



The spatial landscape of gene expression isoforms in tissue sections

Pascal Barbry, Kevin Lebrigand, Joseph Bergensträhle, Kim Thrane, Annelie Mollbrink, Rainer Waldmann, Joakim Lundeberg

► To cite this version:

Pascal Barbry, Kevin Lebrigand, Joseph Bergensträhle, Kim Thrane, Annelie Mollbrink, et al.. The spatial landscape of gene expression isoforms in tissue sections. 2020. <hal-02992334>

HAL Id: hal-02992334

<https://hal.science/hal-02992334v1>

Preprint submitted on 6 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

The spatial landscape of gene expression isoforms in tissue sections

Kevin Lebrigand ^{1§}, Joseph Bergenstråhle ^{2§}, Kim Thrane ^{2§}, Annelie Mollbrink ², Pascal Barbry ^{1*}, Rainer Waldmann ¹, and Joakim Lundeberg ²

¹Université Côte d'Azur, CNRS, Institut de Pharmacologie Moléculaire et Cellulaire, F06560 Sophia Antipolis, France.

²Department of Gene Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Science for Life Laboratory, Solna, Sweden.

[§]These authors contributed equally to this work.

*Corresponding author

ABSTRACT

In situ capturing technologies add tissue context to gene expression data, with the potential of providing a greater understanding of complex biological systems. However, splicing variants and full-length sequence heterogeneity cannot be characterized with current methods. Here, we introduce Spatial Isoform Transcriptomics (SiT), an explorative method for characterizing spatial isoform and sequence heterogeneity in tissue sections, and show how it can be used to profile isoform expression and sequence heterogeneity in a tissue context

Main

Recent advances in the field of transcriptomics have deepened our understanding of tissue organization by integrating global gene expression in a spatial context [1]. Among these approaches, *in situ* capturing technologies produce whole-transcriptome spatial gene expression by tagging spatial barcodes to transcripts after poly(A)-capture. However, current protocols for *in situ* capturing are based on short-read sequencing, which entails fragmentation and makes it impossible to detect alternative splicing events, characterize somatic mutations or allele-specific expression.

While spatial profiling of specific isoforms within a tissue context has been demonstrated previously [2,3], these methods rely on *a priori* knowledge about transcript architecture. Here, we introduce Spatial Isoform Transcriptomics (SiT), an unbiased method based on spatial *in situ* capturing to detect and quantify spatial expression of splicing variants (Methods). Briefly, we fix fresh-frozen tissue samples using methanol. After staining and imaging of the samples, we spatially barcode the mRNA molecules. The resulting sequencing libraries are used for both full-length nanopore sequencing and highly accurate 3' cDNA short-read generation. Nanopore sequencing is well suited for spatial transcriptomics because it generates more reads and information than other long read platforms (Supplementary Table 1). Furthermore, it is fully compatible with already available spatial arrays, which have barcoded poly(T)-primers on a glass slide surface. Spatial barcodes and UMIs were error-corrected and assigned to nanopore reads using the Java toolkit SiceLore [4]. We demonstrate the isoform landscape *in situ* (Fig.1a) in two regions of the mouse brain: the olfactory bulb (MOB) and coronal sections of the left hemisphere (CBS) (Supplementary Fig.1-2). Molecular markers define distinct anatomical regions as landmarks for analysis (Supplementary Table 2). We identify multiple

genes that display alternative isoform expression and sequence heterogeneity in spatially distinct areas of the brain (Supplementary Fig.3). Our findings were confirmed using *in situ* sequencing (ISS).

For the MOB sample, we applied a stringent filter to only retain isoforms that contain all exon-exon junctions (mean exon number 6.7) of the reference transcript. 2,386 unique isoform molecules were found, corresponding to 1,076 distinct isoforms per spatially barcoded spot (55 μ m diameter) across the tissue section (Supplementary Fig.4). Short-read clustering resulted in five anatomically defined areas, as previously demonstrated [5] (Fig.1a, Supplementary Fig.5). Based on this clustering, we identified 19 isoform switching genes out of which *Plp1* and *Myl6* showed the most prominent regional isoform switching (Methods, Supplementary Table 3). Interestingly, *Plp1*, a gene involved in severe pathologies associated with CNS dysmyelination [6] demonstrated a clear regional difference in isoform expression between the outer regions of the olfactory nerve layer and the inner granule cell layer (Fig.1b, Supplementary Fig.6). We validated the identified spatial isoform differences by *in situ* sequencing (ISS) on a tissue section from another individual (Fig.1c, Supplementary Fig.7).

