Single-cell and Spatial Isoform Transcriptomics

Kévin Lebrigand
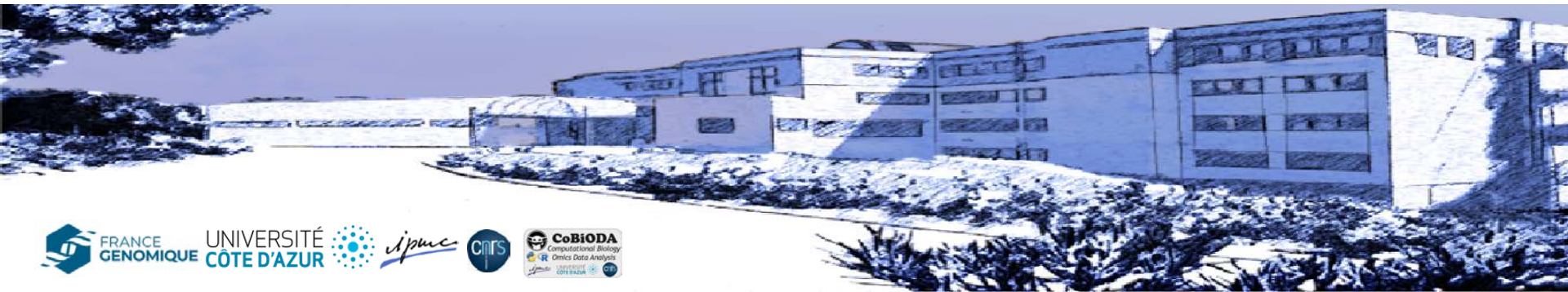Computational Biology and Omics Data Analysis
https://cobioda.github.io
IPMC, CNRS, Côte d'Azur University, France
lebrigand@ipmc.cnrs.fr
@kevinlebrigand

FRANCE GENOMIQUE UNIVERSITÉ CÔTE D'AZUR ipmc cnrs CoBiODA Computational Biology Omics Data Analysis

Institute of Molecular and Cellular Pharmacology

Sophia-Antipolis

20 research teams composed of > 220 members

- Ion channels (pain, perception, epilepsy)
- Molecular signaling (molecular trafficking, lipidomics)
- Neurodegenerative disorders (Alzheimer, Parkinson)
- Neuropsychiatric disorders (nervous breakdown, mental retardation)
- Functional genomics and bioinformatics

15 Engineers running 5 technological platforms

- MICA, Imaging and Flow Cytometry
- CAPABIO, Proteomics and Metabolomic
- ANIPRO, animal care and behavior facility
- CoBiODA, BioinformaticsHub
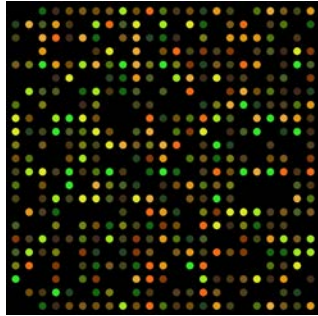- UniCA GenomiX, Functional Genomics platform

Core member of the "France Génomique" network (2008, 60M€) > 500 academics

2015   2016   2018   2022   2024

20 years of transcriptomics

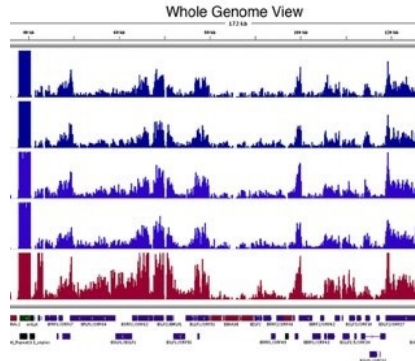Driven by microfluidics technological developments

Early 2000's: DNA microarray

- Large-scale transcriptome
- Oligonucleotide probe tilling
- Fluorochrome signal analysis
- Bulk resolution
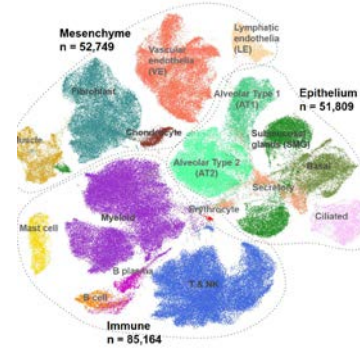
Cost : 4k€
20 samples
25k genes
0,5M matrix

Late 2000's : RNA sequencing

- Whole transcriptome
- Next Generation Sequencing
- Full-transcript coverage
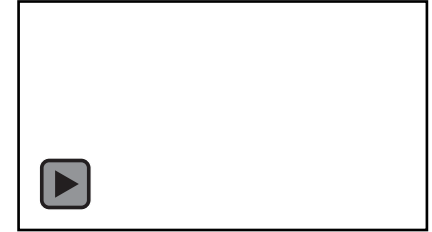- Bulk resolution

Cost : 4k€
20 samples
50k genes
1M matrix

Mid 2010's: Single -cell

- Whole transcriptome
- Microfluidics + NGS
- 3p-end gene signal (UMI)
- Sensitivity (6%)
- Single-cell / state resolution
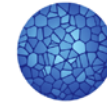
Cost : 4k€
5k cells
50k genes
250M matrix

2020's : Spatial

- 300-1000 gene targets
- Imaging analysis
- Multiplexing FiSH (single molecule)
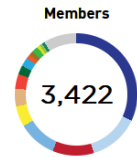- Sensitivity (30-80%)
- Sub-cellular resolution

Cost : 4k€
250k cells
1k genes
250M matrix
+ Spatial dimension

Human Cell Atlas
CZI initiative (2016)

Mission to create comprehensive reference maps of all human cells, the fundamental units of life, as a basis for both understanding human health and diagnosing, monitoring, and treating disease.

Human Cell Atlas
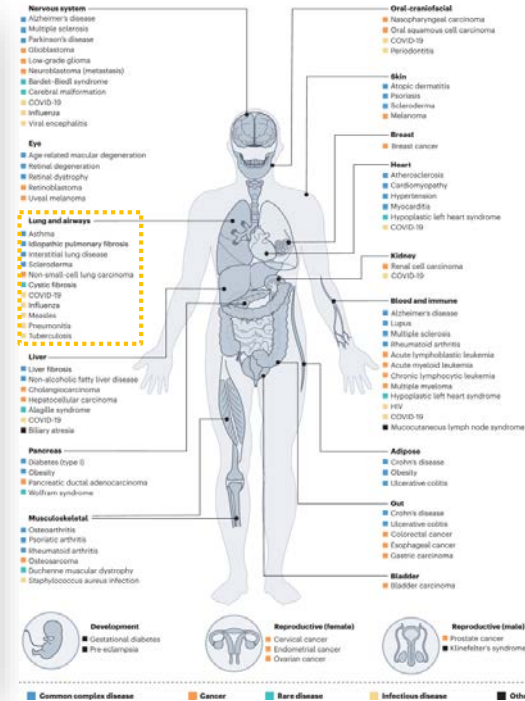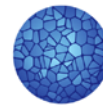## Our contribution

HUMAN CELL ATLAS

80k

10x GENOMICS

Oxford NANOPORE

10x GENOMICS

500k

2,4M

01

Single -Cell isoform Transcriptomics

Single -cell transcriptomics
Context

Bulk mRNA libraries
100ng total RNA
(10,000 cells)

Single cell mRNA libraries
10 pg total RNA
(<<1pg mRNA)
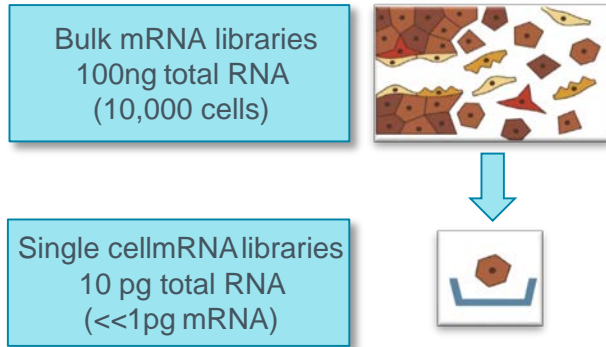
- Highly efficient library preparation techniques

- Elimination of PCR amplification bias and artefacts

- Use of Underline Molecular Identifiers (UMI) to monitor the number of molecules
  - Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Meth 9, 72-74 (2012)
  - Improved accuracy of molecule counting

UMI
Random (N)x sequence introduced during reverse transcription

library preparation, PCR

UMI counting, duplicate removal per transcript

Transcript A

Transcript B

Non amplified cDNA
A/B = 1/2

Pools of UMI

Random labeling

Amplified cDNA   PCR bias
A/B = 1/1

UMI ratio   No PCR bias
A/B = 1/2
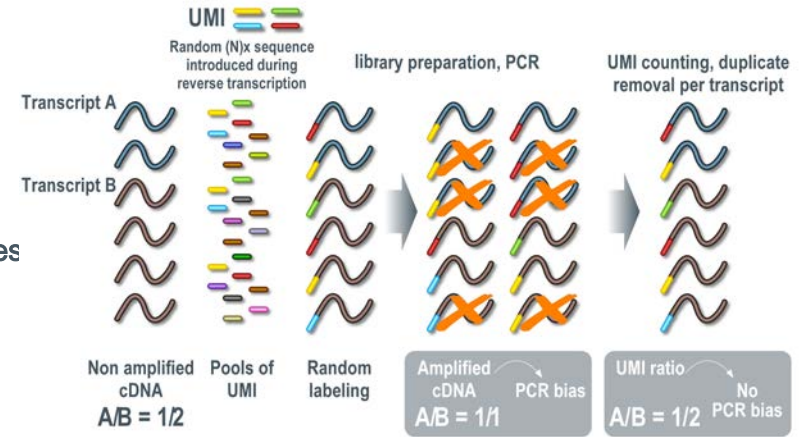
UMI allow a more precise profiling

Single-cell transcriptomics

Evolution of isolation techniques and throughput
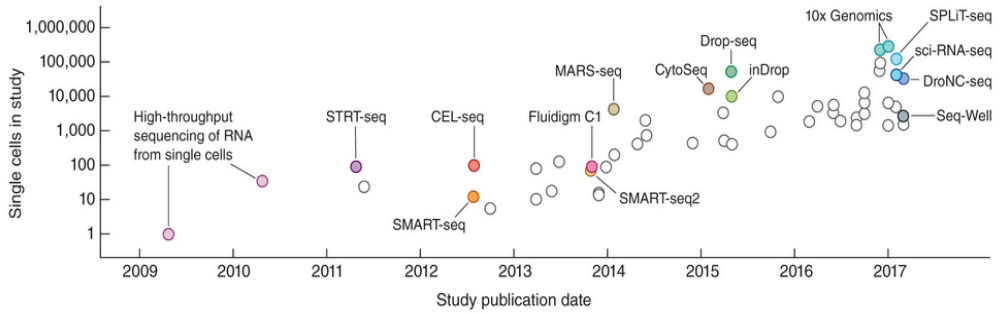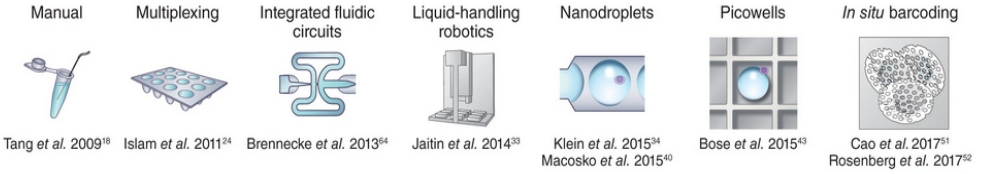
Exponential scaling of single-cell RNA-seq in the past decade
Svensson et al., *Nature Protocols*, 2018

