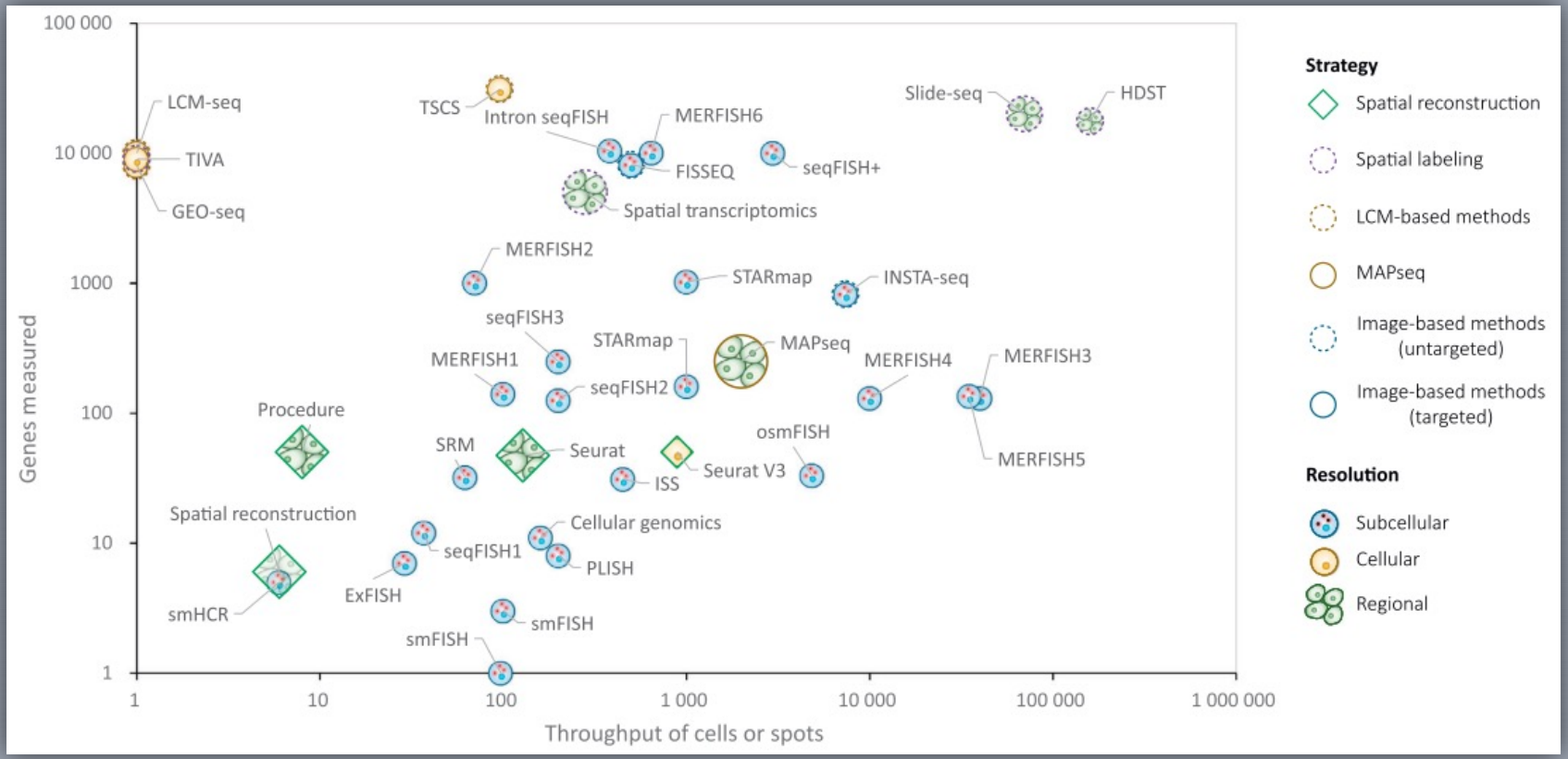
Spatial isoform Transcriptomics (SiT)

Kévin Lebrigand
UCAGenomix, Nice-Sophia-Antipolis
lebrigand@ipmc.cnrs.fr
@kevinlebrigand

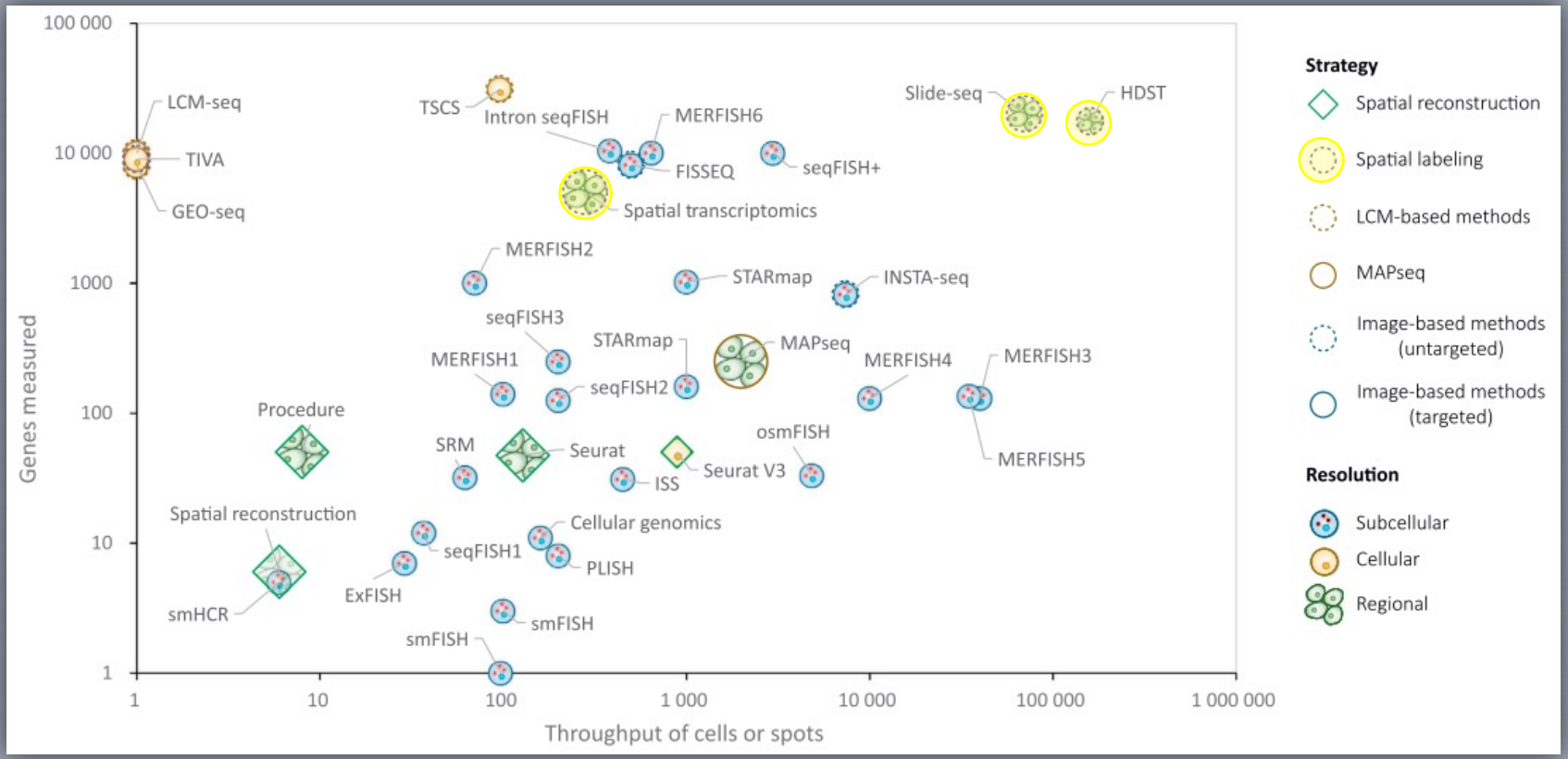
Spatial Transcriptomics approaches



Throughput of Genes and Cells for Spatially Resolved Transcriptomics approaches

Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics
Liao et al., 2020

Spatial Transcriptomics approaches



Throughput of Genes and Cells for Spatially Resolved Transcriptomics approaches

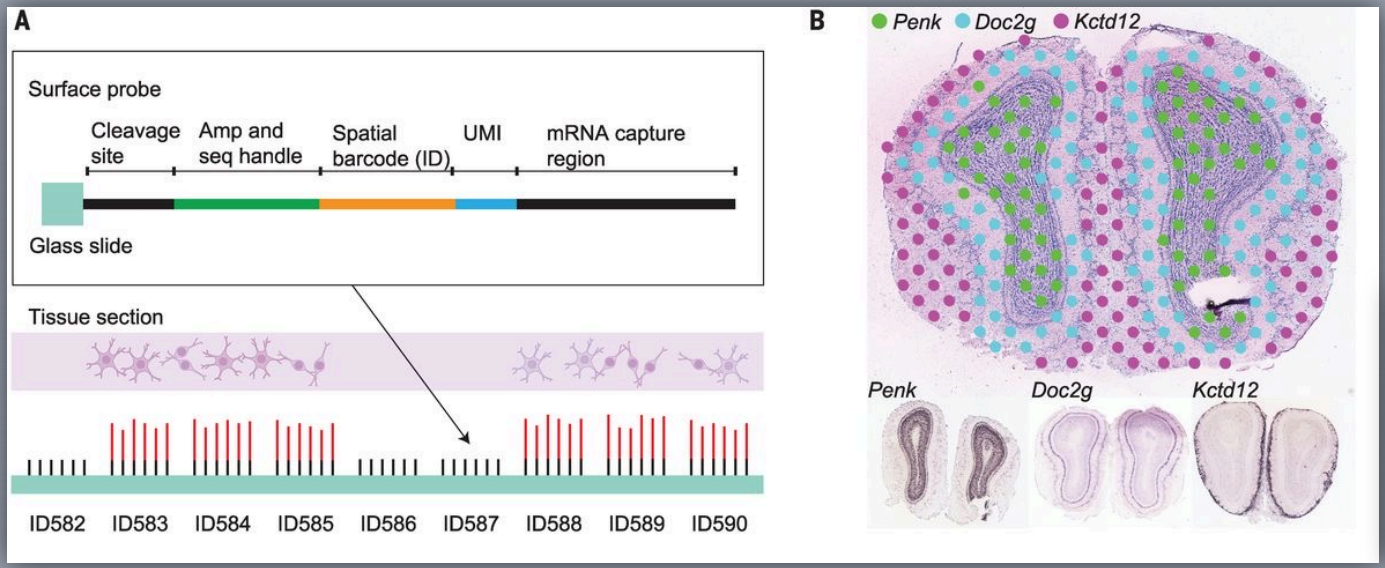
Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics
Liao et al., 2020

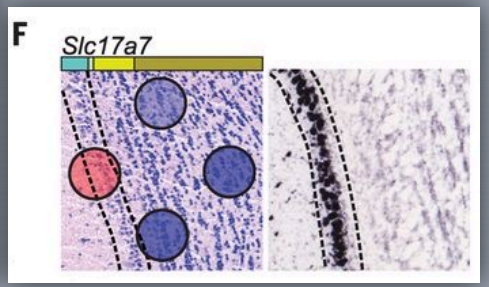
In-situ capture spatial transcriptomics (Stahl et al. Science, 2016)

Visualization and analysis of gene expression in tissue sections by spatial transcriptomics

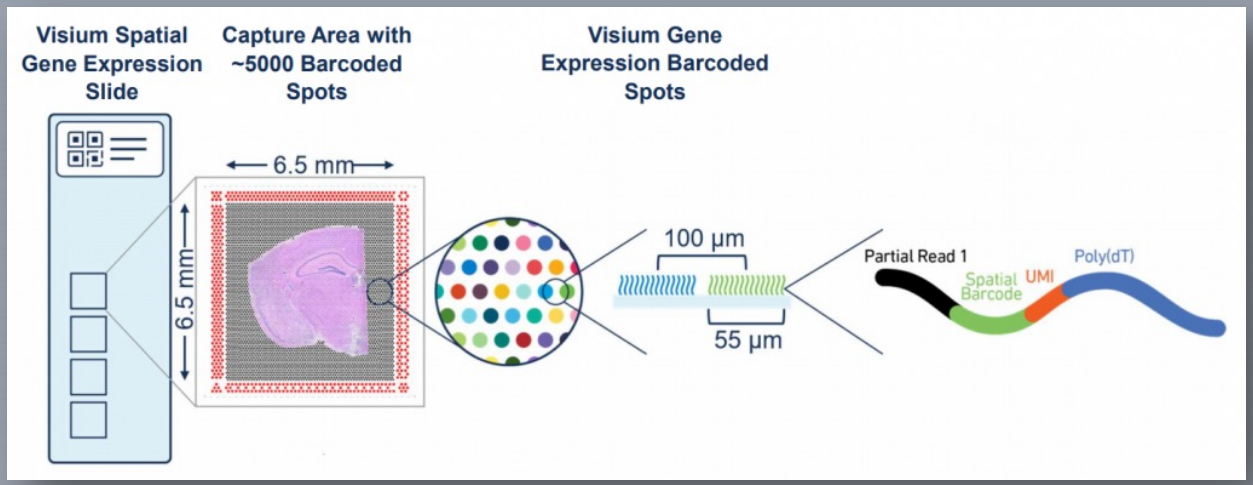
PATRIK L. STÅHL, FREDRIK SALMÉN, SANJA VICKOVIC, ANNA LUNDMARK, JOSÉ FERNÁNDEZ NAVARRO, JENS MAGNUSSON, STEFANIA GIACOMELLO, MICHAELA ASP, JAKUB O. WESTHOLM, MIKAEL HUSS, ANNELIE MOLLBRINK, STEN LINNARSSON, SIMONE CODELUPPI, ÅKE BORG, FREDRIK PONTÉN, PAUL IGOR COSTEA, PELIN SAHLÉN, JAN MULDER, OLAF BERGMANN, JOAKIM LUNDEBERG , AND JONAS FRISÉN fewer Authors Info & Affiliations

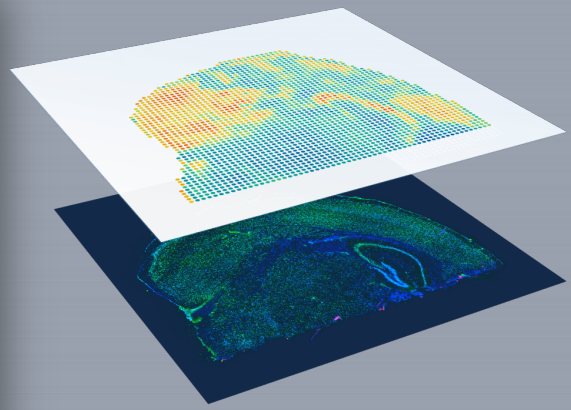
Spatial RNA-seq with position-indexed RT primer arrays on slides (spatial transcriptomics)
200 million oligos, 1007 features (100µm),
~5.000 genes per feature.

