Single-cell to Spatial Isoform Transcriptomics

Kévin LEBRIGAND, PhD
Institut de Pharmacologie Moléculaire et Cellulaire
Université Côte d'Azur, CNRS, France.
lebrigand@ipmc.cnrs.fr
@kevinlebrigand
@kevinlebrigand@genomic.social

FRANCE GENOMIQUE
UNIVERSITÉ CÔTE D'AZUR
ipmc
cnrs

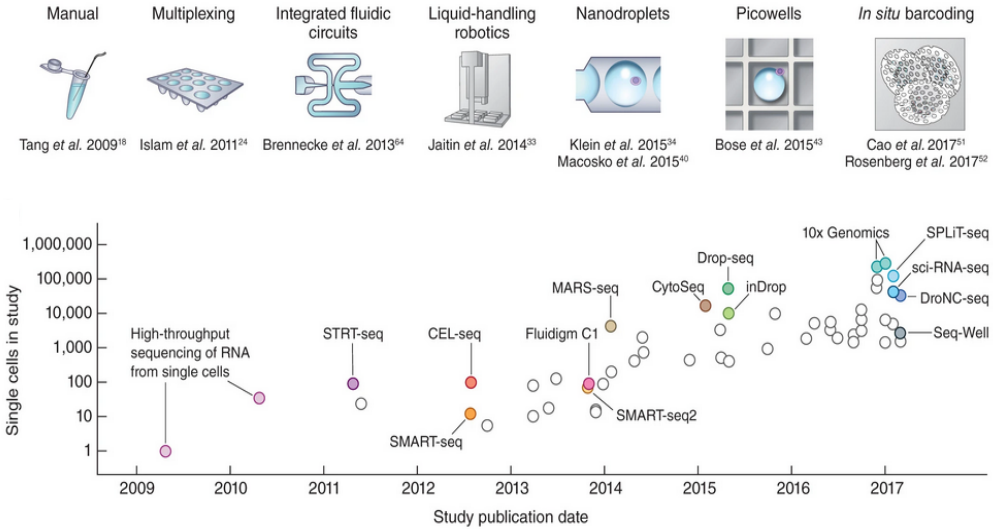
01

Single-cell Transcriptomics (NGS)

Nature Method Of the Year 2013

Single-cell transcriptomics

Evolution of isolation techniques and throughput

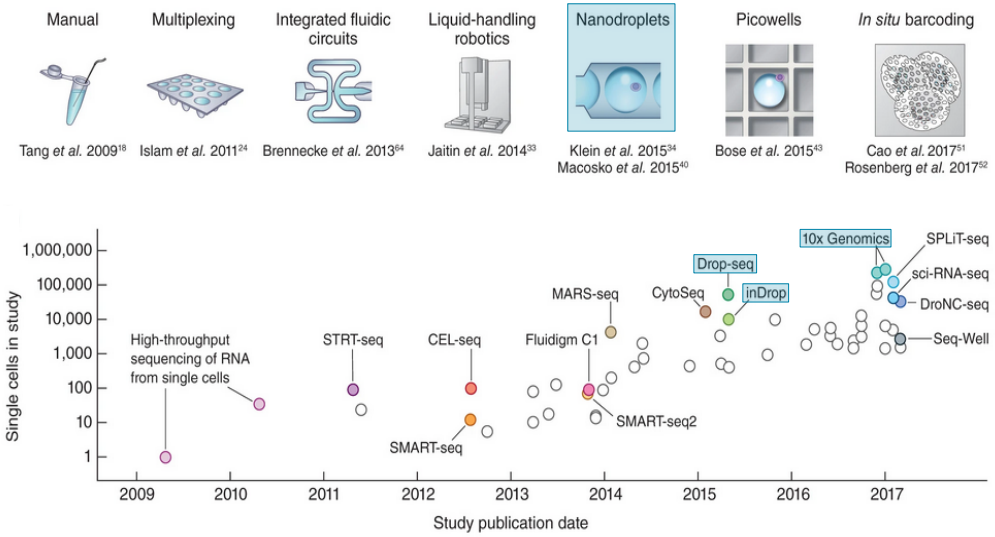


Exponential scaling of single-cell RNA-seq in the past decade
Svensson et al., *Nature Protocols*, 2018