Single -cell t ranscriptomics

Droplet-based approaches
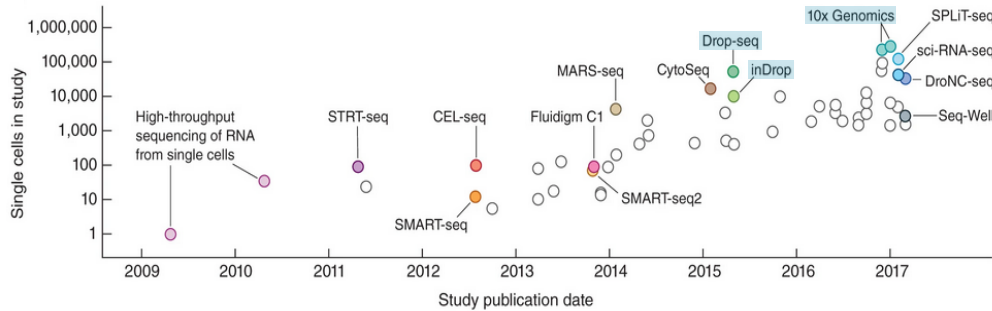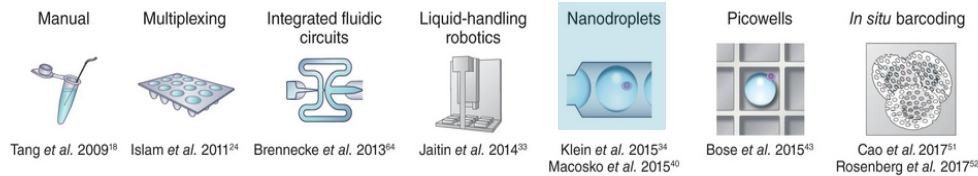
Exponential scaling of single-cell RNA-seq in the past decade
Svensson et al., *Nature Protocols*, 2018

10x Genomics Chromium single cell controller (2016)
- Easy-to-set-up and robust workflow
- Generalize UMI usage
- Shows high scalability (1,3M cells dataset)
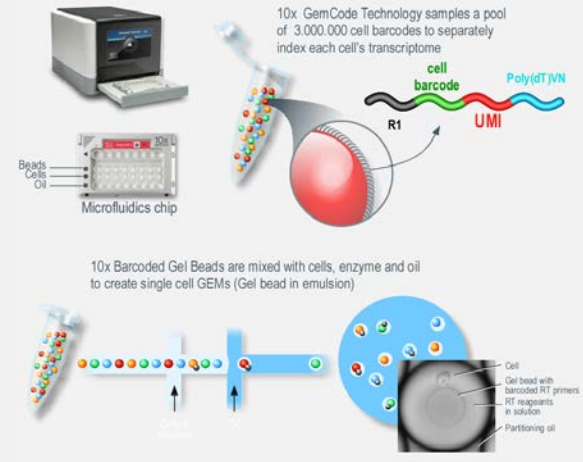
Single -cell t ranscriptomics
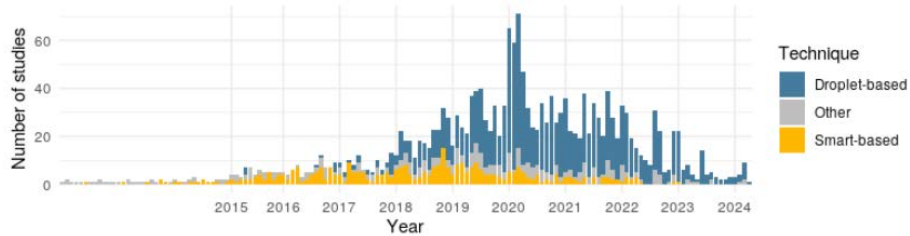
Single cell approaches in publications

https://doi.org/10.1093/database/baaa073

A curated database reveals trends in single cell transcriptomics
Svensson et al.,Database , 2020

- Huge amount of singlecell studies in the past 5 years,

- Droplet-based approaches = 61% (Chromium: 47%)

Technique	Count
Chromium	725
Smart-seq2	177
SMARTer (C1)	124
Drop-seq	74
SMARTer	28
InDrops	23
CITE-seq	18
CEL-seq2	17
STRT-seq	17
MARS-seq	16
Tang	15
CEL-seq	13
STRT-seq (C1)	13
Seq-Well	13
SORT-seq	12
BD Rhapsody	11
BioMark	8
GemCode	7
ICELL8	7
Perturb-seq	7
Patch-seq	6
sc-RT-mPCR	6
MERFISH	5

Droplet-based approaches
- Digital Gene Expression (UMI)
- High cell number throughput
- Limited capture efficiency (<10%)
- 3p or 5p signal (SAGElike)

Single -cell transcriptomics

Single cell approaches in publications

https://doi.org/10.1093/database/baaa073

A curated database reveals trends in single cell transcriptomics
Svenssonet al., Database, 2020

- Huge amount of singlecell studies in the past 5 years,
- Droplet-based approaches = 61% (Chromium: 47%)
- Smart-based approach = 21%, <5% in the last 2 years

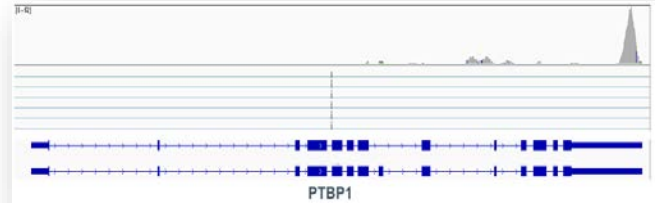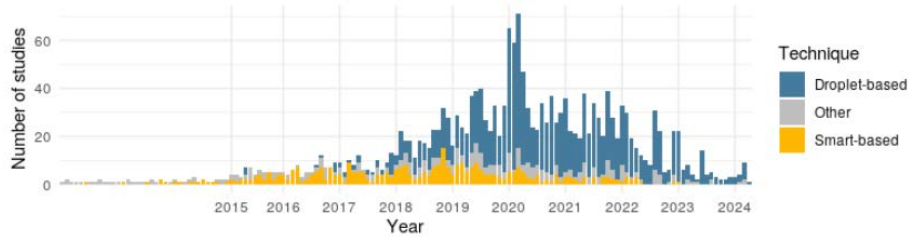| Technique	Count
Chromium	725
Smart-seq2	177
SMARTer (C1)	124
Drop-seq	74
SMARTer	28
InDrops	23
CITE-seq	18
CEL-seq2	17
STRT-seq	17
MARS-seq	16
Tang	15
CEL-seq	13
STRT-seq (C1)	13
Seq-Well	13
SORT-seq	12
BD Rhapsody	11
BioMark	8
GemCode	7
ICELL8	7
Perturb-seq	7
Patch-seq	6
sc-RT-mPCR	6
MERFISH	5

Smart-based approach
- Lower cell number (384plate handling)
- Higher capture efficiency (~30%)
- No UMI before v3 (may 2020)
- Full-length coverage using short reads

Article | Open Access | Published: 30 May 2022
Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress
Michael Hagemann-Jensen, Christoph Ziegenhain & Rickard Sandberg

Brief Communication | Open Access | Published: 30 May 2022
Fast and highly sensitive full-length single-cell RNA sequencing using FLASH-seq
Vincent Hahaut, Dinko Pavlinic, Walter Carbone, Sven Schuierer, Pierre Balmer, Mathieu Quinodoz
Magdalena Renner, Guglielmo Roma, Cameron S. Cowan & Simone Picelli
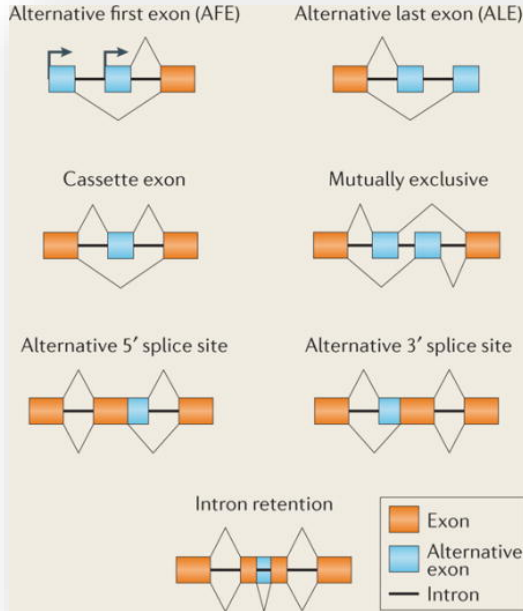
UMIs detected in HEK293 cells
Droplets 10x: 30k (50k reads)
Plate-based : 60k (200k reads)
Smart-seq3: 150k (750k reads)

Mantis Microdispenser

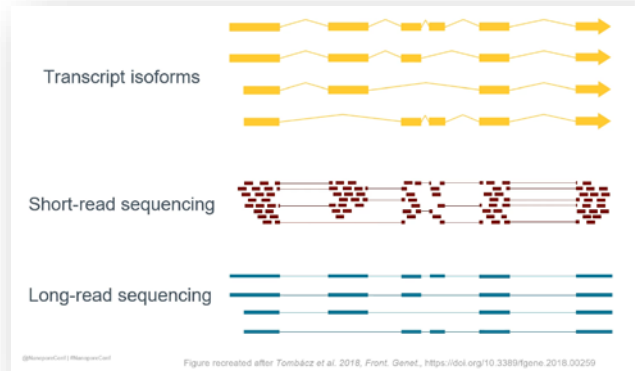# Transcriptomics

Complex outcomes of alternative splicing

Scotti and Swanson, Nat Rev Genet., 2016

- **90%** of the genes are subjected to alternative splicing,
- Gencode v42 : **252,416** distinct isoforms for **62,696** genes,
- On average, a human gene contains **8.8 exons**, mean size of **145 nt**,
- Average encodes mRNA **2,410 nt long** :

5' UTR	Coding sequence	3' UTR
770nt	1,340nt	300nt

Alternative splicing and disease
Tazi et al, 2008

Nature Method of the Year 2022

Inference required

Direct full exonic layout
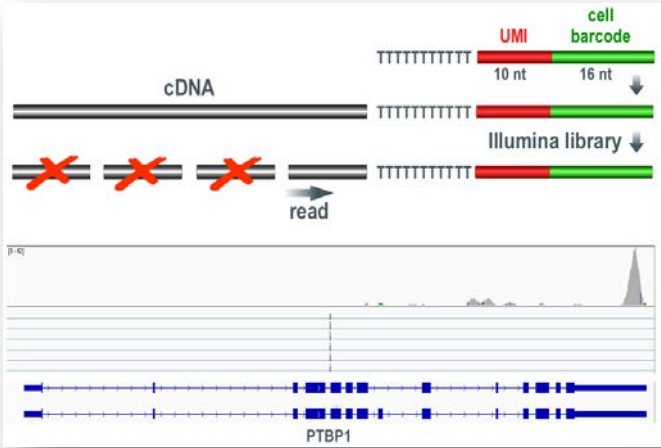One read is one molecule

Single -cell long -read transcriptomics
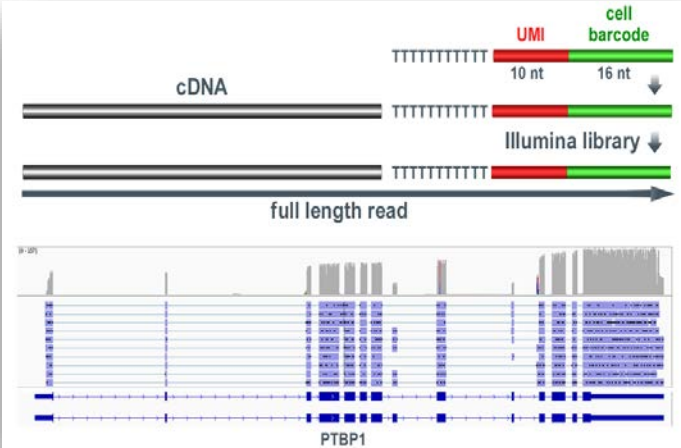
Droplets-based approach short reads vs long reads