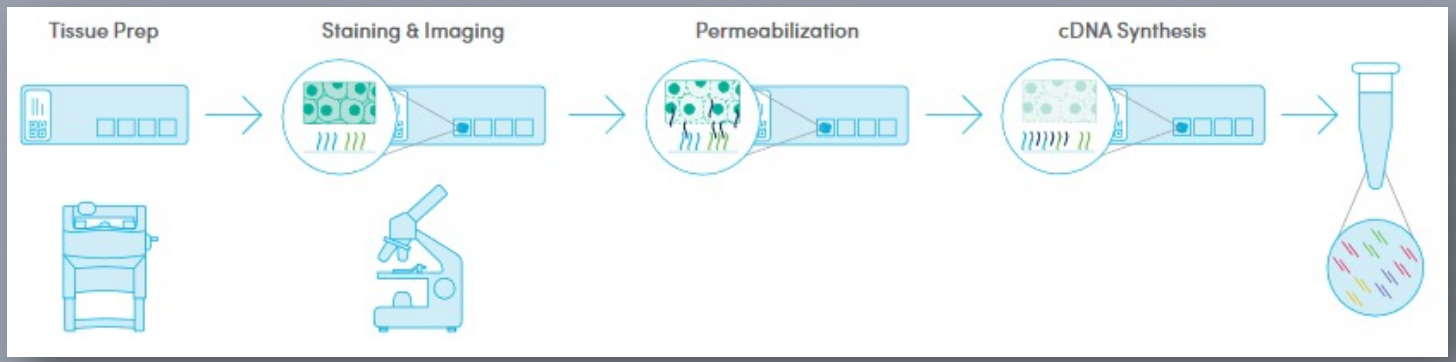




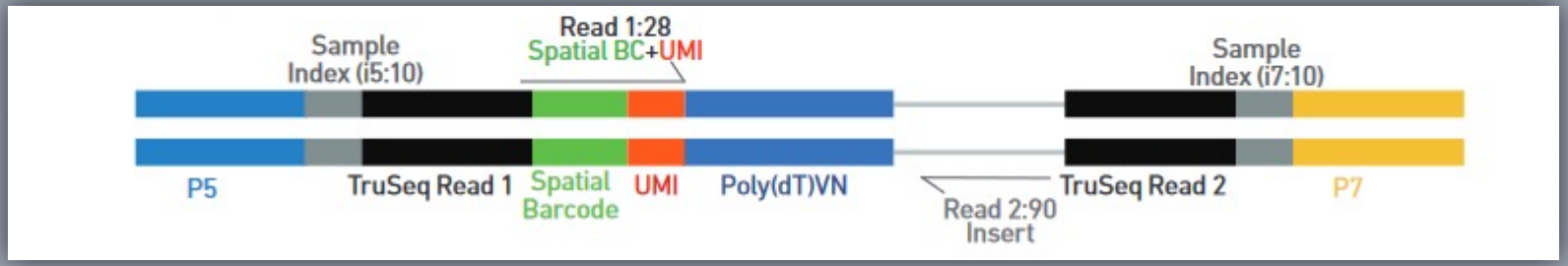
10x Genomics Visium Spatial transcriptomics (2019)

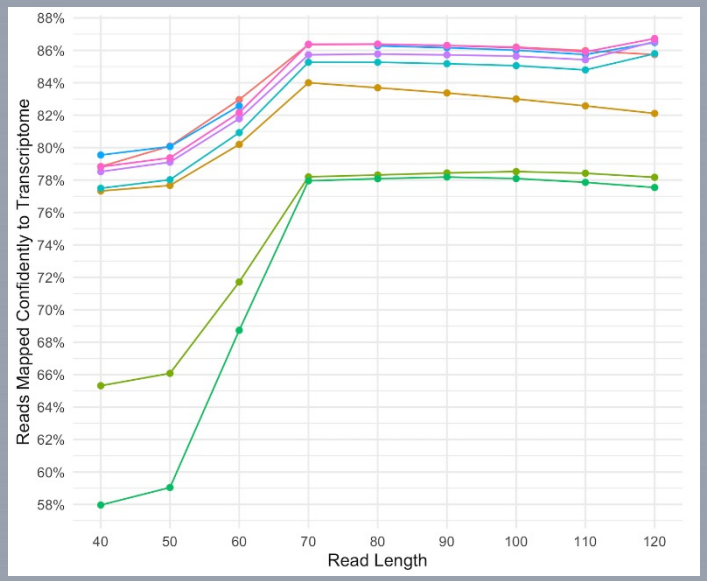






Library sequencing

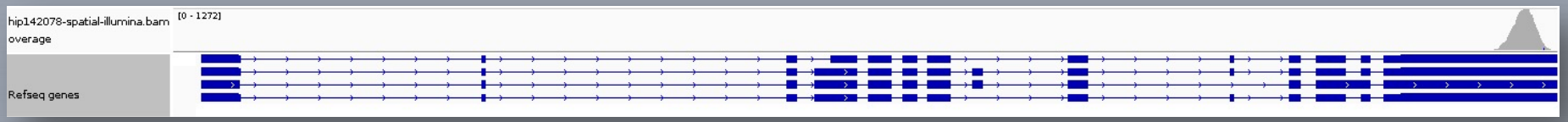




A full length cDNA construct is flanked by the 30 bp template switch oligo (TSO) sequence, AAGCAGTGGTATCAACGCAGAGTACATGGG, on the 5' end and poly-A on the 3' end. Some fraction of sequencing reads are expected to contain either or both of these sequences, depending on the fragment size distribution of the sequencing library. Reads derived from short RNA molecules are more likely to contain TSO and/or poly-A sequence than are longer RNA molecules.

Spatial transcriptomics coverage

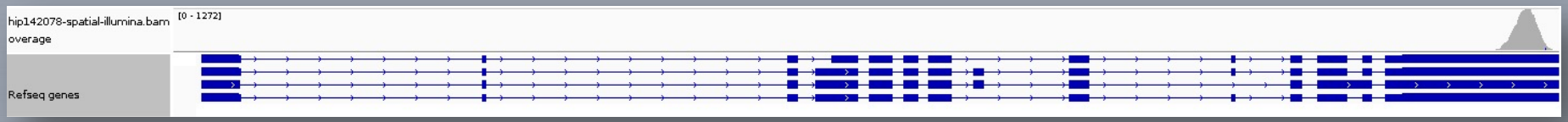
Illumina short-read



PTBP1

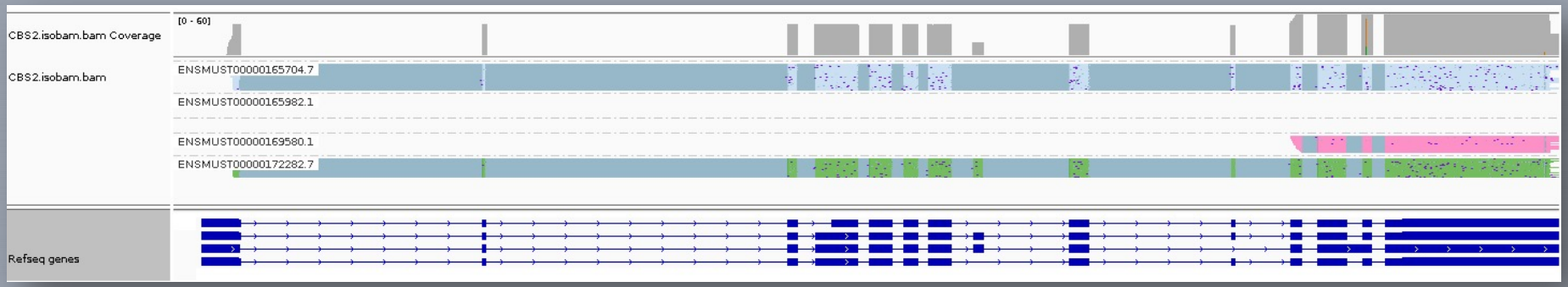
Spatial transcriptomics coverage

Illumina short-read



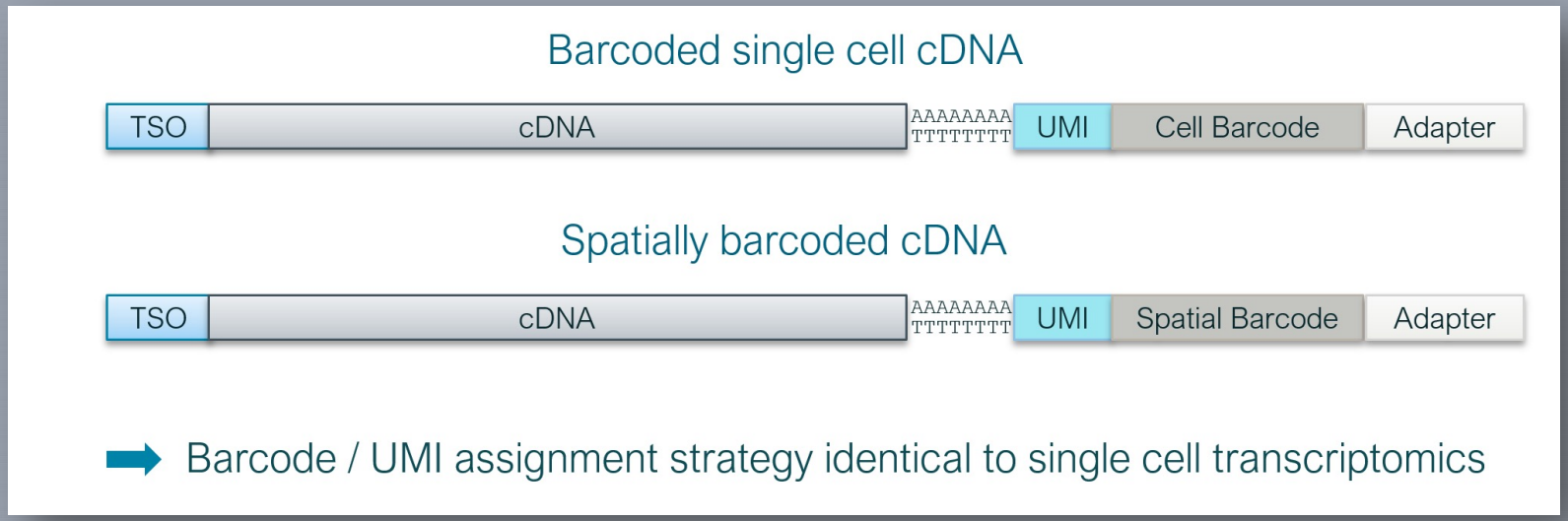
PTBP1

Nanopore long-read



PTBP1

Same library structure as droplet-based single cell transcriptomics



Spatial isoform Transcriptomics (SiT)

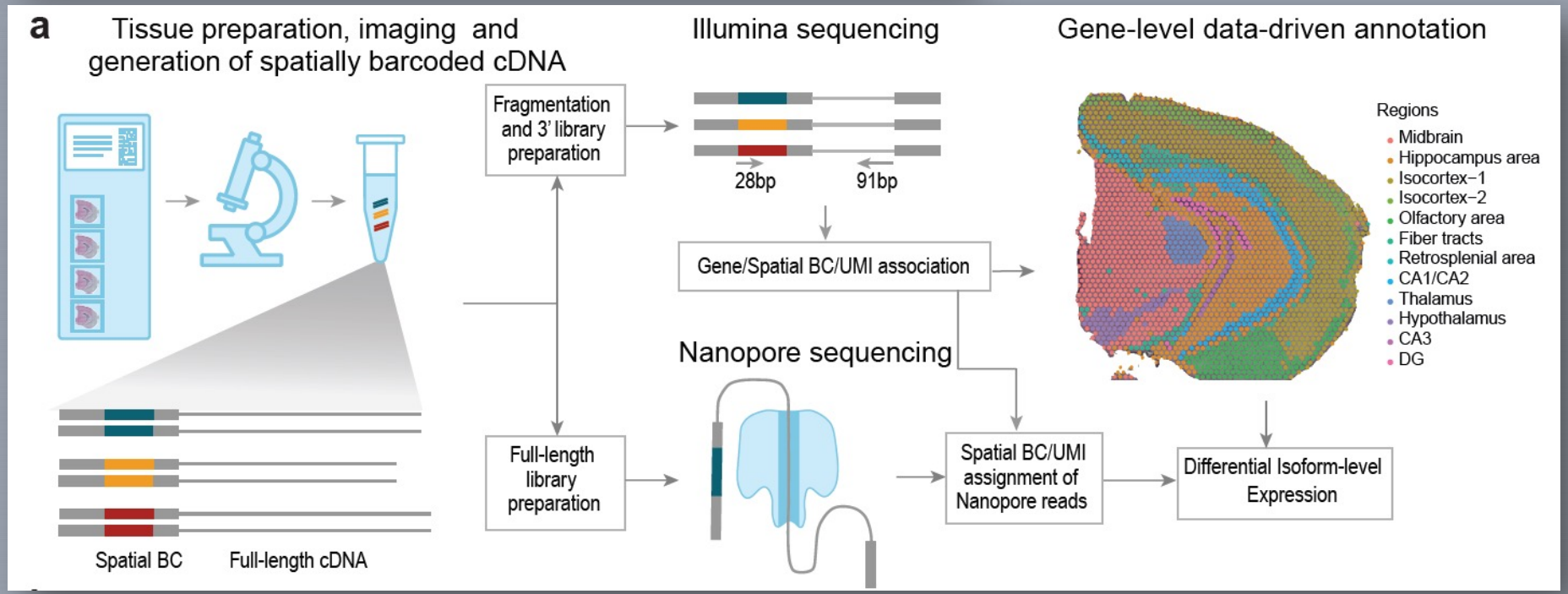
The spatial landscape of gene expression isoforms in tissue sections

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

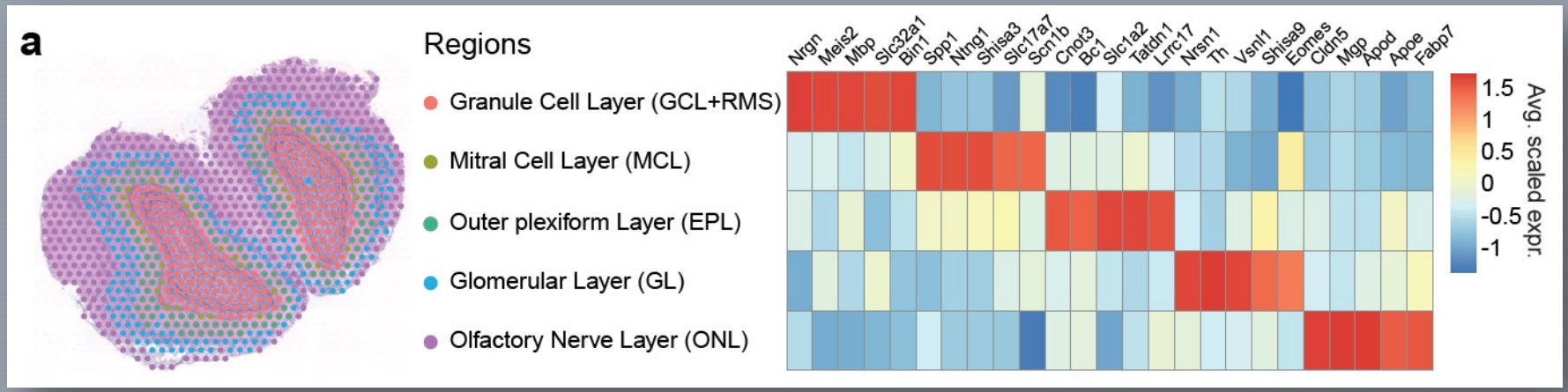
doi: https://doi.org/10.1101/2020.08.24.252296

