Spatial isoform sequencing at sub-micrometer single-cell resolution reveals novel patterns of spatial isoform variability in brain cell types

Lieke Michielsen[1,2,3,4,*], Andrey D. Prjibelski[5,*], Careen Foord[1,2,*], Wen Hu[1,2], Julien Jarroux[1,2], Justine Hsu[1,2], Alexandru I. Tomescu[5], Iman Hajirasouliha[3,4], Hagen U. Tilgner[1,2,+]

[1] Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY, USA.
[2] Center for Neurogenetics, Weill Cornell Medicine, New York, NY, USA.
[3] Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine of Cornell University, New York, NY, USA.
[4] Caryl and Israel Englander Institute for Precision Medicine, The Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA
[5] Department of Computer Science, University of Helsinki, Helsinki, Finland.

* equal contribution
+ corresponding author

Abstract

Spatial long-read technologies are becoming increasingly common but lack nanometer- and therefore often single-cell resolution. This leaves the question unanswered of whether spatially variable isoforms represent spatial variability within one cell type or differences in cell-type composition between different regions. Here, we developed Spl-ISO-Seq2 (220nm spot size and 500nm resolution), and the accompanying software packages Spl-IsoQuant-2 and Spl-IsoFind, enabling long-read sequencing using 140 million barcodes compared to 80,000 previously. Applying this to the adult mouse brain, we compared spatial variability by examining (a) differential isoform abundance between known brain regions and (b) spatial isoform patterns that do not align with predefined regions. While the former revealed more spatial isoform differences, both approaches identified overlapping hits, e.g., *Rps24* in oligodendrocytes. For *Snap25,* previously known to exhibit spatial isoform variation, we now show that this variability occurs in excitatory neurons. The second approach also uncovered patterns not captured by predefined-region comparisons, e.g., *Tnnc1* in excitatory neurons. Furthermore, we show that a surprising number of spatial isoform signals is not driven by cell-type composition alone. Finally, we applied our software to public Visium HD 3' long-read data to demonstrate its applicability and strong reproducibility across protocols and biological replicates. Taken together, our experimental and analytical methods enrich spatial transcriptomics with a so-far elusive isoform view of spatial variation for individual cell types.

Introduction

A fundamental question in spatial isoform biology is whether a cell's spatial position influences its isoform expression. This question can be confounded by distinct areas in the brain being dominated by different cell types. Therefore, single-cell-resolution spatial isoform data is essential for disentangling spatially distributed isoforms from cell-type-derived ones.

Single-cell and single-nucleus long-read sequencing have enabled profiling isoforms in distinct cell types[1–5]. Short-read sequencing-based approaches have also described cell-type-specific splice junctions[6–8], revealing exons that are used in specific cell types or altered in evolution and/or disease[5,9,10]. However, a major shortcoming of these methods is that they lose the spatial location of cells, preventing the study of how isoforms may be spatially regulated at the cell-type-specific level.

Simultaneously, spatial gene-expression profiling has set the field forward by localizing cell types and gene expression across tissue sections[11–18]. Based on 10X Genomics' Visium technology, we and our colleagues simultaneously engineered spatial isoform sequencing, revealing many isoform switches that correlated tightly with brain structures, and other isoforms (e.g., of *Snap25*) that showed isoform changes within a region[4,19]. However, Visium's 55µm spot size exceeds the average cell diameter in the mouse brain, and as such, results in "pseudo-bulk" measurements that likely represent multiple cells and thus potentially multiple cell types.

By adapting Slide-SeqV2[12], we have further developed spatial ISOform sequencing (Spl-ISO-Seq) with 10µm resolution and corresponding software Spl-IsoQuant[20]. Spl-ISO-Seq uses exome-sequencing probes[5] and long-molecule selection[20] to enrich for spliced and (near-)complete cDNAs. Although 10µm resolution is sufficient to identify large cells in the human brain, such as excitatory neurons, it importantly lacks resolution to identify smaller cells, especially when considering other common model organisms as mouse.

Since then, new spatial transcriptomic technologies supporting a higher resolution have been released, including the Visium HD 3' spatial assay[21] and StereoSeq[22]. StereoSeq offers a higher resolution compared to Visium HD (500nm vs 2µm), making it well-suited for single-cell long-read spatial transcriptomics. Although segmentation from ssDNA staining, which labels only nuclei, and ~10µm z-sampling can still lead to doublets[23], the substantial gain in resolution is highly promising for spatially studying individual cells.

Therefore, based on the Stereo-seq approach, coupled with our previously developed exome enrichment[5] and long-molecule selection[20], here we devise Spl-ISO-Seq2 with 20-fold improved spatial resolution (500nm instead of 10µm) for over 100 million barcodes. We show that this spatial isoform technology can work with PacBio as well as with Oxford Nanopore long reads. Along with this wet lab approach, we devise Spl-IsoQuant-2, which adds highly-specific barcode calling to our IsoQuant algorithm[24], which we substantiate by using dry-lab simulation experiments. Furthermore, we show that unique molecules sequenced on both PacBio and Oxford Nanopore reveal high accuracy of transcript assignments.

To detect spatially variable isoforms, we developed Spatial Isoform Finder (Spl-IsoFind). For similar questions regarding gene expression, many analytical frameworks have been developed, e.g. for identifying spatially variable genes[25–32]. However, methods to detect spatially variable isoforms have lagged behind. Spl-IsoFind uses Moran's I[33], a global spatial autocorrelation score, which has been successfully used for gene-expression studies[25,34–36]. We identify cell-type-specific spatially variable isoforms, such as *Snap25* for excitatory neurons and *Rps24* for oligodendrocytes, as well as isoforms with spatial patterns that do not align with predefined regions. Furthermore, using Spl-IsoFind's cell-type constrained

84 permutations, we show that most spatial isoform signals are not solely due to differences in
85 cell-type composition alone. Finally, we demonstrate Spl-IsoQuant-2 and Spl-IsoFind's
86 versitality by applying it to a publicly available Visium HD 3' long-read mouse brain sample.
87 Despite slight differences in slide positioning and brain regions captures, our results show
88 strong reproducibility across protocols and biological replicates.
89 Overall, Spl-ISO-Seq2, Spl-IsoQuant-2, and Spl-IsoFind together provide a long-read
90 compatible spatial transcriptomics framework at a single-cell resolution. Applied to the adult
91 mouse brain, this approach reveals cell-type specific isoform changes both between
92 predefined regions and in previously unrecognized spatial patterns. Altogether, this high-
93 resolution technology can be applied to any tissue type to investigate single-cell isoform
94 patterns across a spatial region.

95 Results

96 Gene-expression patterns in coronal brain slices

97 To define spatially variable isoforms for a given cell type, we used two consecutive coronal
98 slices of an adult mouse brain, covering cortex, hippocampus, thalamus, and midbrain (**Fig**
99 **S1a-b**). We used the Stereo-seq spatial approach to barcode cDNAs at 500nm resolution[22]
100 and performed short-read sequencing as per Stereo-seq guidelines on an aliquot of cDNAs.
101 Another aliquot of cDNAs was used to perform exome enrichment and long-molecule
102 selection, as we have previously done[20], followed by PacBio Kinnex sequencing and Oxford
103 Nanopore long-read sequencing. Short-read sequencing, in conjunction with an ssDNA-stain,
104 was used to segment individual cells, assign cell types and to anchor barcodes to specific
105 spatial locations (Methods). Long reads, both of the PacBio and the Oxford Nanopore
106 platforms, were used to define novel isoforms and to quantify annotated and novel isoforms
107 (**Fig 1a**). In this study, we focused our analysis on the complete right hemisphere. Short-read
108 gene expression clearly indicated biologically meaningful brain regions, including distinct
109 cortical layers (**Fig 1b, S1c**). We used RCTD, a state-of-the-art deconvolution method, to
110 assign cell types to the segmented cells[37–39]. Although Stereo-seq has a high resolution,
111 overlapping cells can still cause doublets[23]. We defined our major cell types using RCTD's
112 singlet spots only (**Fig 1c, S1d, S2**). While excluding doublets reduces the statistical power,
113 especially for smaller cell types such as astrocytes and oligodendrocytes, it provides a more
114 accurate basis for detecting cell-type specific spatially variable isoforms. Furthermore, staining
115 and short-read data were sufficient to align two coronal brain slices (**Fig 1d**). Of note, cortex
116 and above all some hippocampal areas stood out in terms of UMI density, owing to higher cell
117 density (**Fig 1e, S1e**).

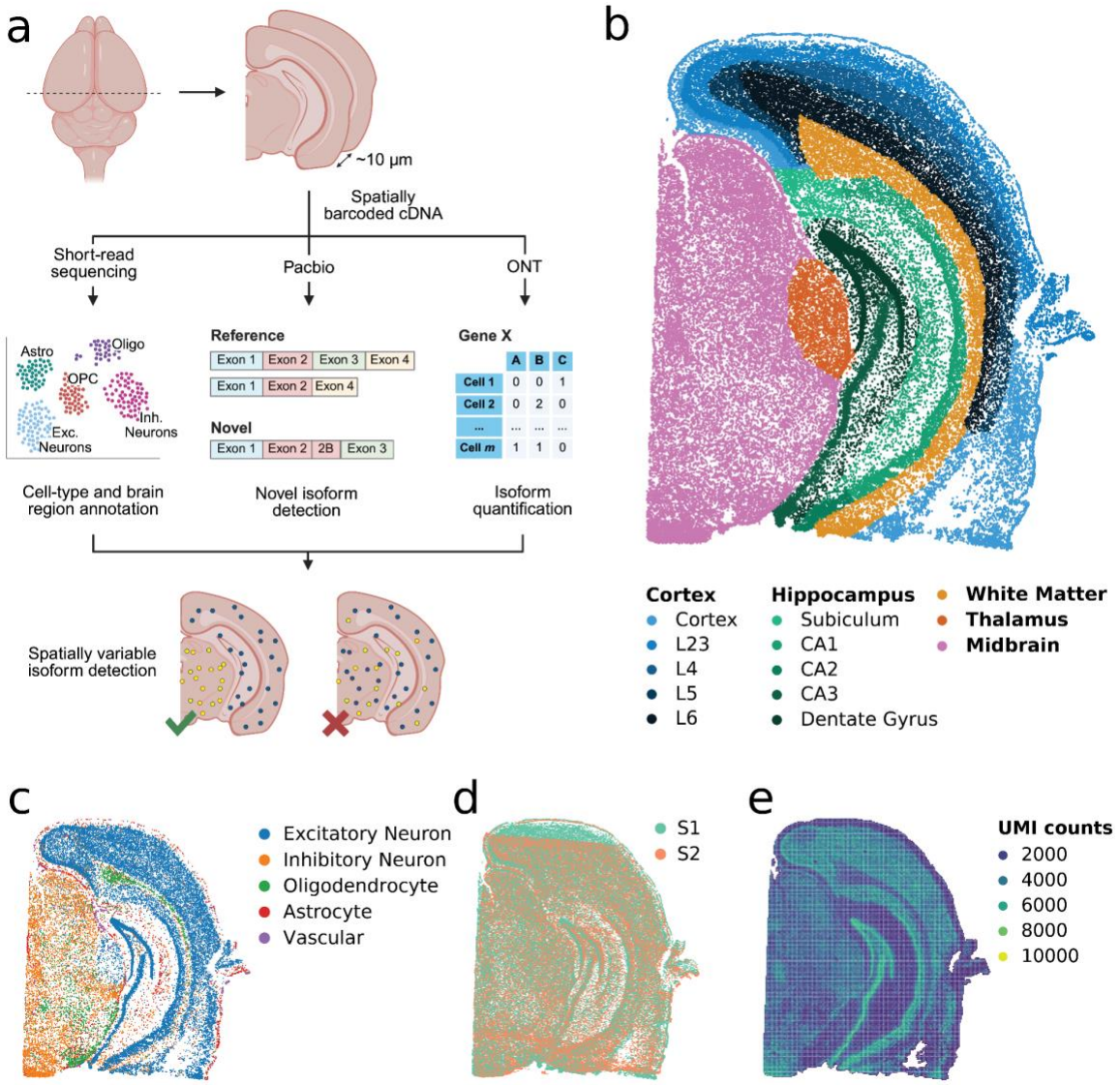

Figure 1. Long-read spatial transcriptomics of two mouse brain slides. a) Schematic overview of the study design. Spatially barcoded cDNA was generated for two coronal sections (10µm apart) of a P56 mouse brain using the Stereo-seq platform[22]. The pool of cDNA is sequenced using 1) short-read technologies, 2) Pacbio, and 3) Oxford Nanopore Technologies. All sequencing data is combined to detect (cell-type-specific) spatially variable (novel) isoforms. **b)** (Sub)region annotation of Sample 1. **c)** Cell-type annotations of Sample 1. Only cells annotated as singlets and cell types with >100 cells are plotted. **d)** Alignment of Sample 1 (S1) and Sample 2 (S2). **e)** UMI counts per bin (50x50spots) for Sample 1.