Standard short-read sequencing

Long-read full-length sequencing

Gene-level matrix
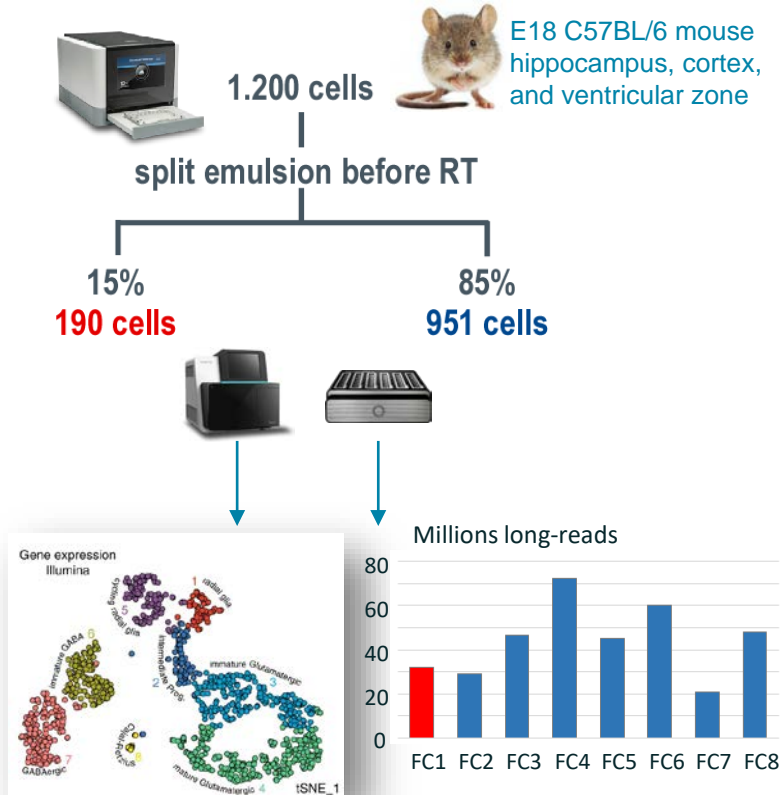
Isoform-level matrix

Information on alternative splicing, fusion transcripts, SNV, editing, imprinting, allelic imbalance

Is lost

Remain accessible

Single-cell long-read transcriptomics
SiCeLoRe, bioinformatics for Single Cell Long Read

E18 C57BL/6 mouse hippocampus, cortex, and ventricular zone

1.200 cells

split emulsion before RT

15%
190 cells

85%
951 cells

Gene expression Illumina

Millions long-reads

190 cells: 32M
951 cells: 322M

R9.4, LSK-109 chemistry (2018)

standard workflow

nature communications

Explore content ∨   About the journal ∨   Publish with us ∨

nature > nature communications > articles > article

Article | Open Access | Published: 12 August 2020

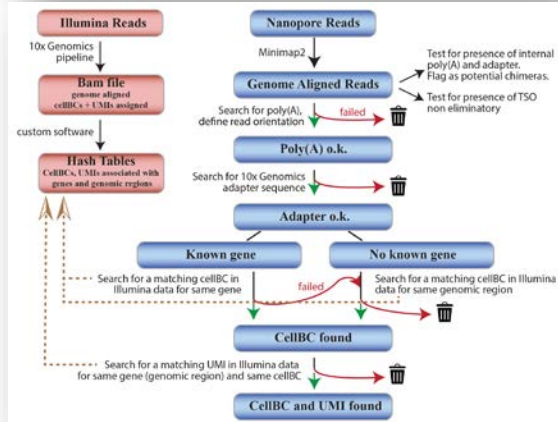**High throughput error corrected Nanopore single cell transcriptome sequencing**

Kevin Lebrigand ✉, Virginie Magnone, Pascal Barbry ✉ & Rainer Waldmann ✉

Nature Communications **11**, Article number: 4025 (2020) | Cite this article

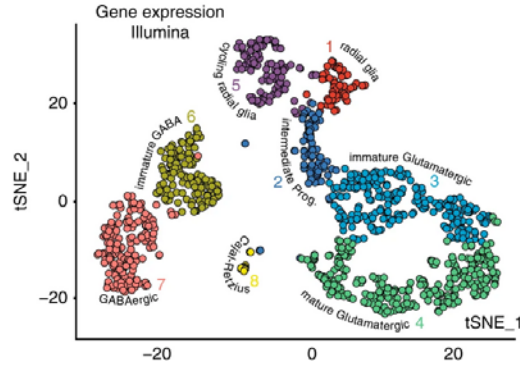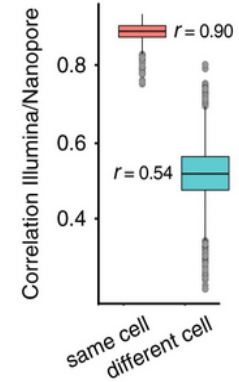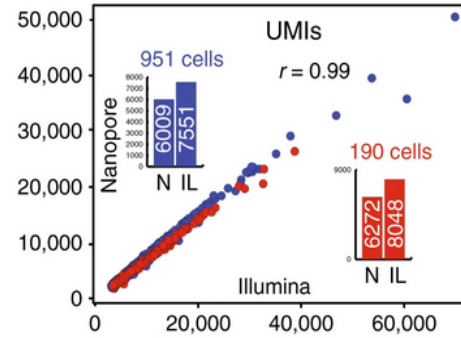**20k** Accesses | **38** Citations | **58** Altmetric | Metrics
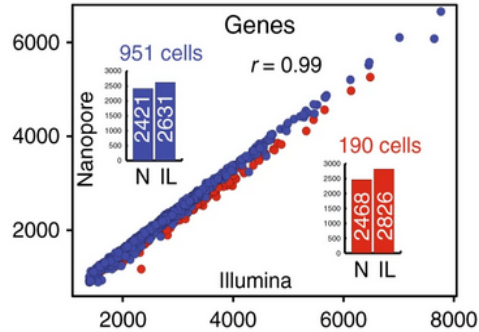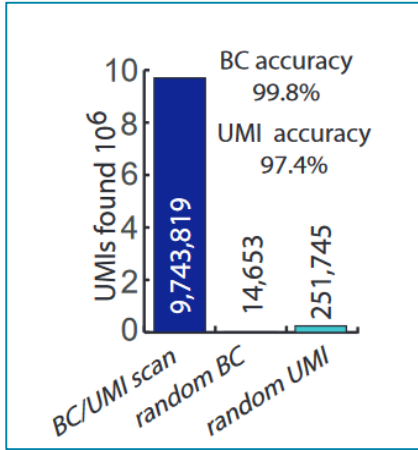
Rainer Waldmann

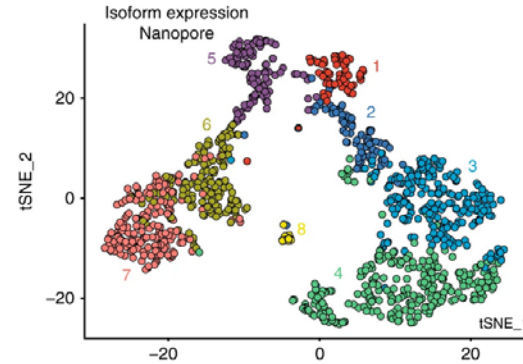https://github.com/ucagenomix/sicelore-2.1

Single-cell long-read transcriptomics

Shows high accuracy, high correlation with short-read and high reproducibility

6,047 Gene-level UMIs / cell
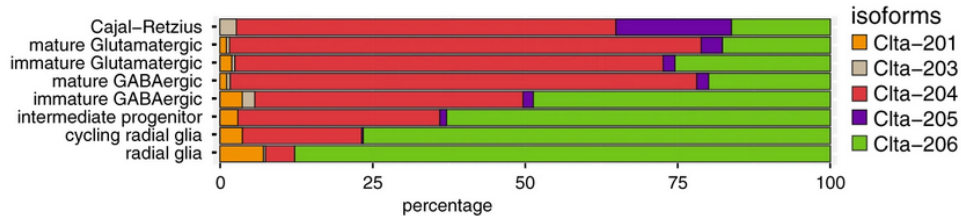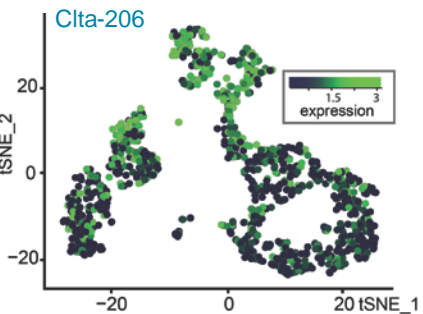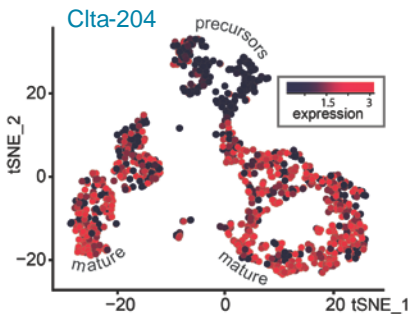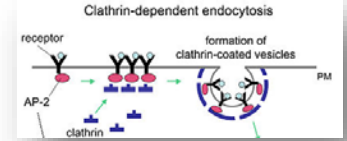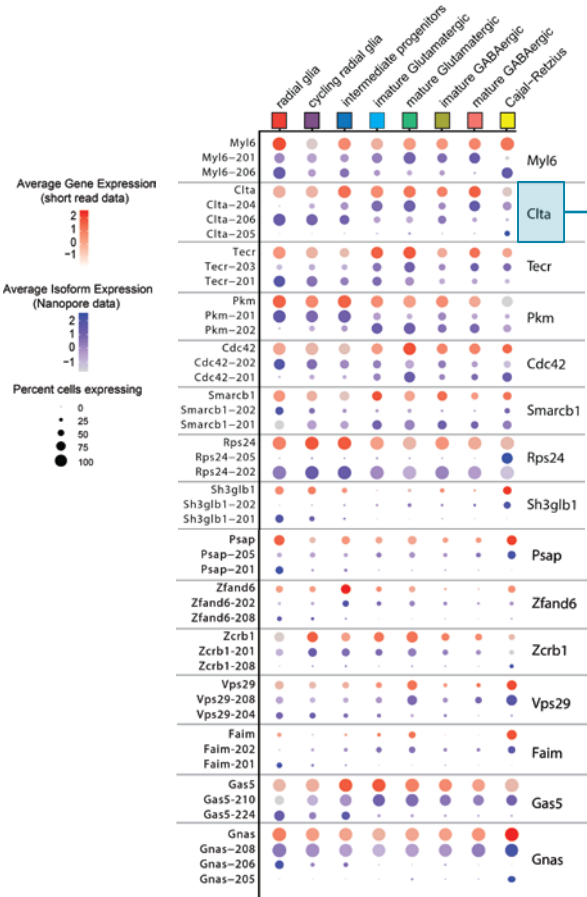
3,795 Isoform-level UMIs / cell
(28,120 Nanopore reads/cell)

Single -cell long -read t ranscriptomics   reveals diversity

76 isoform-switching genes along neuronal maturation

Single -cell long -read t ranscriptomics   reveals sequence heterogeneity

Consensus sequence computation per UMI

UMIs enable elimination of PCR artifacts

Crucial for accurate novel isoform discovery

Identification of 4.388 novel mouse isoforms
- Backed by > 5 UMIs in more than 2 cells
- All SJs confirmed by SR dataset
- 5′ end < 50 nt. from CAGE-seq
- 3'end < 50 nt. from polyA site

