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 Prjibelski[5,*], Careen Foord[1,2,*], Wen Hu[1,2], Julien Jarroux[1,2], Justine Hsu[1,2], Alexandru Tomescu[5], Iman Hajirasouliha[3,4], Hagen 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 more common but lack nanometer- and therefore often single-cell resolution. This leaves the question unanswered whether spatially variable isoforms represent spatial variability within one cell type or differences in cell-type abundance between different regions. Here, we develop Spl-ISO-Seq2 with 220nm spot size and 500nm resolution, and the accompanying software packages Spl-IsoQuant-2 and Spl-IsoFind and apply it to the adult mouse brain. We compare spatial variability within a fixed cell type by examining (a) differential isoform abundance between known brain regions and (b) spatial isoform patterns that do not align with predefined regions. The former reveals larger numbers of spatial isoform differences, e.g. *Rps24* in oligodendrocytes. For the previously appreciated gene with spatially-variable isoforms *Snap25,* we can now show that this variability exists in excitatory neurons. However, the latter approach reveals patterns that the former cannot conceptually model, e.g., *Tnnc1* in excitatory neurons. 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

Single-cell and single-nucleus isoform long-read sequencing for thousands of distinct individual cells has enabled the isoform description of cell types at distinct cell-type-defining granularities[1–5]. Other short-read approaches also enabled the description of cell-type-specific splice junctions[6–8]. These approaches revealed exons that are used cell-type-specifically or altered in evolution and/or disease[5,9,10]. A shortcoming of these methods, however, is that they cannot reveal the location of the cells expressing a specific isoform.

Simultaneously, spatial gene-expression profiling has set the field forward by localizing cell types and gene expression events spatially[11–19]. Based on 10X Genomics' Visium technology, we and our colleagues simultaneously engineered spatial isoform sequencing revealing isoform switches correlating with brain structures and other isoforms (e.g. in *Snap25*) that did not correlate tightly with frontiers between brain structures[3,4]. Next, based on the Slide-SeqV2 method, we engineered spatial ISOform sequencing (Spl-ISO-Seq) with 10µm resolution and corresponding software Spl-ISOquant[20]. Here, we also used our previously developed exome-sequencing probes to remove entirely-intronic cDNAs[5], thus revealing much larger numbers of spliced cDNAs. We revealed that excitatory neurons located in layer 4 of the visual cortex show stronger splicing changes than those of other layers during human adolescent development. In the human brain some neuron types can reach diameters of up to 100µm, so this approach with 10µm resolution largely yields single-cell resolution for these neurons. For smaller cells (such as human microglia) and many mouse-brain cell types this approach however would not yield single-cell resolution.

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 >100 million barcodes. We show that this spatial isoform technology can work with PacBio as well with Oxford Nanopore long reads. Along with this wet lab approach, we devise Spl-Isoquant2, which adds highly-specific barcode calling to our IsoQuant algorithm[21], which we substantiate with using dry-lab simulation experiments. Furthermore, we show that unique molecules sequenced on both PacBio and Oxford Nanopore reveal high accuracy of transcript assignments.

A fundamental question in spatial isoform biology is whether the spatial position of a cell 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 for isoforms in spatial experiments can untangle spatially distributed isoforms from cell-type-derived ones. Here, we approach this question from two distinct angles. First, we use predefined anatomical landmarks to distinguish isoform production in distinct brain areas. While such isoforms are of specific interest, this cannot reveal isoform production that correlates without a priori-defined landmarks. For similar questions regarding gene expression, many analytical frameworks have been developed, e.g. for identifying spatially variable genes[22–29]. However, methods to detect spatially variable isoforms have lagged behind. To this extent, we develop Spatial Isoform Finder (Spl-IsoFind), for which we adapt Moran's I[30]. Moran's I is a global spatial autocorrelation score, which has been successfully used for gene-expression studies[22,31–33]. We successfully identify cell-type-specific spatially variable isoforms, such as *Snap25* for excitatory neurons and *Rps24* for oligodendrocytes, and spatially variable isoforms that do not align with predefined regions.

Results

Gene-expression patterns in coronal brain slices

To define spatially variable isoforms for a given cell type, we used two consecutive coronal slices of an adult mouse brain, covering cortex, hippocampus, thalamus, and midbrain (**Fig S1a-b**). We used the Stereo-seq spatial approach to barcode cDNAs at 500nm resolution[14]. We performed short-read sequencing as per Stereo-seq guidelines on an aliquot of cDNAs. Another aliquot of cDNAs was used to perform exome sequencing and long-molecule selection, followed by PacBio Kinnex sequencing and Oxford Nanopore long-read sequencing. Short-read sequencing, in conjunction with an ssDNA-stain, was used to define individual cells, assign cell types and to anchor barcodes to specific spatial locations. Long reads, both of the PacBio and the Oxford Nanopore platforms, were used to define novel isoforms and to quantify annotated and novel isoforms (**Fig 1a**). In this study, we focused our analysis on the complete right hemisphere. Short-read gene expression clearly indicated biologically meaningful brain regions, including distinct cortical layers (**Fig 1b, S1c**) and within each brain region, we could define major cell types using only singlet spots (**Fig 1c, S1d, S2**). Furthermore, staining and short-read data were sufficient to align two coronal brain slices (**Fig 1d**). Of note, cortex and above all some hippocampal areas stood out in terms of UMI density, owing to higher cell density (**Fig 1e, S1e**).