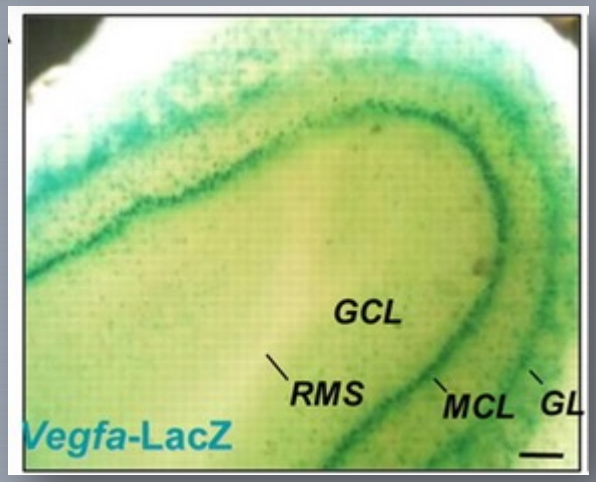
bioRχiv THE PREPRINT SERVER FOR BIOLOGY

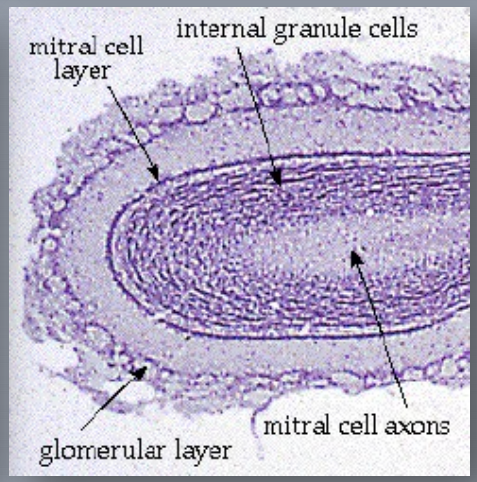
August 24, 2020
January 5, 2022



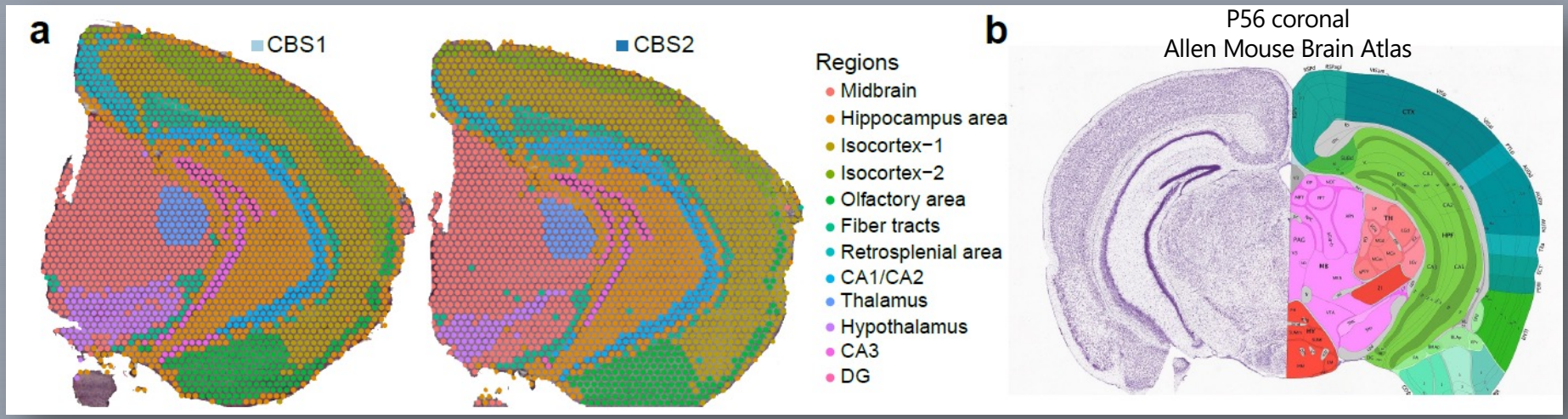
Mouse olfactory Bulb (MOB) region annotation using short-read

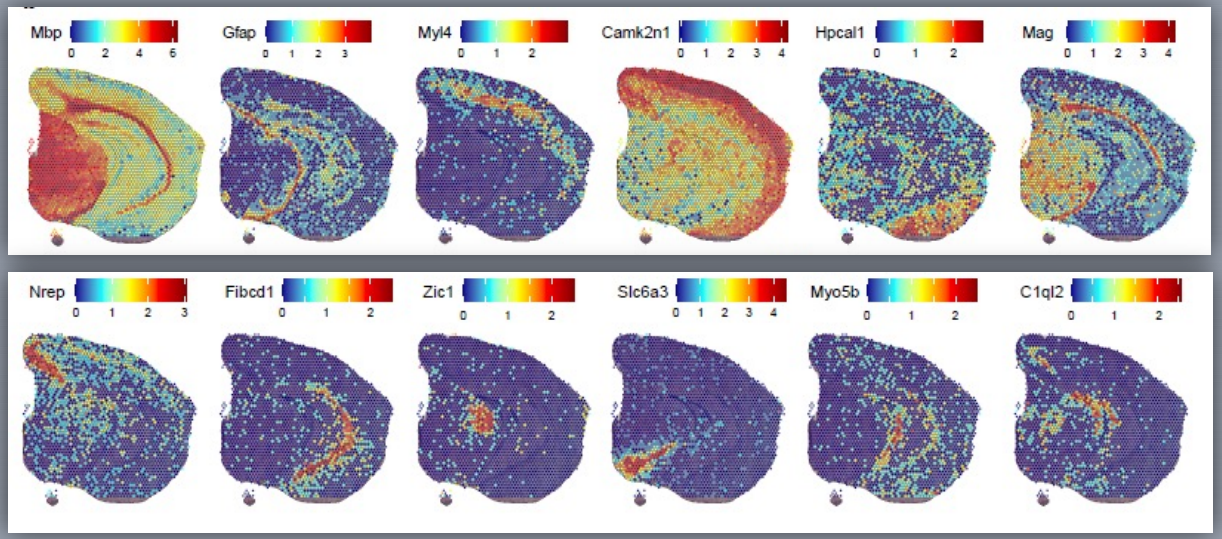




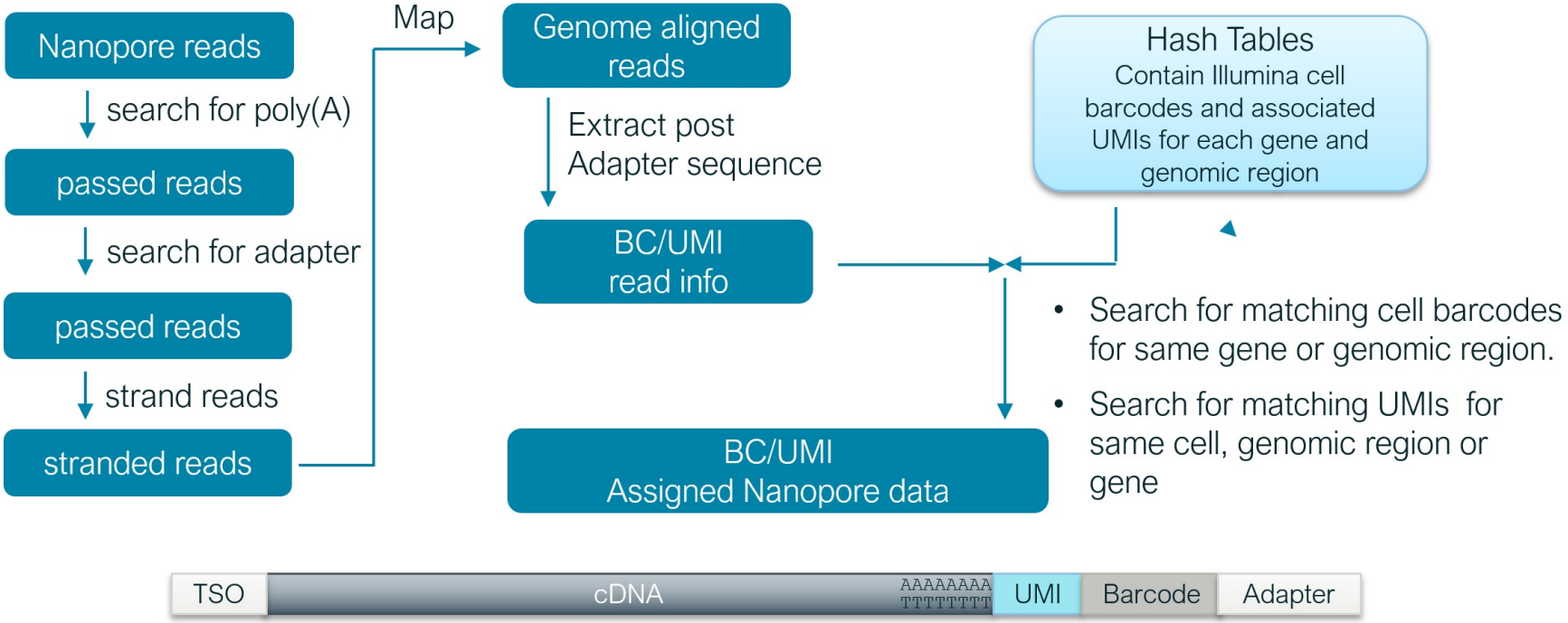


Mouse coronal brain sections (CBS) region annotation using short-read





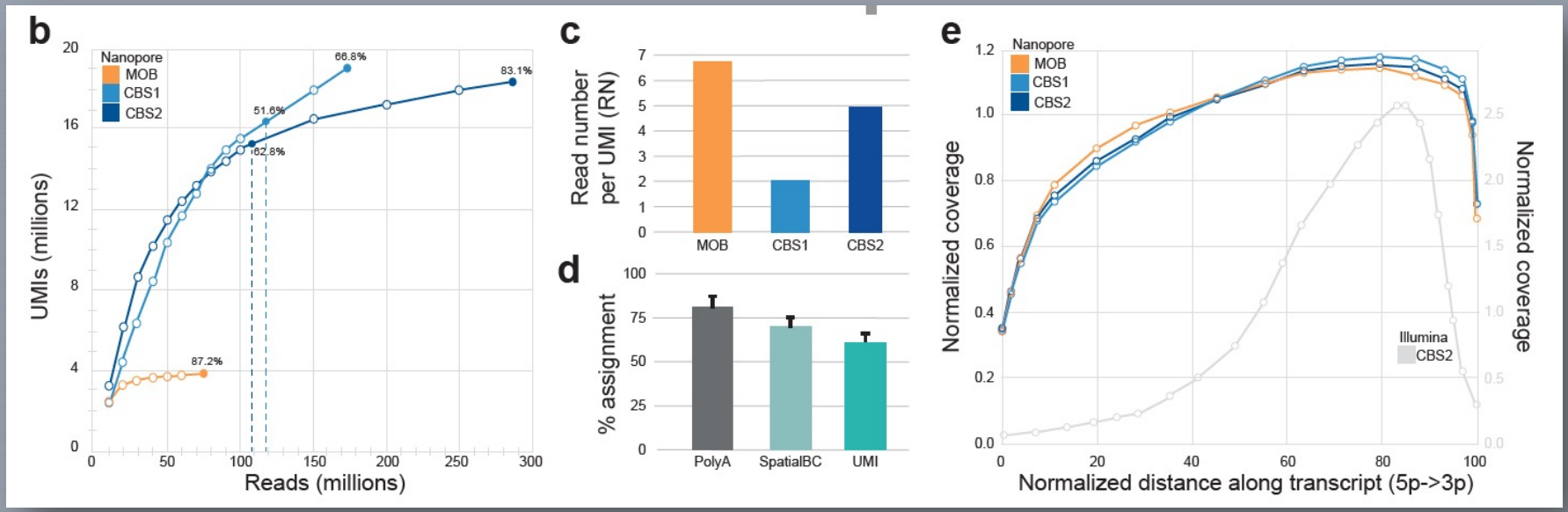
Illumina-guided barcode and UMI assignment



Lebrigand, K., Magnone, V., Barbry, P. & *Waldmann. R.*
High throughput error corrected Nanopore single cell transcriptome sequencing.
Nat Commun **11,** 4025 (2020).

https://github.com/ucagenomix/sicelore

Spatial Isoform Transcriptomics (SiT)