A long-read approach for spatial isoform expression with very high precision

We then engineered a long-read approach for spatial isoform expression using the barcoded, exome-enriched cDNAs after long-molecule selection (Methods). The Stereo-seq protocol, due to the usage of similar primers at 3' and 5' end can allow for the concatenation of multiple barcoded cDNAs. To counter this effect, we used a PCR strategy to append an additional sequence to the barcoded end of the read (Methods). Although this strategy decreased the rate of molecule concatenation, it did not fully eradicate the occurrence. We note that this occurrence affects the cDNA itself, and is unrelated to the combination of reads involved in PacBio library preparation, which can easily be split with existing software (Methods). To account for cDNA read concatenations, Spl-IsoQuant-2 first implements a read 'splitting' strategy to split concatenated molecules followed by a barcode recognition algorithm (Methods). Consequently, extracted cDNAs are assigned to individual genes and isoforms, and PCR duplicates are removed (**Fig 2a**, Methods). To account for possible previously unannotated isoforms we first ran IsoQuant[24] using all sequenced data to generate an annotation containing 28,891 known and 26,445 novel isoforms, which were then used as input to the Spl-IsoQuant-2 pipeline. To generate this annotation, Gencode version 36 mouse gene annotation[40] was used (but other annotations, including RefSeq[41] or Ensembl[42] can be used as well). As a core pipeline integrated into Spl-IsoQuant-2, IsoQuant has demonstrated high precision in novel transcript discovery by previous reports[24,43,44].

Detection of multiple concatenated isoforms is performed by locating the template-switch oligo (TSO) as well as linker and barcode sequences far from read extremities (**Fig 2b**). Although various software for barcode detection in Nanopore reads exists at the moment[45–47], none of these tools is capable of "parsing" the Stereo-seq molecule structure. Indeed using Flexiplex, arguably the most versatile tool, we were not able to obtain any meaningful results (Methods). We thus developed a barcode recognition algorithm specific for the exploited protocol. To match against a set of known barcodes (>400 million barcodes of length 25bp), we first build a k-mer index, which records for every k-mer all the barcodes using this k-mer (k=14 by default). Within a de-concatenated read, potential barcode sequences are extracted from in between the detected primer and linker. If a potential barcode sequence shows at least one common k-mer to a known barcode, the algorithm records it as a potential match. For the top N potential matches a Smith-Waterman alignment is used (N=10 by default). For a given read, Spl-ISO-Seq2 then records the best-matching known barcode as the barcode of the read if and the alignment score is at least S_{min} (22 by default) and no other known barcode has the same score (**Fig 2c**, Methods). While this approach might miss the cases where the true known barcode does not make it to the top N best candidates or has the alignment score lower than S_{min}, it allows to assign known barcodes with high-precision and keep reasonable running time.

To assess recall and precision, we first simulated reads with a 3.8% error rate without simulating read concatenation (Methods). A polyT sequence, representing the polyA tail, could be found in virtually all reads, however primer and linker were detected in only 91.8% of the reads, owing to simulated sequencing reads. A barcode was called in 76.66% of reads and in 76.62% of reads this called barcode corresponded to the known barcode attached during the simulation (**Fig 2d**), thus yielding a precision of 99.95% and a recall of 76.62% (**Fig 2f, top panel**). Similar simulation of sequencing errors along with cDNA concatenation yielded a precision of 99.94% and a recall of 78.07% (**Fig 2e,f**). In summary, Spl-IsoQuant-2 yields close to perfect precision, while delivering high (but not perfect) recall for barcode recognition.

174 We then enriched barcoded cDNA using exome-sequencing probes[4,5,9,20] using also a long-
175 molecule selection step[20]. We sequenced 203.7 and 58.8 million reads on Oxford Nanopore
176 platform for the two coronal slices (from now on called Sample 1 AllExome (AE) and Sample
177 2 (AE) respectively) as well as 19.9 million and 17.6 million PacBio reads. For Sample 1 (AE)
178 we used two ONT flow cells instead of one explaining the difference in read depth. We also
179 performed one Nanopore long-read run on cDNAs enriched for 3378 genes related to various
180 brain diseases and synaptic function (from now on called Sample 1 (4K)), rather than for the
181 whole exome (**Table S1-6**) (Methods). Gene expression as measured using the short-read
182 and long-read sequencing technologies are strongly correlated (**Fig S3**).
183 Of note, due to amplification and splitting of cDNA libraries for distinct sequencing protocols,
184 cDNA copies of the same original molecule can be sequenced on both platforms. We therefore
185 tested reproducibility and the influence of higher error rates in Nanopore sequencing on
186 transcript assignments by taking advantage of such pairs of reads sequenced on both
187 platforms. Indeed 3.6 and 2.3 million unique molecules for Sample 1 (AE) and Sample 2 (AE)
188 respectively were sequenced on both Oxford Nanopore and PacBio (Methods). In both
189 replicates, 99.4% of such read pairs were assigned to the same isoform (**Fig 2g**). This
190 observation supports that errors in Nanopore reads do not lead to widespread mis-
191 assignments of reads to the wrong isoform.

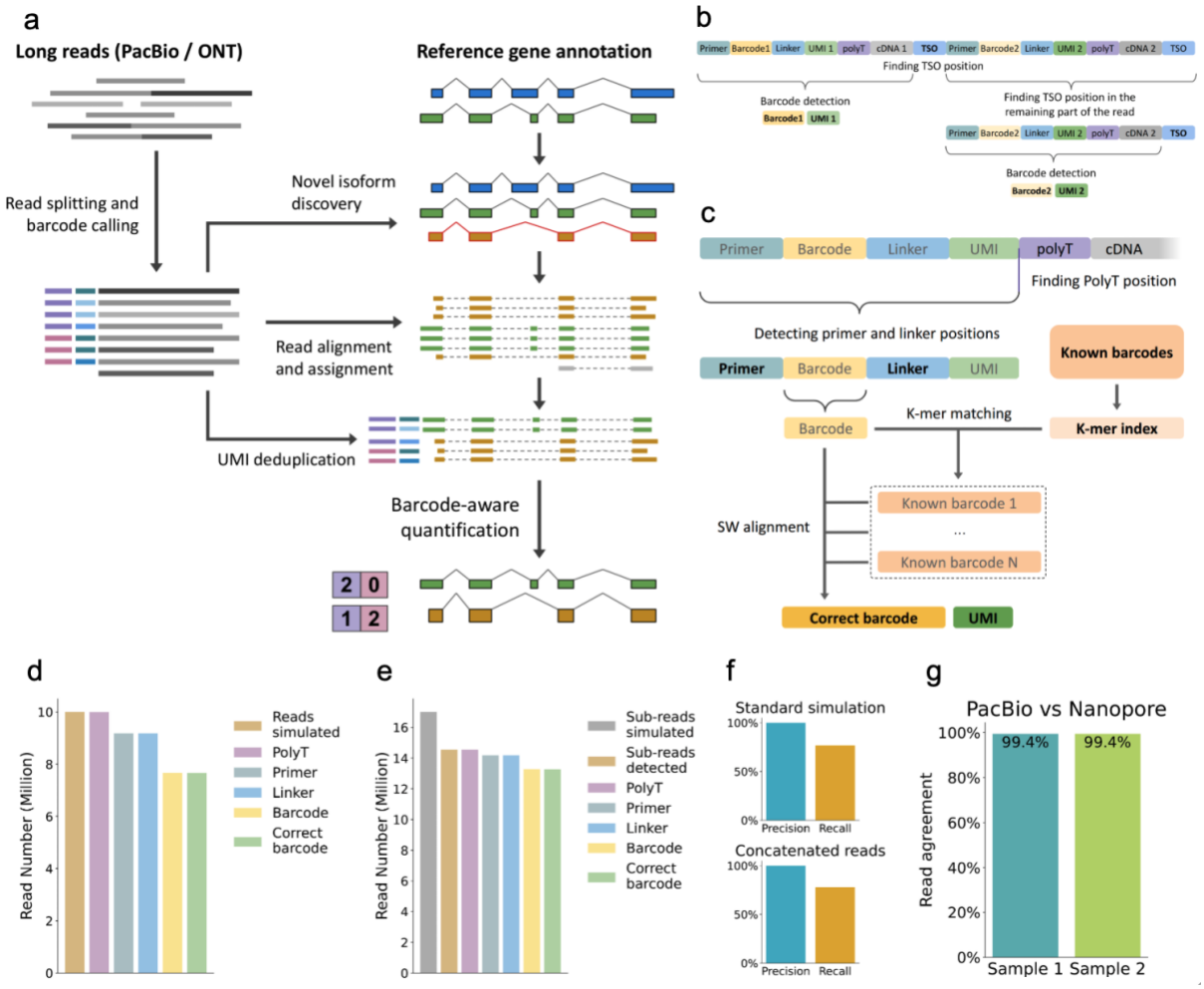

Figure 2. Processing long reads obtained with Spl-ISO-Seq2. a) Spl-IsoQuant-2 pipeline overview. Purple and pink bars next to the reads indicate different barcodes. Blue bars between the reads and the barcodes indicate UMIs. Final diagram shows isoform count per barcode (purple/pink). **b)** Splitting reads into individual cDNAs by locating the template-switch oligos (TSO). **c)** Barcode detection using k-mer indexing and Smith-Waterman local alignment. **d)** Primer, linker and barcode detection statistics for simulated reads without cDNA concatenation. **e)** Primer, linker and barcode detection statistics for simulated reads with cDNA concatenation. A simulated read can consist of multiple concatenated subreads resulting in >16 million subreads in total. **f)** Precision and recall of the designed barcode calling algorithm for simulated data without (top) and with cDNA concatenation (bottom). **g)** Agreement between real PacBio and Oxford Nanopore reads.

Detecting spatially variable isoforms using predefined brain regions

We compared differential relative isoform expression (not confounded by gene expression) between brain region pairs on ONT Sample 1+2 (AE) using our *n* x 2 isoform tests[4,9]. For each gene, we built an *n* × 2 table of isoform counts and applied a chi-squared test with Benjamini-Hochberg (BH) correction (Methods). The midbrain showed the most differentially used isoforms, especially when compared to hippocampus and cortex (**Fig 3a, Table S7-11**). Interestingly, while the percentage of significant genes was similar between the midbrain and thalamus, the total number of significant genes in the thalamus was much lower, which is explained by the lower read depth. Two example genes showing differentially used isoforms between the midbrain and cortex are Neuroplastin (*Nptn*) and Phosphatase and Actin Regulator 1 (*Phactr1*) (BH-corrected p-value = 0.017 and 0.0067 respectively). For both genes, one exon drives the difference between the cortex and midbrain so we used PCR to validate these exons' spatial inclusion (**Fig 3b-g, S4**) (Methods). For *Nptn*, normalized exon inclusion was decreased in the midbrain compared to other regions (p<5.4e-4, Welch's two-sample t-test). Of note, Nptn is a synaptic glycoprotein involved in synaptogenesis, long-term potentiation and memory formation[48]. These two *Nptn* isoforms have been documented by the field as Np55 and Np65, where Np65 contains an additional immunoglobulin domain[49]. For *Phactr1*, which helps to regulate actin cytoskeletal organization[50,51], normalized exon inclusion increased in the midbrain (p<3.4e-4, Welch's two-sample t-test).

We initially expected differences in cell-type composition to drive most significant results. However, restricting the analysis to excitatory neurons alone resulted in a similar percentage of significant isoforms for each comparison (**Fig 3h**). For each comparison, the confidence intervals for all cells and excitatory neurons overlap (**Fig S5**), indicating that cell-type composition alone is unlikely to explain results from all cells.