Each spatially barcoded spot typically captures transcripts from multiple cells. However, using single cell RNA-seq data, it is possible to deconvolute the transcriptional signal into the likely constituent cell types of the spot, and to associate specific cell type(s) to isoform expression. We demonstrate this approach using a previously published MOB scRNA-seq dataset [7] and a deconvolution strategy based on the identification of pairwise cell correspondences [8]. This approach identifies the myelinating-oligodendrocyte-based (Mag+) cell type within the Granule Cell layer as the predominant producer of the *Plp1* standard isoform and the Olfactory Ensheathing Cell-based (Sox10+) cell type within the Olfactory Nerve Layer as the predominant producer of the truncated *Plp1* isoform *DM20* (Supplementary Fig.8).

The same approach of isoform calling was then applied on two independent coronal brain sections. We identified 3,306 unique molecules (UMIs), corresponding to 1,338 distinct isoforms per spatially barcoded spot (Supplementary Fig.4), with an average of 6.7 exons per isoform. Clustering based on short-read gene expression resulted in 12 anatomically defined regions (Fig. 1d, Supplementary Fig.5). We found 64 significant isoform switching genes between regions commonly identified in the two replicates (methods, Supplementary Table 3). The reproducibility of SiT and the provided isoform landscape was demonstrated by taking two CBS sections, located 50 μ m apart in the tissue (Supplementary Fig.9). Our data revealed in both sections a pronounced isoform switching of *Snap25* [9] between the hypothalamus and midbrain (Fig.1e-f, Supplementary Fig.10), and this pattern was confirmed by ISS (Fig.1g). Further validations of the regional isoform switching obtained with SiT were performed by ISS for an additional set of genes (Supplementary Fig.11-12).

Our spatial long read data also identified RNA adenosine-to-inosine (A-to-I) editing events. Such editing has been shown to be essential for neurotransmission and other neuronal functions [10]. While other studies have looked at editing events on bulk samples from mouse brain [11], or spatially resolved by ISS for a limited number of targeted editing sites [2], none has provided an exhaustive spatially-resolved RNA editing map. To this end, we performed additional sequencing for one of the CBS sections (CBS2) to achieve the necessary level of transcript information for robust calls of single nucleotide variants (SNVs) (Supplementary Table 1).