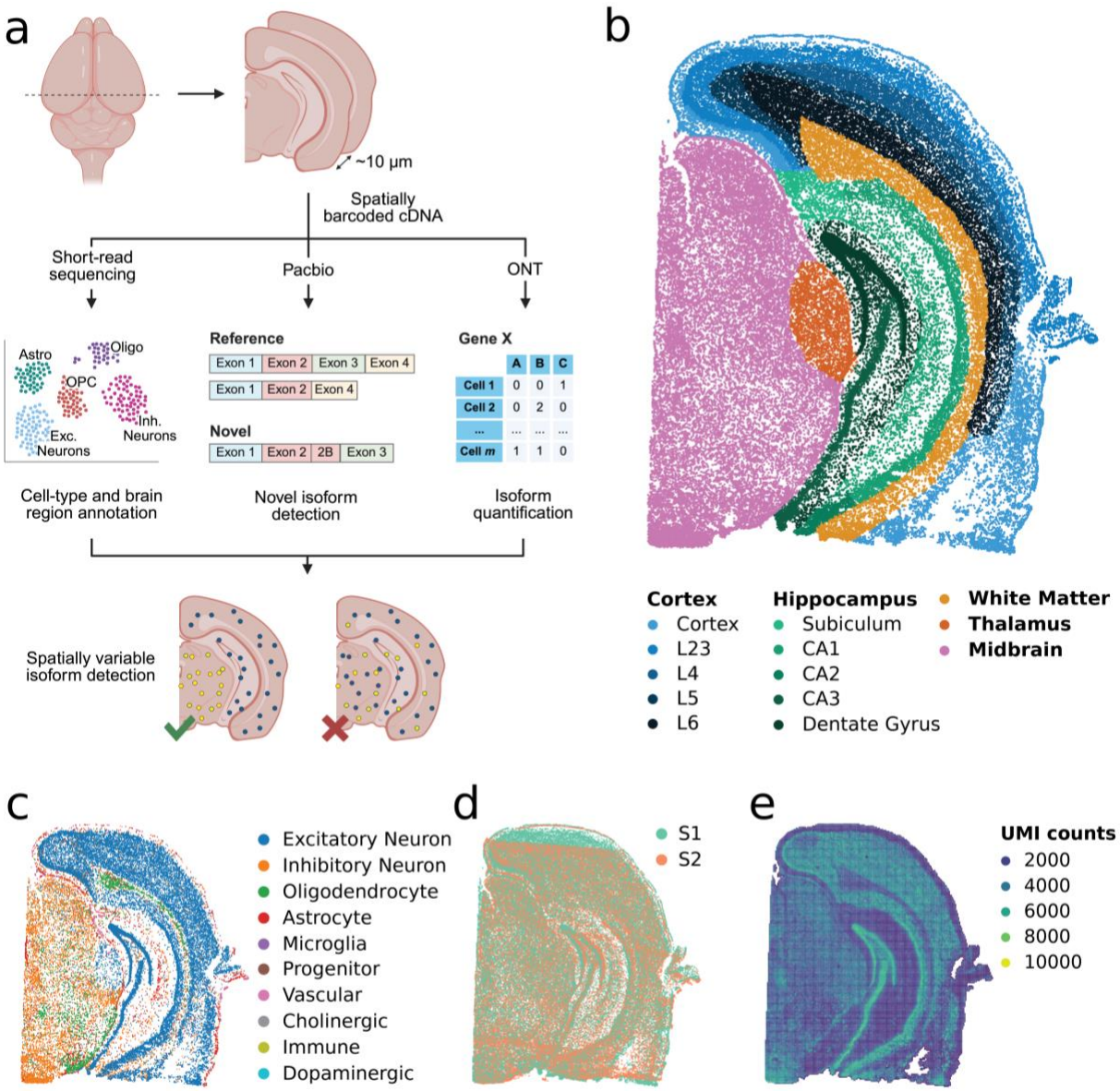

Figure 1. Long-read spatial transcriptomics of two mouse brain slides. a) Schematic overview of the study design. Spatially barcoded cDNA was generated for two coronal sections (10µm apart) of a P56 mouse brain using the Stereo-seq platform[14]. The pool of cDNA is sequenced using 1) short-read technologies, 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 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[21] 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[34] was used (but other annotations, including RefSeq[35] or Ensembl[36] 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[21,37,38].

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[39–41], none of these tools is capable of "parsing" the Stereo-seq molecule structure. 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.

We then enriched barcoded cDNA using exome-sequencing probes as we have done before for single-nuclei[5,9], single-cell[4] and spatial libraries[20] using also a long-molecule selection

step[20]. We sequenced 203.7 and 58.8 million reads on Oxford Nanopore platform for the two coronal slices (from now on called Sample 1 AllExome (AE) and Sample 2 (AE) respectively) as well as 19.9 million and 17.6 million PacBio reads. For Sample 1 (AE) we used two ONT flow cells instead of one explaining the difference in read depth. We furthermore performed one Nanopore long-read run on cDNAs enriched for 3375 genes of interest (from now on called Sample 1 (4K)), rather than for the whole exome (**Table S1-6**) (Methods).

Of note, due to amplification and splitting of cDNA libraries for distinct sequencing protocols, cDNA copies of the same original molecule can be sequenced on both platforms. We therefore tested reproducibility and the influence of higher error rates in Nanopore sequencing on transcript assignments by taking advantage of such pairs of reads sequenced on both platforms. Indeed 3.6 and 2.3 million unique molecules for Sample 1 (AE) and Sample 2 (AE) respectively were sequenced on both Oxford Nanopore and PacBio (Methods). We therefore asked in what fraction of these cases, such read pairs were assigned to the same isoform, which is 99.4% in both replicates (**Fig 2g**). This observation supports that errors in Nanopore reads do not lead to widespread mis-assignments of reads to the wrong isoform.

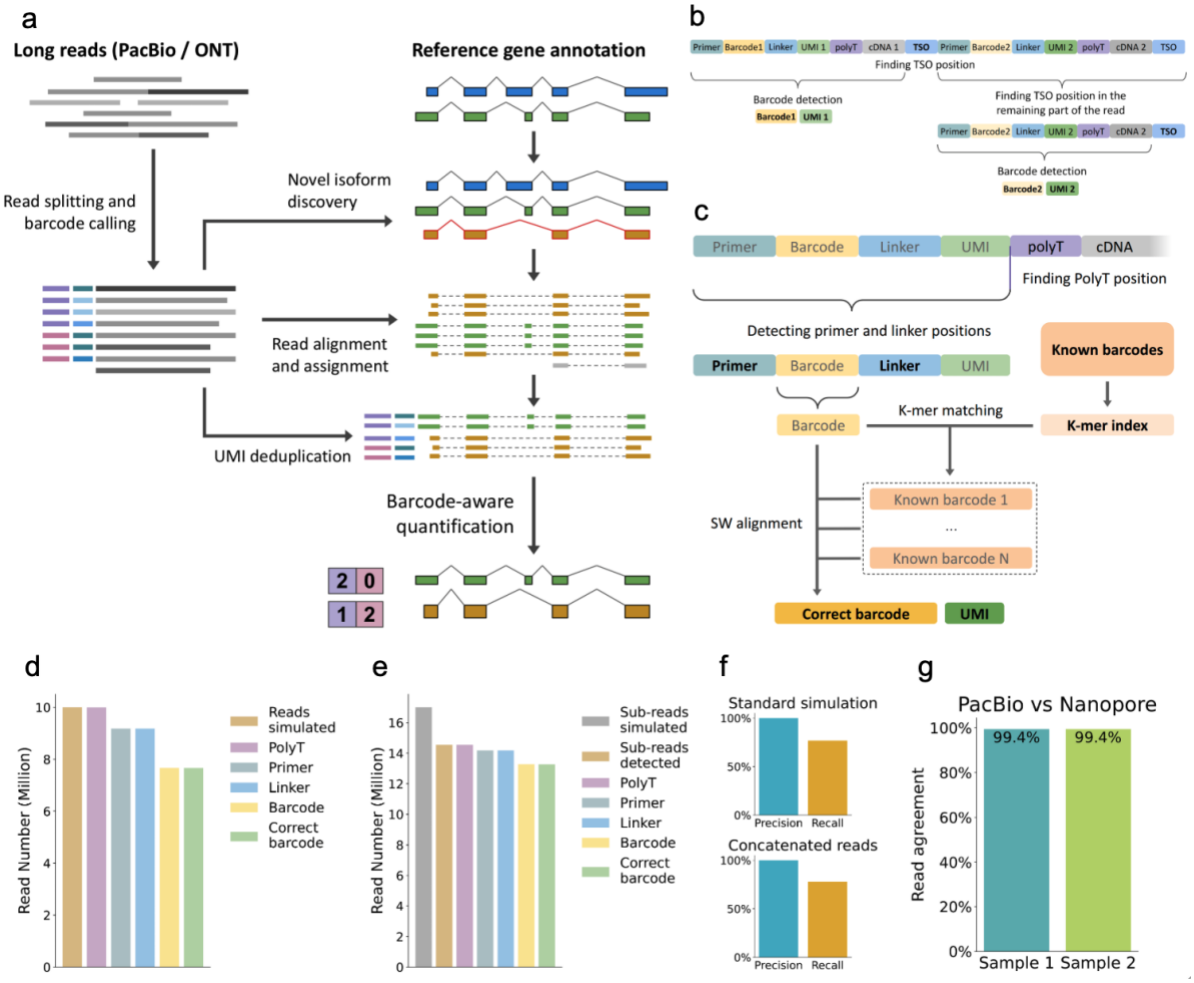

Figure 2. Processing long reads obtained with Spl-ISO-Seq2. a) Spl-IsoQuant-2 pipeline overview. Purple and pink bars next to the reads indicate different barcodes. Blue bars between the reads and the barcodes indicate UMIs. Final diagram shows isoform count per barcode (purple/pink). **b)** Splitting 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. **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 pairs of brain regions using our 2xn isoform tests[4,9] (Methods) using the ONT Sample 1 + 2 (AE) data. The midbrain showed the highest numbers of differentially used isoforms, especially when compared to hippocampus and cortex (**Fig 3a**). 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. An example exon which is almost always excluded in the midbrain compared to the cortex can be seen in the Neuroplastin gene (*Nptn)* (**Fig 3b**). We used PCR to validate this exon's spatial inclusion with primers targeting surrounding exons 1 and 3. Resulting cDNA amplicon length in the midbrain matched with length corresponding to exon skipping (191 bp), whereas combined bulk tissue of other brain regions reflected exon inclusion (539 bp) (**Fig 3c**). We quantified and normalized gel bands to calculate exon inclusion (Methods) which revealed that normalized exon inclusion was decreased in the midbrain compared to other regions (**Fig 3d,** $p<5.4e-4$, Welch's two-sample t-test). Of note, Nptn is a synaptic glycoprotein involved in synaptogenesis, long-term potentiation and memory formation[42]. These two *Nptn* isoforms have been documented by the field as Np55 and Np65, where Np65 contains an additional immunoglobulin domain[43]. We additionally validated an exon in the Phosphatase and Actin Regulator 1 (*Phactr1*) gene, which helps to regulate actin cytoskeletal organization[44,45] (**Fig S3,** $p<3.4e-4$, Welch's two-sample t-test).