Within the hippocampus, especially the CA3 subregion compared to the dentate gyrus (DG) showed differential isoform abundance (**Fig 3i**). Likewise, in the cortex, layer 2-3 showed important changes in isoform regulation especially when compared to layer 5 and layer 6 (**Fig 3j**). *Gria2* shows isoform expression changes between excitatory neurons in the hippocampus and cortex. Isoform 201 is more frequently used in the hippocampus compared to the cortex. However, examining these broader brain regions does not capture all nuances: within the hippocampus, differences are also evident between the CA3 and CA1 subregions, as well as between CA3 and DG (**Fig 3k**). Similarly, in the cortex, notable isoform changes are observed between layers 2-3 and layer 5 (**Fig S6**). *Gria2* encodes the GluA2 protein, one of 4 subunits which constructs the glutamatergic α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic (AMPA) receptor. The two spatially variable *Gria2* isoforms identified here represent the *flip* and *flop* variants, a known alternative-splicing driven variance which alters the channel's rate of desensitization[52–54]. In the other major cell types (inhibitory neurons, oligodendrocytes, and astrocytes), we identified fewer significant genes compared to excitatory neurons, which is mainly explained by a lower read depth (**Fig S7**).

While the previous analysis focused on testing differentially used isoforms, we also tested for alternatively spliced cassette exons (Methods). During these tests, the number of significant results decreases (**Fig S8**), even though a higher number of reads were used during the tests. This decrease is likely due to the isoform tests indirectly capturing a broader range of variations, such as alternative 3' and 5' splice sites, transcription start sites, and polyadenylation sites, which are not considered in the exon-specific tests.

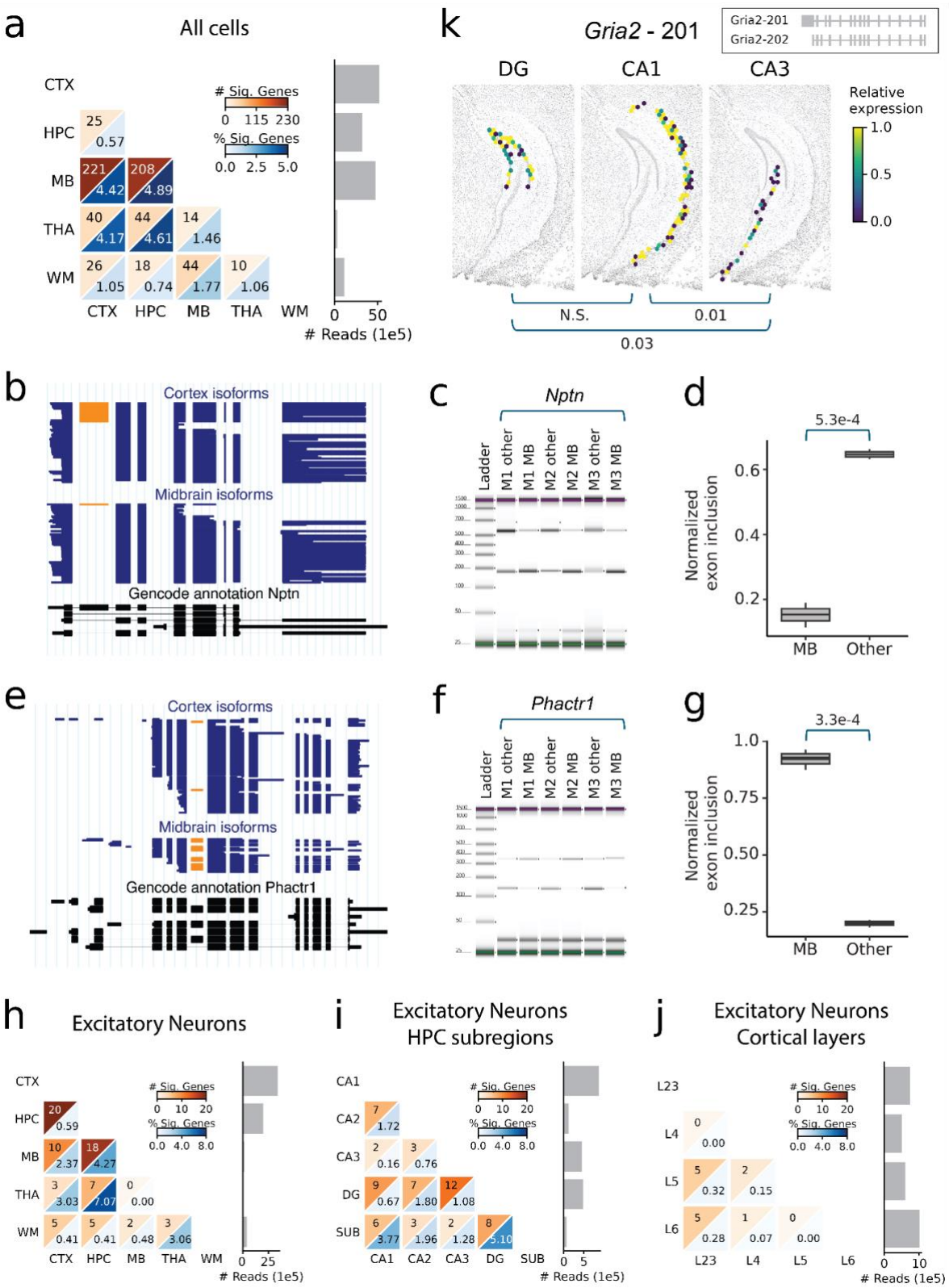

Figure 3. Detecting genes exhibiting differential isoform expression across predefined regions. a) Number and percentage of genes changing isoform expression across predefined brain regions using all cells in combined Sample 1+2 (AE). Barplot on the right shows the number of long reads used during the analysis per brain region. **b,e)** Single-cell long-reads for **b)** *Nptn* and **e)** *Phactr1* measured in the cortex and midbrain. Each line in the top two tracks represents a single cDNA molecule. The orange exon drives the difference between isoforms in the two brain regions. The bottom black track shows the Gencode annotation **b)** (chr9: 58489523-58565238) and **e)** (chr13: 42834099-43292002). **c,f)** Tapestation gel of PCR amplified products using exon-of-interest-spanning **c)** *Nptn* and **f)** *Phactr1*

specific primers. **d,g)** Normalized quantification of tapestation gel for **d)** *Nptn* and **g)** *Phactr1* (Methods). Welch's two-sample t-test was used for statistical testing. **h-j)** Number and percentage of genes changing isoform expression in excitatory neurons across (sub)regions. **k)** Relative expression of *Gria2*-201 in excitatory neurons in different hippocampal subregions of Sample 1 (AE). Every hexagon shows the mean relative expression of the underlying cells. Relative expression indicates the fraction of long reads for that gene (*Gria2*) that are assigned to our isoform of interest (201). A hexagon is only plotted when there are underlying cells expressing *Gria2* (Methods). The ssDNA staining is shown in the background. Numbers indicate BY-corrected p-values. The schematic diagram in the corner shows the gene structure of *Gria2*-201 and *Gria2*-202, the two most common isoforms in the data. CTX = cortex, HPC = hippocampus, MB = midbrain, THA = thalamus, WM = white matter, DG = Dentate Gyrus, SUB = Subiculum, Mx = mouse sample, N.S. = not significant

Spl-IsoFind identifies spatially variable isoforms beyond predefined brain regions

Previously, we compared isoform usage between predefined regions, assuming that the underlying alternative splicing mechanisms respect these borders. In theory, isoform usage might differ within a region or even across regions (**Fig 4a**). Therefore, we developed Spatial Isoform Finder (Spl-IsoFind) to calculate spatial autocorrelation using Moran's I[33] (Methods). Moran's I indicates whether an isoform exhibits a spatial pattern by comparing the values of neighboring cells. This method has been used to detect spatially variable genes[25,34–36], but not for spatially variable isoforms (SVI) yet. Since we will compare the fraction of isoform usage, the latter is slightly more challenging since many cells have missing values as the gene was not measured in the long-read data.

First, we assessed Spl-IsoFind's precision and recall using simulated data (Methods). We converted Moran's I score to a p-value using permutation tests and adjusted for multiple testing using Benjamini-Yekutieli correction. In simulated random data without a spatial pattern, the false discovery rate is low (0.17%) (**Fig S9**). Only the smallest simulated dataset (50 cells) has some false positives (5 out of 165 simulations). The recall, however, greatly depends on the underlying simulated pattern. We detect distinct patterns, such as a border between two areas, during all simulations, regardless of the number of cells. For complex patterns (e.g., a small area of interest which is surrounded by cells with random values), the result depends on the number of cells in our simulation.

Next, we ran Spl-IsoFind on all ONT datasets in a cell-type agnostic and cell-type specific fashion (**Fig 4bc, Table S12-16**). The cell-type agnostic mode results in most SVIs for two reasons: 1) we use more cells during testing, and 2) isoforms can be cell-type-specific and the distribution of cell types can change across regions. As shown with the simulations, having more cells and thus more reads results in more power to detect a pattern. Sample 2 (AE) has the lowest number of reads per isoform (**Table S5**) and as such the lowest number of detected SVIs. The higher percentage of detected isoforms in Sample 1 (4K) is most likely due to our enrichment for neuronal genes.

Most SVIs (61.7%) are detected in two or more datasets (**Fig 4d**). However, 165 isoforms (426 pairwise combinations) are tested in two (or more) datasets, but significant in only one. In 384 cases (90.1%) the isoform is significant in the dataset with the higher read depth (**Fig S10**). Both Sample 1+2 (AE) and Sample 1 (4K) have many unique SVIs. The former highlights the benefit of a higher read depth in general, while the latter demonstrates that targeted sequencing leads to a higher read depth for our genes of interest, leading to many new discoveries.

In previous work, we proposed three models to explain alternative splicing across brain regions: 1) multiple cell types change splicing, 2) one cell type changes splicing, 3) no cell

305 type directly changes splicing, but the cell type composition changes[4]. Cell-type agnostic SVIs
306 can fall into either the first or third category. To distinguish between these, we calculated p-
307 values using cell-type constrained permutations (Methods). Unlike our previous approach,
308 where the values of all cells were shuffled randomly to convert Moran's I to a p-value, we now
309 shuffle values only within the same cell type (**Fig S11**). For Sample 1+2 (AE), the uncorrected
310 p-values of most isoforms remain very similar (**Fig 4e**). After applying multiple testing
311 corrections, 151 out of the 172 isoforms remain significant (FDR < 0.05, BY corrected),
312 suggesting that at least one cell type exhibits spatial variability. Thus, most isoform-level
313 spatial signals cannot be explained by shifts in cell-type composition alone.
314 An example of significant SVIs after applying cell-type constrained permutations are *Snap25*-
315 201 and *Snap25*-202. These isoforms exhibit a spatial pattern across cortical layers and in the
316 midbrain (**Fig 4f-g, S12**). In general, *Snap25*-201 is the dominant isoforms, except for parts of
317 the midbrain and the deeper cortical layers which also show expression of *Snap 25*-202.
318 Previous studies have identified similar spatial patterns in the brain, but their resolution was
319 insufficient to pinpoint individual cell types[4,19]. With our cellular resolution, we can analyze
320 these patterns at the cell type level. Notably, we observe significance in excitatory neurons
321 (where the pattern is present in the cortex) and inhibitory neurons (where the pattern appears
322 in the midbrain) (**Fig S12**). These inhibitory neurons may be dopaminergic, though potentially
323 mislabeled, and the region in the midbrain exhibiting the inclusion of *Snap25*-202 is likely the
324 substantia nigra or VTA.
325 *Snap25* serves as a classic example of two isoforms switching between different conditions.
326 However, for some genes, we observe more complex patterns, with either only one or more
327 than two significant isoforms (**Fig 4h**). For instance, *Arpp19* exhibits four distinct SVIs. Two
328 isoforms (203 and a newly discovered isoform) are specific to the cortex. *Arpp19*-209 is highly
329 expressed across all brain regions except the cortex, while *Arpp19*-205 shows the lowest
330 overall expression but is predominantly used in the hippocampus, midbrain, and thalamus (**Fig
331 4i, S13**).
332 We identified two significant isoforms of *Rps24* across all cell types except astrocytes. For
333 example, in oligodendrocytes, we observed a distinct spatial difference between the midbrain
334 and white matter (**Fig 4j S14**). Isoform *Rps24*-202 was predominantly expressed in the white
335 matter, while the midbrain exhibited a more balanced expression of *Rps24*-202 and *Rps24*-
336 205.
337 We then compared our SVIs to the genes we identified using the predefined regions. This
338 comparison is challenging, as Spl-IsoFind directly pinpoints isoforms, whereas using the
339 predefined regions, we identified genes with splicing changes, but it was not always clear
340 which specific isoforms were affected. In general, using the predefined regions, we identify
341 more significant genes (**Fig. 4k, S15**), as fewer reads are typically required to detect
342 significance. However, by applying Spl-IsoFind, we also uncover new SVIs, such as *Tnnc1*-
343 202 and a newly discovered *Tnnc1* isoform (**Fig. 4l, S16**).