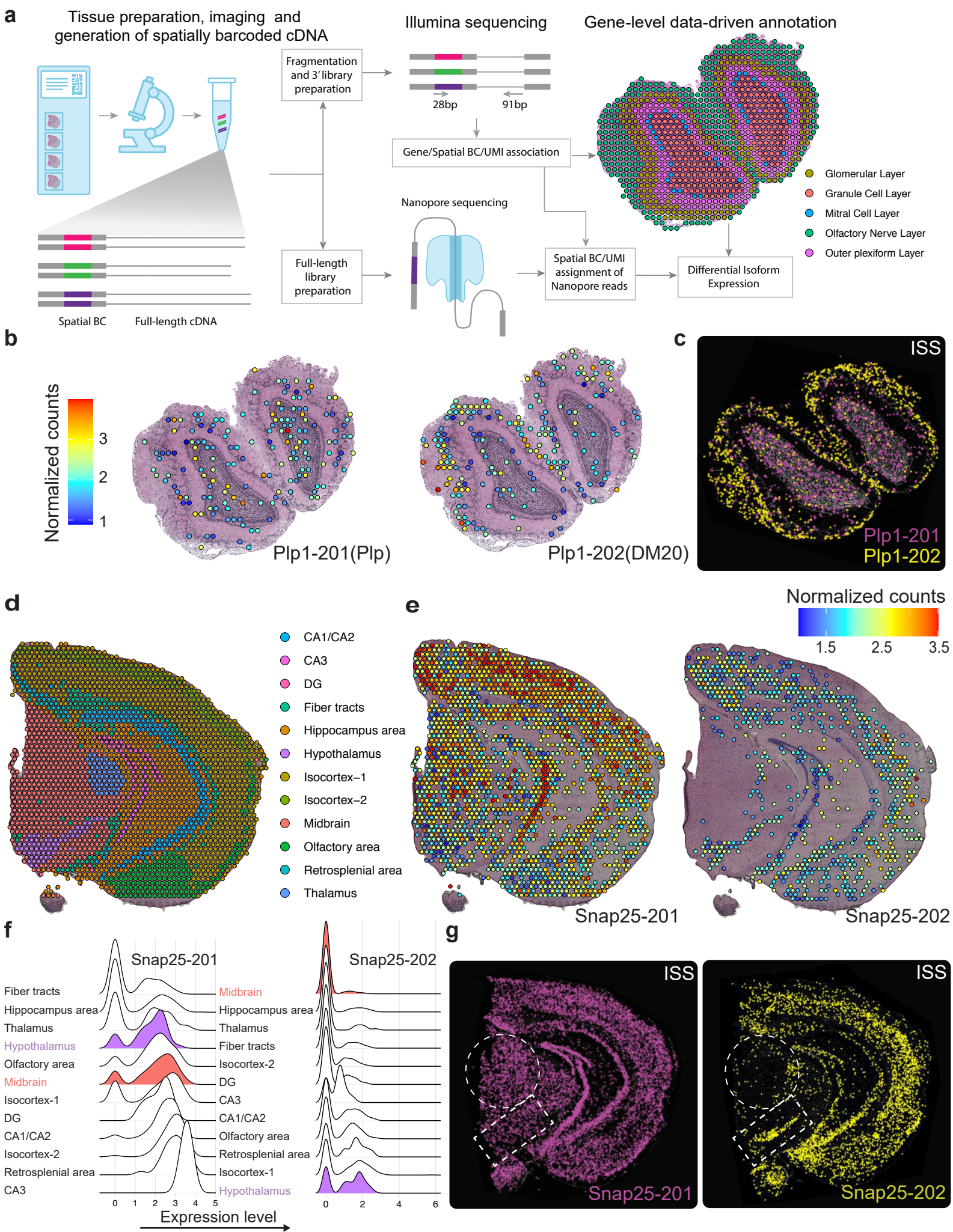


Figure 1. SiT reveals isoform-switch in the mouse brain.

(a) Experimental and computational steps for SiT analysis. Right side shows gene expression clusters in the olfactory bulb. (b-c) Expression of *Plp1* isoforms (*Plp1-201*, *Plp1-202*, also called *DM20*) detected by SiT (b) and ISS (c). (d) Gene expression clusters of the left coronal brain section. (e-g) Expression of *Snap25* isoforms detected by SiT (e), by SiT and visualized as a ridge plot with regions ordered after mean expression (f), and by ISS (g). The dotted circle indicates the midbrain, and the rectangle the hypothalamus.