When focusing on excitatory neurons, however, the comparison of hippocampus and cortex yielded higher numbers (**Fig 3e**). This is explained by the observation that the midbrain contains much larger percentages of inhibitory neurons (**Fig S2**). Thus a hippocampus to midbrain comparison tends to uncover differences between excitatory and inhibitory neurons. Within the hippocampus, especially the CA3 subregion compared to the dentate gyrus (DG) showed differential isoform abundance, possibly owing to ongoing neurogenesis and the presence of immature cell types in the DG (**Fig 3f**). Likewise, in the cortex, layer 2-3 showed important changes in isoform regulation especially when compared to layer 6 and layer 5 (**Fig 3g**). Interestingly, *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 3h**). Similarly, in the cortex, notable isoform changes are observed between layers 2-3 and layer 5 (**Fig S4**). *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[46–48]. 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 S5**).

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 S6**), 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.

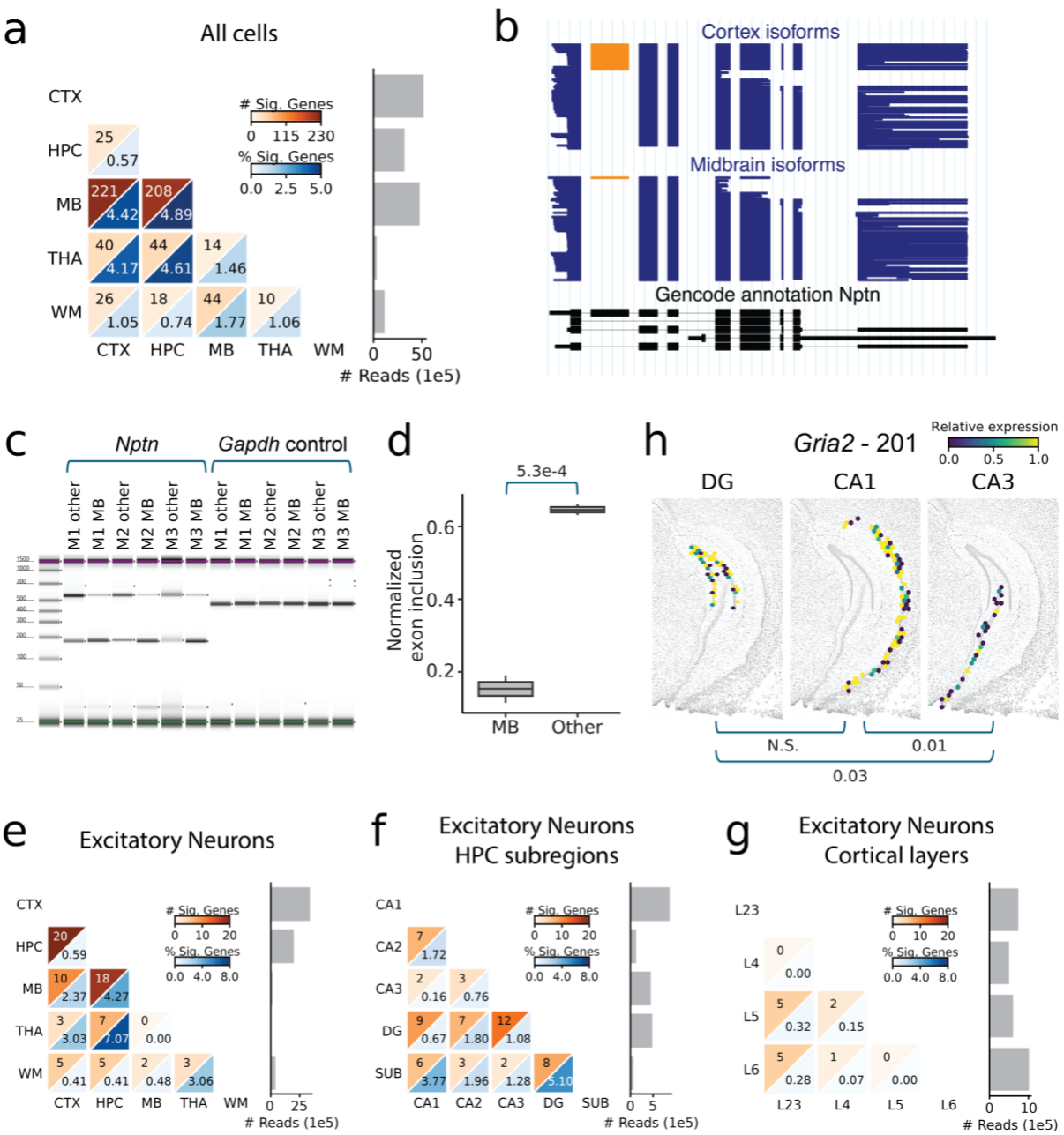

Figure 3. Detecting genes exhibiting differential isoform expression across predefined brain regions. a) The number and percentage of genes changing isoform expression across predefined brain regions using all cells in Sample 1 and Sample 2 combined. Barplot on the right shows the number of long reads used during the analysis per brain region. **b)** Scisorwiz plot showing how different isoforms of *Nptn* are expressed in the cortex and midbrain. **c)** Tapestation gel of PCR amplified products using exon-of-interest-spanning *Nptn* specific primers or *Gapdh* specific primers. Mx = Mouse sample, MB = isolated bulk midbrain, other = bulk brain with midbrain removed. **d)** Normalized quantification of tapestation gel (Methods). Normalized exon inclusion = 539 base pair (bp) band/(191 bp+539 bp band). Bands are normalized to *Gapdh* expression. Welch's two-sample t-test was used for statistical testing. **e-g)** Number and percentage of genes changing isoform expression in excitatory neurons across brain (sub)regions. **h)** Relative expression of *Gria2* isoform 201 in excitatory neurons in different hippocampal subregions of Sample 1. 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). As such, 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. CTX = cortex, HPC = hippocampus, MB = midbrain, THA = thalamus, WM = white matter, DG = Dentate Gyrus, SUB = Subiculum, 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[30] (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[22,31–33], 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 S7**). 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**). The cell-type agnostic mode results in most SVIs for two reasons: 1) we use more cells during testing, 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 (59.5%) are detected in two or more datasets (**Fig 4d**). However, 173 isoforms (444 pairwise combinations) are tested in two (or more) datasets, but significant in only one. In 403 cases (90.8%) the isoform is significant in the dataset with the higher read depth (**Fig S8**). Especially the combined Sample 1 + Sample 2 (AE) dataset 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 type directly changes splicing, but the cell type distribution shifts[4]. The previously identified cell-type agnostic SVIs can fall into either the first or third category. To address this, we calculated p-values using cell-type constrained permutations (Methods). Unlike our previous approach, where the values of all cells were shuffled randomly to convert Moran's I to a p-value, we now shuffle values only within the same cell type. For Sample 1 + Sample 2 (AE), the uncorrected p-values of most isoforms remain very similar (**Fig 4e**). After applying multiple testing corrections, 154 out of the 179 isoforms remain significant (FDR < 0.05, BY corrected), suggesting that at least one cell type exhibits spatial variability for these isoforms.