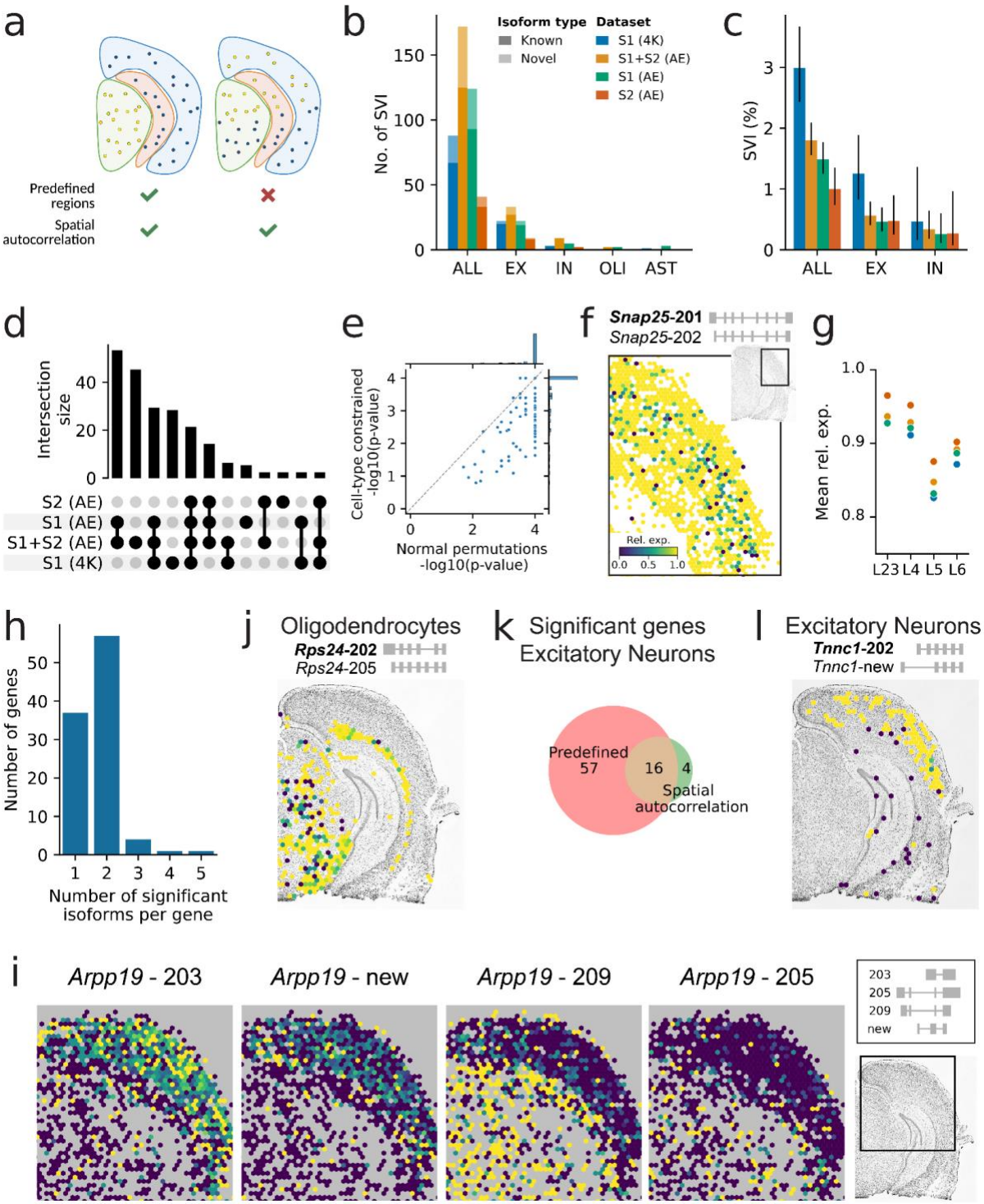

Figure 4. Detecting spatially variable isoforms using spatial autocorrelation. a) Schematic example of a pattern (left) which can be detected using either predefined regions and spatial autocorrelation and a pattern (right) which can only be detected using spatial autocorrelation. **b)** Number and **c)** percentage of significant spatially variable isoforms (FDR < 0.05) per dataset and celltype. Error bars indicate 95% confidence interval. **d)** The overlap of significant isoforms across the datasets. Note: not all isoforms were tested in all datasets due to differences in read depth. **e)** Comparing the uncorrected p-values obtained using normal and cell-type constrained permutations for the isoforms significant in Sample 1+2 (AE) using all cells. **f)** Spatial distribution of *Snap25*-201 across the cortical layers in the Sample 1 (AE). Every hexagon shows the mean relative expression of the underlying cells (Methods). The schematic gene diagram shows the structure of the two *Snap25* isoforms in our data. **g)** Mean relative expression of *Snap25*-201 in different cortical layers. Colors indicate the different datasets. Lower relative expression of *Snap25*-201 in Layer 5 implies higher relative expression of

357 *Snap25*-202. **h)** Number of significant isoforms per gene in Sample 1+2 (AE). **i)** Spatial distribution of
358 the four significant isoforms of *Arpp19* in Sample 1 (AE). Every hexagon shows the mean relative
359 expression of the underlying cells (Methods). The schematic gene diagram shows the structure of the
360 isoforms. **j)** Spatial distribution of *Rps24*-202 in oligodendrocytes of Sample 1 (AE). Every hexagon
361 shows the mean relative expression of the underlying cells (Methods). The gene diagram on top shows
362 the structure of the two most common *Rps24* isoforms in the data. **k)** Overlap between the significant
363 genes found using both methods for the excitatory neurons in the Sample 1 (AE) + Sample 2 (AE) data.
364 **l)** Spatial distribution of *Tnnc1*-202 in excitatory neurons of Sample 1 (AE) which was only discovered
365 using spatial autocorrelation. Every hexagon shows the mean relative expression of the underlying cells
366 (Methods). The gene diagram on top shows the structure of the two most common *Tnnc1* isoforms in
367 the data. SVI = spatially variable isoforms, ALL = all cells, EX = excitatory neurons, IN = inhibitory
368 neurons, OLI = oligodendrocytes, AST = astrocytes, AE = AllExome dataset, 4K = 4K targeted dataset.

369 Spatially variable isoforms strongly replicate across protocols and 370 biological samples

371 To showcase that our software is generalizable to other long-read spatial transcriptomics
372 datasets, we applied Spl-IsoQuant-2 and Spl-IsoFind to a publicly available Visium HD 3' long-
373 read mouse brain dataset. While Visium HD has 2um resolution, the very similar barcodes of
374 neighboring spots makes it challenging to assign reads to the exact location considering the
375 ONT error rate. Using simulated data, we compared the barcode detection performance of
376 Spl-IsoQuant-2 and Space Ranger, the 10x Genomics' software developed to analyze Visium
377 data. The performance is very comparable and grouping the individual spots in 8 or 16um bins
378 greatly improves the precision and recall for both (**Fig 5a**). For Spl-IsoQuant-2, using a minimal
379 alignment score of 13 for each of the two barcode sequences (see Methods) and 8um bins is
380 the best trade off between precision, recall and keeping a sufficient resolution for single-cell
381 analysis. However, when we applied both Spl-IsoQuant-2 and Space Ranger to the Visium
382 HD data, we noticed that Space Ranger uniquely assigns reads to many bins (56,155), which
383 do not overlap with the tissue slice and as such are most likely false positives (**Fig 5b, S17**).
384 In contrast, Spl-IsoQuant-2 only uniquely assigns reads to 5,258 bins. Such a difference
385 between tools' performance on simulated and real data is likely caused by a somewhat
386 simplistic simulation process. In the simulated dataset, every read represents a full-length
387 cDNA molecule constructed according to the protocol with sequencing errors being the only
388 obstacle for detecting the barcodes. Simulation does not account for any kind of
389 contaminations, protocol or sequencing artefacts that may affect barcode recognition. To avoid
390 considering possible false positives and bins outside of the tissue slide, we used the Spl-
391 IsoQuant-2 output for all downstream analysis.
392 The Visium HD long-read data was generated without long-molecule and exome selection,
393 resulting in a shorter overall read length and a lower percentage of spliced reads compared to
394 Spl-ISO-Seq2 (**Fig 5c-d, Tables S17-20**). As a consequence, fewer reads have an isoform
395 assigned (7.8M vs 8.0M in Sample 1 (AE)) even though four ONT flow cells were used
396 compared to two flow cells for Sample 1 (AE). Similar to the Spl-ISO-Seq2 samples, the Visium
397 HD slide is a coronal section of the adult mouse brain, but cut slightly more anterior. As a
398 result, the slide contains some similar brain regions, such as cortex and hippocampus, but the
399 midbrain is excluded while, for example, striatum and hypothalamus are included (**Fig 5e**). As
400 a high-resolution microscope image necessary for segmenting cells was unavailable, we
401 assigned cell types to the 8um bins (**Fig S18**) (see Methods).
402 Next, we applied Spl-IsoFind to the Visium HD data to detect SVIs (**Fig 5f, Table S21-25**).
403 Compared to the Spl-ISO-Seq2 data, relatively more SVIs were detected in astrocytes, which
404 is most likely explained by the larger number of astrocytes in the Visium HD data, while the
405 proportion of SVIs in the other cell types was comparable (**Fig 4c, 5g**). In total, 4,951 isoforms

were tested in both Spl-ISO-Seq2 Sample1+2 AE and Visium HD, of which 144 and 111 were significant, respectively (**Fig S19**). The significant overlap of 52 isoforms between the two experiments indicates strong reproducibility across protocols and biological replicates (p = 2.29e-53, two-sided Fisher's exact test) (**Fig 5h-i**). For instance, *Arpp19* had four significant isoforms in Sample 1+2 (AE), three of which were also detected in Visium HD (**Fig S20**). Different SVIs between the two brain regions can reflect regional specificity. For example, two *Gad2* isoforms were significant only in Visium HD, likely due to their expression in the striatum, which is absent in the Spl-ISO-Seq2 samples (**Fig 5j, S21**).