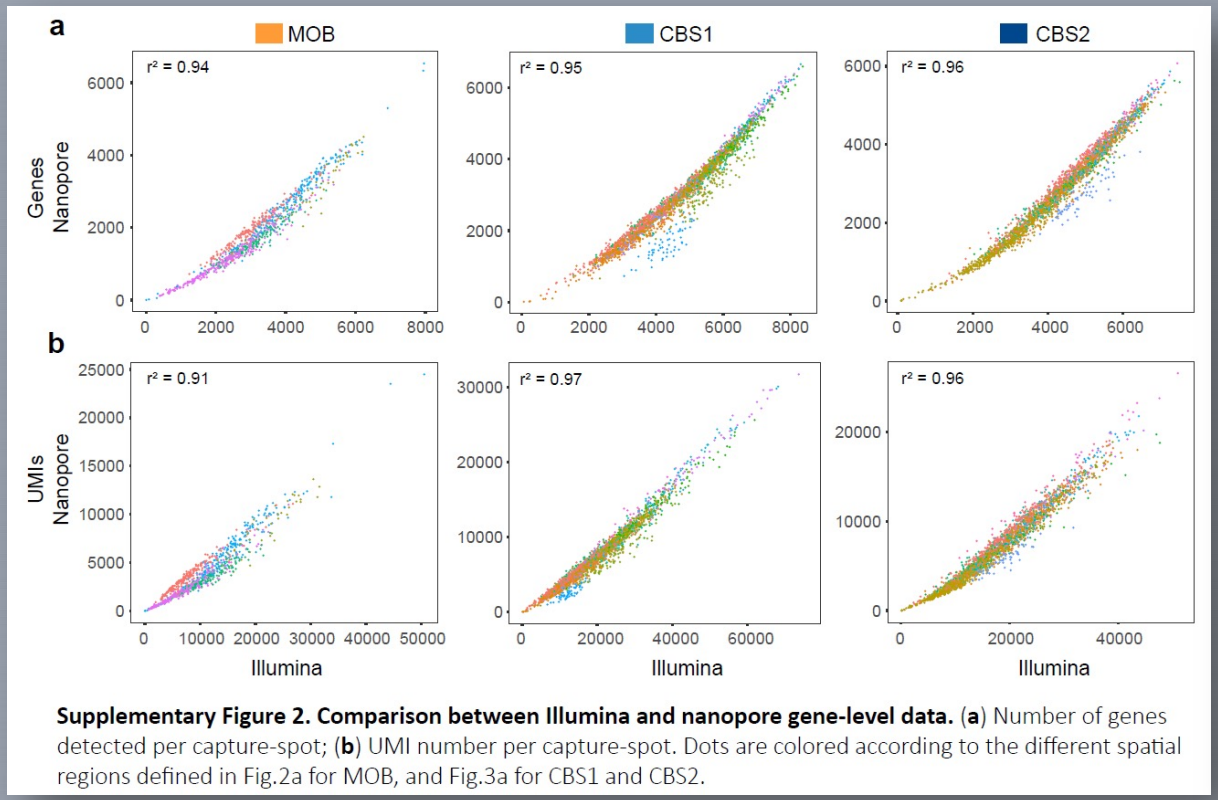


Reads	MOB		CBS1			CBS2											
Date	18 feb. 20	20 mar. 20	18 feb. 20	20 mar. 20	24 feb. 21	12 may 20	13 may 20	19 may 20	25 may 20	25 may 20	26 may 20	27 may 20	09 feb. 21	Total		
Flow cells	PAE06474	PAE59649	PAE01745	PAE59645	PAG52067	PAE59606	PAE59231	PAE32756	PAE32753	PAE31188	PAE21339	PAD99555	PAG56368	13		
Total reads (fastq_pass)	27628000	47272000	24980000	31736000	117280000	22897702	30405384	27492770	18534938	31506774	19108718	25596387	110916000	535354673	%age	
PolyA and Adapter found reads	21318117	47970311	17980183	27286678	80516212	18536047	25199992	22871198	16088962	26777546	15983663	21682530	85837208	428048647	79,96	of Total passed reads
SpatialBC found reads	14506264	29316718	12554655	19051597	54323311	14613934	19867830	14666481	11403706	19099469	11266930	14090779	60154119	294915793	68,90	of PolyA found reads
UMIs found reads	10445006	19328468	7323748	10517081	27584331	8616415	11714126	9347072	7557944	12657620	7448718	9031708	34225619	175797856	59,61	of SpatialBC found reads

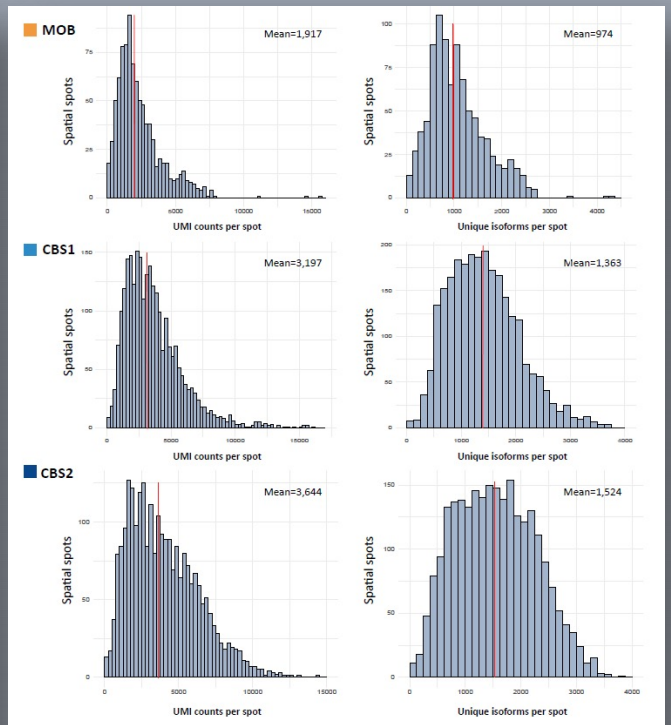
CBS1: One flow cell, 117 M reads → 51.6% sequencing saturation
CBS2: One flow cell, 111 M reads → 62.2% sequencing saturation

→ 1 – 2 Promethion flow cells per slice

Short-read / Long-read profile comparison



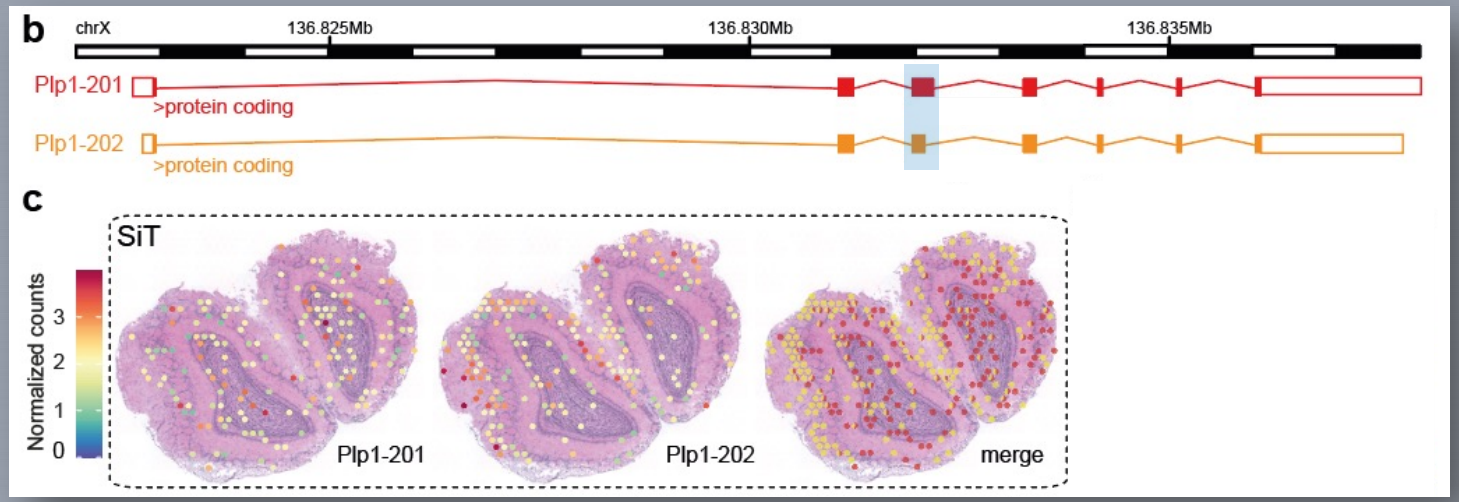
Supplementary Figure 2. Comparison between Illumina and nanopore gene-level data. (**a**) Number of genes detected per capture-spot; (**b**) UMI number per capture-spot. Dots are colored according to the different spatial regions defined in Fig.2a for MOB, and Fig.3a for CBS1 and CBS2.



Supplementary Figure 3. Isoform UMI counts and unique isoforms per capture spot. Means are indicated by red vertical lines.

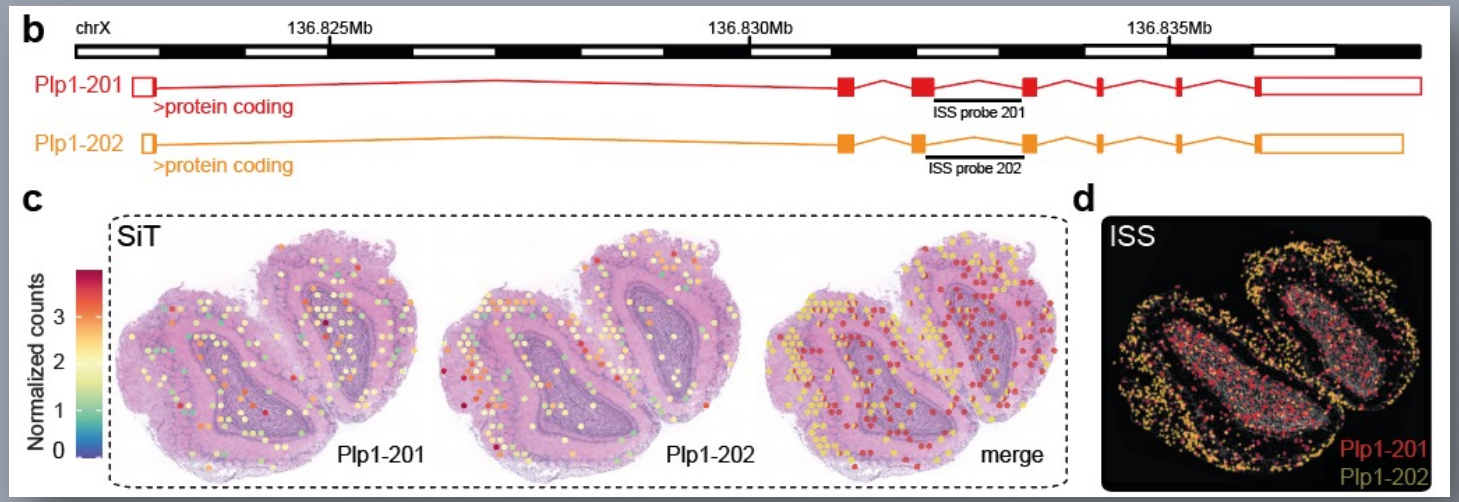
Plp1 Differential Transcript Usage (DTU) across regions (MOB)

Proteolipid protein 1 (Plp1), a gene involved in severe pathologies associated with CNS dysmyelination



Plp1 Differential Transcript Usage (DTU) across regions (MOB)

Proteolipid protein 1 (Plp1), a gene involved in severe pathologies associated with CNS dysmyelination

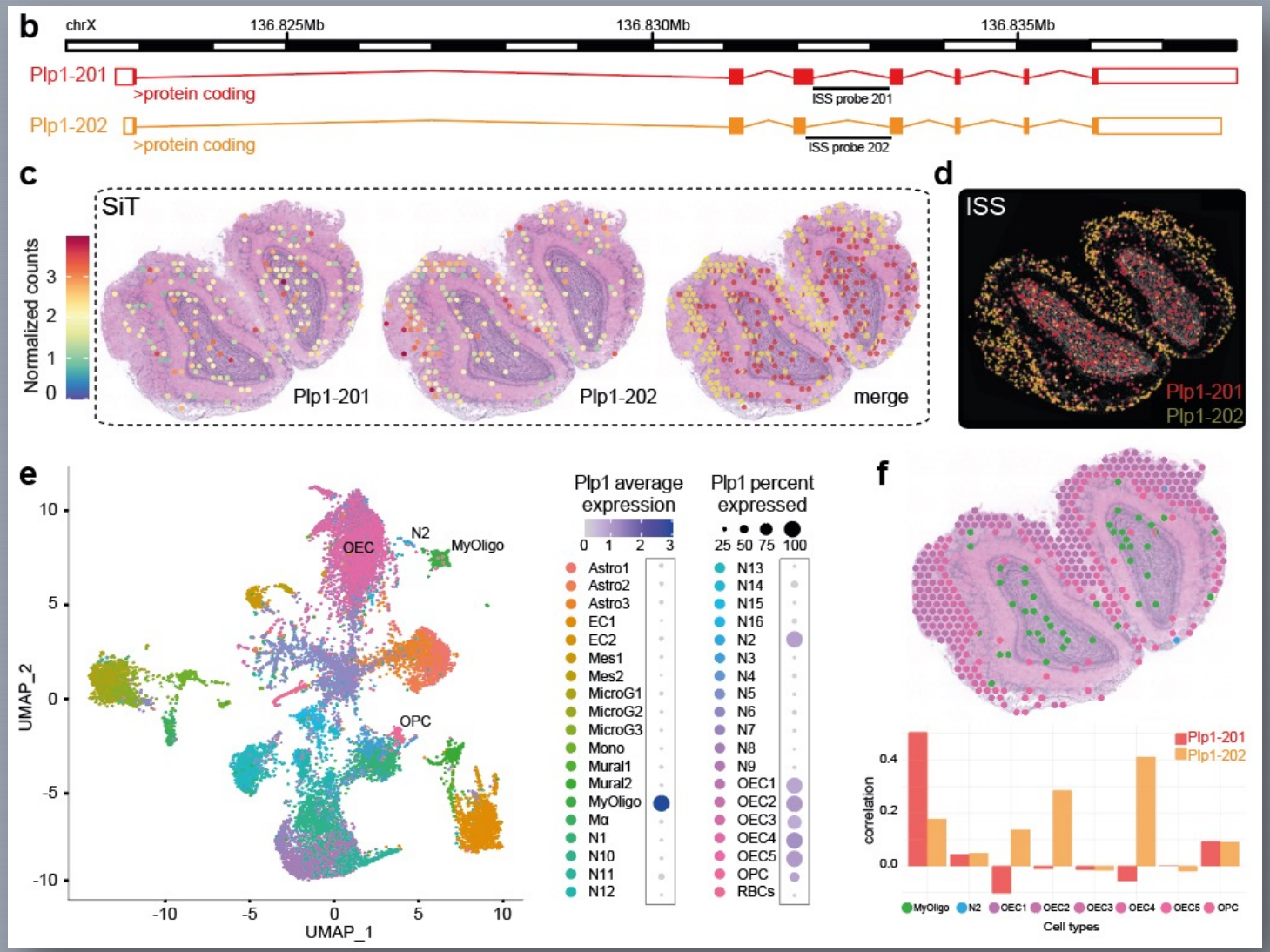


CARTANA
Part of 10x Genomics

***In Situ* Sequencing Data**

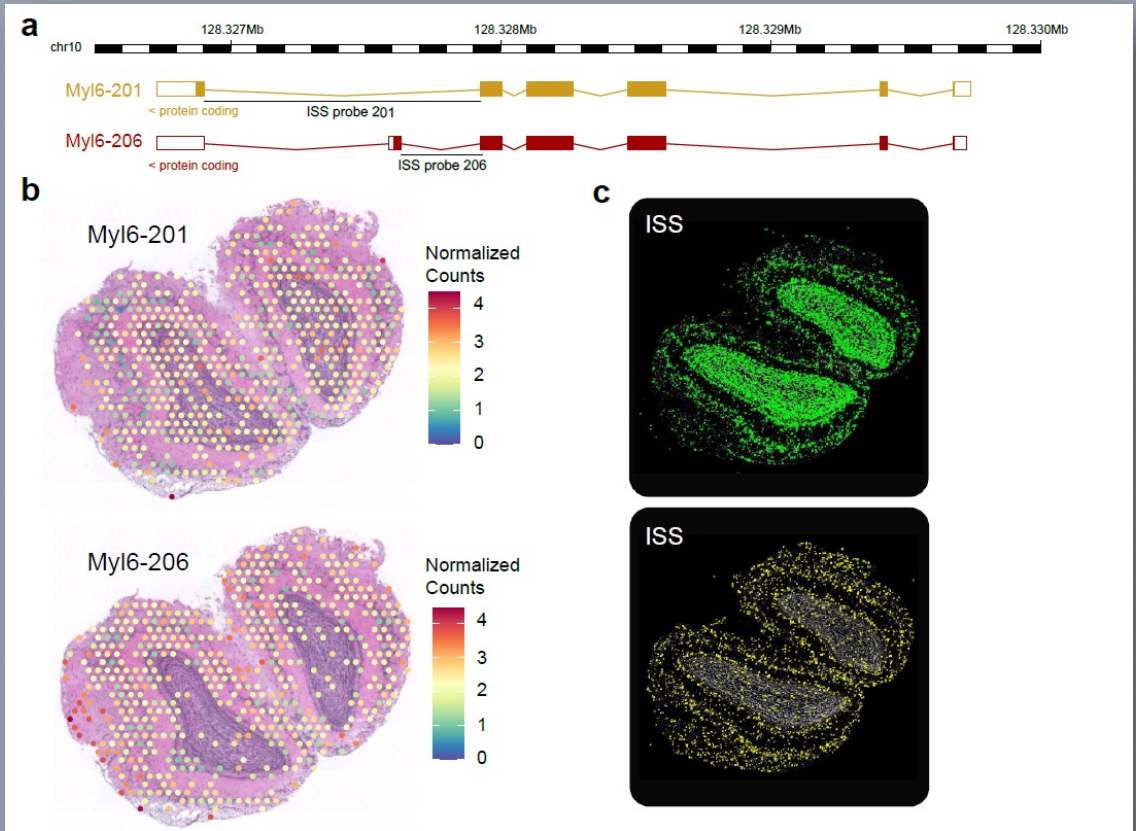
Cell type deconvolution using single cell external dataset (Tepe et al., 2018)

Proteolipid protein 1 (Plp1), a gene involved in severe pathologies associated with CNS dysmyelination



Spatial spot deconvolution of prominent *Plp1* expresser cell types. Satial spot correlation observed between deconvolution score and *Plp1* isoforms expression. Results show that *Plp1* is predominantly expressed by olfactory ensheathing cells (OEC) in the ONL and by myelinating-oligodendrocytes (MyOligo) in the GCL.