An example of significant SVIs after applying cell-type constrained permutations are the isoforms *Snap25* - 201 and *Snap25* - 202. These isoforms exhibit a spatial pattern across

291 cortical layers and in the midbrain (**Fig 4f, S9**). In general, *Snap25* - 201 is the dominant
292 isoforms, except for parts of the midbrain and the deeper cortical layers which also show
293 expression of *Snap 25* - 202. Previous studies have identified similar spatial patterns in the
294 brain, but their resolution was insufficient to pinpoint individual cell types[3,4]. With our cellular
295 resolution, we can analyze these patterns at the cell type level. Notably, we observe
296 significance in excitatory neurons (where the pattern is present in the cortex) and inhibitory
297 neurons (where the pattern appears in the midbrain) (**Fig S9**). These inhibitory neurons may
298 be dopaminergic, though potentially mislabeled, and the region in the midbrain exhibiting the
299 inclusion of *Snap25* - 202 is likely the substantia nigra or VTA.
300 *Snap25* serves as a classic example of two isoforms switching between different conditions.
301 However, for some genes, we observe more complex patterns, with either only one or more
302 than two significant isoforms (**Fig 4g**). For instance, *Arpp19* exhibits four distinct SVIs. Two
303 isoforms (203 and a newly discovered isoform) are specific to the cortex. *Arpp19* - 203 is more
304 concentrated in the upper cortical layers, while the newly identified isoform is less abundant
305 and primarily localized to the deeper layers. *Arpp19* - 209 is highly expressed across all brain
306 regions except the cortex, while *Arpp19* - 205 shows the lowest overall expression but is
307 predominantly used in the hippocampus, midbrain, and thalamus (**Fig 4h, S10**).
308 We identified two significant isoforms of *Rps24* across all cell types. For example, in
309 oligodendrocytes, we observed a distinct spatial difference between the midbrain and white
310 matter (**Fig 4i, S11**). Isoform *Rps24* - 202 was predominantly expressed in the white matter,
311 while the midbrain exhibited a more balanced expression of *Rps24* - 202 and *Rps24* - 210.
312 We then compared our SVIs to the genes we identified using the predefined regions. This
313 comparison is challenging, as Spl-IsoFind directly pinpoints isoforms, whereas using the
314 predefined regions, we identified genes with splicing changes, but it was not always clear
315 which specific isoforms were affected. In general, using the predefined regions, we identify
316 more significant genes (**Fig. 4j, S12**), as fewer reads are typically required to detect
317 significance. However, by applying Spl-IsoFind, we also uncover new SVIs, such as *Tncc1* -
318 202 (**Fig. 4k**).

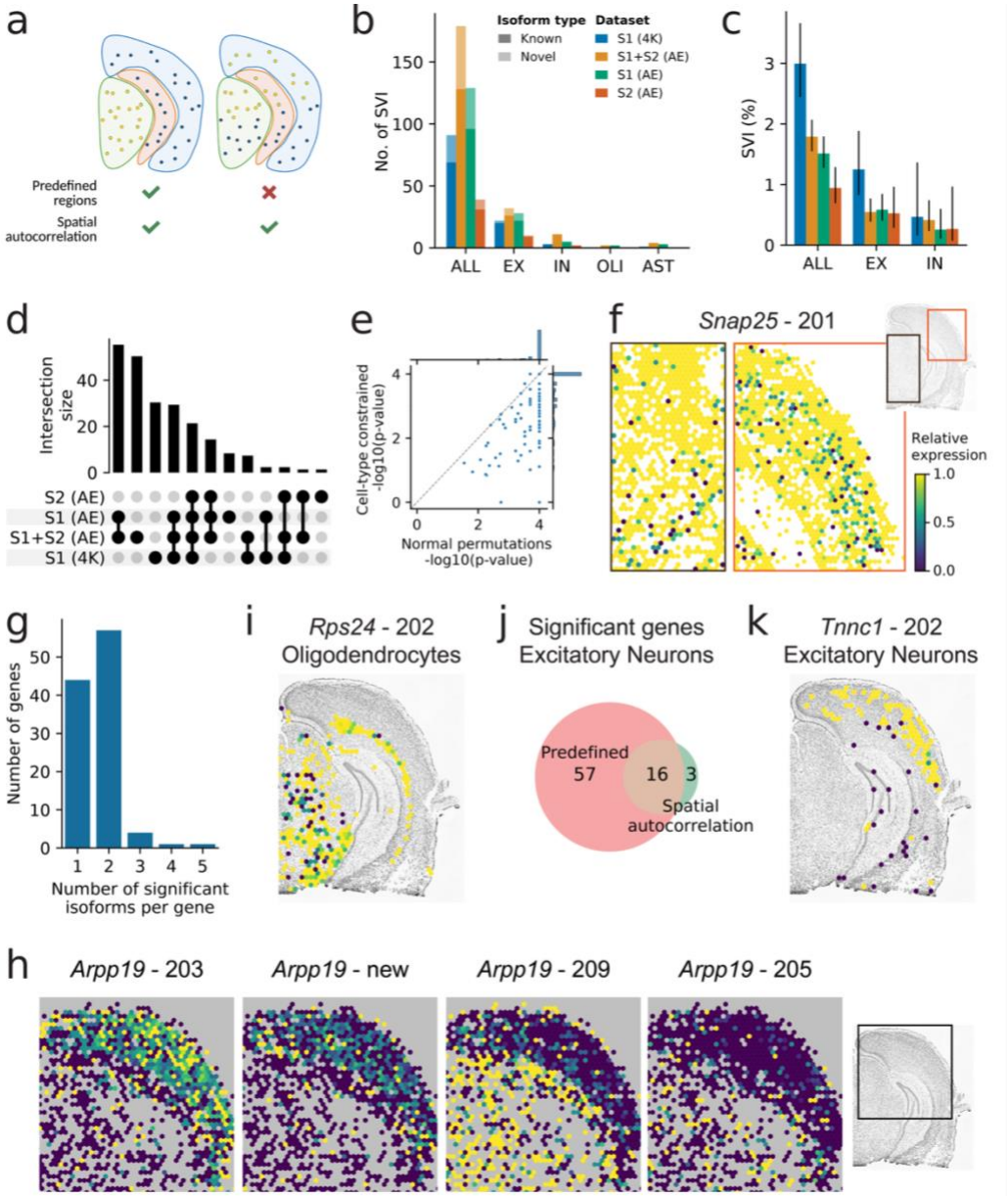

Figure 4. Detecting spatially variable isoforms using spatial autocorrelation. a) Schematic example of a pattern (left) which can be detected using either predefined regions and spatial autocorrelation and another pattern (right) which can only be detected using spatial autocorrelation. **b)** Number and **c)** percentage of significant spatially variable isoforms (FDR < 0.05) per dataset and celltype. Error bars indicate 95% confidence interval. **d)** UpSet Plot showing 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 using normal and cell-type-specific permutations for the isoforms that were significant in the Sample 1 + Sample 2 (AE) dataset using all cells. **f)** Spatial distribution of the *Snap25* - 201 isoform in the midbrain and across the cortical layers in the Sample 1 (AE). Every hexagon shows the mean relative expression of the underlying cells (Methods). **g)** Number of significant isoforms per gene in the Sample 1 + Sample 2 (AE) dataset. **h)** Spatial distribution of the four significant isoforms of *Arpp19* in Sample 1 (AE). Every hexagon shows the mean relative expression of the underlying cells (Methods). **i)** Spatial distribution of *Rps24* - 202 in oligodendrocytes of Sample 1 (AE). Every hexagon shows the mean relative expression of the underlying cells (Methods). **j)** Overlap between the significant genes found using both methods for the excitatory neurons in the Sample 1 (AE) + Sample 2 (AE) data. **k)** Spatial distribution of *Tnnc1* - 202 in excitatory neurons of Sample 1 (AE) which was only discovered using spatial autocorrelation. Every hexagon shows the mean relative expression of the underlying cells (Methods). SVI = spatially variable isoforms, ALL = all cells, EX = excitatory neurons, IN = inhibitory neurons, OLI = oligodendrocytes, AST = astrocytes, AE = AllExome dataset, 4K = 4K targeted dataset.