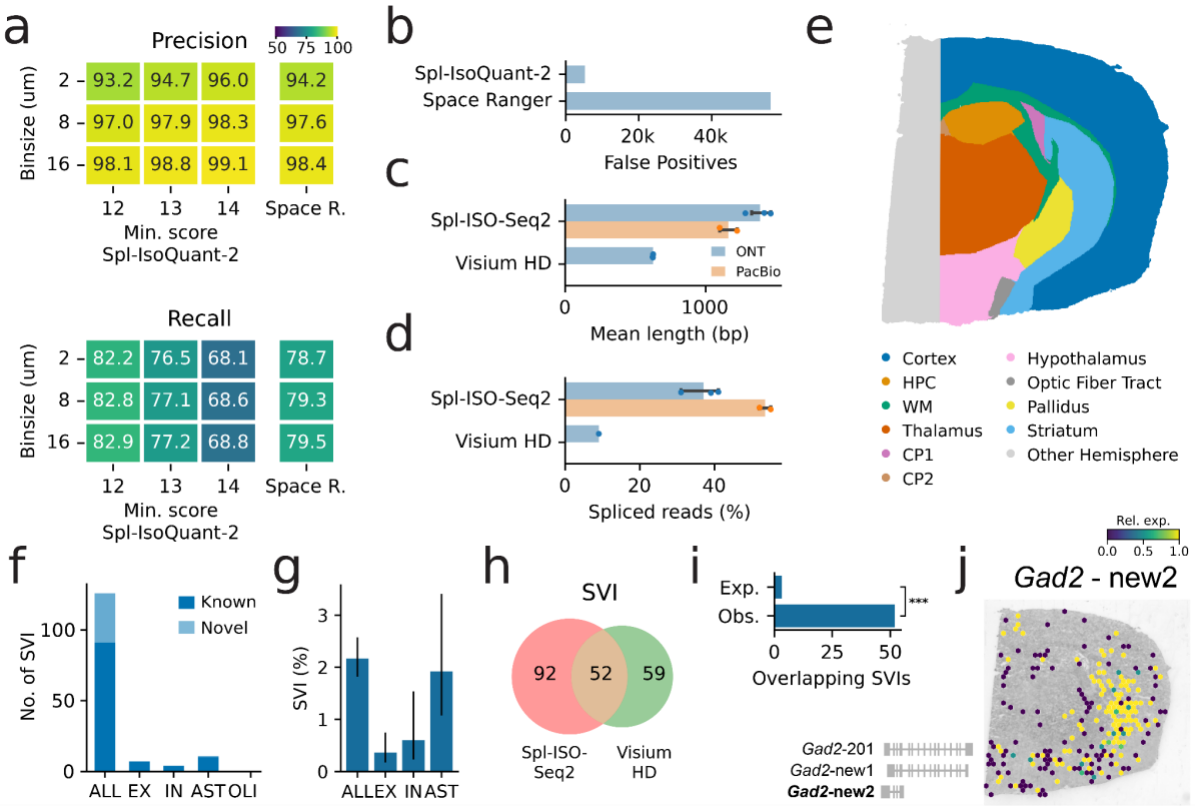

Figure 5 a) Barcode calling precision and recall of Spl-IsoQuant-2 and Space Ranger on simulated Visium HD data using different while grouping the spots in different bin sizes. Different minimal alignment scores (min. score) were tested for Spl-IsoQuant-2. **b)** Number of bins (most likely false positives) that each method uniquely assigned reads to in the Visium HD data. **c)** Average read length and **d)** percentage of spliced reads in the Spl-ISO-Seq2 and Visium HD data. **e)** Region annotation of the Visium HD sample. **f)** Number and **g)** percentage of significant spatially variable isoforms (FDR < 0.05). **h)** Overlap of significant SVIs between Spl-ISO-Seq2 Sample1+2 AE and Visium HD data. Only isoforms tested in both datasets are considered. **i)** Observed and expected overlap of SVIs (p = 2.29e-53, two-sided Fisher's exact test). **j)** Relative expression of *Gad2*-new2. Every hexagon shows the mean relative expression of the underlying cells. Relative expression indicates the fraction of long reads for that gene (*Gad2*) that are assigned to our isoform of interest (new2). A hexagon is only plotted when there are underlying cells expressing *Gad2* (Methods). The CytAssist image is shown in the background. The schematic diagram in the corner shows the gene structure of *Gad2* isoforms. HPC = hippocampus, WM = white matter, SVI = spatially variable isoforms, ALL = all cells, EX = excitatory neurons, IN = inhibitory neurons, OLI = oligodendrocytes, AST = astrocytes.

Discussion

Here, we present Spl-ISO-Seq2, and the accompanying software packages Spl-IsoQuant-2 and Spl-IsoFind, which allows studying spatially variable isoforms (SVIs) at single-cell and thus a cell-type-specific resolution for the first time. Spl-ISO-Seq2 has spot sizes of 220nm and inter-center distances of 500nm, which is far smaller than any known cells in the mouse brain. Thus, the vast majority of long reads can be attributed to a single cell, even achieving single-cell resolution for smaller cell types such as oligodendrocytes, which are thought to be of 6-8µm in the mouse brain[55,56]. As such we can pinpoint isoforms that show cell-type-specific spatial variation, some of which have not been shown before through spatial transcriptomics. Some notable examples of cell-type specific SVIs include *Rps24*, with regional isoform regulation specific to oligodendrocytes, and *Snap25* for excitatory neurons.

Furthermore, our developed software, Spl-IsoQuant-2 and Spl-IsoFind, is versatile and can be applied to data generated using different protocols such as long-read Visium HD data. Even though the Visium HD data was generated from a more anterior coronal cut, many spatially variable isoforms were detected in both datasets. Compared to our Spl-ISO-Seq2 data, however, the Visium HD data contained shorter en fewer spliced reads. In the future, the long-read Visium HD protocol could be enhanced by combining it with our exome enrichment probes and long-molecule selection step.

Previously, we and others showed that two mutually exclusive exons of *Snap25* exhibit spatially defined inclusion[4,19]. However, these initial spatial long-read approaches were based on the 10X Genomics' Visium technology. Due to spot sizes of ~60µm, which exceed the size of most cells in mouse and human brain, we could not conclusively determine the cell-type contribution to these spatially variable isoforms. With our more recent approach of Spl-ISO-Seq[20], using the 10µm spots of Slide-SeqV2[12], we achieved a single-cell resolution for many human excitatory neurons. However, mouse brain cells are usually much smaller, and thus, single-cell resolution requires smaller spot sizes. Therefore, we devised Spl-ISO-Seq2 based on StereoSeq technology with 220 nm spots and single-cell resolution, which revealed that *Snap25* region-specific isoform expression is indeed observed when only considering excitatory-neuron spots.

Moreover, we complement our new wet lab and barcoding methods with a new tool, Spl-IsoFind, to detect splicing regulation that does not conform with well-established anatomical regions. Using our previously developed 2xn isoform tests[4] we could easily test for genes exhibiting differentially used isoforms between predefined brain regions such as cortex vs. hippocampus, or between cortical layers. However, many other spatially-derived isoform patterns may not conform to known brain structures. Here, we implement a region-agnostic detection method, based on Moran's I[33] and contrast it to genes detected by comparing pre-defined brain regions. Using predefined regions increases the number of overall detected events in comparison to the region-agnostic approach. On the other hand, the region-agnostic approach detects events that were not observable with the predefined-regions approach. Indeed, in addition to the previously detected cortical pattern of *Snap25*, this gene also shows a switch in exon inclusion in a restricted area of the midbrain – a region that gave no prior indication of being of specific interest for splicing investigation. Thus, the region-agnostic approach has value in terms of finding isoform regulation specific to unknown areas.

A current limitation of Spl-IsoFind is its low detection rate, which could be improved using a higher read depth. For example, we discovered more than three times as many SVIs in Sample 1, where we used two ONT flow cells, compared to Sample 2, which was sequenced using

477 only one. When we combined both samples, the number of significant hits increased further,
478 with no signs of saturation, suggesting that many SVI patterns remain currently undetected.
479 A higher read depth would also increase the number of cell-type-specific SVIs. This could be
480 further improved by decreasing the number of doublets in the data, thereby increasing the
481 number of cells per cell type. However, due to 10µm slices, it is not always possible to
482 completely exclude the presence of long reads from two distinct cells at the same cartesian
483 x,y coordinate but varying in the "z" direction. Glial cells especially often form doublets[23], which
484 may explain our low number of singlets for oligodendrocytes and astrocytes and the
485 corresponding low number of detected SVIs.
486 Furthermore, cell segmentation is currently based on nuclear staining only, expanding each
487 segmented nucleus outward to a set radius or until it touched a neighboring cell. Because cells
488 vary in size and morphology, this approach has clear limitations. Using additional stains to
489 mark full cell boundaries could improve segmentation accuracy and thus reduce the number
490 of doublets in future work.
491 In summary, our wet lab and dry lab methods increase the resolution of spatial and single-cell
492 isoform research. By combining high-resolution spatial data with isoform detection at the cell-
493 type level, our approach represents a crucial step forward in understanding the spatially
494 regulated mechanisms of splicing in complex tissues at a real single-cell resolution.

Methods

Ethics Statement

All experiments were conducted in accordance with relevant NIH guidelines and regulations, related to the Care and Use of Laboratory Animals tissue. Animal procedures were performed according to protocols approved by the Research Animal Resource Center at Weill Cornell College of Medicine.

Animals and Tissue Processing

1 male C57BL P56 mouse was anesthetized with isoflurane and sacrificed via decapitation following IACUC protocols for Stereo-seq and subsequent sequencing. The brain was quickly removed, embedded in a cryocube with Tissue-Tek OCT(VWR 25608-930) and frozen at -80C. The brain was cut at 10um thick using a Dakewe 6250 Cryostat Microtome and immediately mounted onto the Stereo-seq Chip Slide.

Stereo-seq Protocol

After tissue sections (10um) were adhered to the Stereo-seq chip surface they were incubated at 37C for 5 minutes. Then, the sections were fixed in methanol and incubated for 30 minutes at -20C. Subsequently, the same sections were stained with nucleic acid dye (Thermo fisher, Q10212) and imaging was performed with a EVOS m7000 microscope prior to in situ capture at the channel of GFP. Tissue permeabilization, tissue removal, reverse transcription, and cDNA release was then performed based on Stereo-seq transcriptomics set for chip-on-a-slide user manual. cDNA was then amplified for 15 cycles and cleaned up using SPRI beads (Beckman Coulter, California, USA). cDNA quality was assessed by TapeStation High Sensitivity D5000 (Agilent Technologies Inc., California, USA) and quantified by Qubit™ dsDNA Quantification Kit (Thermofisher, Massachusetts, USA). Final library was constructed based on the manufacturer's protocol. The quality was then assessed by TapeStation High Sensitivity D1000 (Agilent Technologies Inc., California, USA) and quantified by Qubit™ dsDNA Quantification Kit.

DNB Sequencing

Short-read sequencing of final libraries was performed on the DNBSEQ-T7RS using the DNBSEQ-T7RS Stereo-seq Visualization Reagent Set (T7 STO FCL PE100) with a configuration of Read1: 50 (26-40 dark) cycles, Read2: 100 cycles, Barcode: 10 cycles.

Linear PCR of cDNA

We first performed a linear/asymmetric PCR on the cDNA aliquoted for long-read experiments with the purpose of appending additional basepairs to the 3' end of reads in order to avoid any TSO-TSO artifacts. The first asymmetric PCR used primer 5'-CTTCCGATCTATGGCGACCTTATCAG-3' with the following settings: 95 °C for 3 min, 12 cycles of 98 °C for 20 s, 64°C for 30 s and 72 °C for 60 s. The resulting cDNA product was cleaned up with 0.8× SPRIselect beads (Beckman Coulter, B23318) and washed with 80% ethanol twice. This cDNA was then inputted into a symmetric PCR again using primer 5'-CTTCCGATCTATGGCGACCTTATCAG-3' and TSO primer 5'-CTGCTGACGTACTGAGAGGC-3' with 95 °C for 3 min, 8 cycles of 98 °C for 20 s, 64°C for

535 30 s and 72 °C for 60 s. Amplified cDNA was again purified with 0.8× SPRIselect beads
536 (Beckman Coulter, B23318) and washed with 80% ethanol twice.

537 Exome Enrichment of cDNA

538 500 ng of the cDNA collected from the previous step was used as input for the exome
539 enrichment. Exome capture was performed as previously described to selectively amplify
540 spliced molecules[5]. Agilent probe kit SureSelectXT Mouse All Exon (Agilent, 5190-4641) and
541 the reagent kit SureSelectXT Mouse All Exon library (Agilent, G9611A) were used according
542 to the manufacturer's manual. First, the block oligonucleotide mix was made by combining 1
543 ul of LPCR primer (5'-CTTCCGATCTATGGCGACCTTATCAG-3') and 1 ul of TSO primer (5'-
544 CTGCTGACGTACTGAGAGGC-3') at 200 ng/ul. We next combined 5 ul of 50-100 ng/ul of
545 linearly amplified cDNA (see section above) and diluted it with 2 ul blockmix and 2 ul nuclease
546 free water (NEB, AM9937). This mix was incubated on a thermocycler at 95 °C for 5 min, 65 °C
547 for 5 min and 65 °C on hold. Next, 20 ul of SureSelect Hyb1, 0.8 ul of SureSelect Hyb2, 8.0 ul
548 of SureSelect Hyb3 and 10.4 ul of SureSelect Hyb4 were combined at room temperature to
549 make hybridization mix. Once the cDNA block mix reached to 65 °C on hold, 5 µl of probe mix,
550 1.5 µl of nuclease-free water, 0.5 µl of 1:4 diluted RNase Block and 13 µl of the hybridization
551 mix were added to the cDNA block oligo mix and incubated for 24 h at 65 °C. M-270
552 Streptavidin Dynabeads (Thermo Fisher Scientific, 65305) were washed three times and
553 resuspended with 200 µl of binding buffer after 24 hours of incubation. The incubated cDNA-
554 block mix was mixed with the M270 Dynabeads and placed on a Hula mixer for 30 minutes.
555 After this incubation time, the bead buffer was replaced with 200 µl of wash buffer 1 (WB1)
556 followed by another 15-min incubation with low speed. Next, the WB1 was replaced with WB2,
557 and the tube was transferred to the thermocycler for the next round of incubation (10 minutes).
558 WB2 was replaced 2 more times at 10 minute intervals with the pre-heated WB2. Following
559 this last incubation, beads were resuspended in 18 ul nuclease free H20 and then the spliced
560 cDNA, which is attached to the beads, was amplified as described below.