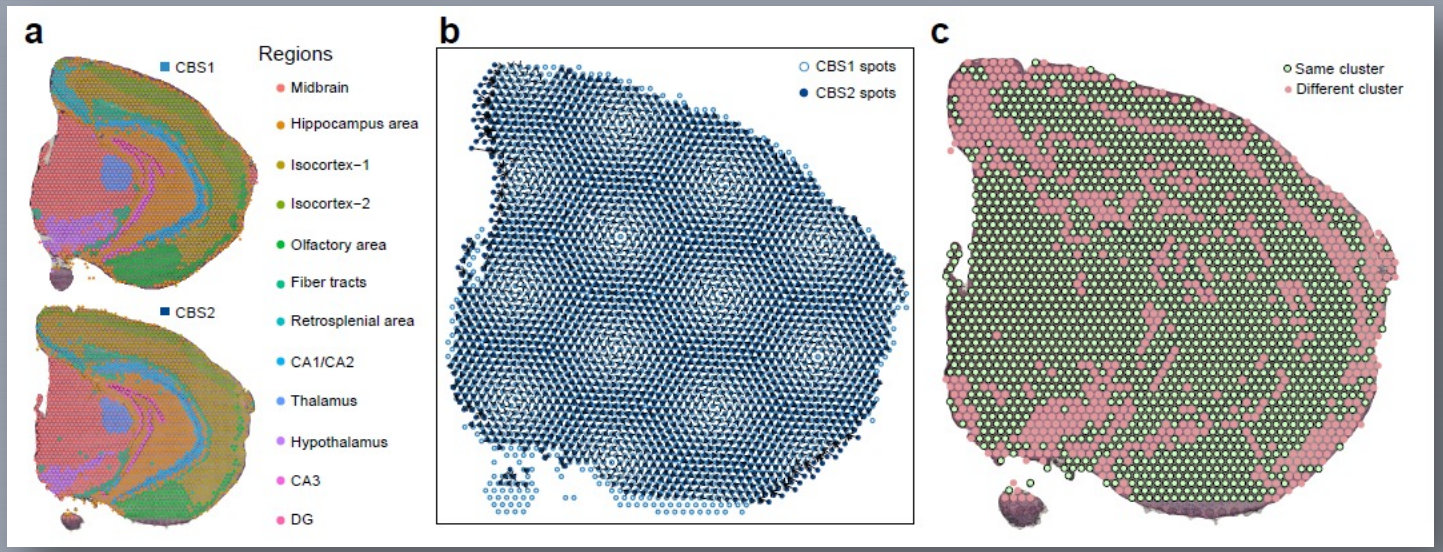
Myl6 Differential Transcript Usage (DTU) across regions (MOB)

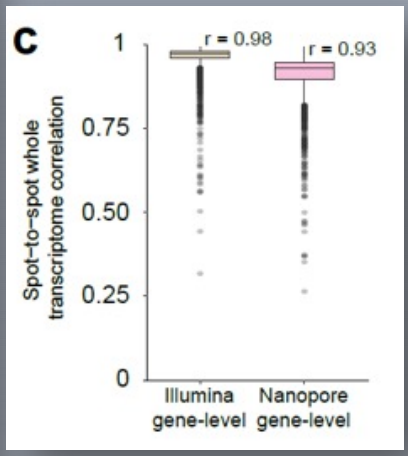


Supplementary Figure 4. Mouse Olfactory Bulb (MOB) Myl6 isoform expression. (**a**) Schematic view of *Myl6* gene locus (mm10 build coordinates). (**b,c**) Expression of *Myl6* isoforms detected by SiT (b) and ISS (c). ISS isoform specific probes are indicated in panel (a).

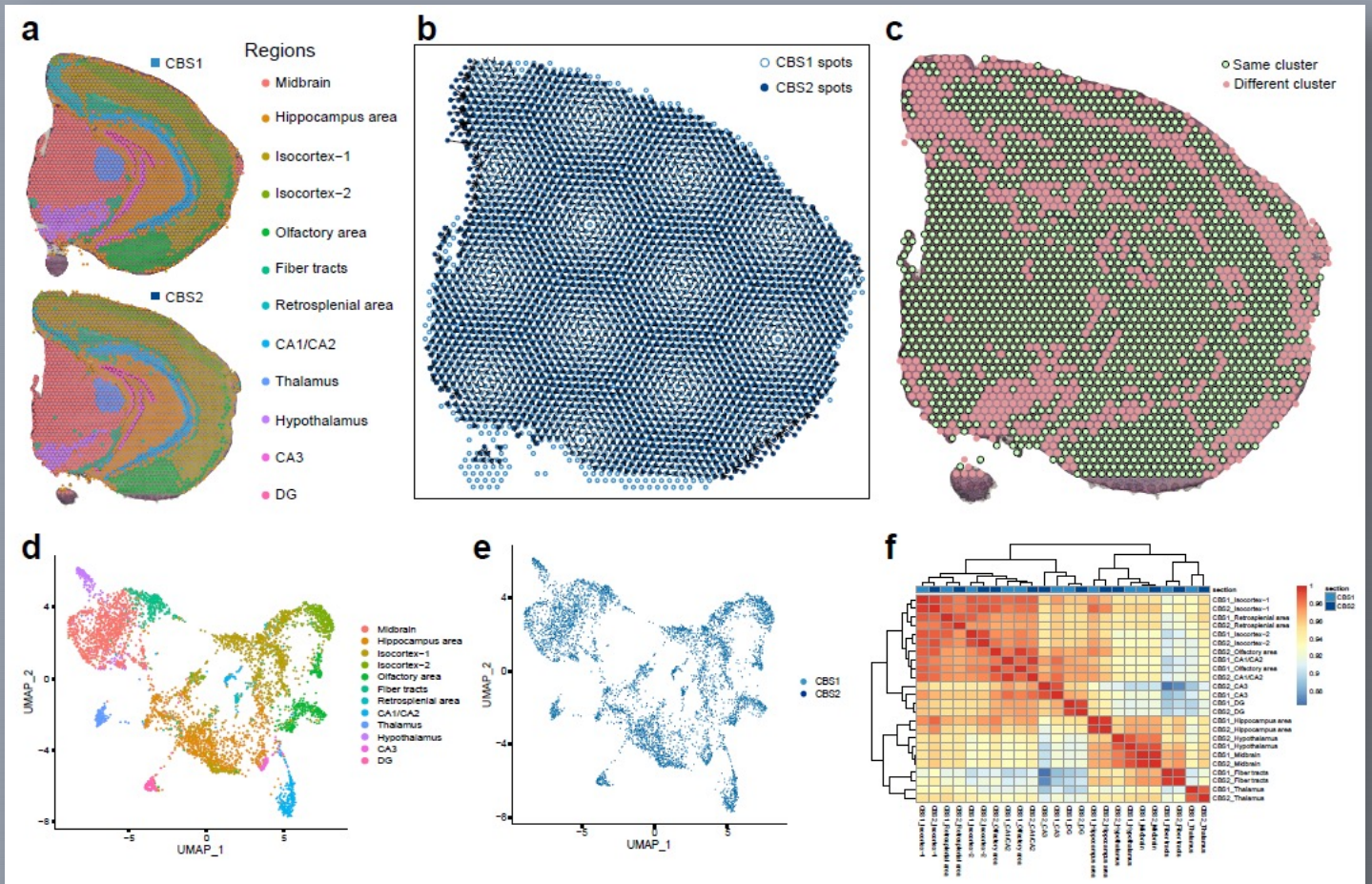
Myosin Light Chain 6 (Myl6), codes for the non-phosphorylatable alkali light chain component of the hexameric Myosin motor protein, that has been shown to be involved in neuronal migration and synaptic remodeling in immature and mature neurons

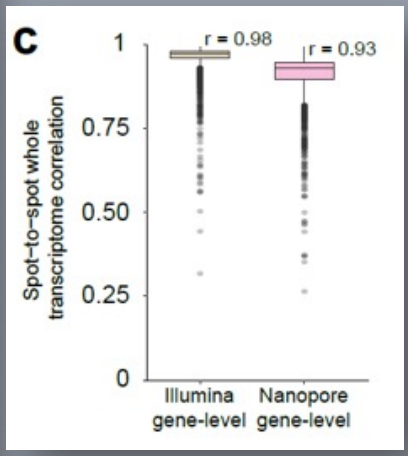
SiT robustness assessment using two coronal brain section





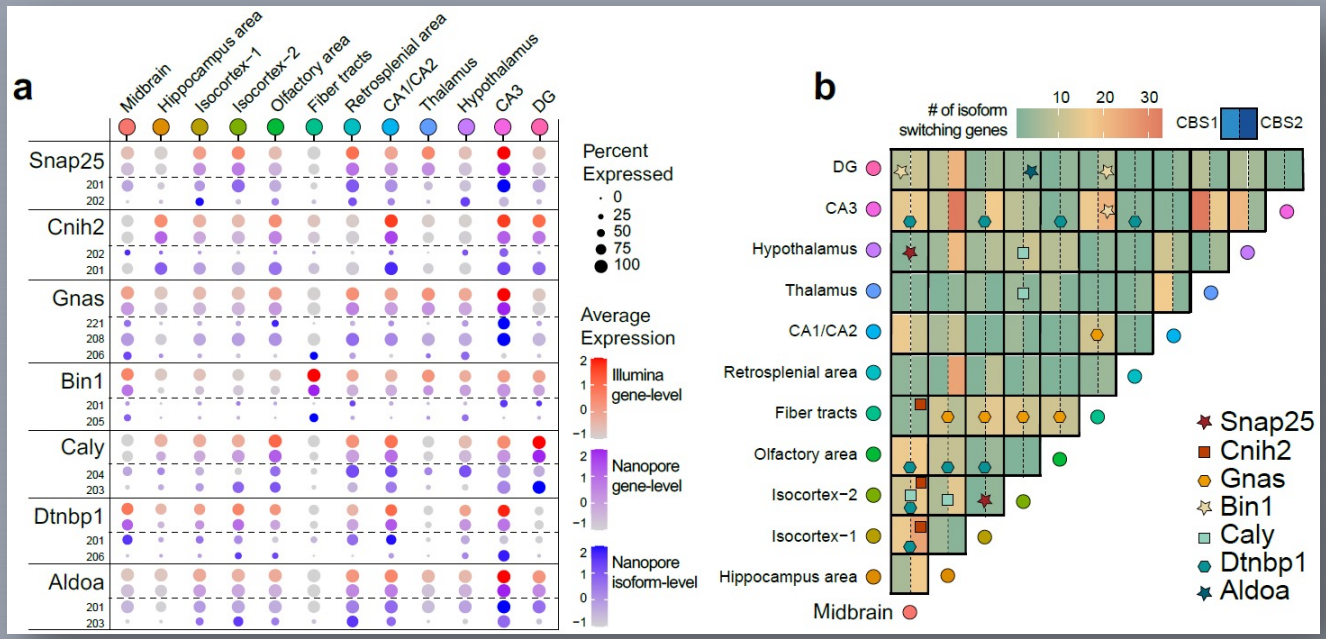
SiT robustness assessment using two coronal brain section

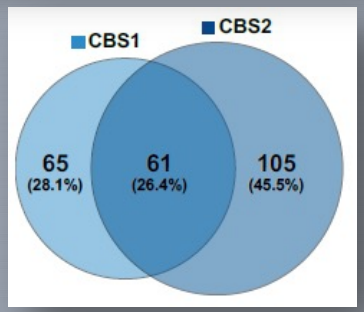




Supplementary Figure 7. (**a**) Data-driven annotation of the two coronal brain sections through transcriptome clustering of short-read data. (**b**) Correspondance between pairs of spatial spots from CBS1 and CBS2 after image alignment and distance minimization. (**c**) Visualization of pairs of spots belonging to same and different annotated regions in the two coronal brain sections.UMAP plot coronal brain section CBS1 and CBS2 integration using Nanopore isoform-level assay (ISO) colored by cluster labels (**d**) and section of origin (**e**). (**f**) Heatmap of Pearson correlation score between cells grouped by coronal brain section and clusters.

