles/10.3389/fncel.2021.806618/full

1292 54. Jones-Tabah J, Mohammad H, Hadj-Youssef S, Kim LEH, Martin RD, Benaliouad F, et al.
1293 Dopamine D1 receptor signalling in dyskinetic Parkinsonian rats revealed by fiber photometry
1294 using FRET-based biosensors. Sci Rep. 2020;10:14426.

1295 55. Sun H, Monenschein H, Schiffer HH, Reichard HA, Kikuchi S, Hopkins M, et al. First-Time
1296 Disclosure of CVN424, a Potent and Selective GPR6 Inverse Agonist for the Treatment of
1297 Parkinson's Disease: Discovery, Pharmacological Validation, and Identification of a Clinical
1298 Candidate. J Med Chem. 2021;64:9875–90.

1299 56. Brice NL, Schiffer HH, Monenschein H, Mulligan VJ, Page K, Powell J, et al. Development of
1300 CVN424: A Selective and Novel GPR6 Inverse Agonist Effective in Models of Parkinson
1301 Disease. J Pharmacol Exp Ther. 2021;377:407–16.

1302 57. Barekatain M, Johansson LC, Lam JH, Chang H, Sadybekov AV, Han GW, et al. Structural
1303 insights into the high basal activity and inverse agonism of the orphan receptor GPR6 implicated
1304 in Parkinson's disease. Sci Signal. 2024;17:eado8741.

1305 58. Margolin DH, Brice NL, Davidson AM, Matthews KL, Carlton MBL. A Phase I, First-in-
1306 Human, Healthy Volunteer Study to Investigate the Safety, Tolerability, and Pharmacokinetics of

1307 CVN424, a Novel G Protein-Coupled Receptor 6 Inverse Agonist for Parkinson's Disease. J
1308 Pharmacol Exp Ther. 2022;381:33–41.

1309 59. Kieburtz K, Brice N, Carlton M, Thompson C, Dubow J. CVN424, A novel GPR6 inverse
1310 agonist for the treatment of Parkinson disease (PD): Phase 2 clinical trial. Parkinsonism Relat
1311 Disord [Internet]. 2024 [cited 2025 Apr 20];122. Available from: https://www.prd-
1312 journal.com/article/S1353-8020%2824%2900136-6/fulltext?utm_source=chatgpt.com

1313 60. Brice NL, Carlton M, Margolin DH, Bexon M, Matthews KL, Dawson LA, et al. CVN424, a
1314 GPR6 inverse agonist, for Parkinson's disease and motor fluctuations: a double-blind,
1315 randomized, phase 2 trial. eClinicalMedicine [Internet]. 2024 [cited 2025 Apr 17];77. Available
1316 from: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370%2824%2900461-
1317 9/fulltext?utm_source=chatgpt.com

1318 61. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. Cell.
1319 2020;181:936-953.e20.

1320 62. Heng N, Malek N, Lawton MA, Nodehi A, Pitz V, Grosset KA, et al. Striatal Dopamine Loss
1321 in Early Parkinson's Disease: Systematic Review and Novel Analysis of Dopamine Transporter
1322 Imaging. Mov Disord Clin Pract. 2023;10:539–46.

1323 63. van Vliet EF, Knol MJ, Schiffelers RM, Caiazzo M, Fens MHAM. Levodopa-loaded
1324 nanoparticles for the treatment of Parkinson's disease. J Controlled Release. 2023;360:212–24.

1325 64. Bogetofte H, Alamyar A, Blaabjerg M, Meyer M. Levodopa Therapy for Parkinson's Disease:
1326 History, Current Status and Perspectives. CNS Neurol Disord - Drug Targets Former Curr Drug
1327 Targets - CNS Neurol Disord. 2020;19:572–83.

1328 65. Delafontaine P, Song Y-H, Li Y. Expression, Regulation, and Function of IGF-1, IGF-1R, and
1329 IGF-1 Binding Proteins in Blood Vessels. Arterioscler Thromb Vasc Biol. 2004;24:435–44.

1330 66. Dallinga MG, Yetkin-Arik B, Kayser RP, Vogels IMC, Nowak-Sliwinska P, Griffioen AW, et al.
1331 IGF2 and IGF1R identified as novel tip cell genes in primary microvascular endothelial cell
1332 monolayers. Angiogenesis. 2018;21:823–36.

1333 67. Imrie H, Viswambharan H, Sukumar P, Abbas A, Cubbon RM, Yuldasheva N, et al. Novel
1334 Role of the IGF-1 Receptor in Endothelial Function and Repair: Studies in Endothelium-
1335 Targeted IGF-1 Receptor Transgenic Mice. Diabetes. 2012;61:2359–68.

1336 68. Liang W, Zheng Y, Zhang J, Sun X. Multiscale modeling reveals angiogenesis-induced drug
1337 resistance in brain tumors and predicts a synergistic drug combination targeting EGFR and
1338 VEGFR pathways. BMC Bioinformatics. 2019;20:203.

1339 69. Le X, Nilsson M, Goldman J, Reck M, Nakagawa K, Kato T, et al. Dual EGFR-VEGF
1340 Pathway Inhibition: A Promising Strategy for Patients With *EGFR*-Mutant NSCLC. J Thorac
1341 Oncol. 2021;16:205–15.

1342 70. Milner CS, Hansen TM, Singh H, Brindle NPJ. Roles of the receptor tyrosine kinases Tie1
1343 and Tie2 in mediating the effects of angiopoietin-1 on endothelial permeability and apoptosis.
1344 Microvasc Res. 2009;77:187–91.