561 4K Gene Set Enrichment of cDNA

562 We also enriched linear-PCR amplified cDNA with a custom gene set of 3,378 mouse ortholog
563 genes related to various brain diseases and synaptic function, as we have published
564 previously (**Table S26**)[10]. All experimental procedures remained the same as the exome
565 enrichment and the "SureSelectXT Mouse All Exon" probes were replaced with the custom
566 set (Agilent Part No. 5191-6900).

567 Long-molecule Capture and Amplification

568 The long-molecule selection of spliced cDNA was performed as previously published[20].
569 Spliced cDNA bound with the M-270 Dynabeads was amplified with primers (LPCR: 5'-
570 CTTCCGATCTATGGCGACCTTATCAG-3'; TSO: 5'-CTGCTGACGTACTGAGAGGC-3') and
571 KAPA-HiFi enzyme by using the following PCR protocol: 95°C for 3 min, 8 cycles of 98°C for
572 20 s 64°C for 60 s and 72°C for 3 min. The amplified cDNA was isolated from M-270 beads
573 as supernatant and followed by a size selection/purification with 0.48× SPRIselect beads (in
574 1.25M NaCl-20% PEG buffer) and then eluted in 25ul EB buffer. All 25 ul size-selected spliced
575 cDNA was used as template for the second round amplification of 6-8 cycles with the same
576 PCR conditions suggested above. The product of the second round PCR was size

577 selected/purified with 0.48× SPRIselect beads (in 1.25M NaCl-20% PEG buffer) and then
578 eluted in 50 ul EB buffer.

579 Oxford Nanopore Library Preparation

580 For the 2 samples, ~75 fmol exome enriched and long-molecule selected cDNA was used as
581 input into the ONT Ligation Sequencing Kit (ONT, SQK-LSK114), and Sample 1 (AE) was
582 sequenced twice. Library construction was performed according to the manufacturer's
583 protocol (Nanopore Protocol, Amplicons by Ligation, version
584 ACDE_9163_v114_revO_29Nov2022). The ONT library was loaded onto a PromethION
585 sequencer by using PromethION Flow Cell (ONT, FLO-PR114M) and sequenced for 72 h.
586 Base-calling was performed with high-accuracy settings on the Oxford Nanopore MinKNOWUI
587 dorado basecaller.

588 PacBio Library Preparation

589 Exome enriched and long-molecule selected cDNA was also sequenced with PacBio using
590 the PacBio Mas-Seq for 10X 3' Concatenation kit (102-659-600). As this kit was meant for
591 cDNA coming from the 10X Chromium Next GEM Single Cell 3' kit, we performed an initial
592 PCR such that the correct sequences would be added to the ends of each molecule such that
593 the 3' and 5' ends resembled those from 10X (10x_PR1_Stereo: 5'-
594 CTACACGACGCTCTTCCGATCTCTTCCGATCTATGGCGACCTTATCAG-3';
595 10x_PR2_Stereo: 5'-AAGCAGTGGTATCAACGCAGAGCTGCTGACGTACTGAGAGGC-3').
596 The PacBio Mas-Seq for 10X 3' Concatenation kit was used following PacBio's published
597 protocol. The resulting library was sequenced using the PacBio Revio using 1 SMRT cell per
598 sample. Reads were split by identifying known read-linking sequences using skera (version
599 1.4.0).

600 Tissue isolation and reverse transcription for exon validation

601 To validate spatially variable isoforms found by comparing two regions, we isolated the
602 midbrain region from the remaining portion of the brain from three P56 mice. RNA was isolated
603 from 30-50mg of these regions using the Qiagen miRNA Tissue/Cells Advanced Mini Kit
604 (Qiagen Cat. no. 217684) following the established protocol. RNA was frozen at -80C. Next,
605 we performed reverse transcription using the SuperScript III First-strand Synthesis System
606 (Invitrogen, 18080051) for RT-PCR kit following the manufacturer's protocol. We combined
607 5uL RNA with 0.5uL random hexamers, 0.5uL oligodT, 1uL dNPTs, and 3uL H20. This mix
608 was incubated at 65C for 5 minutes followed by being placed on ice for 1 minute. We next
609 added 2 uL 10x Buffer, 4 uL MgCl2, 2uL DTT, 1uL RNAse Out, and 1 uL SuperScriptIII to the
610 samples and incubated at 25C for 10 minutes, 50C for 50 minutes, and 85C for 5 minutes.
611 The cDNA was then stored at -20 until further use.
612

613 PCR exon validation

614 cDNA was amplified using target primers (see below) using the following settings: 95 °C for
615 3 min, 26 cycles of 98 °C for 20 s, 64°C for 30 s and 72 °C for 60 s. Resulting cDNA was
616 inputted onto an Agilent Tapestation System 4200 using D1000 ScreenTapes following the
617 manufacturer's protocol.

Primers used for QPCR Exon Validation

Nptn
5'-CTCTTGCTGGTCTCTGGCTC-3' and 5'-GTGGCAGTGAGTTCTACCCC-3'
Phactr1
5'-CTGTCGGAAATGGAGCCTGT-3' and 5'-CGGAGGCGACAGTTTCACT-3'
Gapdh
ReadyMade_GAPDH_Fwd (IDT catalog #51-01-07-12): 5'-ACCACAGTCCATGCCATCAC-3'
ReadyMade_GAPDH_Rvs (IDT catalog #51-01-07-13): 5'-TCCACCACCCTGTTGCTGTA-3'

Tapestation Gel Quantifications

Tapestation gel images were saved and quantified in FIJI using established gel quantification methods[57]. Values per sample were normalized to that sample's Gapdh levels. Exon inclusion values were calculated using $\frac{Normalized\ large\ band}{Normalized\ large\ band + Normalized\ small\ band}$ per sample and brain region. For *Nptn*, normalized exon inclusion = 539 base pair (bp) band/(191 bp+539 bp band). For *Phactr1*, normalized exon inclusion = 345 base pair (bp) band/(345 bp+138 bp band).

Long-read processing

For upstream analysis of PacBio and ONT reads we implemented Spl-IsoQuant-2 — a new version of the Spl-IsoQuant pipeline (based on IsoQuant[24]). Spl-IsoQuant-2 is capable of processing long-read data obtained with the Stereo-seq protocol. It starts with splitting reads into individual cDNAs, further identifying barcode and UMI sequences in every cDNA sequence. Then, these cDNAs are mapped to the reference genome with minimap2[58] in spliced mode and assigned to known isoforms and genes using existing IsoQuant algorithms. Finally, Spl-IsoQuant removes PCR duplicates and performs gene and transcript quantification. Below we provide details on every step of the pipeline.

To account for unannotated isoforms, in this work we first ran standard bulk IsoQuant pipeline using Gencode version 36 mouse basic reference annotation and combination of all Stereo-seq long-read data (3 ONT and 2 PacBio samples, 391 million reads in total). Default settings specific for ONT data were chosen to avoid potential false positives. The generated GTF, which contained 28'891 known and 26'445 novel isoforms, was further provided into the Spl-IsoQuant-2 pipeline as the reference annotation.

Stereo-seq barcode detection

Due to similarity of primers at 3' and 5' ends Stereo-seq protocol multiple barcoded cDNAs may be concatenated. To account for this fact, Spl-IsoQuant-2 implements an algorithm that is capable of detecting multiple barcoded cDNA within a single read. The algorithm first identifies the poly(T) sequence, the 3' primer (STOmics sequencing primer in Stereo-seq notation) and the linker (fixed sequence between the barcode and the UMI sequence, see protocol). Further, it detects the closest TSO primer following the detected poly(T) stretch (**Fig 2b**) and repeats this procedure iteratively for the remaining part of the read. Iterative primer detection is then repeated for the reverse-complemented read sequences, thus allowing for reverse concatenation. The resulting set of non-overlapping sequences located between 3' primer and 5' TSO are outputted as individual cDNA sequences.

660 These individual subsequences are then used for detecting barcode and UMI sequences
661 (CIDs and MIDs in Stereo-seq notation respectively) (**Fig 2c**), as well as the remaining pipeline
662 steps, such as read alignment and feature quantification. Read barcode sequence (located
663 between the 3' primer and the linker, see **Fig 2c**) is matched against the set of barcodes from
664 the input whitelist. Only cDNA with barcodes with the alignment score above a certain
665 threshold S_{min} are reported (see details below); cDNAs with non-matching barcodes are
666 rejected. UMI sequence is reported as a read subsequence located between the linker and
667 the start of the poly(T) stretch.
668
669 Here we exploit the same pattern matching as in the previous version of Spl-IsoQuant[20], which
670 is based on short k-mer indexing and conventional Smith-Waterman local alignment (via the
671 ssw-py library[59]). To efficiently detect sequences from a given set (e.g. barcodes) or a fixed
672 sequence (e.g. linker or primer), we use k-mer indexing: a hash table where each key is a k-
673 mer and the value is the set of known sequences containing it. Given a read, we extract all of
674 its k-mers and use the index to retrieve potential matches. To "parse" the read according to
675 the molecule structure, we exploit three separate indices: for the TSO primer (23 bp, k=8), the
676 3' primer (26 bp, k=6) and the linker (15 bp, k=5). These indices are used to quickly detect
677 potential positions of a fixed sequence in a read, which are then verified via Smith-Waterman
678 local alignment.
679
680 Once the positions of the primer and the linker are identified, the barcode sequence is
681 extracted (between the primer and the linker) and potential matches are identified among the
682 known barcodes using a k-mer index constructed for the input barcode whitelist (25 bp, k=14).
683 While smaller k-mer length increases sensitivity for detecting the correct known barcode even
684 in the presence of sequencing errors, it also raises the chance of spurious matches, especially
685 in large barcode sets (>400 million for Stereo-seq protocol). Therefore, k-mer matching serves
686 only to shortlist potential matches. Top N (N=10 by default) potential matches with the highest
687 number of shared k-mers are validated using Smith-Waterman local alignment. Extracted
688 barcode sequence is required to have an alignment score at least S_{min} (22 by default, classic
689 alignment scoring system: +1 for match, -1 for indel/mismatch) with the closest candidate from
690 the whitelist. If two or more known barcodes have the same highest alignment score or all
691 potential matches have score below S_{min}, no barcode is reported for this cDNA. Thus, the k-
692 mer index reduces the search space and limits the number of costly alignment operations
693 needed for barcode identification, while the alignment itself helps to filter out possible false
694 positive matches and uncertain candidates.

695 Benchmarking Stereo-seq barcode detection

696 To evaluate the barcode calling algorithm, we simulated Nanopore data using a modified
697 version of Trans-NanoSim[60], available at https://github.com/andrewprzh/lrgasp-simulation.
698 This version improves read truncation procedures, enabling realistic ONT truncation
699 probabilities and non-truncated reads, and was previously used for IsoQuant benchmarking[24].
700 The error model was trained with actual sequencing data from this project (ONT sequencing,
701 Sample 2), showing a sequencing error rate of 3.8% (1.5% mismatches, 1.1% insertions, 1.2%
702 deletions), ensuring accurate recall and precision estimates. No read truncation was induced
703 during the simulation process.
704

705 In this work we assessed the barcode calling method using two simulated datasets: (1) the
706 one where each read contains a single cDNA sequence, and (2) another one where each read
707 can contain from 1 to 4 concatenated cDNAs. Molecule templates were created using
708 Gencode reference transcript sequences according to the Stereo-seq molecule structure (**Fig**
709 **2c**). Reference transcripts were randomly selected based on the expression profile derived
710 from the Sample 2. For the second dataset, we randomly concatenated between 1 and 4 cDNA
711 templates. The probability of each cDNA benign reverse-complemented prior to concatenation
712 is 0.4. While these parameters might not exactly match the ones for real data, the resulting
713 simulated dataset contained a substantial fraction of concatenated reads similar to the reads
714 observed in real data, thus allowing to accurately assess the algorithm performance. In both
715 experiments we simulated 10 million reads.
716
717 We ran Spl-IsoQuant-2 on the simulated datasets and compared the detected barcodes
718 against the true barcodes attached during simulation. Correctly identified barcodes were
719 considered true positives, incorrectly reported barcodes as false positives, and undetected
720 barcodes as false negatives. For reads containing multiple cDNAs, each cDNA was counted
721 individually (**Fig 2e**).
722
723 We also tried running Flexiplex on our data as it allows to specify custom molecule structure
724 with user-defined primers. Unfortunately, after running for 5 days in 16 threads on a dataset
725 containing just 1 million Nanopore reads it was able to detect less than 20 thousand barcodes
726 (<2% of reads with detected barcodes). Similarly, it was not able to provide any reasonable
727 result for simulated data (2 detected barcodes for the entire dataset). In contrast, our algorithm
728 processed a full dataset with 120 million real ONT reads in under 11 hours with 8 threads
729 detecting the barcode in ~55% of reads.