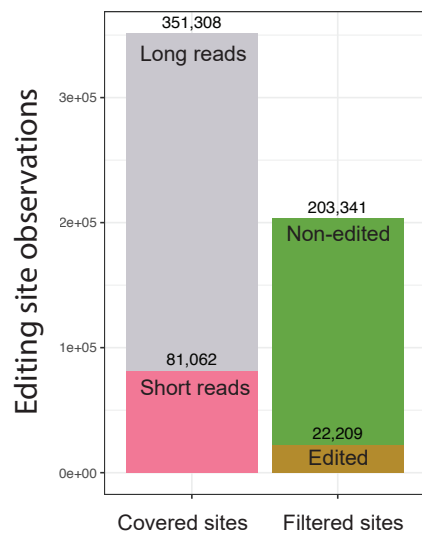
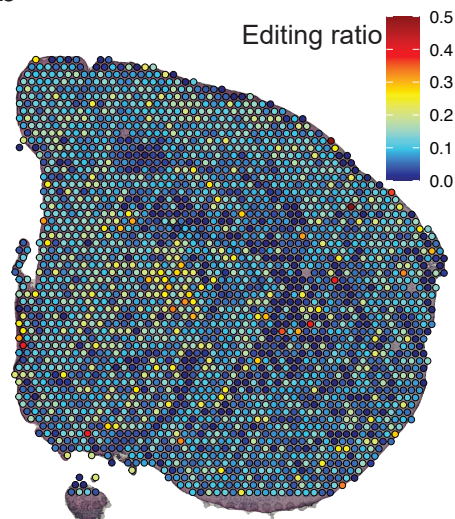
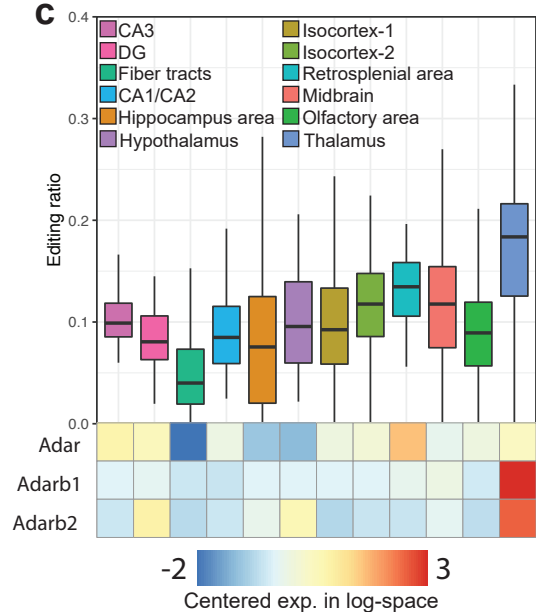
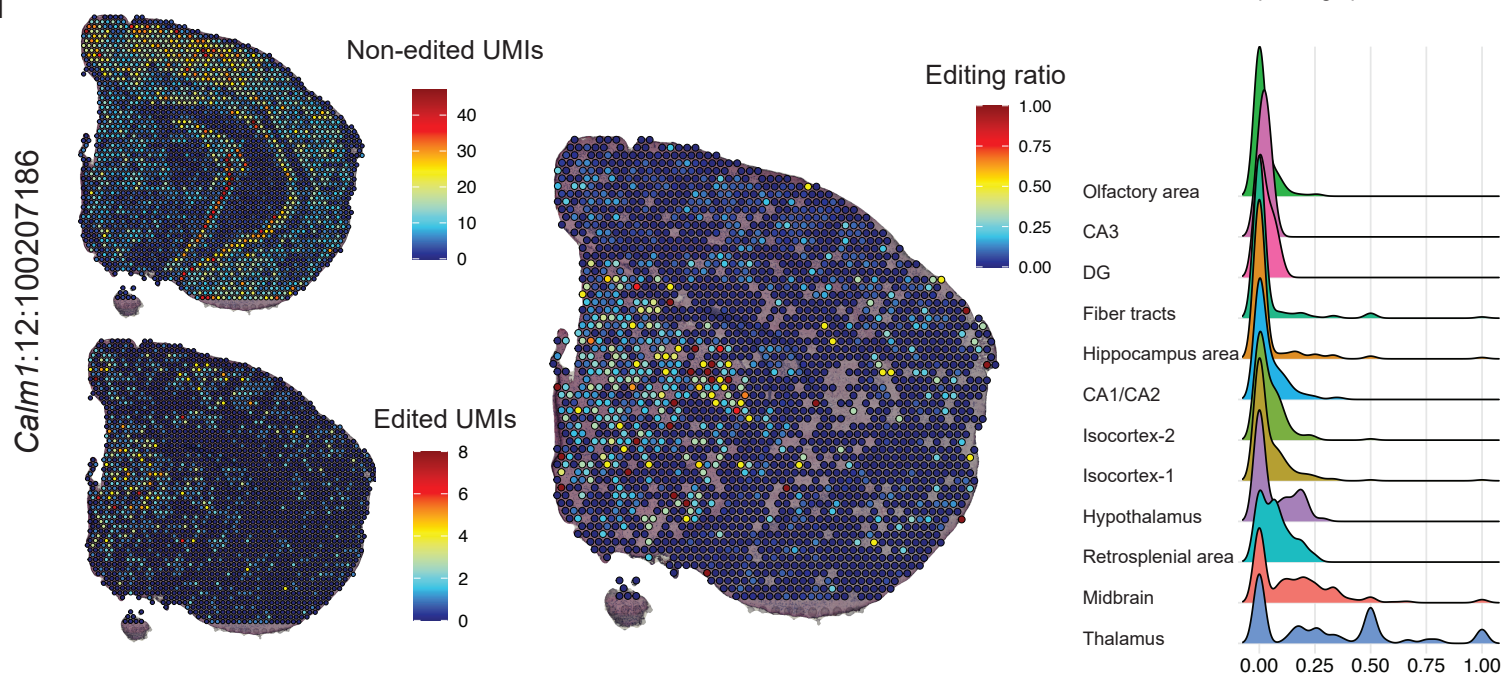
a**b****c****d**

Figure 2. SiT reveals A-to-I RNA editing specificity in the mouse brain.