UMIs enable correction of sequencing errors

Crucial for high accuracy SNVcall

RNA A-to-I editing of the AMPA receptor Gria2

Q/R site regulates AMPA receptor Ca^{2+}-permeability
R/G site is involved in desensitization and recovery of the receptor
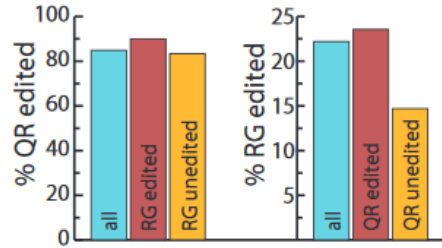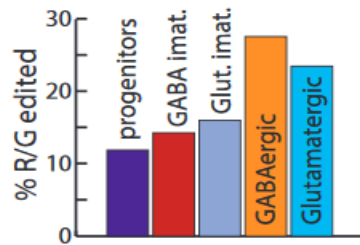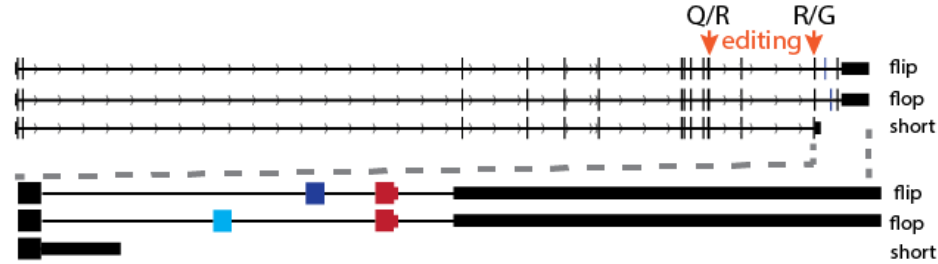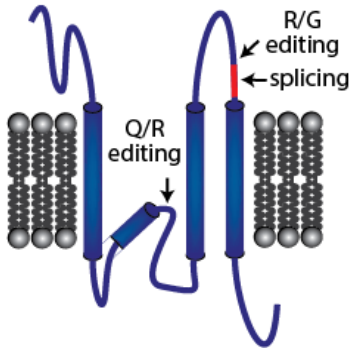
RG site less edited in non-edited QR molecules
→ Editing process synchronization ?

02

Spatial isoform   Transcriptomics

Spatial Transcriptomics approaches
Historical timeline

- Spatial transcriptomics aims at directly visualize gene expression in their original environment
- Tackle the main limitation of single cell experiment missing the spatial organization
- A lot of developments in the last years thanks to recent advances in different fields
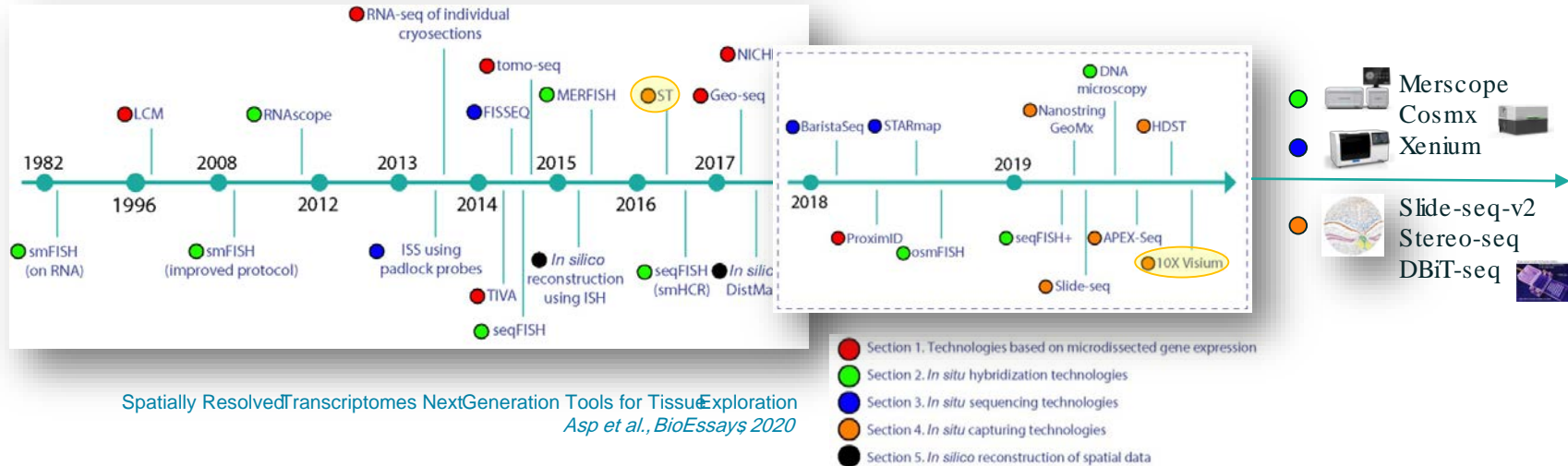
Spatially Resolved Transcriptomes Next Generation Tools for Tissue Exploration
Asp et al., BioEssays 2020

In-situ capture spatial transcriptomics

Ståhl et al. (2016); 10x Genomics Visium(2019)
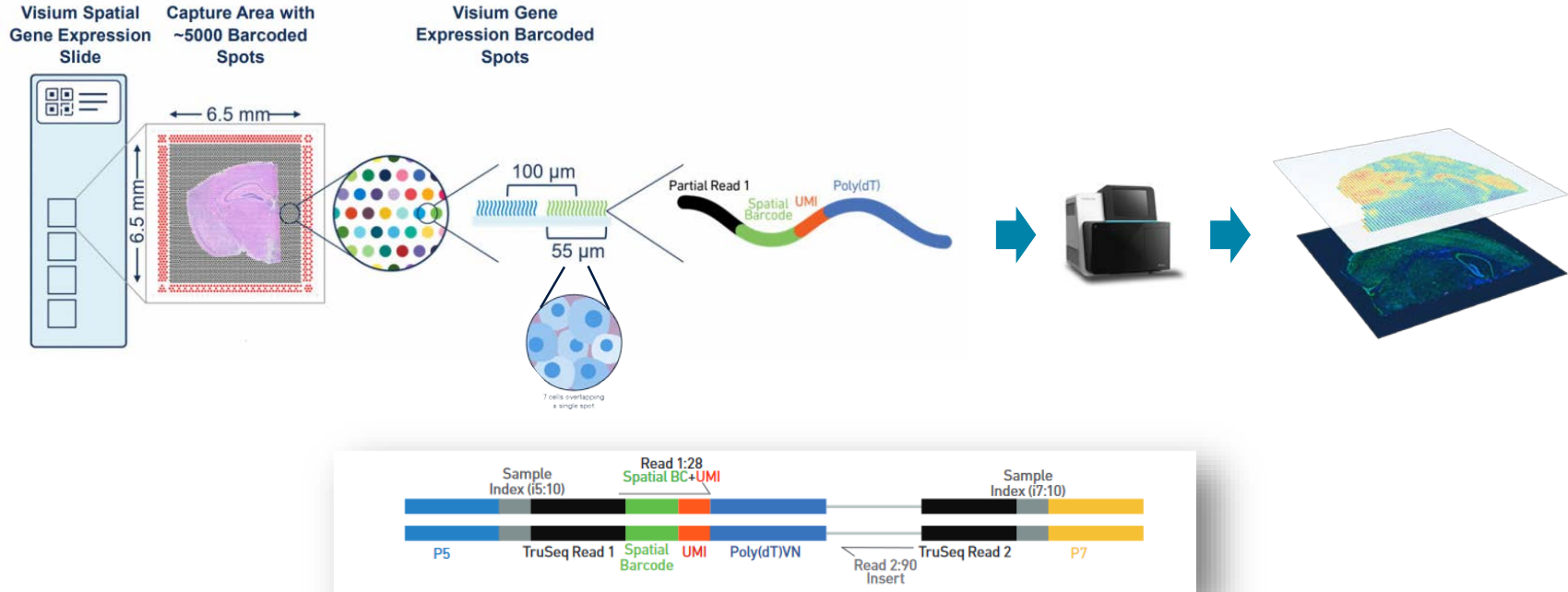
→ Spatial barcode / UMI assignment strategy identical as the droplet-based single-cell approach

Spatial isoform Transcriptomics (SiT)

Nucleic Acids Research, 2023

Nanopore promethION long-read sequencing

Provides isoform-level spatial transcriptomics

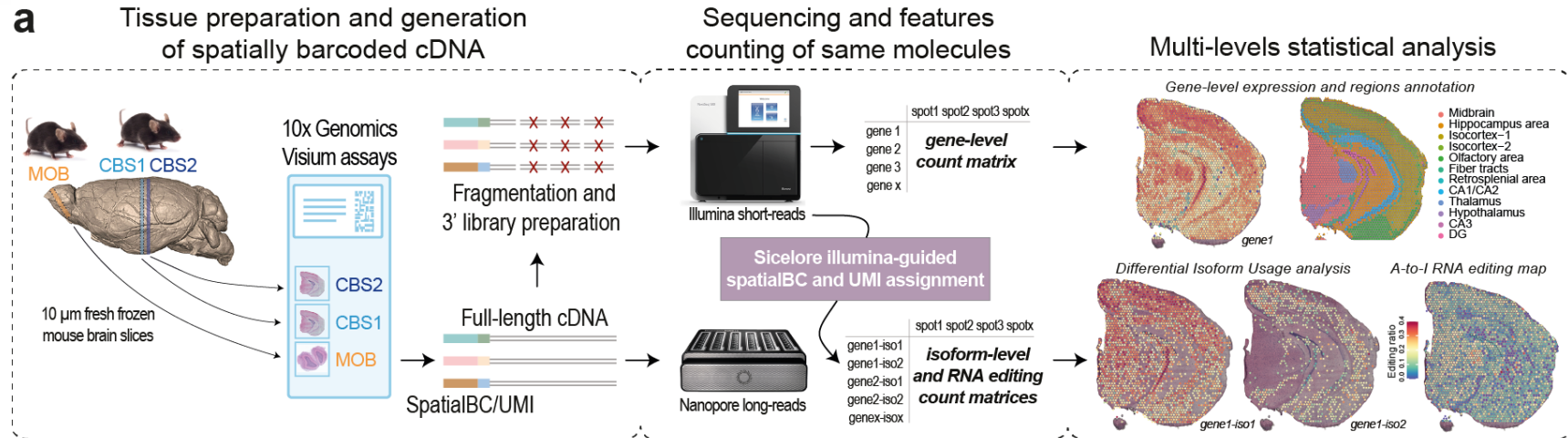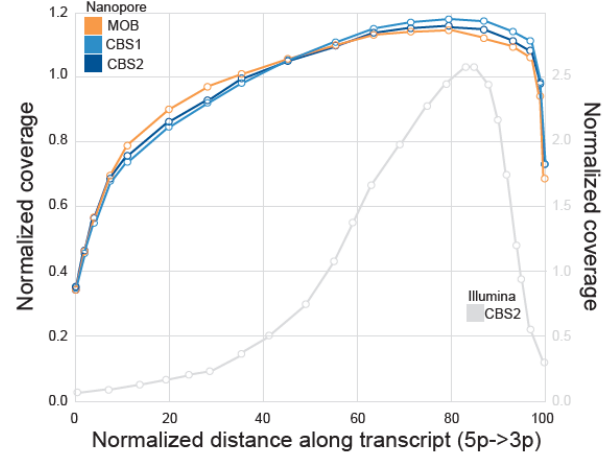### Sequencing saturation curves per sample
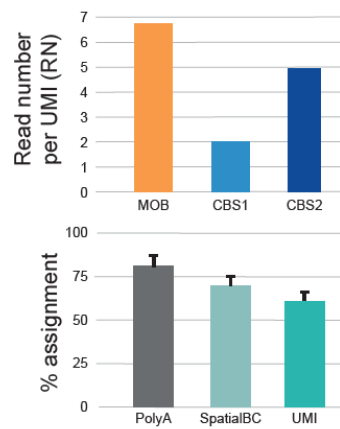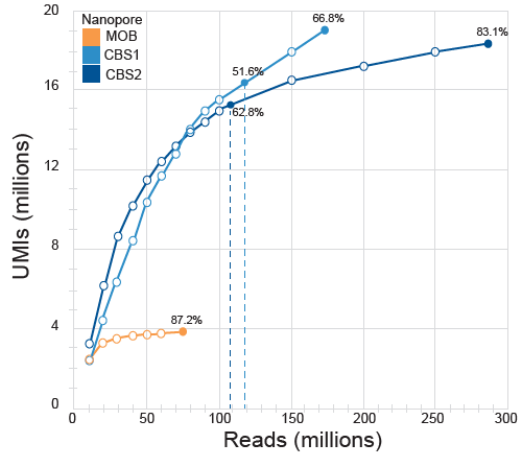
Transcripts full-length coverage