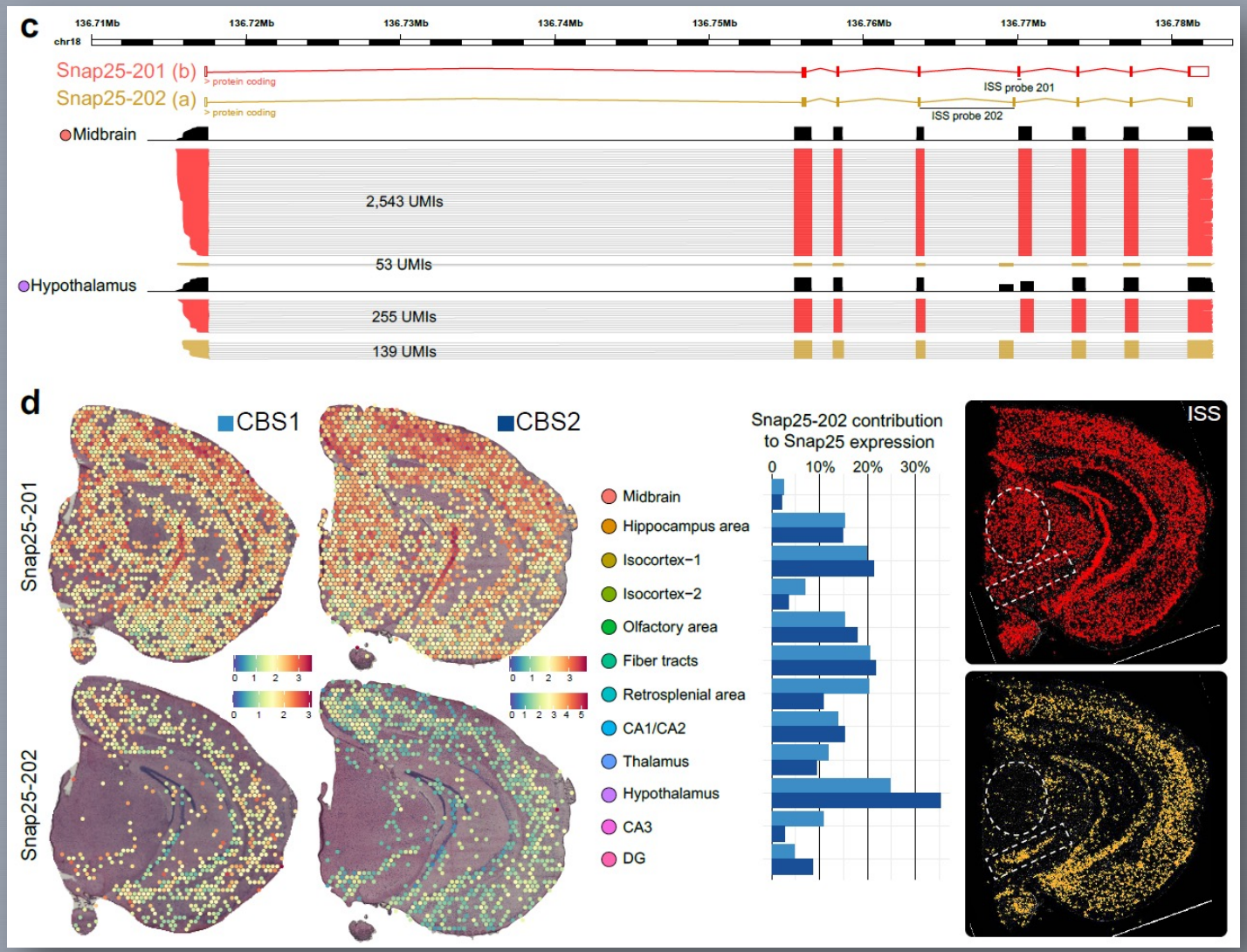
Differential Transcript/Isoform Usage (DTU) across regions (CBS)



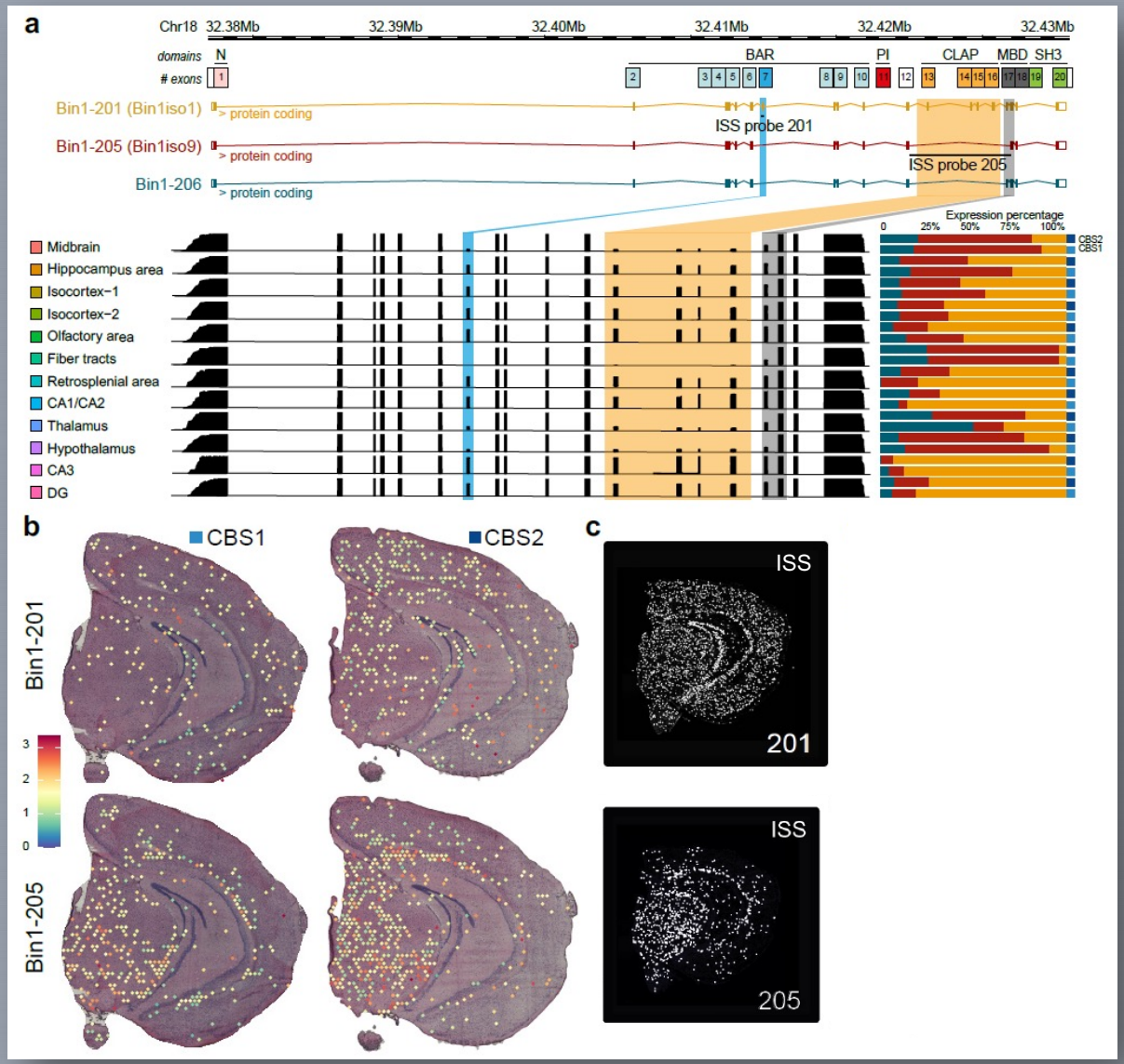


61 common switching genes in "CBS1" and "CBS2"					
Anapc11	Capzb	Eno2	Kctd13	Pcp4	Rpn1
Ap2m1	Cck	Ensa	Kctd17	Pnkd	Rps24
App	Cdc42	Faim	Mff	Polr2g	Rps6
Arl2bp	Clta	Fam173a	Mrpl48	Polr2h	S100a16
Atp5g1	Cltb	Fis1	Mrpl55	Ppp1r1a	Sept8
Bbip1	Cnih1	Fkbp8	Myl6	Ppp3ca	Sft2d1
BC031181	Cspg5	Ftl1	Nbdy	Psme2	Slc3a2
Bdnf	Dbndd2	Gap43	Ndrg4	Rexo2	**Snap25**
Bin1	Dctn6	**Gnas**	Ngrn	Rpl13a	Tpm1
Caly	Dtnbp1	Hsd11b1	Nkain4	Rpl5	Tsc22d3
					U2af1

Snap25 DTU across regions (CBS)

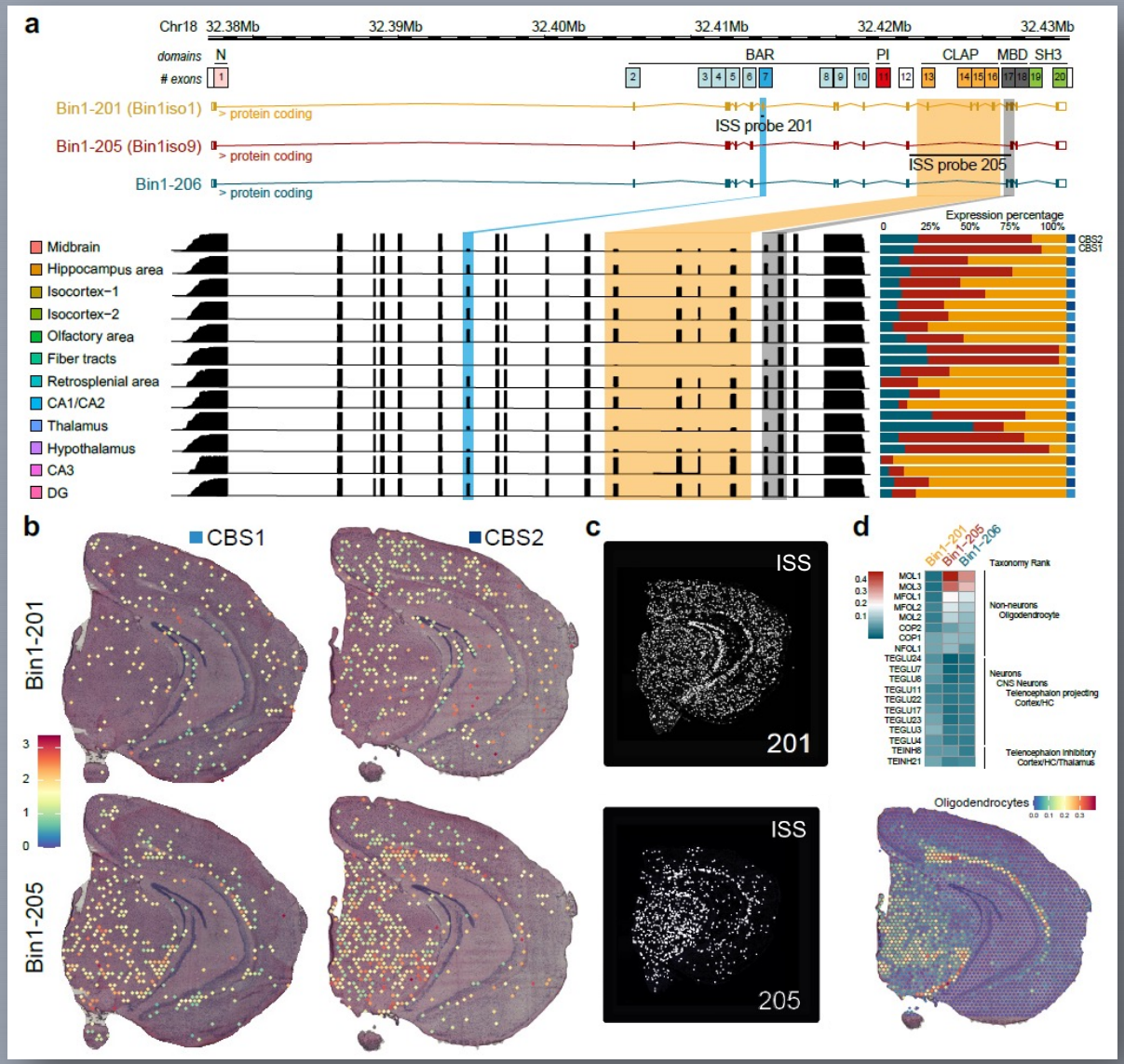


Bin1 DTU across regions (CBS)



Bin1 locus has been identified as a leading modulator of genetic risk in Alzheimer's disease.

Bin1 DTU across regions (CBS)



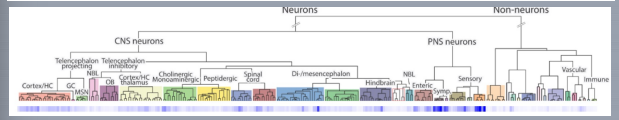
Bin1 locus has been identified as a leading modulator of genetic risk in Alzheimer's disease.

Cell
Volume 174, Issue 4, 9 August 2018, Pages 999-1014.e22
CellPress

Resource

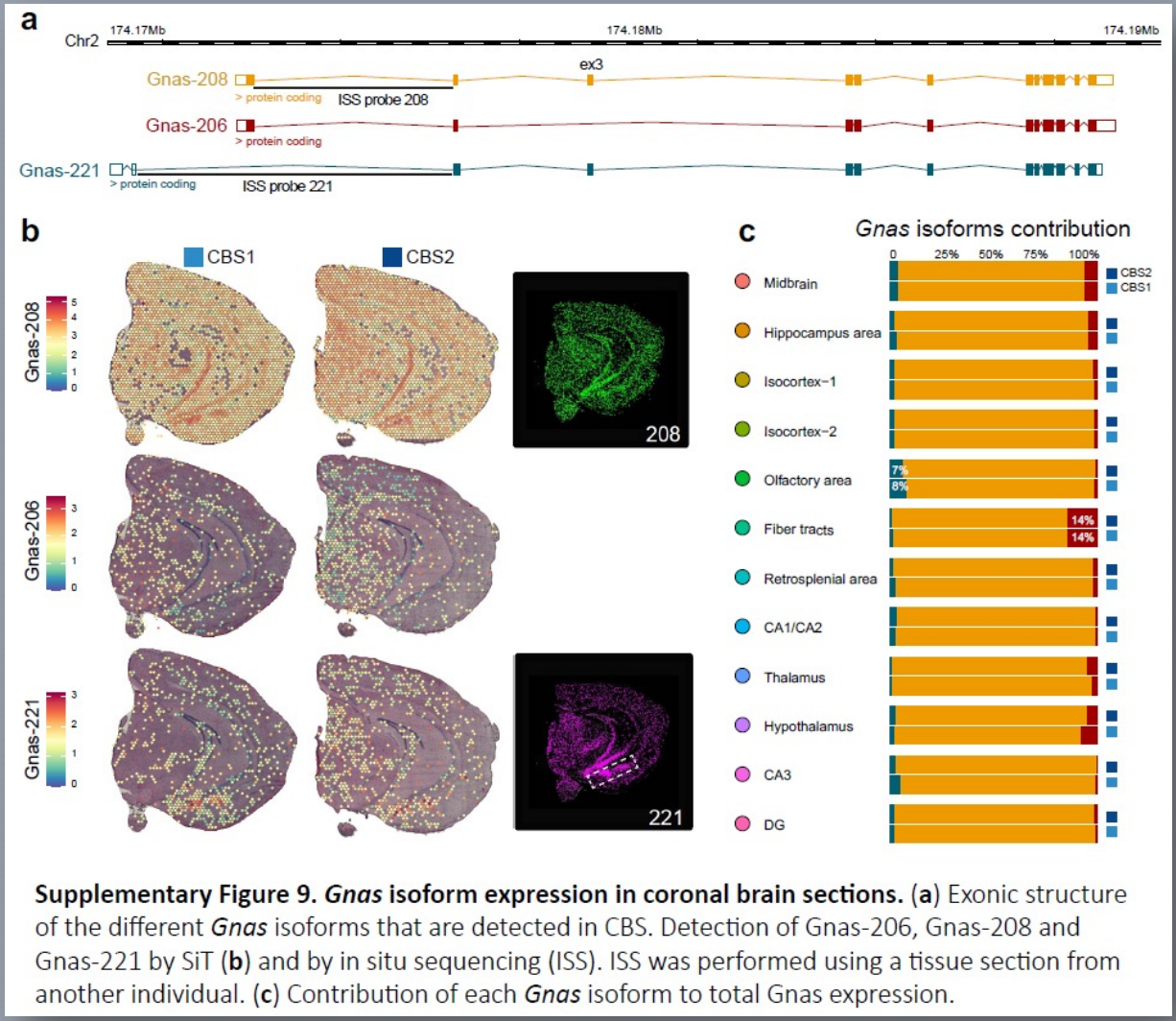
Molecular Architecture of the Mouse Nervous System