(a) Summary of editing sites obtained from the literature covered by our long- and short-read data (left bar) and the profile of those sites (edited vs non-edited) after high confidence call filtration (right bar). (b) Spatial map of editing ratios. (c) Editing ratio per region and the average gene-level expression value of ADAR enzymes. The mean editing ratio (10.9%) is indicated in the figure. (d) Spatial map of *Calm1* editing site (chr12:100,207,186) and editing ratio per region ordered by increasing means.

We explored a total of 100,838 A-to-I RNA editing sites described in the literature [11,12]. To reach high confidence calls with long-read data, we defined *ad hoc* thresholds for the number of reads per UMI and consensus base quality by initially looking at the agreement between long and short read base calls for 81,062 shared UMIs (Supplementary Fig.13, Methods). We kept 57.9% of UMIs that passed this filtering for downstream analysis (Fig.2a). Globally, we observed an editing ratio of 10.9% for 7,635 editing sites covered by at least one UMI (Supplementary Table 4). Interestingly, editing ratios displayed a non-uniform spatial distribution (Fig.2b). Thalamus had significantly higher editing ratios while fiber tracts had significantly lower editing ratios (Monte Carlo permutation test p -value ≤ 0.05). Consistent with this finding, we observed a positive correlation between adenosine deaminases acting on RNAs (ADARs) and editing ratios (Fig.2c, Supplementary Fig.14).

Among all editing sites, we noticed a site within *Calm1* that displayed a particularly robust variation across regions. It showed a particularly high editing ratio in the thalamus compared to the other regions (mean per spot 30.6 % vs. 4.9 %; p -value $9.19\text{e-}33$) and in the Midbrain (15.1% vs 6.3%; p -value $1.19\text{e-}49$) (Supplementary Fig.15). The gene *Calm1* encodes calmodulin, which acts as a major intracellular " Ca^{2+} -receptor" that controls cellular responses to modifications of cytoplasmic Ca^{2+} [13]. The editing site in question is located 2,638 bp from the 3' end within the UTR of *Calm1* and has previously been characterized as an allele-specific editing site [14].

In conclusion, we provide, to our knowledge, the first genome-wide approach to explore and discover isoform expression and sequence heterogeneity in a tissue context. The SiT methodology is based on easily available reagents and enables a deepened investigation of the isoform landscape, including studies of bi-allelic expression, fusion transcripts, and SNP expression in a spatial context, which we believe will be helpful to understand biological systems and provide an additional layer of information to Cell Atlas initiatives.

Methods

Mouse Brain Samples

Olfactory bulbs were isolated from C57BL/6 mice (>2 months old), snap-frozen in Isopentane (Sigma-Aldrich) and embedded in cold optimal cutting temperature (OCT, Sakura) before sectioning. The left hemisphere was isolated from an C57BL/6J (8-12 weeks old) mouse and processed in the same way. Olfactory bulbs from two different individuals were used for the Visium and ISS experiments, whereas the same sample of the left hemisphere was used for both methods.

10x Genomics Visium experiments

The Visium Spatial Tissue Optimization Slide & Reagent kit (10X Genomics, Pleasanton, CA, USA) was used to optimize permeabilization conditions for mouse brain tissue. Two coronal sections of the left hemisphere (IDs: CBS1 and CBS2) and one section of olfactory bulb (ID: "MOB") were processed according to the manufacturer's protocol. Spatially barcoded full-length cDNA was generated using Visium Spatial Gene Expression Slide & Reagent kit (10X Genomics) following the manufacturer's protocol. Tissue permeabilization was performed for 6 and 9 min (CBS1, CBS2) and 12 min (MOB). cDNA amplification was conducted with 12 (CBS) and 17 (MOB) cycles. A fraction of each cDNA library was used for nanopore sequencing, whereas 10 μl was used in the 10X Genomics Visium library preparation protocol of fragmentation, adapter ligation, and indexing. The libraries were sequenced on a NextSeq500 (Illumina), with 28 bases from read 1 and 91 from read 2, and at a depth of 253, 217, and 210 million reads for MOB, CBS1, and CBS2 samples, respectively. The raw sequencing data was processed with a pre-launch of the Space Ranger pipeline (10X Genomics) and mapped to the mm10 genome assembly.