Single-cell transcriptomics

Droplet-based approaches

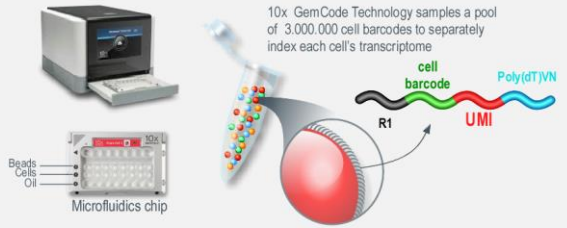
InDrop, Klein et al, 2015
Drop-seq, Macosko et al, 2015
10x Genomics, Zheng et al, 2016

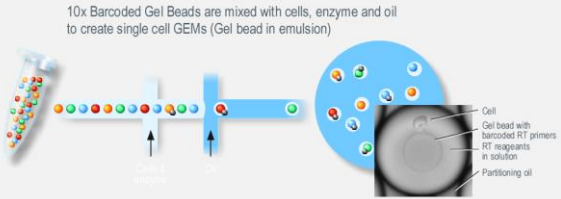


Exponential scaling of single-cell RNA-seq in the past decade
Svensson et al., *Nature Protocols*, 2018

10x Genomics Chromium single cell controler (2016)

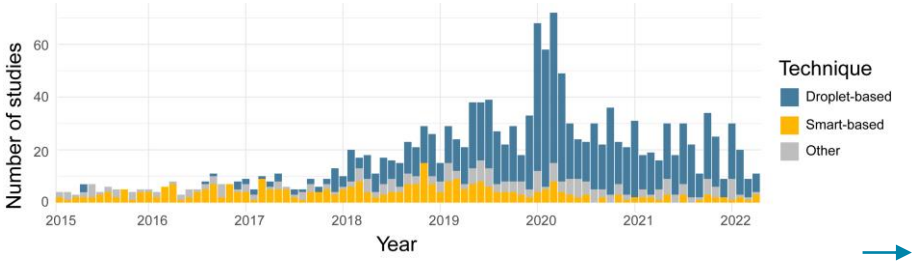
- Easy-to-set-up and robust workflow
- High scalability (1,3M cells dataset)





Single-cell transcriptomics

Single cell approaches in publications

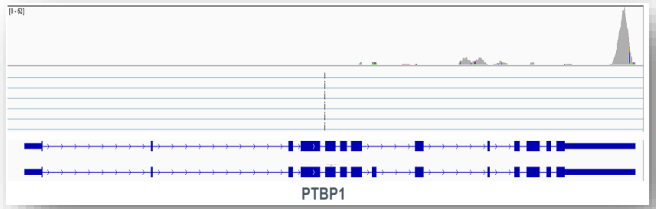


A curated database reveals trends in single cell transcriptomics
Svensson et al., Database , 2020

- Huge amount of single-cell studies in the past 5 years,
- Droplet-based approaches = 61% (Chromium: 47%)

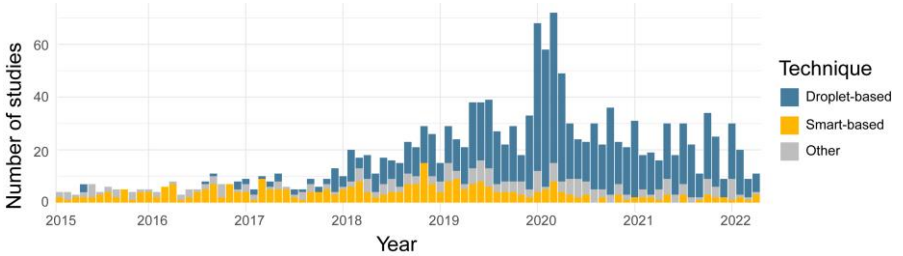
Technique	Count
Chromium	725
Smart-seq2	177
SMARTer (C1)	124
Drop-seq	74
SMARTer	28
InDrops	23
CITE-seq	18
CEL-seq2	17
STRT-seq	17
MARS-seq	16
Tang	15
CEL-seq	13
STRT-seq (C1)	13
Seq-Well	13
SORT-seq	12
BD Rhapsody	11
BioMark	8
GemCode	7
ICELL8	7
Perturb-seq	7
Patch-seq	6
sc-RT-mPCR	6
MERFISH	5