Reads	MOB		CBS1			CBS2							Total		
Date	18 feb. 20	20 mar. 20	18 feb. 20	20 mar. 20	24 feb. 20	12 may 20	13 may 20	19 may 20	25 may 20	25 may 20	26 may 20	27 may 20	09 feb. 21		
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 or 2 Promethion flow cells per Visium slice

SiT reveals specific splicing pattern across MOB regions
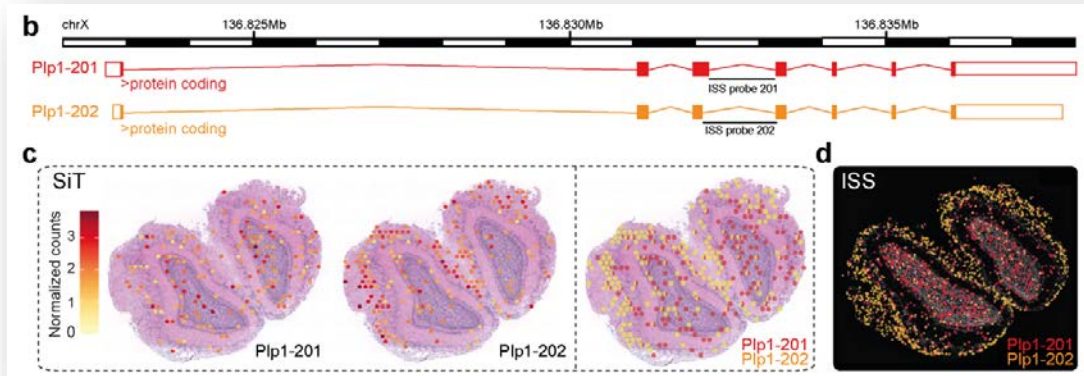
Plp1 Differential Transcript Usage (DTU)
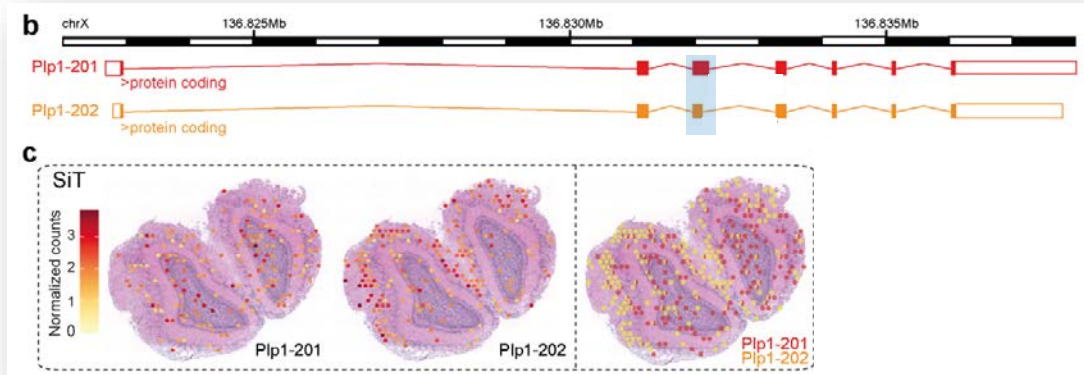
Proteolipid Protein 1 (Plp1) is a gene involved in severe pathologies associated with CNS dysmyelination

SiT reveals specific splicing pattern across MOB regions
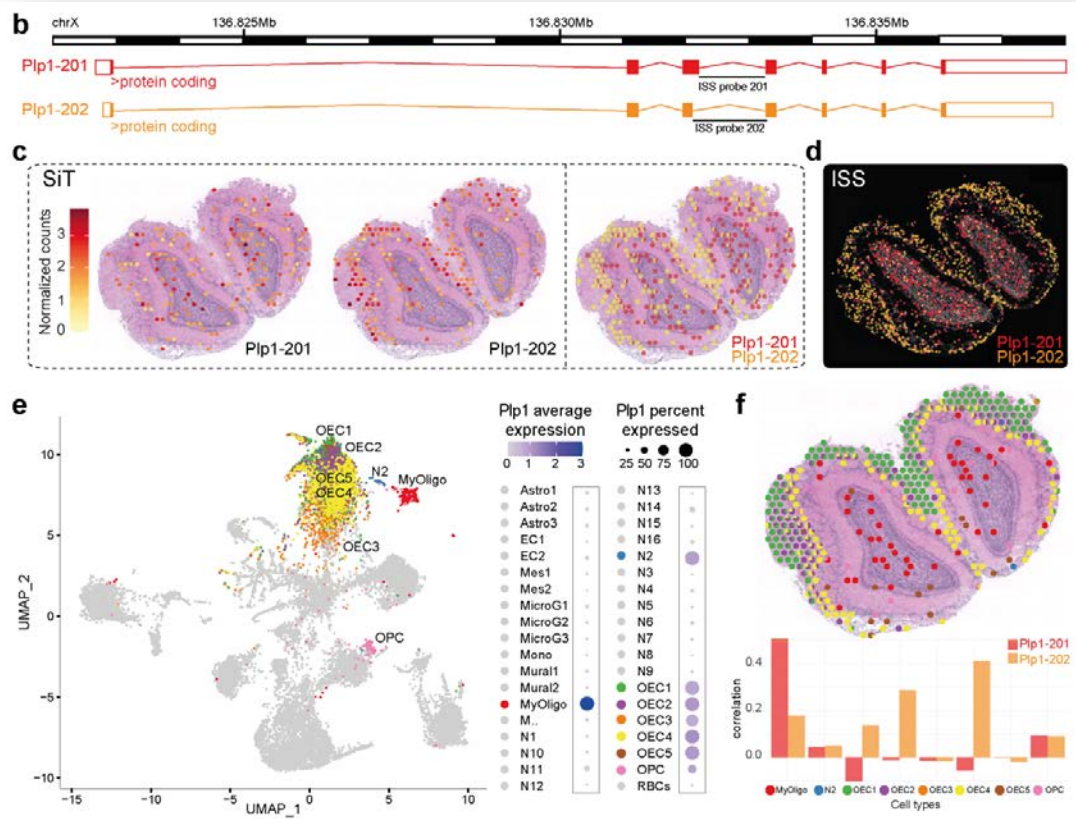
Plp1 Differential Transcript Usage (DTU)

Proteolipid Protein 1 (Plp1) is a gene involved in severe pathologies associated with CNS dysmyelination

In Situ Sequencing Data

SiT reveals specific splicing pattern across MOB regions

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

Proteolipid Protein 1 (Plp1) is a gene involved in severe pathologies associated with CNS dysmyelination

In Situ Sequencing Data

Spatial spot deconvolution of prominent *Plp1* expresser cell types. Correlation Deconvolution score / *Plp1* isoforms expression correlation shows that *Plp1* is predominantly expressed as Plp1-202 by olfactory ensheathing cells (OEC) in the ONL and as Plp1-201 isoform by myelinating-oligodendrocytes (MyOligo) in the GCL.

SiT reveals full-length sequence heterogeneity

Global A-to-I RNA editing spatial map

- Exploration of 5,817 Ato-I RNA editing sites described in the literature (Ramaswami et al., 2013 (RADAR), Licht et al., 2019)
- Long read high confidence call thresholding, looking at agreement between long and short read base calls for 88,175 shared UMIs
 — number of reads per UMI >= 3
  — consensus Phred score QV >= 6
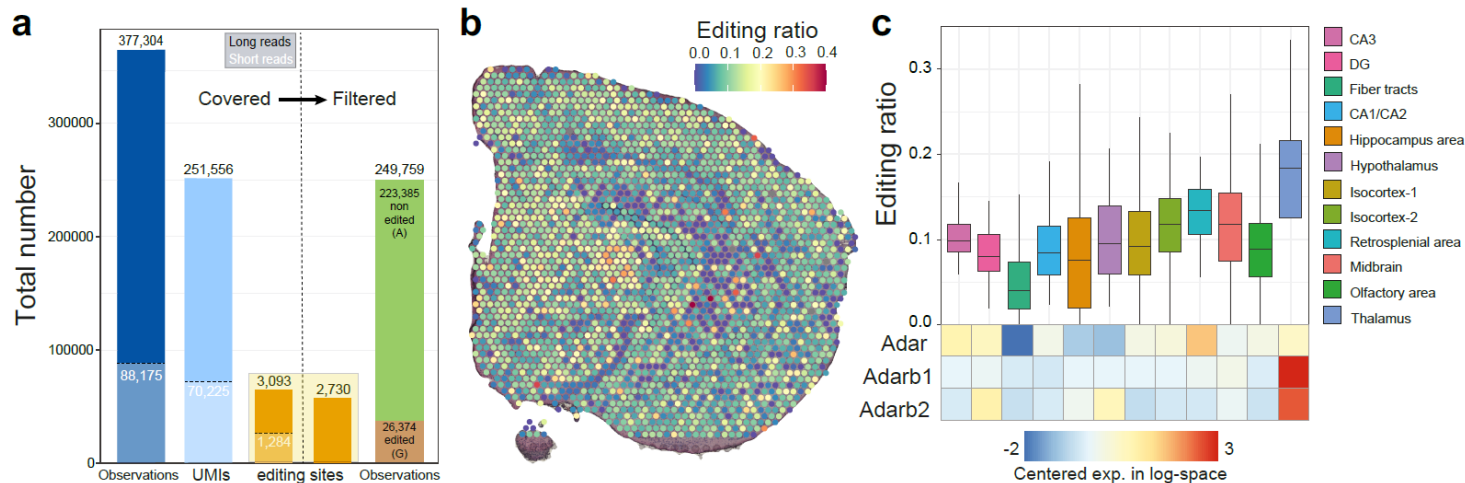
SiT reveals full -length sequence heterogeneity

Global A-to-I RNA editing spatial map

- Exploration of 5,817 A-to-I RNA editing sites described in the literature (Ramaswami et al., 2013 (RADAR), Licht et al., 2019)
- Long read high confidence call thresholding, looking at agreement between long and short read base calls for 88,175 shared UMIs
  — number of reads per UMI >= 3
  — consensus Phred score QV >= 6
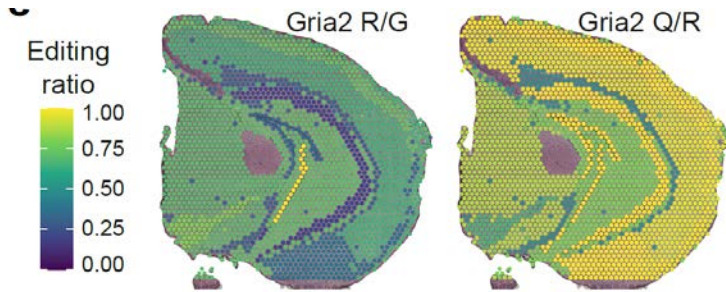
Individual A to-I editing site editing ratio per region