Oxford Nanopore sequencing

Nanopore sequencing of libraries prepared with cDNA from the 10x Genomics workflow yield 20 – 50% reads without the 3' adapter sequence and thus lack the spatial barcode and UMI [4]. To deplete such DNA, we initially selected for cDNA that contains a biotinylated 3' primer. 10 ng of the 10x Genomics Visium PCR product were amplified for 5 cycles with 5'-AAGCAGTGGTATCAACGCAGAGTACAT-3' and 5' Biotine-AAAACTACACGACGCTCTCCGATCT 3'. Excess biotinylated primers were removed by 0.55x SPRIselect (Beckman Coulter) purification and the biotinylated cDNA (in 40 µl EB, Qiagen) was bound to 15 µl 1x SSPE washed Dynabeads™ M-270 Streptavidin beads (Thermo) in 10 µl 5x SSPE for 15 min at room temperature on a shaker. Beads were washed twice with 100 µl 1x SSPE and once with 100 µl EB. The beads were suspended in 100 µl 1x PCR mix and amplified for 8 cycles with the primers NNNAAGCAGTGGTATCAACGCAGAGTACAT and NNNCTACACGACGCTCTCCGATCT to generate enough material (1 – 2 µg) for nanopore sequencing library preparation. To deplete small fragments which are typically of little interest for transcript isoform analysis (cDNA from degraded RNA, ribosomal RNAs), small cDNA (< 1 kB) was depleted with a 0.5x SPRI select purification. If fragments between 0.5 and 1 kB need to be retained, SPRIselect concentration should be increased to 0.8x. nanopore sequencing libraries were prepared with the LSK-109 kit from Oxford nanopore (1 µg cDNA) following the instructions from the manufacturer. PromethION flowcells were loaded with 200 ng library each. PCR amplifications for nanopore library preparations were made with Kapa Hifi Hotstart polymerase (Roche Sequencing Solutions): initial denaturation, 3 min at 95°C; cycles: 98°C for 30 sec, 64°C for 30 sec, 72°C for 5 min; final elongation: 72°C for 10 min, primer concentration was 1 µM.

Oxford Nanopore data processing

Nanopore reads were processed according to the scNaUmi-seq protocol [4] (<https://github.com/ucagenomix/sicelore>) with slight modifications. Briefly, to eliminate reads that originate from chimeric cDNA generated during library preparation, we initially scanned reads for internal (> 200 nt from end) Template Switching Oligonucleotide (TSO, AAGCAGTGGTATCAACGCAGAGTACAT) and 3' adapter sequences (CTACACGACGCTCTCCGATCT) flanked by a poly(T) (poly(T)-adapter). When two adjacent poly(T)-adapters, two TSOs or one TSO in proximity of a poly(T)-adapter were found, the read was split into two separate reads. Next the reads were scanned for poly(A/T) tails and the 3' adapter sequence to define the orientation of the read and strand-specificity. Scanned reads were then aligned to *Mus musculus* mm10 with minimap2 v2.17 in spliced alignment mode. Spatial BCs and UMIs were then assigned to nanopore reads using the strategy and software previously described for single cell libraries [8]. SAM records for each spatial spot and gene were grouped by UMI after removal of low-quality mapping reads (mapqv=0) and potentially chimeric reads (terminal Soft/Hard-clipping of > 150 nt). A consensus sequence per molecule (UMI) was computed depending on the number of available reads for the UMI using the *ComputeConsensus* sicelore-2.0 method. For molecules supported by more than two reads (RN > 2), a consensus sequence was computed with the SPOA software [15] using the sequence between the end of the TSO (SAM Tag: TE) and the base preceding the polyA sequence (SAM Tag: PE). Phred scores for consensus nucleotides were assigned as $-10 \cdot \log_{10}(\text{n Reads not conform with consensus nucleotide} / \text{n Reads total})$, (Phred score maximum set to 20). Consensus cDNA sequences were aligned to the *Mus musculus* mm10 build with minimap2 v2.17 in spliced alignment mode. SAM records matching known genes were analyzed for matching Gencode vM24 transcript isoforms (same exon makeup) as described [4]. To assign a UMI to a Gencode transcript, we required a full match between the UMI and the Gencode transcript exon-exon junction layout authorizing a two-base margin of added or lacking sequences at exon boundaries, to allow for indels at exon junctions and imprecise mapping by minimap2. Detailed statistics of each step of nanopore read processing are provided in Supplementary Table 1.