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



Single-cell transcriptomics

Single cell approaches in publications



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

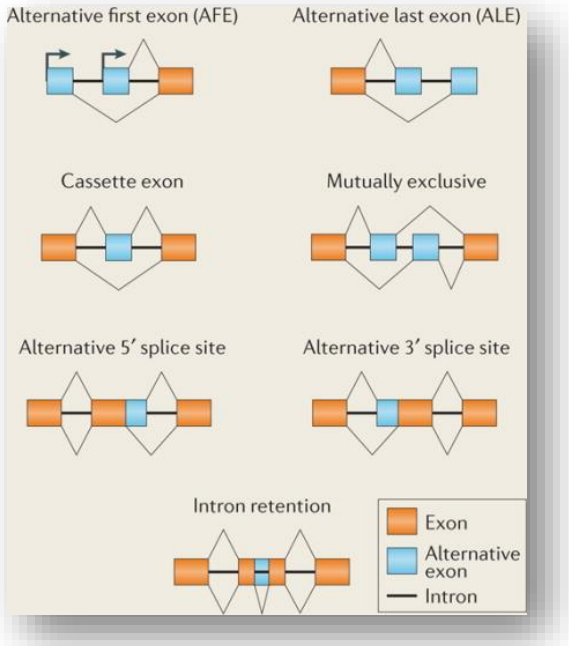
- Huge amount of single-cell studies in the past 5 years,
- Droplet-based approaches = 61% (Chromium: 47%)
- Smart-based approach = 21%, <5% in the last 2 years

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

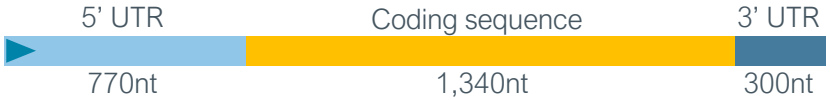
Transcriptomics

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** :