Gria2
- R/G site is involved in desensitization and recovery of the receptor
- Q/R site regulates AMPA receptor Ca^{2+}-permeability
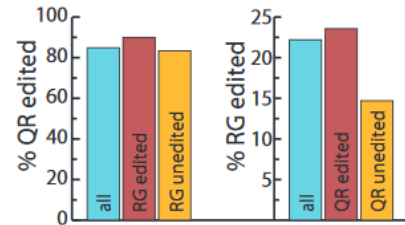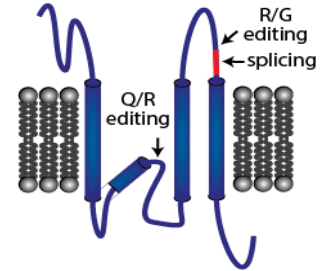
*Single-cell (lebrigand et al., 2020)*

Single cell and Spatial isoform transcriptomics

Summary

- Accurate single-cell and spatial transcriptomics using Nanopore long-read sequencing is feasible

- Long reads sequencing reveals transcript diversity that is missed with standard short reads workflows

- Single Nucleotide Variation calls (SNV, editing) in single-cell and in a spatial context can be achieve

- Sicelore-2.1 : we don't need short reads anymore

Raw read accuracy over two years
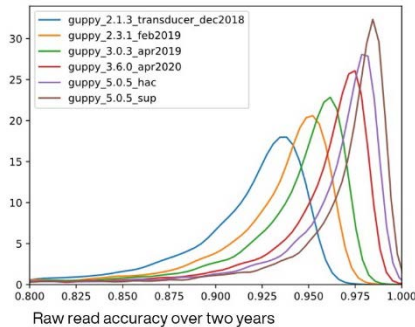
Nanopore PromethION sequencing

2018: 20M reads/FC, 92% raw read accuracy

2022: 150M reads/FC, 98% raw read accuracy

https://github.com/ucagenomix/sicelore2.1
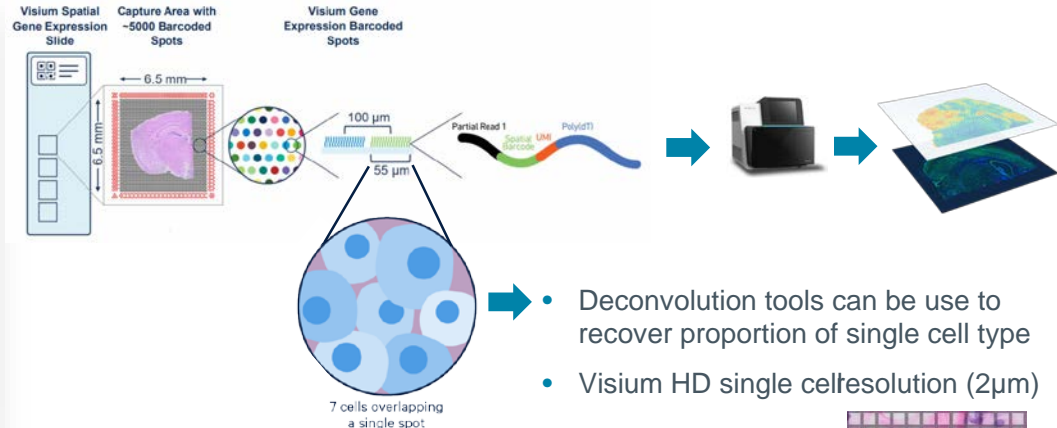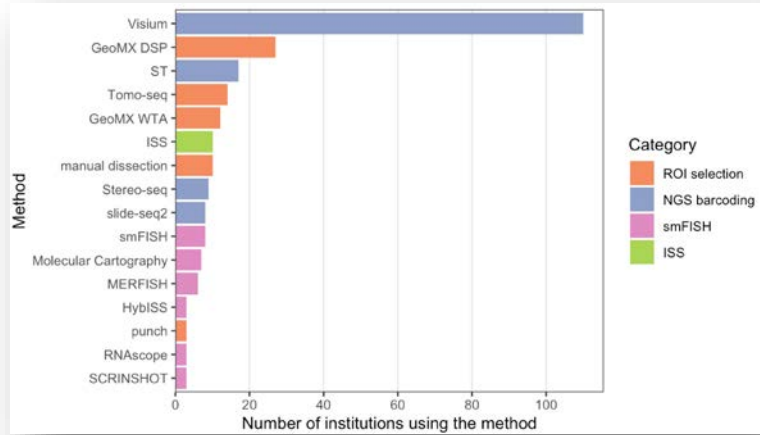- Visium and single cell 3' and 5' libraries
- Illumina-free profiling available

03

Spatial imaging -based Transcriptomics

In-situ capture Spatial Transcriptomics (2017 -2022)

Visium is widely adopted by academics

- Deconvolution tools can be use to recover proportion of single cell type
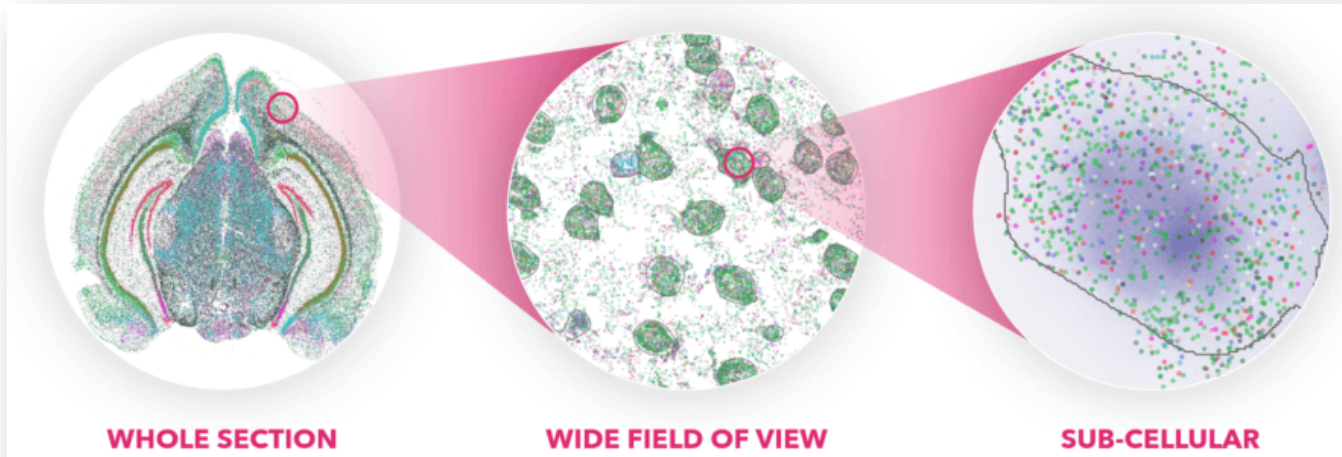
- Visium HD single cell resolution (2μm)
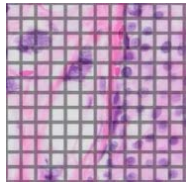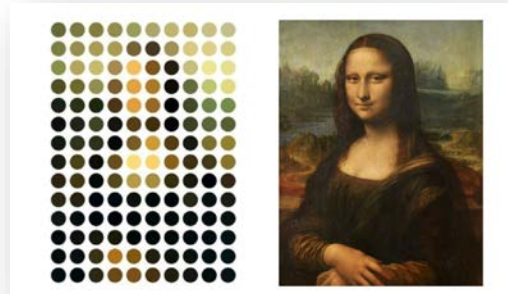
But is not the ideal readout for spatial biology
(Akoya credit rough caricature)

Imaging -based Spatial Transcriptomics (2022)

No more sequencing for direct single cell resolution

- **Lower gene panel targets** (from whole transcriptome to ~1,000 genes)
- **Higher sensitivity** (from ~6% to 30-80%)
- **Larger imaging area** (42 to 236 mm2)
- **Higher resolution** (from 55 µm to subcellular)

WHOLE SECTION   WIDE FIELD OF VIEW   SUB-CELLULAR

Imaging -based Spatial Transcriptomics   (2022)

No more sequencing for direct single-cell resolution

Nanostring CosMx

- 960 targets (panel 20k, AGBT24)
- Sensitivity : << 30-80% (+)
- Imaging area:16 mm2 (2 days)
- Resolution:200 nm
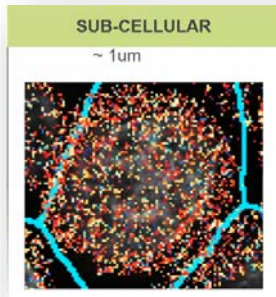
Vizgen Merscope

- 1.000 targets
- Sensitivity: 30-80% (+++)
- Imaging area: 100 mm2 (2 days)
- Resolution 100 nm
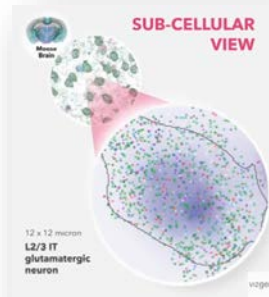
10xGenomics Xenium

- 400 - 6,000 targets
- Sensitivity : 530% (++)
- Imaging area:236 mm2 (4 days)
- Resolution 200 nm
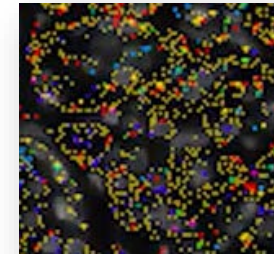
Imaging -based Spatial Transcriptomics   (2022)

No more sequencing for direct single-cell resolution

Nanostring CosMx
ISH-based

Vizgen Merscope
Multiplex Error-Robust FISH
Available (oct.2022)

10xGenomics Xenium
Cartana ISS, padlock probes / RCA
Available (jan.2024)

x15-50

x1-5

x4-8 / target gene
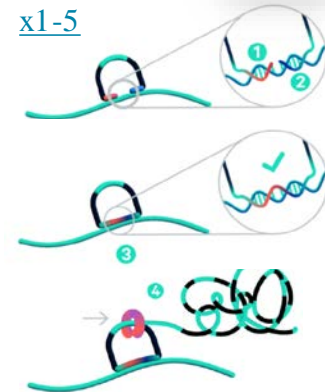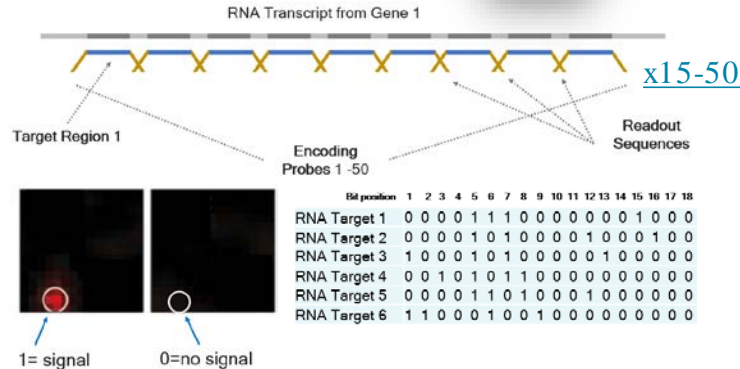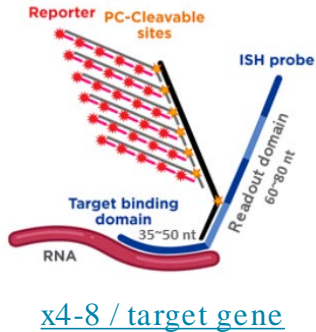
Cyclic *in situ* Hybridization Chemistries