1345 71. Niu Q, Perruzi C, Voskas D, Lawler J, Dumont D, Benjamin LE. Inhibition of Tie-2 Signaling
1346 Induces Endothelial Cell Apoptosis, Decreases Akt Signaling and Induces Endothelial Cell
1347 Expression of the Endogenous Anti-Angiogenic Molecule, Thrombospondin-1. Cancer Biol Ther.
1348 2004;3:402–5.

1349 72. Savant S, La Porta S, Budnik A, Busch K, Hu J, Tisch N, et al. The Orphan Receptor Tie1
1350 Controls Angiogenesis and Vascular Remodeling by Differentially Regulating Tie2 in Tip and
1351 Stalk Cells. Cell Rep. 2015;12:1761–73.

1352 73. Zhang Y, Xu D, Huang H, Chen S, Wang L, Zhu L, et al. Regulation of Glucose Homeostasis
1353 and Lipid Metabolism by PPP1R3G-mediated Hepatic Glycogenesis. Mol Endocrinol.
1354 2014;28:116–26.

1355 74. Aaron PA, Jamklang M, Uhrig JP, Gelli A. The blood–brain barrier internalises via the
1356 EphA2-tyrosine kinase receptor. Cell Microbiol. 2018;20:e12811.

1357 75. Phillips RA, Oh S, Bach SV, Du Y, Miller RA, Kleinman JE, et al. Transcriptomic
1358 characterization of human lateral septum neurons reveals conserved and divergent marker
1359 genes across species. iScience. 2025;28:111820.

1360 76. Chi Y, Qi R, Zhou Y, Tong H, Jin H, Turck CW, et al. scBrainMap: a landscape for cell types
1361 and associated genetic markers in the brain. Database. 2023;2023:baad035.

1362 77. Langlieb J, Sachdev NS, Balderrama KS, Nadaf NM, Raj M, Murray E, et al. The molecular
1363 cytoarchitecture of the adult mouse brain. Nature. 2023;624:333–42.

1364 78. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular
1365 segmentation. Nat Methods. 2021;18:100–6.

1366 79. Fu X, Lin Y, Lin DM, Mechtersheimer D, Wang C, Ameen F, et al. BIDCell: Biologically-
1367 informed self-supervised learning for segmentation of subcellular spatial transcriptomics data.
1368 Nat Commun. 2024;15:509.

1369 80. Ma J, Xie R, Ayyadhury S, Ge C, Gupta A, Gupta R, et al. The multimodality cell
1370 segmentation challenge: toward universal solutions. Nat Methods. 2024;21:1103–13.

1371 81. Choudhary S, Satija R. Comparison and evaluation of statistical error models for scRNA-
1372 seq. Genome Biol. 2022;23:27.

1373 82. Yu Q, Jiang M, Wu L. Spatial transcriptomics technology in cancer research. Front Oncol
1374 [Internet]. 2022 [cited 2024 Aug 21];12. Available from:
1375 https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2022.1019111/full

1376 83. Chen W-T, Lu A, Craessaerts K, Pavie B, Frigerio CS, Corthout N, et al. Spatial
1377 Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. Cell. 2020;182:976-
1378 991.e19.

1379 84. Gurkar AU, Gerencser AA, Mora AL, Nelson AC, Zhang AR, Lagnado AB, et al. Spatial
1380 mapping of cellular senescence: emerging challenges and opportunities. Nat Aging.
1381 2023;3:776–90.

85. Lundberg E, Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. Nat Rev Mol Cell Biol. 2019;20:285–302.

86. Zeng H, Huang J, Ren J, Wang CK, Tang Z, Zhou H, et al. Spatially resolved single-cell translatomics at molecular resolution. Science. 2023;380:eadd3067.

87. Russell AJC, Weir JA, Nadaf NM, Shabet M, Kumar V, Kambhampati S, et al. Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. Nature. 2024;625:101–9.

88. Hu C, Borji M, Marrero GJ, Kumar V, Weir JA, Kammula SV, et al. Scalable spatial transcriptomics through computational array reconstruction. Nat Biotechnol. 2025;1–7.

89. Bentley JL. Multidimensional binary search trees used for associative searching. Commun ACM. 1975;18:509–17.

90. Arya S, Mount DM, Netanyahu NS, Silverman R, Wu AY. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. J ACM. 1998;45:891–923.

91. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57:289–300.

92. Borm LE, Mossi Albiach A, Mannens CCA, Janusauskas J, Özgün C, Fernández-García D, et al. Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. Nat Biotechnol. 2023;41:222–31.

93. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42:293–304.

94. Puchades MA, Csucs G, Ledergerber D, Leergaard TB, Bjaalie JG. Spatial registration of serial microscopic brain images to three-dimensional reference atlases with the QuickNII tool. PLOS ONE. 2019;14:e0216796.

95. Allen WE, Blosser TR, Sullivan ZA, Dulac C, Zhuang X. Molecular and spatial signatures of mouse brain aging at single-cell resolution. Cell. 2023;186:194-208.e18.

96. Csárdi G, Nepusz T, Müller K, Horvát S, Traag V, Zanini F, et al. igraph for R: R interface of the igraph library for graph theory and network analysis [Internet]. Zenodo; 2024 [cited 2024 Aug 21]. Available from: https://zenodo.org/records/10681749

97. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019;9:5233.

98. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2:193–218.

99. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

100. The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224:iyad031.

101. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLOS ONE. 2011;6:e21800.

102. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003;13:2498–504.

103. Morris JH, Kuchinsky A, Ferrin TE, Pico AR. enhancedGraphics: a Cytoscape app for enhanced node graphics [Internet]. F1000Research; 2014 [cited 2024 Aug 21]. Available from: https://f1000research.com/articles/3-147

104. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008;2008:P10008.

105. Borchers HW. pracma: Practical Numerical Math Functions [Internet]. 2023 [cited 2025 Apr 20]. Available from: https://cran.r-project.org/web/packages/pracma/index.html