Amit Zeisel [1,3], Hannah Hochgerner [1,3], Peter Lönnerberg [1], Anna Johnsson [1], Fatima Memic [1], Job van der Zwan [1], Martin Häring [1], Emelie Braun [1], Lars E. Borm [1], Gioele La Manno [1], Simone Codeluppi [1], Alessandro Furlan [1,4], Kawai Lee [1], Nathan Skene [1], Kenneth D. Harris [2], Jens Hjerling-Leffler [1], Ernest Arenas [1], Patrik Ernfors [1] ... Sten Linnarsson [1,5]



133 samples
509,876 cells
>100 cell types

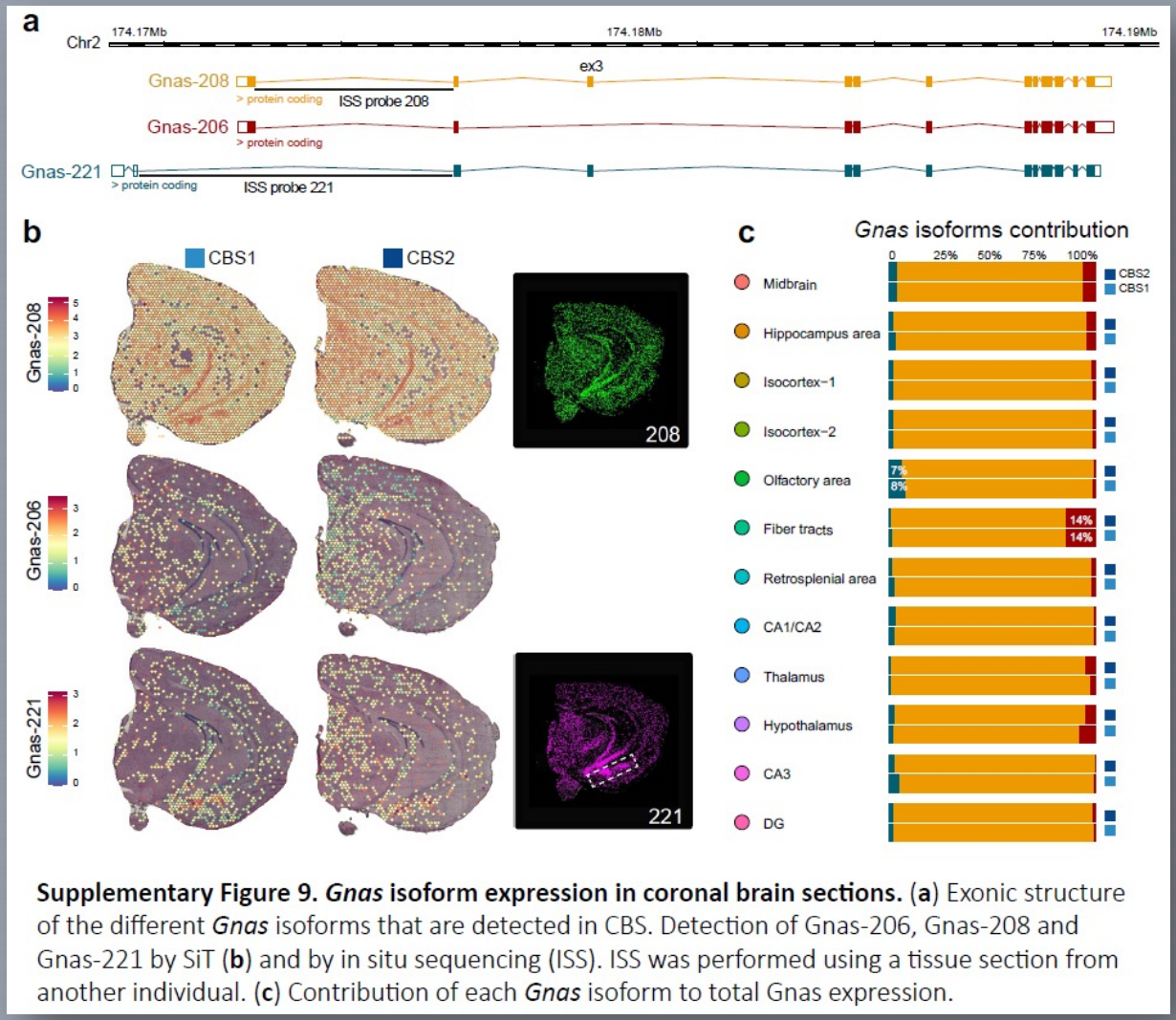
Gnas DTU across regions (CBS)



Supplementary Figure 9. *Gnas* isoform expression in coronal brain sections. (**a**) Exonic structure of the different *Gnas* isoforms that are detected in CBS. Detection of Gnas-206, Gnas-208 and Gnas-221 by SiT (**b**) and by in situ sequencing (ISS). ISS was performed using a tissue section from another individual. (**c**) Contribution of each *Gnas* isoform to total Gnas expression.

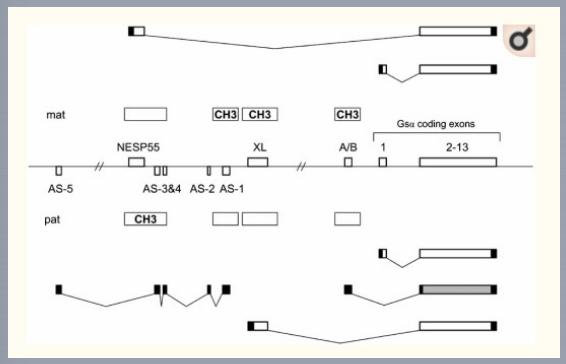
Gnas is an important component of the cyclic AMP signaling pathway

Gnas DTU across regions (CBS)



Supplementary Figure 9. *Gnas* isoform expression in coronal brain sections. (**a**) Exonic structure of the different *Gnas* isoforms that are detected in CBS. Detection of Gnas-206, Gnas-208 and Gnas-221 by SiT (**b**) and by in situ sequencing (ISS). ISS was performed using a tissue section from another individual. (**c**) Contribution of each *Gnas* isoform to total Gnas expression.

Gnas is an important component of the cyclic AMP signaling pathway



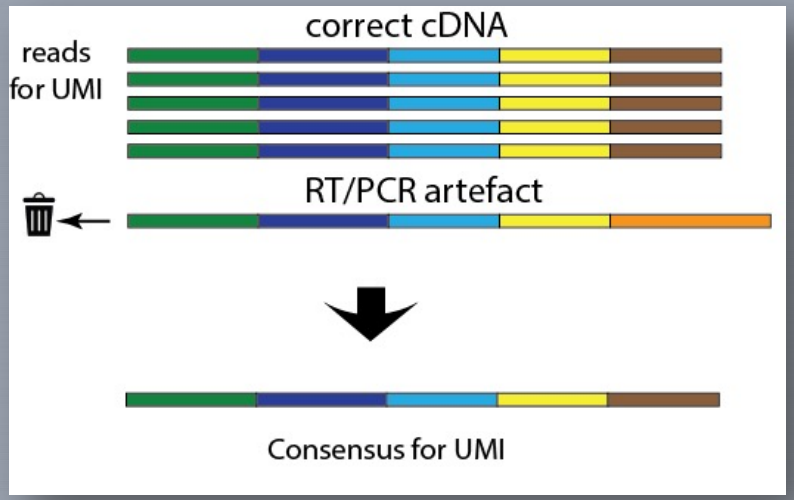
The *GNAS* Locus: Quintessential Complex Gene Encoding Gsα, XLαs, and other Imprinted Transcripts
Curr Genomics. 2007 Sep; 8(6): 398–414.
doi: 10.2174/138920207783406488

Long read sequencing identifies SNVs

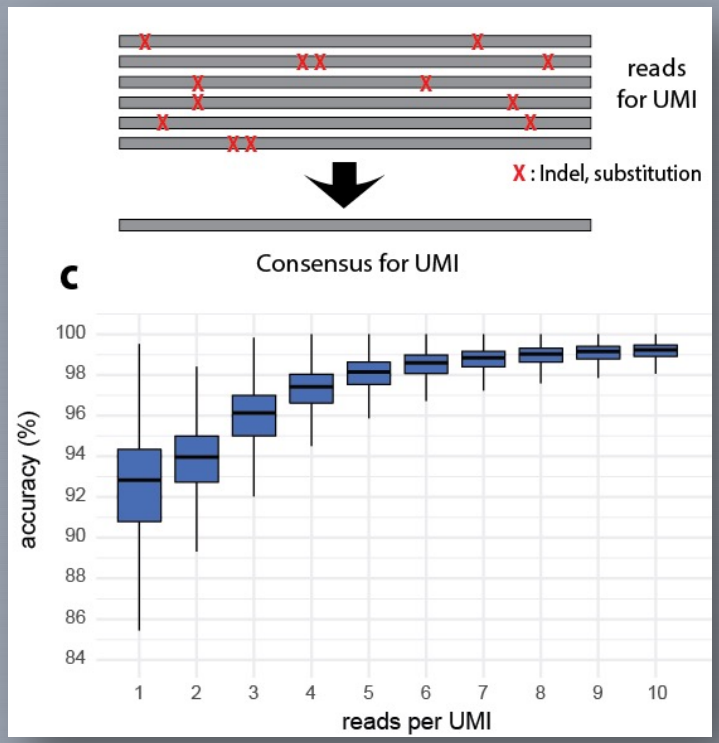
RNA editing in mouse brain regions (CBS)

UMIs are crucial for long read single cell RNA-seq

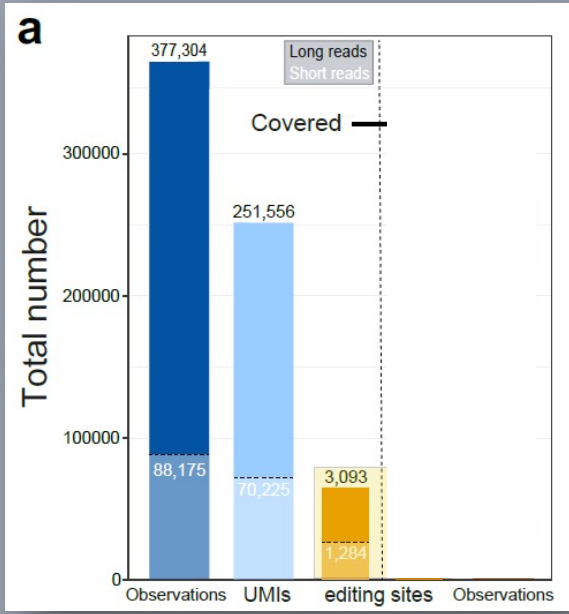
UMIs enable elimination of PCR artifacts

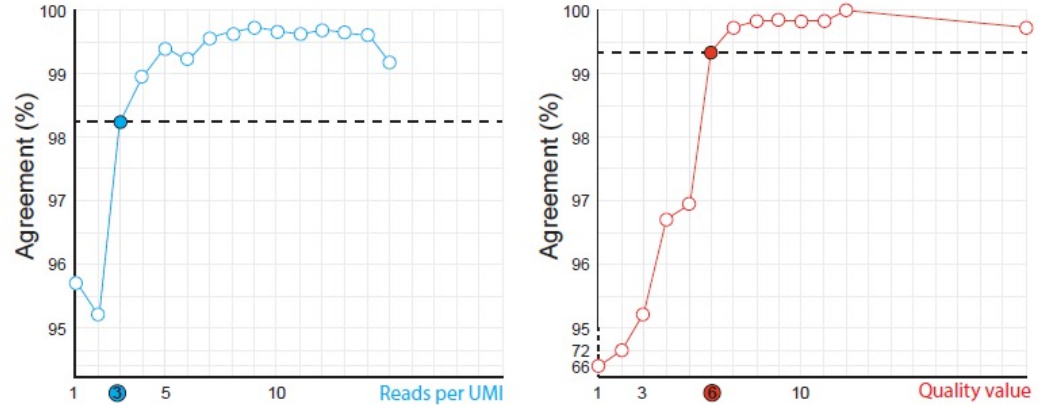


UMIs enable correction of sequencing errors



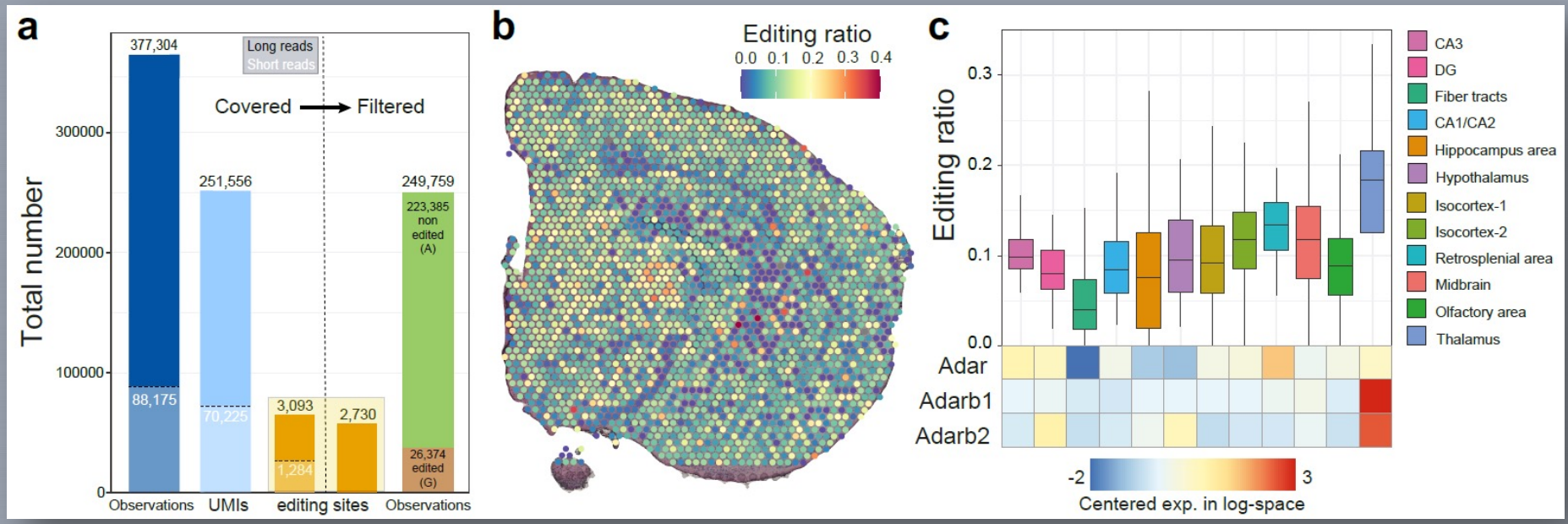
High confidence SNV calling calibration using short-read