Imaging-based Spatial Transcriptomics platforms comparison

2 recent bioRxiv comparative studies

A Comparative Analysis of Imaging-Based Spatial Transcriptomics Platforms

David P. Cook[1], Kirk B. Jensen[2,3,4], Kellie Wise[2,3], Michael J. Roach[2,3], Felipe Segato Dezem[6,7], Natalie K. Ryan[3,5], Michel Zamojski[9], Ioannis S. Vlachos[10,11,12], Simon R. V. Knott[13,14], Lisa M. Butler[3,5], Jeffrey L. Wrana[1,15], Nicholas E. Banovich[16], Jasmine T. Plummer[6,7,8*], Luciano G. Martelotto[2,3*]

Systematic benchmarking of imaging spatial transcriptomics platforms in FFPE tissues

Huan Wang[1,*], Ruixu Huang[2,*], Jack Nelson[1,*], Ce Gao[3], Miles Tran[3], Anna Yeaton[4], Kristen Felt[5], Kathleen L. Pfaff[6], Teri Bowman[7], Scott J. Rodig[6,7], Kevin Wei[3,7], Brittany A. Goods[2,**], Samouil L. Farhi[1,**]
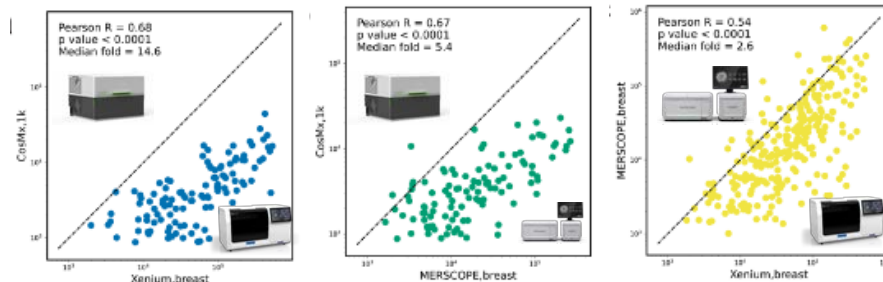
- CosMx is much less sensitive (high FPR)
- Merscope / Xenium for Fresh frozen slice
- Xenium optimal for FFPE slice

Gene targets panel design
Depending on the biological question !

Depending of your specific scientific focus

- Identify all major cell types, resolve cell subtypes
- Explore functional information
- Investigate interactions between cell types
- Ligand-receptors analysis
- Explore canonical signaling pathways
- Profile immune checkpoint molecules
- …

Satisfy technological system limitations

- Number of targets available
- Range of gene targets expression
- Total gene targets expression
- Budget around 15 k€ for 10 reactions

https://portal.vizgen.com/
https://cloud.10xgenomics.com/xeniumpanel-designer
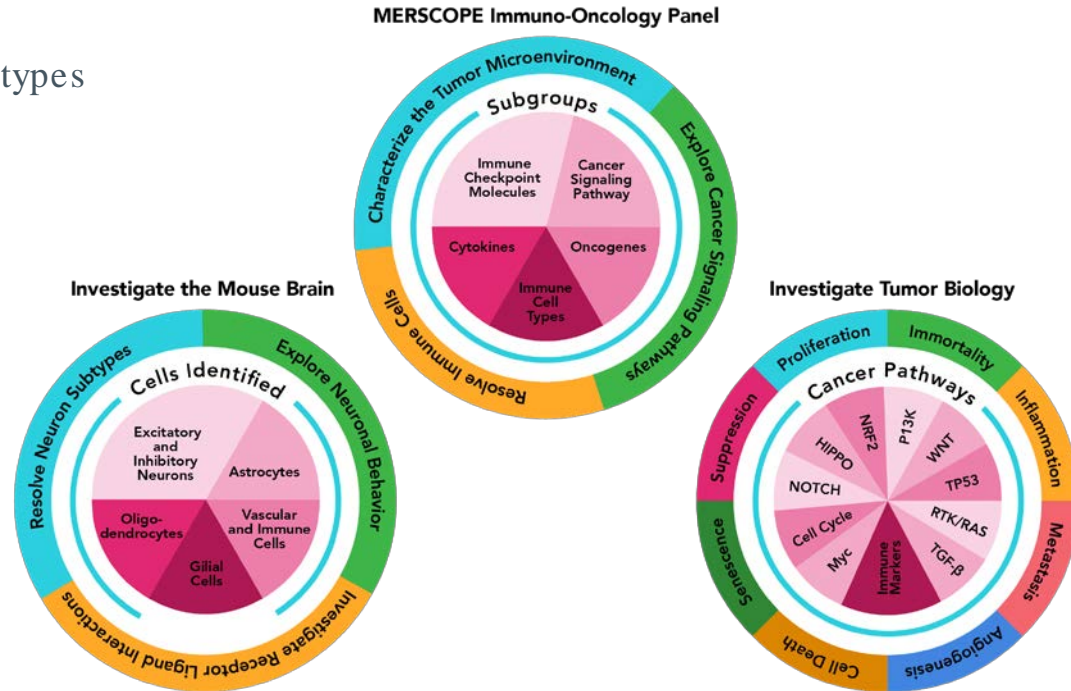
Experimental design
Take advantage of the large imaging area

Image-based spatial transcriptomics identifies molecular niche dysregulation associated with distal lung remodeling in pulmonary fibrosis

Annika Vannan[1,a], Ruqian Lyu[2,3,a], Arianna L. Williams[1], Nicholas M. Negretti[4], Evan D. Mee[1], Joseph Hirsh[4], Samuel Hirsh[4], David S. Nichols[5], Carla L. Calvi[5], Chase J. Taylor[5], Vasiliy. V. Polosukhin[5], Ana PM Serezani[5], A. Scott McCall[5], Jason J. Gokey[5], Heejung Shim[3], Lorraine B. Ware[5,7], Matthew J. Bacchetta[8], Ciara M. Shaver[5], Timothy S. Blackwell[5,9,10], Rajat Walia[11], Jennifer MS Sucre[4,9], Jonathan A. Kropski[5,9,10,b], Davis J McCarthy[2,3,b], Nicholas E. Banovich[1,b,*]

https://www.ihcworld.com/products/Quick-Ray-Mold.htm
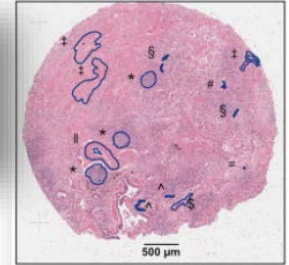
Each slide cost around 5 k€
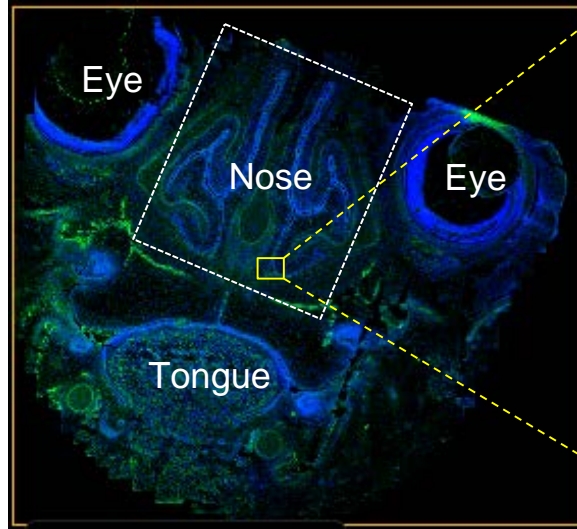multiplexing to remove batch effect and increase replicates for robust statistical analysis

Data acquisition
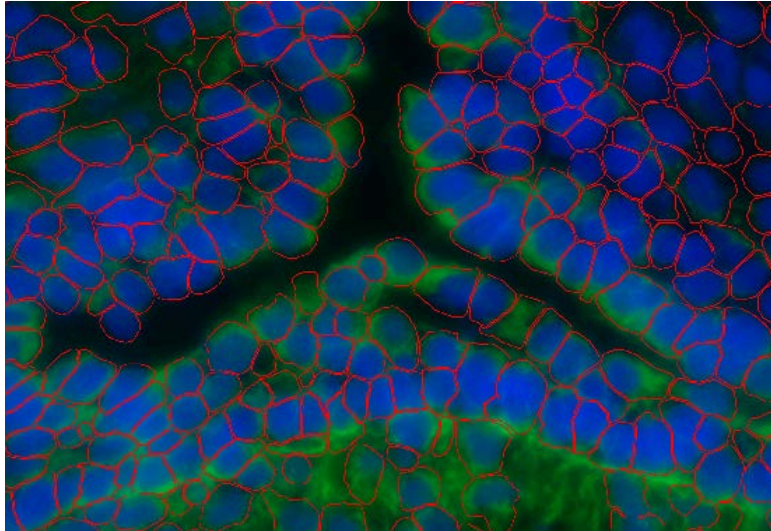DAPI and cell boundaries staining for cell segmentation
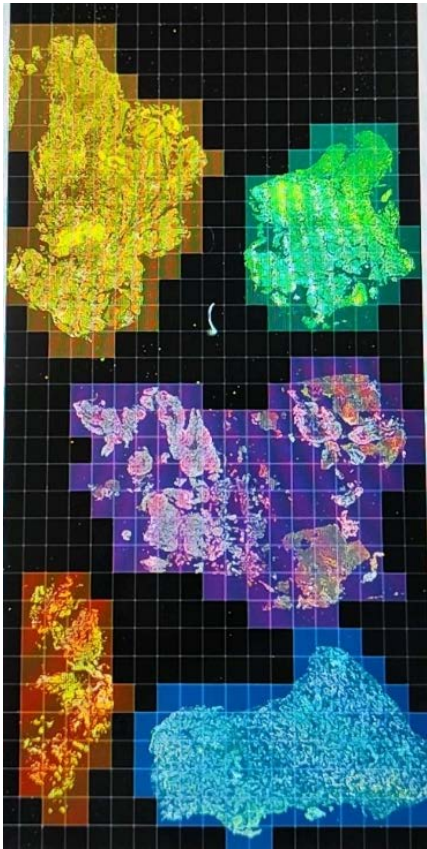
Human fetal head section (PCW9)
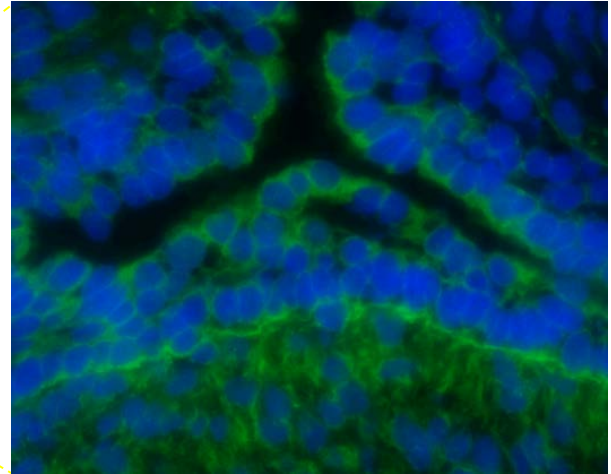
DAPI chanel
Cell boundaries chanel

Data acquisition
Cell segmentation

Cell segmentation is crucial to ensure cell x gene matrix purity for good subsequent biology