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 a cell-type-specific resolution for the first time. Spl-ISO-Seq2 has spot sizes of 220nm –far smaller than any known cells in the mouse brain– and inter-center distances of 500nm. 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[49,50]. As such we can pinpoint isoforms that show cell-type-specific spatial variation, such as *Rps24* for oligodendrocytes and *Snap25* for excitatory neurons.

Previously, we and others showed that two mutually exclusive exons of *Snap25* exhibit spatially defined inclusion[3,4]. 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, revealing amongst others that for *Snap25* the 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[30] 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 detectable 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 potential explanation for the lower detection rate of Spl-IsoFind is its need for 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 only one. When we combined both samples, the number of significant hits increased further, with no signs of saturation, suggesting that many SVI patterns remain undetected.

A higher read depth would also increase the number of cell-type-specific SVIs. This could be further improved by decreasing the number of doublets in the data, thereby increasing the number of cells per cell type. However, due to 10µm slices, it is not always possible to completely exclude the presence of long reads from two distinct cells at the same coordinate. Especially glial cells often form doublets[51], which may explain our low number of singlets for oligodendrocytes and astrocytes and the corresponding low number of detected SVIs.

In summary, our wet lab and dry lab methods increase the resolution of spatial and single-cell isoform research. By combining high-resolution spatial data with isoform detection at the cell-

type level, our approach represents a crucial step forward in understanding the spatially 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

428 30 s and 72 °C for 60 s. Amplified cDNA was again purified with 0.8× SPRIselect beads
429 (Beckman Coulter, B23318) and washed with 80% ethanol twice.

430 Exome Enrichment of cDNA

431 500 ng of the cDNA collected from the previous step was used as input for the exome
432 enrichment. Exome capture was performed as previously described to selectively amplify
433 spliced molecules[5]. Agilent probe kit SureSelectXT Mouse All Exon (Agilent, 5190-4641) and
434 the reagent kit SureSelectXT Mouse All Exon library (Agilent, G9611A) were used according
435 to the manufacturer's manual. First, the block oligonucleotide mix was made by combining 1
436 ul of LPCR primer (5'-CTTCCGATCTATGGCGACCTTATCAG-3') and 1 ul of TSO primer (5'-
437 CTGCTGACGTACTGAGAGGC-3') at 200 ng/ul. We next combined 5 ul of 50-100 ng/ul of
438 linearly amplified cDNA (see section above) and diluted it with 2 ul blockmix and 2 ul nuclease
439 free water (NEB, AM9937). This mix was incubated on a thermocycler at 95 °C for 5 min, 65 °C
440 for 5 min and 65 °C on hold. Next, 20 ul of SureSelect Hyb1, 0.8 ul of SureSelect Hyb2, 8.0 ul
441 of SureSelect Hyb3 and 10.4 ul of SureSelect Hyb4 were combined at room temperature to
442 make hybridization mix. Once the cDNA block mix reached to 65 °C on hold, 5 µl of probe mix,
443 1.5 µl of nuclease-free water, 0.5 µl of 1:4 diluted RNase Block and 13 µl of the hybridization
444 mix were added to the cDNA block oligo mix and incubated for 24 h at 65 °C. M-270
445 Streptavidin Dynabeads (Thermo Fisher Scientific, 65305) were washed three times and
446 resuspended with 200 µl of binding buffer after 24 hours of incubation. The incubated cDNA-
447 block mix was mixed with the M270 Dynabeads and placed on a Hula mixer for 30 minutes.
448 After this incubation time, the bead buffer was replaced with 200 µl of wash buffer 1 (WB1)
449 followed by another 15-min incubation with low speed. Next, the WB1 was replaced with WB2,
450 and the tube was transferred to the thermocycler for the next round of incubation (10 minutes).
451 WB2 was replaced 2 more times at 10 minute intervals with the pre-heated WB2. Following
452 this last incubation, beads were resuspended in 18 ul nuclease free H20 and then the spliced
453 cDNA, which is attached to the beads, was amplified as described below.

454 4K Gene Set Enrichment of cDNA

455 We also enriched linear-PCR amplified cDNA with a custom gene set of 3375 mouse ortholog
456 genes related to various brain diseases and synaptic function, as we have published
457 previously[10]. All experimental procedures remained the same as the exome enrichment and
458 the "SureSelectXT Mouse All Exon" probes were replaced with the custom set (Agilent Part
459 No. 5191-6900).

460 Long-molecule Capture and Amplification

461 The long-molecule selection of spliced cDNA was performed as previously published[20].
462 Spliced cDNA bound with the M-270 Dynabeads was amplified with primers (LPCR: 5'-
463 CTTCCGATCTATGGCGACCTTATCAG-3'; TSO: 5'-CTGCTGACGTACTGAGAGGC-3') and
464 KAPA-HiFi enzyme by using the following PCR protocol: 95°C for 3 min, 8 cycles of 98°C for
465 20 s 64°C for 60 s and 72°C for 3 min. The amplified cDNA was isolated from M-270 beads
466 as supernatant and followed by a size selection/purification with 0.48× SPRIselect beads (in
467 1.25M NaCl-20% PEG buffer) and then eluted in 25ul EB buffer. All 25 ul size-selected spliced
468 cDNA was used as template for the second round amplification of 6-8 cycles with the same
469 PCR conditions suggested above. The product of the second round PCR was size

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

472 Oxford Nanopore Library Preparation

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

481 PacBio Library Preparation

482 Exome enriched and long-molecule selected cDNA was also sequenced with PacBio using
483 the PacBio Mas-Seq for 10X 3' Concatenation kit (102-659-600). As this kit was meant for
484 cDNA coming from the 10X Chromium Next GEM Single Cell 3' kit, we performed an initial
485 PCR such that the correct sequences would be added to the ends of each molecule such that
486 the 3' and 5' ends resembled those from 10X (10x_PR1_Stereo: 5'-
487 CTACACGACGCTCTTCCGATCTCTTCCGATCTATGGCGACCTTATCAG-3';
488 10x_PR2_Stereo: 5'-AAGCAGTGGTATCAACGCAGAGCTGCTGACGTACTGAGAGGC-3').
489 The PacBio Mas-Seq for 10X 3' Concatenation kit was used following PacBio's published
490 protocol. The resulting library was sequenced using the PacBio Revio using 1 SMRT cell per
491 sample. Reads were split by identifying known read-linking sequences using skera (version
492 1.4.0).

493 Tissue isolation and reverse transcription for exon validation

494 To validate spatially variable isoforms found by comparing two regions, we isolated the
495 midbrain region from the remaining portion of the brain from three P56 mice. RNA was isolated
496 from 30-50mg of these regions using the Qiagen miRNA Tissue/Cells Advanced Mini Kit
497 (Qiagen Cat. no. 217684) following the established protocol. RNA was frozen at -80C. Next,
498 we performed reverse transcription using the SuperScript III First-strand Synthesis System
499 (Invitrogen, 18080051) for RT-PCR kit following the manufacturer's protocol. We combined
500 5uL RNA with 0.5uL random hexamers, 0.5uL oligodT, 1uL dNPTs, and 3uL H20. This mix
501 was incubated at 65C for 5 minutes followed by being placed on ice for 1 minute. We next
502 added 2 uL 10x Buffer, 4 uL MgCl2, 2uL DTT, 1uL RNAse Out, and 1 uL SuperScriptIII to the
503 samples and incubated at 25C for 10 minutes, 50C for 50 minutes, and 85C for 5 minutes.
504 The cDNA was then stored at -20 until further use.
505

506 PCR exon validation

507 cDNA was amplified using target primers (see below) using the following settings: 95 °C for
508 3 min, 26 cycles of 98 °C for 20 s, 64°C for 30 s and 72 °C for 60 s. Resulting cDNA was
509 inputted onto an Agilent Tapestation System 4200 using D1000 ScreenTapes following the
510 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[52]. 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.