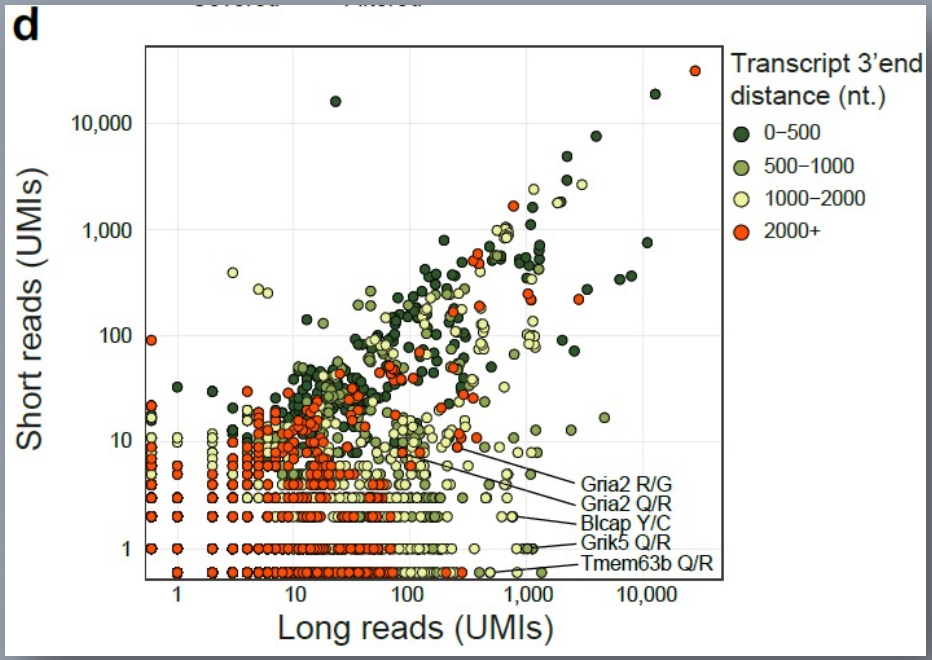


Supplementary Figure 11. Agreement of editing rates between long-read and short-read data for the genomic regions that are detected by both approaches. The percentage of agreement between the two sequencing approaches is plotted as a function of nanopore read numbers per UMI (left plot) and nanopore consensus base quality value (right plot). The highlighted thresholds were used for editing site calling with nanopore reads.

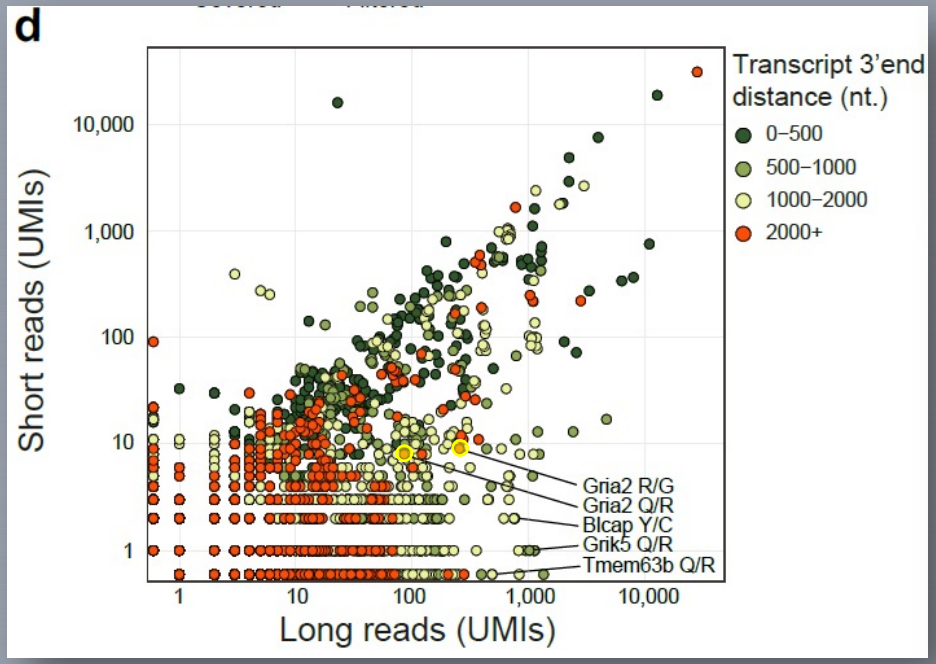
In-depth A-to-I RNA editing map of adult mouse brain

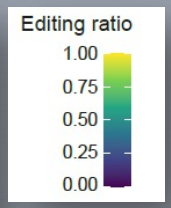


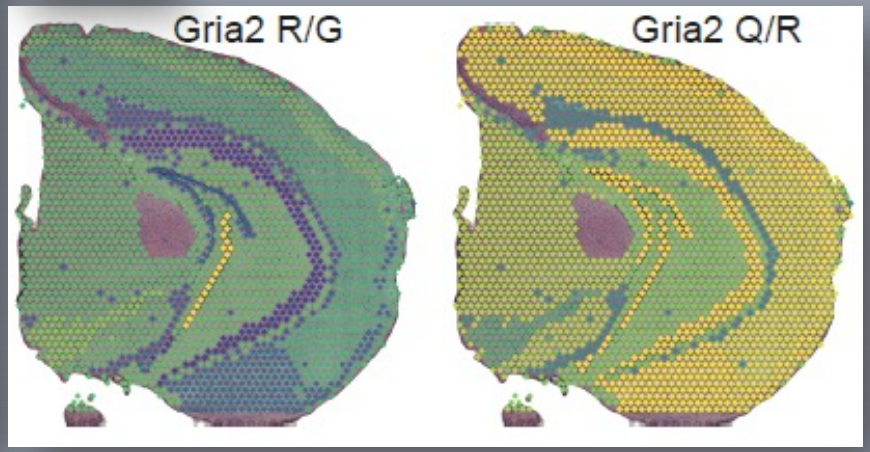
SiT reveals A-to-I RNA editing specificity in the mouse brain

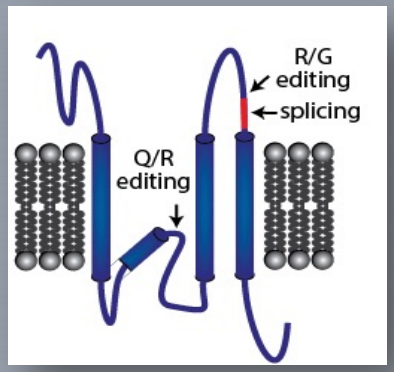


SiT reveals A-to-I RNA editing specificity in the mouse brain









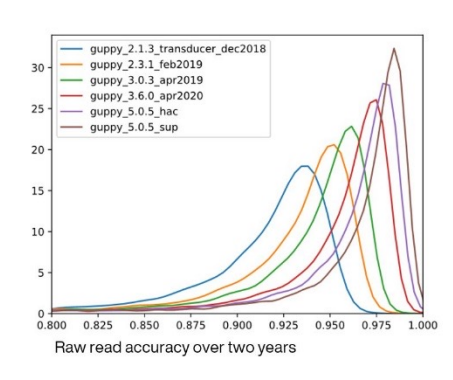
AMPA receptors (AMPARs) mediate most of the fast excitatory neurotransmission in the brain. Gria2 (GluA2) subunit is known to be edited at two positions: the R/G site in the ligand-binding domain where editing causes faster desensitization and recovery from desensitization, and the Q/R site within the channel pore, which when edited renders AMPARs virtually Ca2+-impermeable and thereby affects a key aspect of **neurotransmission**.

Conclusion

- Accurate in-situ capture spatial transcriptomics with Nanopore sequencing is feasible.
- Nanopore sequencing yields spatially resolved information on splicing and SNVs (a priori free discovery)

Conclusion

- Accurate in-situ capture spatial transcriptomics with Nanopore sequencing is feasible.
- Nanopore sequencing yields spatially resolved information on splicing and SNVs (a priori free discovery)
- In the near future we should get rid of Illumina sequencing



Current PromethION flow cell:

- **> 98% modal accuracy**
- **> 100 million reads per Promethion flow cell**

Early Access:

- **> 99 % accuracy**

Institut de Pharmacologie Moléculaire et Cellulaire

Nice-Sophia-Antipolis

Pascal Barbry Lab (IPMC, CNRS, France)

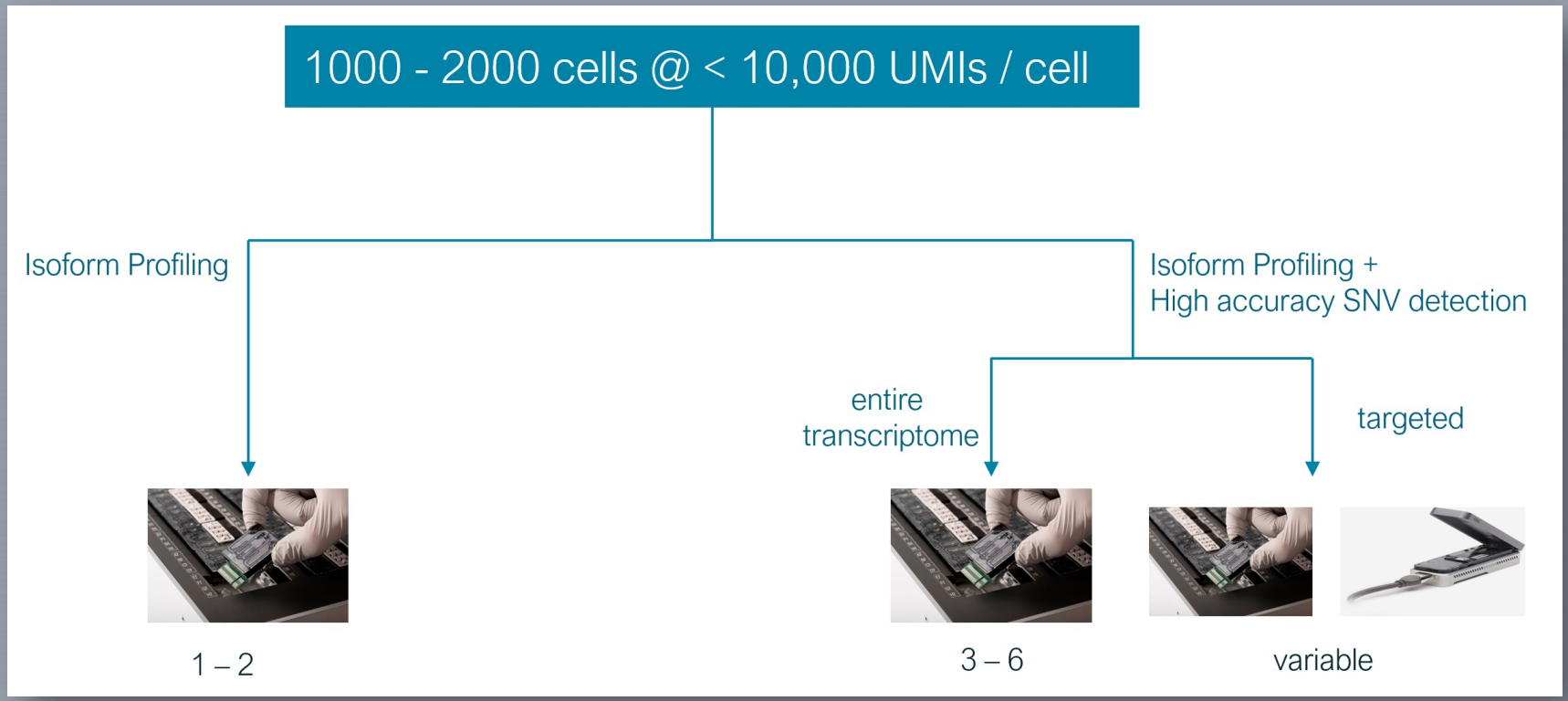
- Rainer Waldmann

Joakim Lundeberg Lab (KTH Royal Institute of Technology, Sweden)

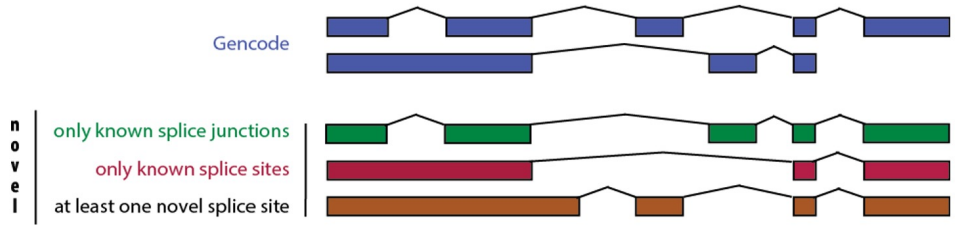
- Joseph Bergenstråhle
- Kim Thrane
- Annelie Mollbrink

How many long reads do we need ?

Depends on number of cells and mRNA content of cells (complexity)



Novel isoforms detection

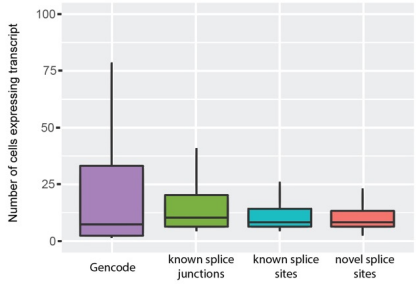


	Total	All splice junction in Illumina data	CAGE peak	polyA site	Final	% of tot
Known Gencode	33,002	33,002	20,533	14,908	11,186	34%
Novel	10,681	8,134	9,164	6,858	4,388	41%
Only known junctions	3,063	3,063	2,644	1,939	1,696	55%
Only known splice sites	2,111	1,905	1,906	1,366	1,115	53%
At least one novel splice site	5,507	3,166	4,614	3,553	1,577	29%

Novel isoforms are expressed at low level

3795 Gencode UMIs/cell
60 novel isoform UMIs/cell

Novel isoforms are expressed in fewer cells



Leakiness of the splicing machinery or physiologically relevant ?