Count matrices and data analysis

Raw gene expression matrices generated by Space Ranger were processed using R/Bioconductor (version 3.5.2) and the Seurat R package (version 3.1.4). Visualizations of spatial data were generated with the STUtility R package (version 1.0.0). We created Seurat objects for each of the three samples (MOB, CBS1 and CB2) with different assays for the analysis as follows: (i) “Spatial” containing gene-level raw short read data from the Space Ranger output, (ii) “ISOG” containing the gene-level nanopore long read data, (iii) “ISO” containing isoform-level transcript information where only the molecules where all exons are observed are kept, (iv) “JUNC” containing each individual exon-exon junction observation per isoform, and (v) “AtoI” containing editing sites from the RADAR database (mm9 UCSC liftover to mm10) and from the Licht study [11], for which we observed at least one UMI in our dataset. The “AtoI” assay stored non edited UMI count (@counts slot), edited UMI count (@data slot), and the editing ratio (@scale.data slot) per editing site. RDS files with all the assays stored are available on demand.

10x Genomics Visium data-driven annotation of anatomical regions

The Spatial assay was normalized with *SCTransform* using standard parameters. The first 30 principal components of the assay were used for UMAP representation and clustering (resolution = 0.4). Brain regions defined by clustering were assigned to known anatomical regions based on the Allen Mouse Brain Atlas. Spot clusterings were similar between short and long read data (Supplementary Figure 16). As short read data contains more UMIs per spot, our different representations are based on short read data.

Differential splicing detection

The *FindMarkers* function in Seurat (logfc.threshold = 0.25, test.use = “wilcox”, min.pct = 0.1) was used to detect genes showing at least two isoforms as markers of different brain regions with a *p*-value ≤ 0.05 using the nanopore isoform-level “ISO” assay.

Long-read calibration for spatial gene editing events

To only keep high confidence base calls, the nanopore data were filtered by exploring the percentage of agreement between both sequencing methods as a function of long read number (RN) per molecule and nanopore consensus base quality value (Supplementary Table 5). Long-read molecules having a minimum read number of 3 (MINRN=3) and a base quality value at the editing position of 6 (MINQV=6) were chosen to be of sufficient quality for editing sites calling using the *SNPMatrix* sicelore-2.0 method.

Editing ratios

Samples CBS2 and MOB were used for calculating global editing ratios. Four capture spots in CBS2 (out of 2,499) and five capture spots in MOB (out of 918) with zero UMIs across all editing sites were excluded from the analysis. To test the significance of our findings, resampling of capture-spots across the sample were performed. Observed editing ratios per spot was kept and each spot were randomly assigned a region-label from the pool of original labels without replacement 10k times. A normal distribution was fitted to the simulated editing ratios to calculate the probability of observing a value equal to, or more extreme, than the observed value. For *Calm1*, the same approach was used with spots from CBS2.

In situ sequencing validation

Two 10 μm cryosections each of the olfactory bulb and coronal sections of the left hemisphere were placed on SuperFrost Plus microscope slides (ThermoFisher Scientific), stored at -80 °C and shipped on dry ice to CARTANA for library preparation, probe hybridization, probe ligation, rolling circle amplification, and fluorescence labeling using the HS Library Preparation Kit (P/N 1110) and for the in situ sequencing using the ISS kit (P/N 3110) and sequential imaging using a 20x objective. The result table of the spatial coordinates of each molecule of all targets together with the reference DAPI image per sample were provided by CARTANA. The list of transcripts that were investigated is listed in Supplementary Table 6.

Data availability

All relevant data have been deposited in Gene Expression Omnibus under accession number GSE153859 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153859>).

Code availability

All custom software used is available on Github (<https://github.com/ucagenomix/sicelore>). R figures and analysis scripts are available on Github (<https://github.com/ucagenomix/SiT>). Nanopore long reads .bam files are available within an interactive IGV web browser for visual inspection (<https://www.genomique.info/SiT/>). Seurat object .RDS files for the three samples are available on demand.