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[21]). 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[53] 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 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.

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.

552 These individual subsequences are then used for detecting barcode and UMI sequences
553 (CIDs and MIDs in Stereo-seq notation respectively) (**Fig 2c**), as well as the remaining pipeline
554 steps, such as read alignment and feature quantification. Read barcode sequence (located
555 between the 3' primer and the linker, see **Fig 2c**) is matched against the set of barcodes from
556 the input whitelist. Only cDNA with barcodes with the alignment score above a certain
557 threshold S_{min} are reported (see details below); cDNAs with non-matching barcodes are
558 rejected. UMI sequence is reported as a read subsequence located between the linker and
559 the start of the poly(T) stretch.
560
561 Here we exploit the same pattern matching as in the previous version of Spl-IsoQuant[20], which
562 is based on short k-mer indexing and conventional Smith-Waterman local alignment (via the
563 ssw-py library[54]). To efficiently detect sequences from a given set (e.g. barcodes) or a fixed
564 sequence (e.g. linker or primer), we use k-mer indexing: a hash table where each key is a k-
565 mer and the value is the set of known sequences containing it. Given a read, we extract all of
566 its k-mers and use the index to retrieve potential matches. To "parse" the read according to
567 the molecule structure, we exploit three separate indices: for the TSO primer (23 bp, k=8), the
568 3' primer (26 bp, k=6) and the linker (15 bp, k=5). These indices are used to quickly detect
569 potential positions of a fixed sequence in a read, which are then verified via Smith-Waterman
570 local alignment.
571
572 Once the positions of the primer and the linker are identified, the barcode sequence is
573 extracted (between the primer and the linker) and potential matches are identified among the
574 known barcodes using a k-mer index constructed for the input barcode whitelist (25 bp, k=14).
575 While smaller k-mer length increases sensitivity for detecting the correct known barcode even
576 in the presence of sequencing errors, it also raises the chance of spurious matches, especially
577 in large barcode sets (>400 million for Stereo-seq protocol). Therefore, k-mer matching serves
578 only to shortlist potential matches. Top N (N=10 by default) potential matches with the highest
579 number of shared k-mers are validated using Smith-Waterman local alignment. Extracted
580 barcode sequence is required to have an alignment score at least S_{min} (22 by default, classic
581 alignment scoring system: +1 for match, -1 for indel/mismatch) with the closest candidate from
582 the whitelist. If two or more known barcodes have the same highest alignment score or all
583 potential matches have score below S_{min}, no barcode is reported for this cDNA. Thus, the k-
584 mer index reduces the search space and limits the number of costly alignment operations
585 needed for barcode identification, while the alignment itself helps to filter out possible false
586 positive matches and uncertain candidates.

587 Benchmarking barcode detection

588 To evaluate the barcode calling algorithm, we simulated Nanopore data using a modified
589 version of Trans-NanoSim[55], available at https://github.com/andrewprzh/lrgasp-simulation.
590 This version improves read truncation procedures, enabling realistic ONT truncation
591 probabilities and non-truncated reads, and was previously used for IsoQuant benchmarking[21].
592 The error model was trained with actual sequencing data from this project (ONT sequencing,
593 Sample 2), showing a sequencing error rate of 3.8% (1.5% mismatches, 1.1% insertions, 1.2%
594 deletions), ensuring accurate recall and precision estimates. No read truncation was induced
595 during the simulation process.
596

597 In this work we assessed the barcode calling method using two simulated datasets: (1) the
598 one where each read contains a single cDNA sequence, and (2) another one where each read
599 can contain from 1 to 4 concatenated cDNAs. Molecule templates were created using
600 Gencode reference transcript sequences according to the Stereo-seq molecule structure (**Fig**
601 **2c**). Reference transcripts were randomly selected based on the expression profile derived
602 from the Sample 2. For the second dataset, we randomly concatenated between 1 and 4 cDNA
603 templates. The probability of each cDNA benign reverse-complemented prior to concatenation
604 is 0.4. While these parameters might not exactly match the ones for real data, the resulting
605 simulated dataset contained a substantial fraction of concatenated reads similar to the reads
606 observed in real data, thus allowing to accurately assess the algorithm performance. In both
607 experiments we simulated 10 million reads.
608
609 We ran Spl-IsoQuant-2 on the simulated datasets and compared the detected barcodes
610 against the true barcodes attached during simulation. Correctly identified barcodes were
611 considered true positives, incorrectly reported barcodes as false positives, and undetected
612 barcodes as false negatives. For reads containing multiple cDNAs, each cDNA was counted
613 individually (**Fig 2e**).

614 Read processing and UMI deduplication

615 Overall, the Spl-IsoQuant-2 pipeline is similar to its previous version. Individual cDNA
616 sequences extracted from raw reads are mapped to the reference genome using minimap2
617 with IsoQuant default settings (minimap2 -x splice --junc-bed annotation.bed -a --MD -k 14 -Y
618 --secondary=yes). Spl-IsoQuant then assigns these mapped reads to reference genes and
619 isoforms using the IsoQuant standard algorithm (Prjibelski et al., 2024). Multimapped reads
620 are resolved by selecting the most consistent alignment with the reference annotation or
621 discarding them. Reads uniquely assigned to a known gene undergo further processing to
622 remove PCR duplicates.
623
624 To eliminate PCR duplicates, reads are grouped by gene and barcode. Within each group,
625 reads with identical or highly similar UMI sequences (edit distance ≤ 4) are considered
626 duplicates. We use an iterative algorithm to identify and assign reads to representative UMIs,
627 selecting the read covering the most splice junctions as the representative and ignoring others.
628 These UMI-filtered reads are then used for downstream analyses, such as exon and transcript
629 quantification and differential splicing analysis. While this method may occasionally remove
630 non-duplicate reads due to similar UMIs, the randomness of UMI sequences minimizes this
631 impact on the analysis.

632 Comparison PacBio and ONT isoform assignment

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

636 Cell segmentation and barcode to cell assignment

637 Cells were segmented based on the ssDNA staining and short-read gene expression using
638 the SAW[56] pipeline (version 7). This generates several outputs including a masking image and
639 a GEF (gene expression file) that contains the CellID for each x and y coordinate combination.

640 However, this GEF only includes coordinates that have at least one short-read assigned. If we
641 used this file directly to map coordinates (and thus barcodes) to the cells, we would lose the
642 coordinates that only have long-reads but no short-reads assigned. To overcome this, we
643 generated an artificial GEM (gene expression matrix) with one short-read for each x and y
644 coordinate. We then converted this artificial GEM into a new cell-binned GEF using the
645 previously obtained masking image using the SAW package. This new GEF includes the
646 CellID for all coordinates on the chip. Barcodes are assigned to coordinates using the Stereo-
647 seq mask files.