730 Read processing and UMI deduplication

731 Overall, the Spl-IsoQuant-2 pipeline is similar to its previous version. Individual cDNA
732 sequences extracted from raw reads are mapped to the reference genome using minimap2
733 with IsoQuant default settings (minimap2 -x splice --junc-bed annotation.bed -a --MD -k 14 -Y
734 --secondary=yes). Spl-IsoQuant then assigns these mapped reads to reference genes and
735 isoforms using the IsoQuant standard algorithm[20]. Multimapped reads are resolved by
736 selecting the most consistent alignment with the reference annotation or discarding them.
737 Reads uniquely assigned to a known gene undergo further processing to remove PCR
738 duplicates.
739
740 To eliminate PCR duplicates, reads are grouped by gene and barcode. Within each group,
741 reads with identical or highly similar UMI sequences (edit distance ≤ 4) are considered
742 duplicates. We use an iterative algorithm to identify and assign reads to representative UMIs,
743 selecting the read covering the most splice junctions as the representative and ignoring others.
744 These UMI-filtered reads are then used for downstream analyses, such as exon and transcript
745 quantification and differential splicing analysis. While this method may occasionally remove
746 non-duplicate reads due to similar UMIs, the randomness of UMI sequences minimizes this
747 impact on the analysis.

Comparison PacBio and ONT isoform assignment

For reads with a barcode-UMI-gene combinations sequenced with both PacBio and ONT, we compared the assigned isoforms by Spl-ISO-Seq2 only when both platforms assigned an isoform.

Cell segmentation and barcode to cell assignment

Cells were segmented based on the ssDNA staining and short-read gene expression using the SAW[61] pipeline (version 7). This generates several outputs including a masking image and a GEF (gene expression file) that contains the CellID for each x and y coordinate combination. However, this GEF only includes coordinates that have at least one short-read assigned. If we used this file directly to map coordinates (and thus barcodes) to the cells, we would lose the coordinates that only have long-reads but no short-reads assigned. To overcome this, we generated an artificial GEM (gene expression matrix) with one short-read for each x and y coordinate. We then converted this artificial GEM into a new cell-binned GEF using the previously obtained masking image using the SAW package. This new GEF includes the CellID for all coordinates on the chip. Barcodes are assigned to coordinates using the Stereo-seq mask files.

(Sub)region annotation

Regions and Subregions were annotated manually by using a combination of gene expression markers, clustering analyses, cell type distribution, and known brain region architecture. Regions based on these criteria were outlined using the Polygon function, turned into masks, and uploaded such that all spots within the masked region were annotated as so.

Cell type identification

Gene expression matrices were inputted into Robust Cell Type Decomposition (V2.2.1)[37] program with a reference[9] file of including 6 P56 mouse single cell samples from multiple brain regions including hippocampus, visual cortex, striatum, and thalamus. Previously annotated "doublets" from the single cell data were removed prior to analysis. Basic RCTD parameters were used in "doublet" mode following published tutorials (https://github.com/dmcable/spacexr/blob/master/vignettes/spatial-transcriptomics.Rmd). Spots designated as "singlets" were assigned as cell types per sample.

Slide alignment

We aligned Sample 1 and Sample 2 using the `spateo.align.morpho_align()` function from the spateo[28] package (version 1.1.0) with parameters beta=0.1 and lambdaVF=1.0. This greatly improved the alignment compared to naively overlapping the slides (**Fig 1d, S22**)

Spatially variable isoform tests using predefined regions

We used the isoform assignments from Spl-ISO-Seq2 to detect spatially variable isoforms instead of using full-length reads only. For each region of interest, we counted the number of unique reads per isoform. Afterwards, we used the `DiffSplicingAnalysis()` function from the ScisorSeqR[4] package (version 0.1.9) to identify genes that exhibit differential isoform usage between the predefined brain regions using the default settings (minNumReads = 25,

numIsoforms = 10). Using these default settings, genes with less than 25 reads are filtered out. NumIsoforms means that we tested a maximum of 10 isoforms per gene and grouped the remaining isoforms in a rest group, which results in an 11x2 count matrix at most. The returned p-values are calculated using a chi-squared test and corrected for multiple testing using Benjamini Hochberg correction. Genes with FDR <= 0.05 and delta Pi >= 0.1 are considered significant. Delta Pi was calculated for each gene as the sum of change in percent isoform from the top two isoforms detected. We tested all pairwise combinations for the general brain regions, hippocampal subregions, and cortical layers using all cells and per cell type.

Differential exon testing using predefined regions

We called differentially expressed exons as we have previously[10,20] done. After executing SpliceIsoQuant on the long-read sequencing data, which mapped, barcoded, and spliced reads while correcting for UMIs, we generated an AllInfo file. This file contains detailed metadata for each read, including: read identifier, gene name, comparison group (such as specific cell type or brain region), barcode, UMI, intron chain, transcription start site (TSS), polyadenylation site, exon chain, transcript classification (known or novel), and intron count. You can find further information on AllInfo files at: https://github.com/noush-joglekar/scisorseqr [4].
AllInfo files were split based on the specific comparisons being analyzed. These were analyzed using the casesVcontrols() function in the scisorATAC R package (https://github.com/careenfoord/scisorATAC)[10], with a threshold of 10 minimum reads per exon, to identify differentially included exons using standard parameters.
The process to detect alternative exons involved the following steps:

1. Read Selection
   We used reads from both comparison groups (e.g., cortex barcodes vs. midbrain barcodes) that contain introns and adhere to canonical splice site rules.
2. Alternative Exon Identification
   For each candidate exon, we tallied:
   A: Reads supporting both splice junctions of the exon
   B: Reads supporting the left splice site, ending within the exon on the right
   C: Reads supporting the right splice site, starting within the exon on the left
   D: Reads skipping the exon (but assigned to the same gene)
   E: Reads overlapping the exon in any fashion
3. Filtering Criteria
   - Exclude exons where A + B + C + D < 10, due to insufficient read coverage
   - Exclude exons where (A + B + C + D)/(A + B + C + D + E) < 0.8, indicating insufficient clarity between inclusion and exclusion
   - Calculate PSI (Percent Spliced In):
     - leftPSI = (A + B) / (A + B + D)
     - rightPSI = (A + C) / (A + C + D)
     - Keep exons where both PSI values are within [0.05, 0.95]
4. Cluster-Specific PSI Calculation
   For exons that passed all prior filters:
   - Calculate group-specific PSI = (A + B + C) / (A + B + C + D), only if A + B + C + D ≥ 10
   - Construct a 2x2 contingency table using exon inclusion and skipping counts for each group (e.g., layer 4 old vs. young)

832 ○ Retain tables where at least 3 of the 4 cells have expected values ≥ 5

833 5. Statistical Testing
834 Each retained exon is tested for association between group identity and exon inclusion
835 using a two-sided Fisher's exact test. Resulting p-values are adjusted for multiple
836 comparisons via the Benjamini-Yekutieli correction.

837 Spatially variable isoform tests using Moran's I

838 We used the `esda.Moran` function from the ESDA[62] python package (version 2.4.3) to
839 calculate Moran's I and the corresponding p-value using a one-tailed permutation test based
840 on 1e5 permutations. The input data for this analysis consists of the relative expression of the
841 tested isoform in every cell. For the spatial weights, we focused only on cells that express the
842 corresponding gene in the long-read data. The weight for the 10 nearest neighbors of each
843 cell was set to 1, while all other neighbors were assigned a weight of 0.
844 To reduce the number of tested isoforms, we only included isoforms that met the following
845 criteria:
846 - The corresponding gene is expressed in at least 50 cells in the long-read data.
847 - At least 20 cells express the tested isoform and at least 20 cells express another
848 isoform (to ensure some minimum expression of the isoform of interest and variance
849 in the data).
850 - The fraction of cells expressing the tested isoform should be between 0.05 and 0.95
851 (to ensure some variance in the data).
852 The p-values are corrected for multiple testing using the Benjamini-Yekutieli method. Isoforms
853 are considered significant if FDR <= 0.05.

854 Using simulations to assess Moran's I precision and recall

855 For all simulations, we used the original coordinates of the cells from Sample 1. To assess
856 precision, we generated random data by assigning values of 0 and 1 randomly to all cells. We
857 simulated different imbalances in the distribution of 0s and 1s (50-50, 40-60, 30-70, 20-80, 10-
858 90, 5-95). To evaluate precision, we simulated relative expression values based on four
859 patterns:

860 1. *Frontiers*: A clear border is simulated between cells with a value of 1 and cells with a
861 value of 0.
862 2. *Square*: A square region within the tissue is filled with 1s, with surrounding cells
863 assigned 0. Different square sizes (200, 500, 1000, 2000 μm) were tested.
864 3. *Square - surrounded by random points*: Similar to the square pattern, but the square
865 is surrounded by cells with random values (either 0 or 1).
866 4. *Line - surrounded by random points*: Instead of a square, a horizontal or vertical line
867 of 1s is simulated, surrounded by random points on both sides. Different line widths
868 (100, 250, 500 μm) were tested.

869 Each scenario was simulated 100 times and downsampled to different cell numbers (50, 100,
870 500, 1000, 5000, 10000 cells). After downsampling, Moran's I and the corresponding p-value
871 were calculated as described earlier. If downsampling resulted in violating any of the
872 predefined criteria for Moran's I (e.g., fewer than 20 cells per category), we did not calculate
873 the Moran's I score or p-value for those cases.

Cell-type constrained permutations

By default, we permute all cells randomly, which means that the relative expression values of each cell are shuffled across all cells, regardless of their assigned cell type. To investigate which SVIs are caused by differences in cell-type abundances, we performed cell-type constrained permutations. In this approach, only cells with assigned cell-type labels could be used. We used cells that RCTD labeled as 'singlet' or 'doublet_certain'. For these doublets, we assigned a label based on the RCTD assigned doublet probabilities For singlets, we used the assigned cell-type labels, and for doublets, we included only those for which both cell types were clearly identified. During each permutation, the doublets were assigned a cell-type label with probabilities based on their assigned doublet probabilities. Next, we randomly shuffle the values of the cells within each specific cell type. Due to the increased computational time, we performed 1e4 permutations (1e4) instead of the usual 1e5.
In Figure 4e, we compared the p-values from this cell-type-constrained permutation method to those obtained using normal random permutations. Since the original p-values were calculated using all cells and 1e5 permutations, we recalculated the normal p-values as well using 1e4 permutations and only cells with an assigned cell type.

Spatial hexagon plots

To avoid overlapping cells or very small cells, we decided to plot spatial isoform expression using spatial hexagon plots. Every hexagon shows the mean relative expression of the underlying cells. Relative expression indicates the fraction of long reads for a gene that are assigned to our isoform of interest. As such, a hexagon is only plotted when there are underlying cells expressing the corresponding gene. A relative expression of 1 thus indicates that all underlying long reads express the isoform of interest, while a relative expression of 0 indicates that all underlying long reads express another isoform.

Visium HD 3' long-read data

We used publicly available Visium HD 3' long-read data from a P56 male mouse brain (https://epi2me.nanoporetech.com/visium_hd_2025.06/).

The algorithm for barcode detection for Visium HD 3' ONT data was implemented based on the Visium HD 3' molecule structure in a similar manner as the described Stereo-seq barcode calling method, but without read deconcatenation (see above). K-mer size of 7 was used for detecting potential positions of R1 primer (22 bp) and both barcode parts (14-16 bp each). For alignment-based barcode validation, the default minimal score S_{min}=13 was chosen for each of the barcode parts.

The implemented barcode finding algorithm was benchmarked and compared against Space Ranger on Visium HD 3' mouse simulated data. The data was generated using Trans-NanoSim[60] in a similar way as Stereo-seq simulated data (3.8% error rate, see above), but without cDNA concatenation (i.e. one read contains a single cDNA molecule). The reported barcodes were converted to Visium HD slide spot ids and assessed at three resolution levels: 2 um, 8 um and 16 um (**Fig 5a**).