Acknowledgements

This project was supported by Institut National contre le Cancer (PLBIO2018-156), FRM (DEQ20180339158), the Inserm Cross-cutting Scientific Program HuDeCA 2018, the National Infrastructure France Génomique (Commissariat aux Grands Investissements, ANR-10-INBS-09-03, ANR-10-INBS-09-02), the Swedish Research Council, Swedish Foundation for Strategic Research, Horizon2020 HCA discovAIR, Knut and Alice Wallenberg Foundation (2018.0172), Erling-Persson Family Foundation (HDCA), and Science for Life Laboratory. We would like to thank the National Genomics Infrastructure (NGI) Sweden for providing infrastructure support. We thank Ludvig Bergenstråhle and Alma Andersson for advice and helpful discussions.

Contributions

K.T., A.M., R.W. performed the experiments. K.L, J.B. analyzed the data. P.B., R.W. and J.L. supervised the research. All authors contributed to the writing of the manuscript.

Competing Interests

J.L., J.B. and K.T. are advisors to 10x Genomics Inc, which holds IP rights to the ST technology. J.B. is a shareholder of Cartana AB.

References

1. Asp M, Bergenstråhle J, Lundeberg J. **Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration.** *BioEssays*, 1900221.
2. Lundin E, Wu C, Widmark A, Behm M, Hjerling-Leffler J, Daniel C, Öhman M, Nilsson M. **Spatiotemporal mapping of RNA editing in the developing mouse brain using in situ sequencing reveals regional and cell-type-specific regulation.** *BMC Biology* 2020, 18:6.
3. Erben L, He MX, Laeremans A, Park E, Buonanno A. **A Novel Ultrasensitive In Situ Hybridization Approach to Detect Short Sequences and Splice Variants with Cellular Resolution.** *Mol Neurobiol* 2018, 55:6169-6181.
4. Lebrigand K, Magnone V, Barbry P, Waldmann R. **High throughput, error corrected Nanopore single cell transcriptome sequencing.** *Nature Communications.* 2020, 11: 4025.

5. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, et al. **Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.** *Science* 2016, 353:78.
6. Nave KA. **Myelination and support of axonal integrity by glia.** *Nature* 2010 468:244-52.
7. Tepe B, Hill MC, Pekarek BT, Hunt PJ, Martin TJ, Martin JF, Arenkiel BR. **Single-Cell RNA-Seq of Mouse Olfactory Bulb Reveals Cellular Heterogeneity and Activity-Dependent Molecular Census of Adult-Born Neurons.** *Cell Reports* 2018, 25:2689-2703.e2683.
8. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. **Comprehensive Integration of Single-Cell Data.** *Cell* 2019, 177:1888-1902.e1821.
9. Irfan M, Gopaul KR, Miry O, Hökfelt T, Stanton PK, Bark C. **SNAP-25 isoforms differentially regulate synaptic transmission and long-term synaptic plasticity at central synapses.** *Scientific Reports* 2019, 9:6403.
10. Behm M, Öhman M. **RNA Editing: A Contributor to Neuronal Dynamics in the Mammalian Brain.** *Trends in Genetics* 2016, 32:165-175.
11. Licht K, Kapoor U, Amman F, Picardi E, Martin D, Bajad P, Jantsch MF. **A high resolution A-to-I editing map in the mouse identifies editing events controlled by pre-mRNA splicing.** *Genome Res* 2019, 29:1453-1463.
12. Ramaswami G, Li JB. **RADAR: a rigorously annotated database of A-to-I RNA editing.** *Nucleic Acids Research* 2013, 42:D109-D113.
13. Hanson PI, Schulman H. **Neuronal Ca²⁺/Calmodulin-Dependent Protein Kinases.** *Annu. Rev. Biochem.* 1992. 61:559-601.
14. Zhou Z-Y, Hu Y, Li A, Li Y-J, Zhao H, Wang S-Q, Otecko NO, Zhang D, Wang J-H, Liu Y, et al. **Genome wide analyses uncover allele-specific RNA editing in human and mouse.** *Nucleic Acids Research* 2018, 46:8888-8897.
15. Vaser R, Sović I, Nagarajan N, Šikić M. **Fast and accurate de novo genome assembly from long uncorrected reads.** *Genome research* 2017, 27:737-746.