Article | Published: 14 December 2020

Cellpose: a generalist algorithm for cellular segmentation

Carsen Stringer, Tim Wang, Michalis Michaelos & Marius Pachitariu ✉

Nature Methods 18, 100–106 (2021) | Cite this article

10 μm

Z
0
1
2
3
4
5
6
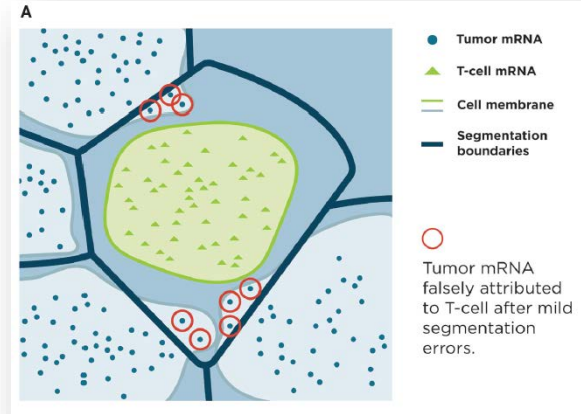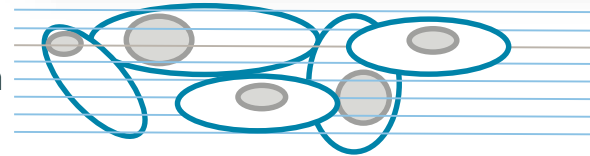
3D segmentation required, actually not used, 2D segmentation per Z then harmonizing and summing the detected transcripts for all Z into the harmonized segmentation mask (nuclei of full cell)

Raw data

Cell x genematrix
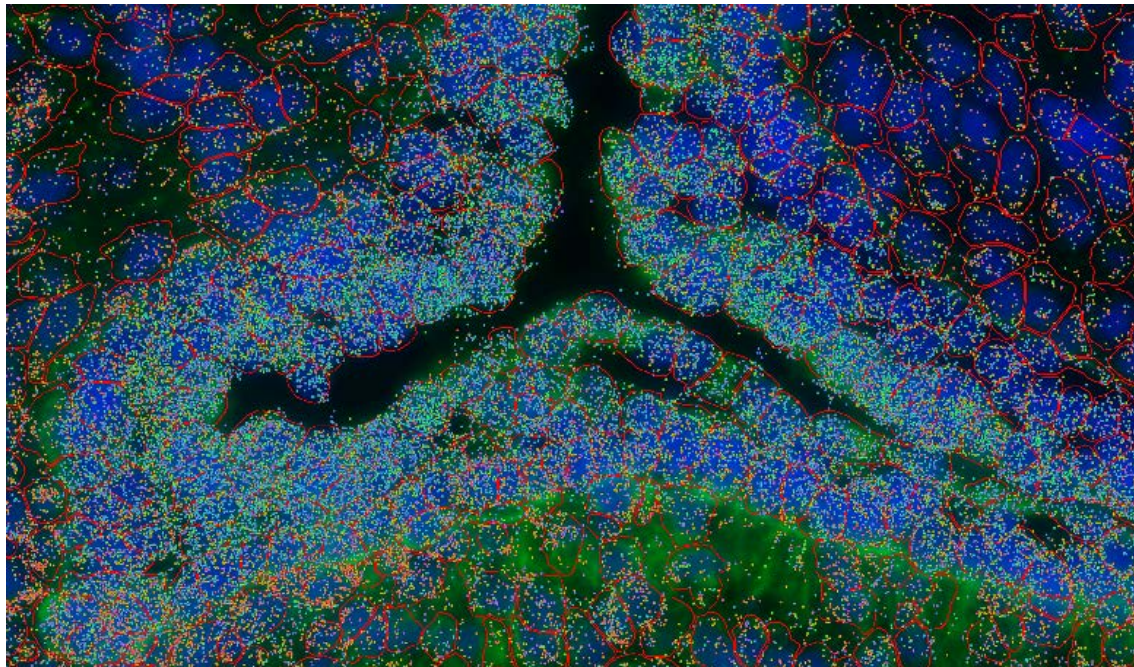
Gene-level matrix

100k's cells →

```
Clta  5  4 6 7   1 4  3  9  5  5 . 1 4 3   6 4 . 7  5 2  4
My16  5  2 5 1   2 4 13  2  5  2 4 4 1 4   8 4 2 3  3 1  1
Pkm   3  2 1 .   . 8  3  9  4 10 5 1 . 4 12 5 1 3   9 1  7
Tecr  3  1 5 2   4 6   5  9  6  3 2 2 1 4   7 1 1 1  3 2  8
Meis2 8 29 3 .   . 24 6  6 21 25 16 6 1 . 2 29 . .  6 18 . 10
...
```

1.000 Genes
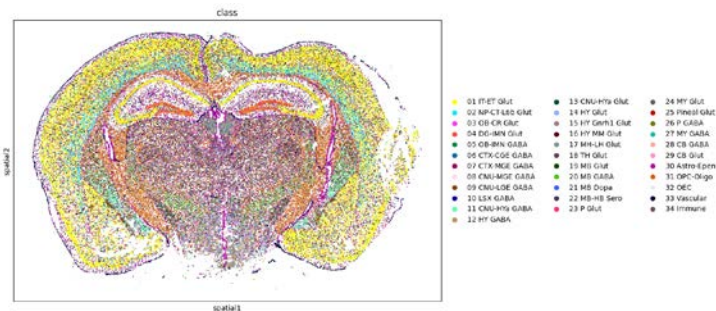
Statistical data analysis
Standardized workflows + packages development

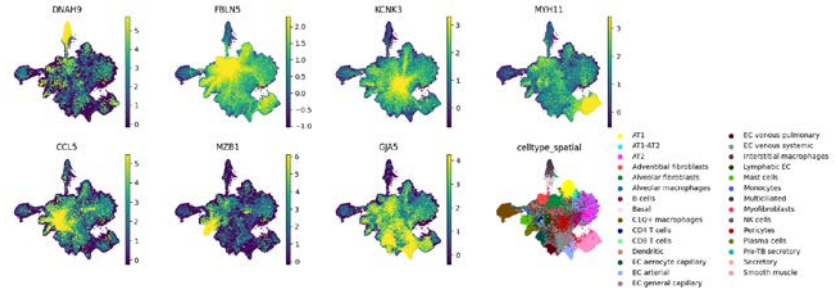

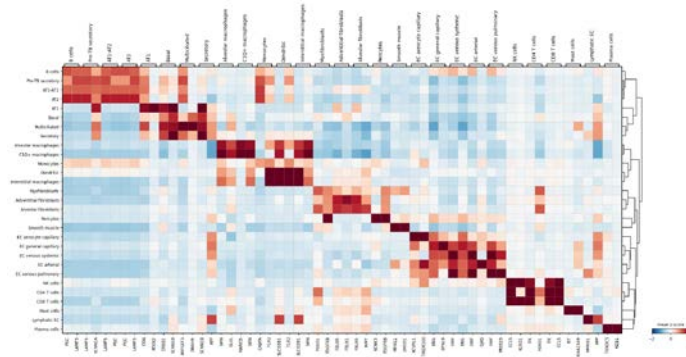Satija's lab, NYGC

Theis's lab, h...

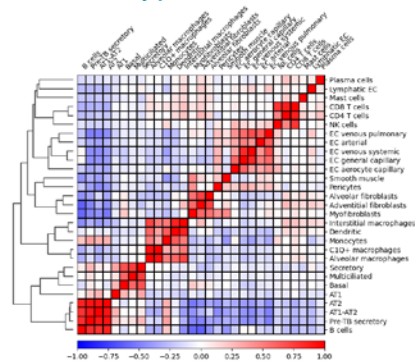Scverse ecosystem, Oliver Stegle & Fabian J. Theis

Single -cell data analysis
Scanpy and Squidpy toolkits

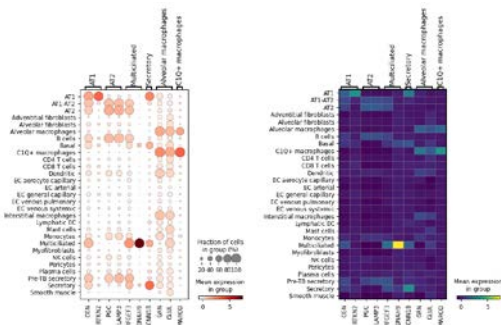## Gene marker detection, manual or automatic cell type identification

Batch effect correction, sample integration, cell type labeling transfer from single-cell references dataset
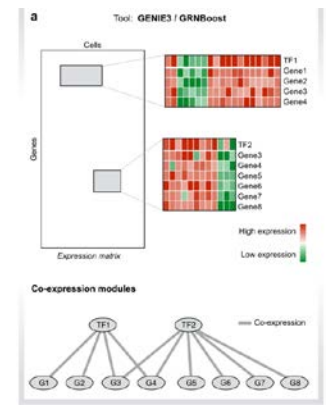
Cell type correlation

Differential expression analysis
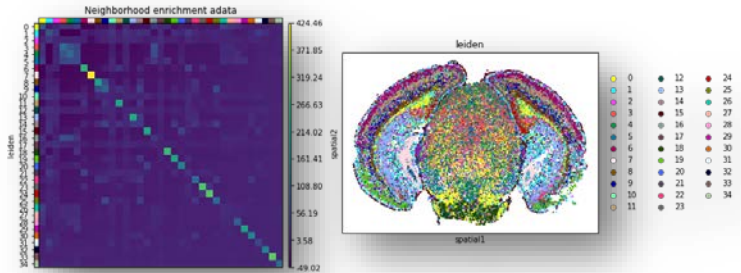## Gene set functional enrichment
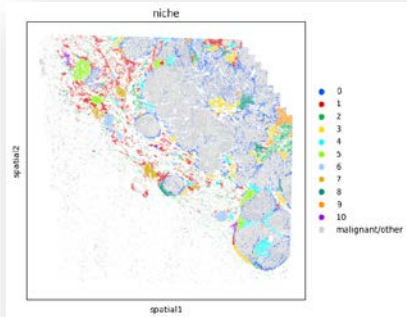
Transcription Regulatory Network

Single -cell data analysis including spatial resolution

New vast area for computational biologists (just like single-cell 5 years ago)


Neighbors enrichment analysis

Test if cells belonging to 2 clusters are close to each other more often than expected (co-occurrence probability)


Cellular niches analysis

for each cell, we count the number of neighbors that are of each cell type thus forming a "neighborhood profile" vector of length C, where C is the number of cell types. We then cluster all neighborhood profiles and call each cluster a "niche".
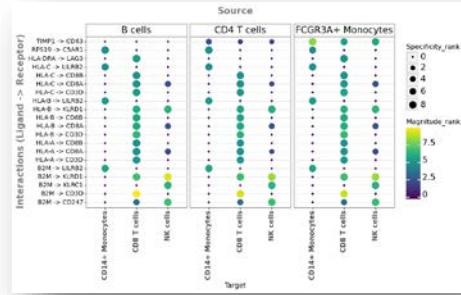
## Cell-cell communication Ligand-Receptor analysis

- Need to be in gene panel or inferred
- CellPhoneDB [Efremova et al., 2020]
- Omnipath [Türei et al., 2016].

Machine Learning & Statistical Analysis

Geospatial Statistics | Molecular Gradients | Subcellular Compartments | Data Visualization
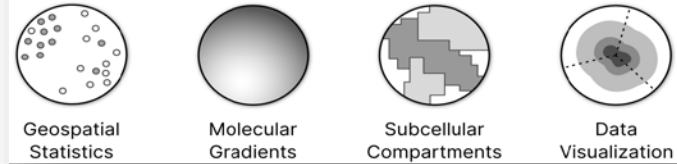
Sub-cellular exploration

Bento is a Python toolkit for performing subcellular analysis of spatial transcriptomics
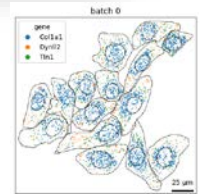
Acknowledgments

Institut de Pharmacologie Moléculaire et Cellulaire