A high-resolution microscope image and short-read sequencing data are not available for this dataset. As an alternative for cell segmentation, we used the 8µm bins and used those as a

proxy for cells. Next, we ran RCTD on long-read gene expression data. Afterwards, we applied Spl-IsoFind as described above.

Code availability

Spl-IsoQuant-2 is an open source software and is openly available at https://github.com/algbio/spl-IsoQuant.
Spl-IsoFind is available at https://github.com/tilgnerlab/Spl-IsoFind

Funding statement and acknowledgements

We thank Adrian Tan, Chendong Pan, Aihong Liu, Seongeun Oh and Jenny Xiang from the Genomics Resources Core Facility at Weill Cornell Medicine for performing RNA sequencing. We also thank Weill Cornell Medicine Scientific Computing Unit (SCU) for use of their computational resources.
Figure 1a, 4a, and S10 were created using BioRender.com.
Supported by: MIRA R35 GM152101-01 (H.U.T), Brain Initiative grant 1RF1MH121267-01 (H.U.T.), NIDA U01 DA053625-01 (H.U.T among others), NIDA grant 2T32DA039080 (J.H.), NSF GRFP # 2139291 (C.F.), the Feil Family Foundation (H.U.T.), European Union (European Research Council, SCALEBIO, 101169716. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them) (A.D.P, A.I.T.), Research Council of Finland grants No. 322595, 346968, 358744 (A.D.P, A.I.T.), NIGMS Maximizing Investigators' Research Award (MIRA) R35 GM138152 (I.H.).

Competing interests

H.U.T. has presented at user meetings of 10× Genomics, Oxford Nanopore Technologies, and Pacific Biosciences, which in some cases included payment for travel and accommodations. H.U.T. has recently agreed to consult for ISOgenix Ltd., for work unrelated to the present manuscript. The other authors declare no competing interests.

944 ## References

945 1. Gupta, I. *et al.* Single-cell isoform RNA sequencing characterizes isoforms in thousands
946 of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018).
947 2. Al'Khafaji, A. M. *et al.* High-throughput RNA isoform sequencing using programmed
948 cDNA concatenation. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-023-01815-7.
949 3. Lebrigand, K., Magnone, V., Barbry, P. & Waldmann, R. High throughput error corrected
950 Nanopore single cell transcriptome sequencing. *Nat. Commun.* **11**, 4025 (2020).
951 4. Joglekar, A. *et al.* A spatially resolved brain region- and cell type-specific isoform atlas
952 of the postnatal mouse brain. *Nat. Commun.* **12**, 463 (2021).
953 5. Hardwick, S. A. *et al.* Single-nuclei isoform RNA sequencing unlocks barcoded exon
954 connectivity in frozen brain tissue. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-
955 01231-3.
956 6. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution
957 using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).
958 7. Hagemann-Jensen, M., Ziegenhain, C. & Sandberg, R. Scalable single-cell RNA
959 sequencing from full transcripts with Smart-seq3xpress. *Nat. Biotechnol.* **40**, 1452–1457
960 (2022).
961 8. Salmen, F. *et al.* High-throughput total RNA sequencing in single cells using VASA-seq.
962 *Nat. Biotechnol.* **40**, 1780–1793 (2022).
963 9. Joglekar, A. *et al.* Single-cell long-read sequencing-based mapping reveals specialized
964 splicing patterns in developing and adult mouse and human brain. *Nature Neuroscience*
965 **27**, 1051–1063 (2024).
966 10. Hu, W. *et al.* Combined single-cell profiling of chromatin-transcriptome and splicing
967 across brain cell types, regions and disease state. *Nat. Biotechnol.* 1–13 (2025)
968 doi:10.1038/s41587-025-02734-5.
969 11. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide
970 expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

971 12. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution
972 with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
973 13. Liu, Y. *et al.* High-spatial-resolution multi-omics sequencing via deterministic barcoding
974 in tissue. *Cell* **183**, 1665–1681.e18 (2020).
975 14. Poovathingal, S. *et al.* Nova-ST: Nano-patterned ultra-dense platform for spatial
976 transcriptomics. *Cell Rep. Methods* **4**, 100831 (2024).
977 15. Schott, M. *et al.* Open-ST: High-resolution spatial transcriptomics in 3D. *Cell* **187**, 3953–
978 3972.e26 (2024).
979 16. Langlieb, J. *et al.* The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**,
980 333–342 (2023).
981 17. Liu, Y. *et al.* High-plex protein and whole transcriptome co-mapping at cellular resolution
982 with spatial CITE-seq. *Nat. Biotechnol.* **41**, 1405–1409 (2023).
983 18. Zhang, D. *et al.* Spatial epigenome-transcriptome co-profiling of mammalian tissues.
984 *Nature* **616**, 113–122 (2023).
985 19. Lebrigand, K. *et al.* The spatial landscape of gene expression isoforms in tissue
986 sections. *Nucleic Acids Res.* **51**, e47 (2023).
987 20. Foord, C. *et al.* A spatial long-read approach at near-single-cell resolution reveals
988 developmental regulation of splicing and polyadenylation sites in distinct cortical layers
989 and cell types. *Nat. Commun.* **16**, 8093 (2025).
990 21. Oliveira, M. F. de *et al.* High-definition spatial transcriptomic profiling of immune cell
991 populations in colorectal cancer. *Nat. Genet.* **57**, 1512–1523 (2025).
992 22. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA
993 nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
994 23. Tiesmeyer, S. *et al.* 2D, or not 2D? Investigating vertical signal integrity of tissue slices.
995 *bioRxiv* 2025.01.13.632601 (2025) doi:10.1101/2025.01.13.632601.
996 24. Prjibelski, A. D. *et al.* Accurate isoform discovery with IsoQuant using long reads. *Nat.*
997 *Biotechnol.* **41**, 915–918 (2023).
998 25. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods*

999 **19**, 171–178 (2022).
1000 26. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially
1001 variable genes. *Nat. Methods* **15**, 343–346 (2018).
1002 27. Marconato, L. *et al.* SpatialData: an open and universal data framework for spatial
1003 omics. *Nat. Methods* **22**, 58–62 (2025).
1004 28. Qiu, X. *et al.* Spatiotemporal modeling of molecular holograms. *Cell* **187**, 7351–
1005 7373.e61 (2024).
1006 29. Fang, S. *et al.* Stereopy: modeling comparative and spatiotemporal cellular
1007 heterogeneity via multi-sample spatial transcriptomics. *Nat. Commun.* **16**, 3741 (2025).
1008 30. Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial
1009 expression data. *Genome Biol.* **22**, 78 (2021).
1010 31. Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially
1011 resolved transcriptomic studies. *Nat. Methods* **17**, 193–200 (2020).
1012 32. Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and
1013 robust detection of spatial expression patterns for large spatial transcriptomic studies.
1014 *Genome Biol.* **22**, 184 (2021).
1015 33. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23
1016 (1950).
1017 34. Miller, B. F., Bambah-Mukku, D., Dulac, C., Zhuang, X. & Fan, J. Characterizing spatial
1018 gene expression heterogeneity in spatially resolved single-cell transcriptomic data with
1019 nonuniform cellular densities. *Genome Res.* **31**, 1843–1855 (2021).
1020 35. Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution.
1021 *Nature* **622**, 552–561 (2023).
1022 36. Ninh, V. K. *et al.* Spatially clustered type I interferon responses at injury borderzones.
1023 *Nature* **633**, 174–181 (2024).
1024 37. Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics.
1025 *Nat. Biotechnol.* **40**, 517–526 (2022).
1026 38. Tao, Q. *et al.* Benchmarking mapping algorithms for cell-type annotating in mouse brain

1027 by integrating single-nucleus RNA-seq and Stereo-seq data. *Brief. Bioinform.* **25**,
1028 bbae250 (2024).
1029 39. Li, B. *et al.* Benchmarking spatial and single-cell transcriptomics integration methods for
1030 transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670
1031 (2022).
1032 40. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res.* **49**, D916–D923 (2021).
1033 41. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status,
1034 taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45
1035 (2016).
1036 42. Dyer, S. C. *et al.* Ensembl 2025. *Nucleic Acids Res.* **53**, D948–D957 (2025).
1037 43. Su, Y. *et al.* Comprehensive assessment of mRNA isoform detection methods for long-
1038 read sequencing data. *Nat. Commun.* **15**, 3972 (2024).
1039 44. Pardo-Palacios, F. J. *et al.* Systematic assessment of long-read RNA-seq methods for
1040 transcript identification and quantification. *Nat. Methods* **21**, 1349–1363 (2024).
1041 45. You, Y. *et al.* Identification of cell barcodes from long-read single-cell RNA-seq with
1042 BLAZE. *Genome Biol.* **24**, 66 (2023).
1043 46. Cheng, O. *et al.* Flexiplex: a versatile demultiplexer and search tool for omics data.
1044 *Bioinformatics* **40**, btae102 (2024).
1045 47. Ebrahimi, G. *et al.* Fast and accurate matching of cellular barcodes across short-reads
1046 and long-reads of single-cell RNA-seq experiments. *iScience* **25**, 104530 (2022).
1047 48. Ilic, K., Mlinac-Jerkovic, K., Sedmak, G., Rosenzweig, I. & Kalanj-Bognar, S.
1048 Neuroplastin in human cognition: review of literature and future perspectives. *Transl.*
1049 *Psychiatry* **11**, 394 (2021).
1050 49. Langnaese, K., Beesley, P. W. & Gundelfinger, E. D. Synaptic membrane glycoproteins
1051 gp65 and gp55 are new members of the immunoglobulin superfamily. *J. Biol. Chem.*
1052 **272**, 821–827 (1997).
1053 50. Mouilleron, S., Wiezlak, M., O'Reilly, N., Treisman, R. & McDonald, N. Q. Structures of
1054 the Phactr1 RPEL domain and RPEL motif complexes with G-actin reveal the molecular

1055 basis for actin binding cooperativity. *Structure* **20**, 1960–1970 (2012).
1056 51. Wiezlak, M. *et al.* G-actin regulates the shuttling and PP1 binding of the RPEL protein
1057 Phactr1 to control actomyosin assembly. *J. Cell Sci.* **125**, 5860–5872 (2012).
1058 52. Pei, W. *et al.* Flip and flop: a molecular determinant for AMPA receptor channel
1059 opening. *Biochemistry* **48**, 3767–3777 (2009).
1060 53. Koike, M., Tsukada, S., Tsuzuki, K., Kijima, H. & Ozawa, S. Regulation of kinetic
1061 properties of GluR2 AMPA receptor channels by alternative splicing. *J. Neurosci.* **20**,
1062 2166–2174 (2000).
1063 54. Wen, W., Lin, C.-Y. & Niu, L. R/G editing in GluA2Rflop modulates the functional
1064 difference between GluA1 flip and flop variants in GluA1/2R heteromeric channels. *Sci.*
1065 *Rep.* **7**, 13654 (2017).
1066 55. Karasek, M., Swiltosławski, J. & Zielińiska, A. Ultrastructure of the central nervous
1067 system: the basics. *Folia Neuropathol.* **42 Suppl B**, 1–9 (2004).
1068 56. Edgar, J. M. & Griffiths, I. R. White Matter Structure. in *Diffusion MRI* 127–153 (Elsevier,
1069 2014). doi:10.1016/b978-0-12-396460-1.00007-x.
1070 57. Stael, S., Miller, L. P., Fernández-Fernández, Á. D. & Van Breusegem, F. Detection of
1071 damage-activated metacaspase activity by Western blot in plants. *Methods Mol. Biol.*
1072 **2447**, 127–137 (2022).
1073 58. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–
1074 3100 (2018).
1075 59. Zhao, M., Lee, W.-P., Garrison, E. P. & Marth, G. T. SSW library: an SIMD Smith-
1076 Waterman C/C++ library for use in genomic applications. *PLoS One* **8**, e82138 (2013).
1077 60. Hafezqorani, S. *et al.* Trans-NanoSim characterizes and simulates nanopore RNA-
1078 sequencing data. *Gigascience* **9**, giaa061 (2020).
1079 61. Gong, C. *et al.* SAW: an efficient and accurate data analysis workflow for Stereo-seq
1080 spatial transcriptomics. *GigaByte* **2024**, gigabyte111 (2024).
1081 62. Rey, S. J. & Anselin, L. PySAL: A python library of spatial analytical methods. in
1082 *Handbook of Applied Spatial Analysis* 175–193 (Springer Berlin Heidelberg, Berlin,

1083 Heidelberg, 2010). doi:10.1007/978-3-642-03647-7_11.