648 Slide alignment

649 We aligned Sample 1 and Sample 2 using the `spateo.align.morpho_align()` function
650 from the spateo[25] package (version 1.1.0) with parameters beta=0.1 and lambdaVF=1.0. This
651 greatly improved the alignment compared to naively overlapping the slides (**Fig 1d, S13**)

652 Spatially variable isoform tests using predefined regions

653 We used the isoform assignments from Spl-ISO-Seq2 to detect spatially variable isoforms
654 instead of using full-length reads only. For each region of interest, we counted the number of
655 unique reads per isoform. Afterwards, we used the `DiffSplicingAnalysis()` function
656 from the ScisorSeqR[4] package (version 0.1.9) to identify genes that exhibit differential isoform
657 usage between the predefined brain regions using the default settings (minNumReads = 25,
658 numIsoforms = 10). Using these default settings, genes with less than 25 reads are filtered
659 out. NumIsoforms means that we tested a maximum of 10 isoforms per gene and grouped the
660 remaining isoforms in a rest group, which results in an 11x2 count matrix at most. The returned
661 p-values are calculated using a chi-squared test and corrected for multiple testing using
662 Benjamini Hochberg correction. Genes with FDR <= 0.05 and delta Pi >= 0.1 are considered
663 significant. Delta Pi was calculated for each gene as the sum of change in percent isoform
664 from the top two isoforms detected. We tested all pairwise combinations for the general brain
665 regions, hippocampal subregions, and cortical layers using all cells and per cell type.

666 Differential exon testing using predefined regions

667 We called differentially expressed exons as we have previously[10,20] done. After executing Spl-
668 IsoQuant on the long-read sequencing data, which mapped, barcoded, and spliced reads
669 while correcting for UMIs, we generated an AllInfo file. This file contains detailed metadata for
670 each read, including: read identifier, gene name, comparison group (such as specific cell type
671 or brain region), barcode, UMI, intron chain, transcription start site (TSS), polyadenylation site,
672 exon chain, transcript classification (known or novel), and intron count. You can find further
673 information on AllInfo files at: https://github.com/noush-joglekar/scisorseqr [4].
674 AllInfo files were split based on the specific comparisons being analyzed. These were
675 analyzed using the casesVcontrols() function in the scisorATAC R package
676 (https://github.com/careenfoord/scisorATAC) [10], with a threshold of 10 minimum reads per
677 exon, to identify differentially included exons using standard parameters.
678
679 The process to detect alternative exons involved the following steps:
680 1. Read Selection
681 We used reads from both comparison groups (e.g., cortex barcodes vs. midbrain
682 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)
   - Retain tables where at least 3 of the 4 cells have expected values ≥ 5
5. Statistical Testing
   Each retained exon is tested for association between group identity and exon inclusion using a two-sided Fisher's exact test. Resulting p-values are adjusted for multiple comparisons via the Benjamini-Yekutieli correction.

Spatially variable isoform tests using Moran's I

We used the `esda.Moran` function from the ESDA[57] python package (version 2.4.3) to calculate Moran's I and the corresponding p-value, based on 1e6 permutations. The input data for this analysis consists of the relative expression of the tested isoform in every cell. For the spatial weights, we focused only on cells that express the corresponding gene in the long-read data. The weight for the 10 nearest neighbors of each cell was set to 1, while all other neighbors were assigned a weight of 0.
To reduce the number of tested isoforms, we only included isoforms that met the following criteria:

- The corresponding gene is expressed in at least 50 cells in the long-read data.
- At least 20 cells express the tested isoform and at least 20 cells expressed another isoform (to ensure some minimum expression of the isoform of interest and variance in the data).
- The fraction of cells expressing the tested isoform should be between 0.05 and 0.95 (to ensure some variance in the data).

The p-values are corrected for multiple testing using the Benjamini-Yekutieli method. Isoforms are considered significant if FDR <= 0.05.

Using simulations to assess Moran's I precision and recall

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

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

Each scenario was simulated 100 times and downsampled to different cell numbers (50, 100, 500, 1000, 5000, 10000 cells). After downsampling, Moran's I and the corresponding p-value were calculated as described earlier. If downsampling resulted in violating any of the predefined criteria for Moran's I (e.g., fewer than 20 cells per category), we did not calculate 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 10,000 permutations (1e4) instead of the usual 1 million (1e6).
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 1e6 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

768 that all underlying long reads express the isoform of interest, while a relative expression of 0
769 indicates that all underlying long reads express another isoform.

770 Code availability

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

774 Funding statement and acknowledgements

775 We thank Adrian Tan, Chendong Pan, Aihong Liu, Seongeun Oh and Jenny Xiang from the
776 Genomics Resources Core Facility at Weill Cornell Medicine for performing RNA sequencing.
777 We also thank Weill Cornell Medicine Scientific Computing Unit (SCU) for use of their
778 computational resources.
779 Figure 1a and 4a were created using BioRender.com.
780 Supported by: MIRA R35 GM152101-01 (H.U.T), Brain Initiative grant 1RF1MH121267-01
781 (H.U.T.), NIDA U01 DA053625-01 (H.U.T among others), NIDA grant 2T32DA039080 (J.H.),
782 NSF GRFP # 2139291 (C.F.), the Feil Family Foundation (H.U.T.), European Research
783 Council (ERC) European Union's Horizon 2020 research and innovation programme (grant
784 agreement No. 851093, SAFEBIO) (A.D.P, A.I.T.), Research Council of Finland grants No.
785 322595, 346968, 358744 (A.D.P, A.I.T.), NIGMS Maximizing Investigators' Research Award
786 (MIRA) R35 GM138152 (I.H.).

787 Conflict of interest

788 H.U.T. has presented at user meetings of 10x Genomics, Oxford Nanopore Technologies, and
789 Pacific Biosciences, which in some cases included payment for travel and accommodations.
790 Additionally the Tilgner lab has received discounted reagents from Agilent, Pacific
791 Biosciences, Illumina and Oxford Nanopore. Last but not least H.U.T. will likely become a paid
792 consultant for isogenix, Inc. The other authors declare no competing interests.
793

794 ## References

795 1. Gupta, I. *et al.* Single-cell isoform RNA sequencing characterizes isoforms in thousands
796 of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018).
797 2. Al'Khafaji, A. M. *et al.* High-throughput RNA isoform sequencing using programmed
798 cDNA concatenation. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-023-01815-7.
799 3. Lebrigand, K., Magnone, V., Barbry, P. & Waldmann, R. High throughput error corrected
800 Nanopore single cell transcriptome sequencing. *Nat. Commun.* **11**, 4025 (2020).
801 4. Joglekar, A. *et al.* A spatially resolved brain region- and cell type-specific isoform atlas
802 of the postnatal mouse brain. *Nat. Commun.* **12**, 463 (2021).
803 5. Hardwick, S. A. *et al.* Single-nuclei isoform RNA sequencing unlocks barcoded exon
804 connectivity in frozen brain tissue. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-
805 01231-3.
806 6. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution
807 using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).
808 7. Hagemann-Jensen, M., Ziegenhain, C. & Sandberg, R. Scalable single-cell RNA
809 sequencing from full transcripts with Smart-seq3xpress. *Nat. Biotechnol.* **40**, 1452–1457
810 (2022).
811 8. Salmen, F. *et al.* High-throughput total RNA sequencing in single cells using VASA-seq.
812 *Nat. Biotechnol.* **40**, 1780–1793 (2022).
813 9. Joglekar, A. *et al.* Single-cell long-read sequencing-based mapping reveals specialized
814 splicing patterns in developing and adult mouse and human brain. *Nature Neuroscience*
815 **27**, 1051–1063 (2024).
816 10. Hu, W. *et al.* ScISOr-ATAC reveals convergent and divergent splicing and chromatin
817 specificities between matched cell types across cortical regions, evolution, and in
818 Alzheimer's Disease. *bioRxiv* 2024.02.24.581897 (2024)
819 doi:10.1101/2024.02.24.581897.
820 11. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide

821 expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
822 12. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution
823 with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
824 13. Liu, Y. *et al.* High-spatial-resolution multi-omics sequencing via deterministic barcoding
825 in tissue. *Cell* **183**, 1665–1681.e18 (2020).
826 14. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA
827 nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
828 15. Poovathingal, S. *et al.* Nova-ST: Nano-patterned ultra-dense platform for spatial
829 transcriptomics. *Cell Rep. Methods* **4**, 100831 (2024).
830 16. Schott, M. *et al.* Open-ST: High-resolution spatial transcriptomics in 3D. *Cell* **187**, 3953–
831 3972.e26 (2024).
832 17. Langlieb, J. *et al.* The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**,
833 333–342 (2023).
834 18. Liu, Y. *et al.* High-plex protein and whole transcriptome co-mapping at cellular resolution
835 with spatial CITE-seq. *Nat. Biotechnol.* **41**, 1405–1409 (2023).
836 19. Zhang, D. *et al.* Spatial epigenome-transcriptome co-profiling of mammalian tissues.
837 *Nature* **616**, 113–122 (2023).
838 20. Foord, C., Prjibelski, A. D., Hu, W., Michielsen, L. & Vandelli, A. A spatial long-read
839 approach at near-single-cell resolution reveals developmental regulation of splicing and
840 polyadenylation sites in distinct cortical layers and cell types. *bioRxiv*
841 2025.06.10.658877 (2025) doi:10.1101/2025.06.10.658877.
842 21. Prjibelski, A. D. *et al.* Accurate isoform discovery with IsoQuant using long reads. *Nat.*
843 *Biotechnol.* **41**, 915–918 (2023).
844 22. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods*
845 **19**, 171–178 (2022).
846 23. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially
847 variable genes. *Nat. Methods* **15**, 343–346 (2018).
848 24. Marconato, L. *et al.* SpatialData: an open and universal data framework for spatial

849 omics. *Nat. Methods* **22**, 58–62 (2025).
850 25. Qiu, X. *et al.* Spatiotemporal modeling of molecular holograms. *Cell* **187**, 7351–
851 7373.e61 (2024).
852 26. Fang, S. *et al.* Stereopy: modeling comparative and spatiotemporal cellular
853 heterogeneity via multi-sample spatial transcriptomics. *Nat. Commun.* **16**, 3741 (2025).
854 27. Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial
855 expression data. *Genome Biol.* **22**, 78 (2021).
856 28. Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially
857 resolved transcriptomic studies. *Nat. Methods* **17**, 193–200 (2020).
858 29. Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and
859 robust detection of spatial expression patterns for large spatial transcriptomic studies.
860 *Genome Biol.* **22**, 184 (2021).
861 30. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23
862 (1950).
863 31. Miller, B. F., Bambah-Mukku, D., Dulac, C., Zhuang, X. & Fan, J. Characterizing spatial
864 gene expression heterogeneity in spatially resolved single-cell transcriptomic data with
865 nonuniform cellular densities. *Genome Res.* **31**, 1843–1855 (2021).
866 32. Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution.
867 *Nature* **622**, 552–561 (2023).
868 33. Ninh, V. K. *et al.* Spatially clustered type I interferon responses at injury borderzones.
869 *Nature* **633**, 174–181 (2024).
870 34. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res.* **49**, D916–D923 (2021).
871 35. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status,
872 taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45
873 (2016).
874 36. Dyer, S. C. *et al.* Ensembl 2025. *Nucleic Acids Res.* **53**, D948–D957 (2025).
875 37. Su, Y. *et al.* Comprehensive assessment of mRNA isoform detection methods for long-
876 read sequencing data. *Nat. Commun.* **15**, 3972 (2024).

877 38. Pardo-Palacios, F. J. *et al.* Systematic assessment of long-read RNA-seq methods for
878 transcript identification and quantification. *Nat. Methods* **21**, 1349–1363 (2024).
879 39. You, Y. *et al.* Identification of cell barcodes from long-read single-cell RNA-seq with
880 BLAZE. *Genome Biol.* **24**, 66 (2023).
881 40. Cheng, O. *et al.* Flexiplex: a versatile demultiplexer and search tool for omics data.
882 *Bioinformatics* **40**, btae102 (2024).
883 41. Ebrahimi, G. *et al.* Fast and accurate matching of cellular barcodes across short-reads
884 and long-reads of single-cell RNA-seq experiments. *iScience* **25**, 104530 (2022).
885 42. Ilic, K., Mlinac-Jerkovic, K., Sedmak, G., Rosenzweig, I. & Kalanj-Bognar, S.
886 Neuroplastin in human cognition: review of literature and future perspectives. *Transl.*
887 *Psychiatry* **11**, 394 (2021).
888 43. Langnaese, K., Beesley, P. W. & Gundelfinger, E. D. Synaptic membrane glycoproteins
889 gp65 and gp55 are new members of the immunoglobulin superfamily. *J. Biol. Chem.*
890 **272**, 821–827 (1997).
891 44. Mouilleron, S., Wiezlak, M., O'Reilly, N., Treisman, R. & McDonald, N. Q. Structures of
892 the Phactr1 RPEL domain and RPEL motif complexes with G-actin reveal the molecular
893 basis for actin binding cooperativity. *Structure* **20**, 1960–1970 (2012).
894 45. Wiezlak, M. *et al.* G-actin regulates the shuttling and PP1 binding of the RPEL protein
895 Phactr1 to control actomyosin assembly. *J. Cell Sci.* **125**, 5860–5872 (2012).
896 46. Pei, W. *et al.* Flip and flop: a molecular determinant for AMPA receptor channel
897 opening. *Biochemistry* **48**, 3767–3777 (2009).
898 47. Koike, M., Tsukada, S., Tsuzuki, K., Kijima, H. & Ozawa, S. Regulation of kinetic
899 properties of GluR2 AMPA receptor channels by alternative splicing. *J. Neurosci.* **20**,
900 2166–2174 (2000).
901 48. Wen, W., Lin, C.-Y. & Niu, L. R/G editing in GluA2Rflop modulates the functional
902 difference between GluA1 flip and flop variants in GluA1/2R heteromeric channels. *Sci.*
903 *Rep.* **7**, 13654 (2017).
904 49. Karasek, M., Swiltosławski, J. & Zielińiska, A. Ultrastructure of the central nervous

905 system: the basics. *Folia Neuropathol.* **42 Suppl B**, 1–9 (2004).
906 50. Edgar, J. M. & Griffiths, I. R. White Matter Structure. in *Diffusion MRI* 127–153 (Elsevier,
907 2014). doi:10.1016/b978-0-12-396460-1.00007-x.
908 51. Tiesmeyer, S. *et al.* 2D, or not 2D? Investigating vertical signal integrity of tissue slices.
909 *bioRxiv* 2025.01.13.632601 (2025) doi:10.1101/2025.01.13.632601.
910 52. Stael, S., Miller, L. P., Fernández-Fernández, Á. D. & Van Breusegem, F. Detection of
911 damage-activated metacaspase activity by Western blot in plants. *Methods Mol. Biol.*
912 **2447**, 127–137 (2022).
913 53. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–
914 3100 (2018).
915 54. Zhao, M., Lee, W.-P., Garrison, E. P. & Marth, G. T. SSW library: an SIMD Smith-
916 Waterman C/C++ library for use in genomic applications. *PLoS One* **8**, e82138 (2013).
917 55. Hafezqorani, S. *et al.* Trans-NanoSim characterizes and simulates nanopore RNA-
918 sequencing data. *Gigascience* **9**, giaa061 (2020).
919 56. Gong, C. *et al.* SAW: an efficient and accurate data analysis workflow for Stereo-seq
920 spatial transcriptomics. *GigaByte* **2024**, gigabyte111 (2024).
921 57. Rey, S. J. & Anselin, L. PySAL: A python library of spatial analytical methods. in
922 *Handbook of Applied Spatial Analysis* 175–193 (Springer Berlin Heidelberg, Berlin,
923 Heidelberg, 2010). doi:10.1007/978-3-642-03647-7_11.

924
925