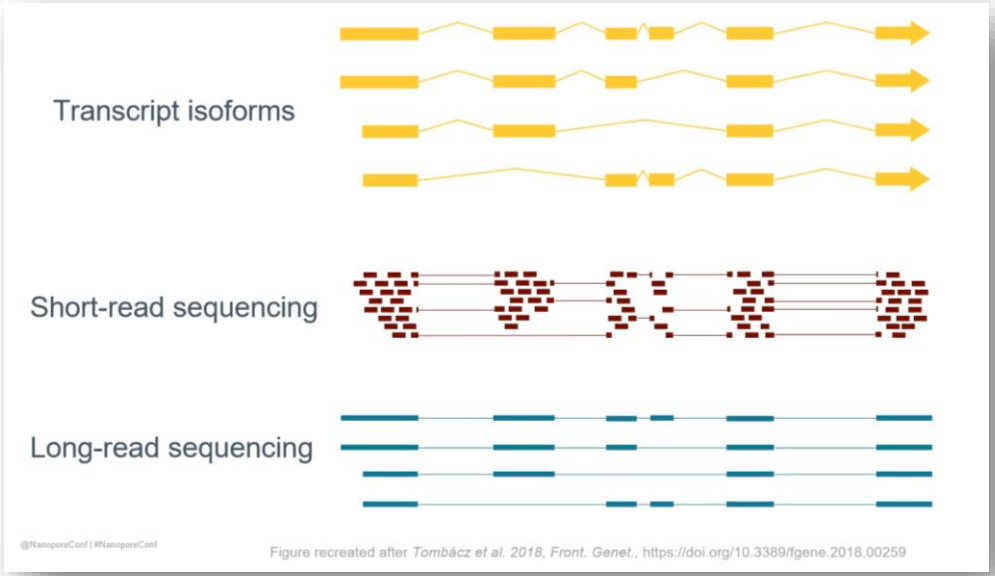


Alternative splicing and disease
Tazi et al., 2008

Long-read sequencing identifies isoforms efficiently

Nature Method Of The Year 2022

Nature Method of the Year 2022



➡ Inference required

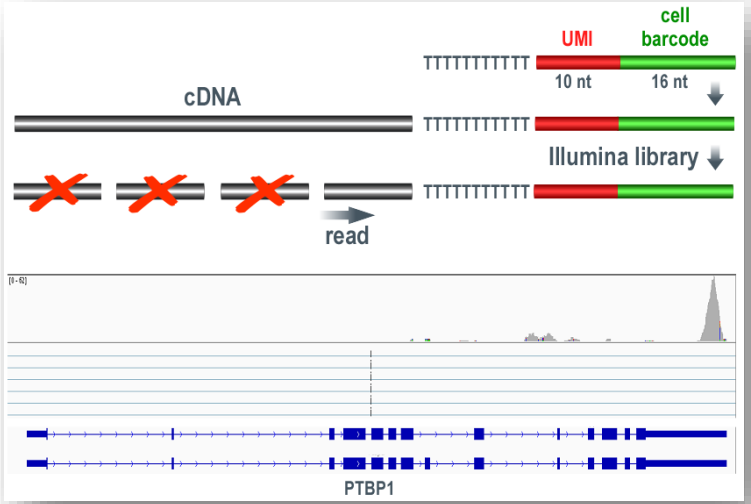
➡ Direct full exonic layout
One read is one molecule

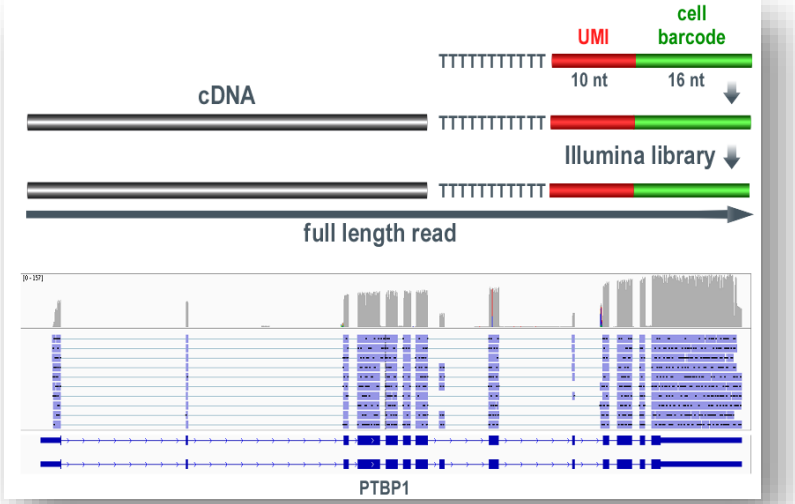
Single-cell long-read transcriptomics

Droplets-based approach short reads vs long reads

Illumina Nextseq500

Nanopore PromethION





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

is lost

remain accessible

Single-cell long-read transcriptomics

SiCeLoRe, bioinformatics for Single Cell Long Read

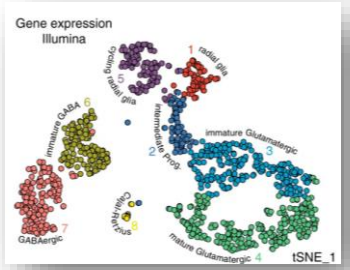
1.200 cells

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

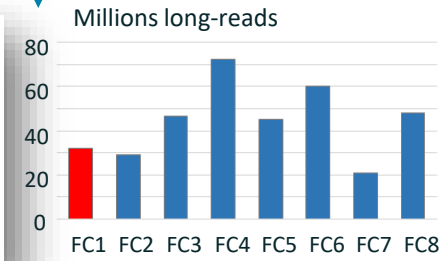
split emulsion before RT

15%
190 cells

85%
951 cells

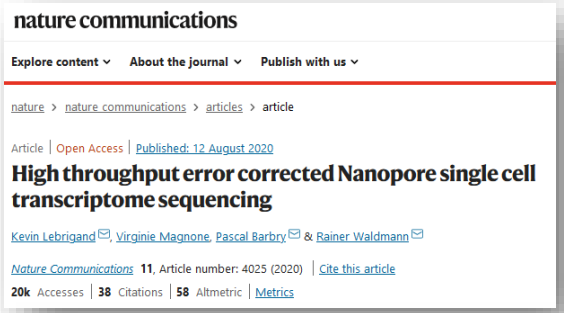


standard workflow



190 cells: 32M
951 cells: 322M

R9.4, LSK-109 chemistry (2018)



nature communications

Explore content · About the journal · Publish with us

nature › nature communications › articles › article

Article | Open Access | Published: 12 August 2020

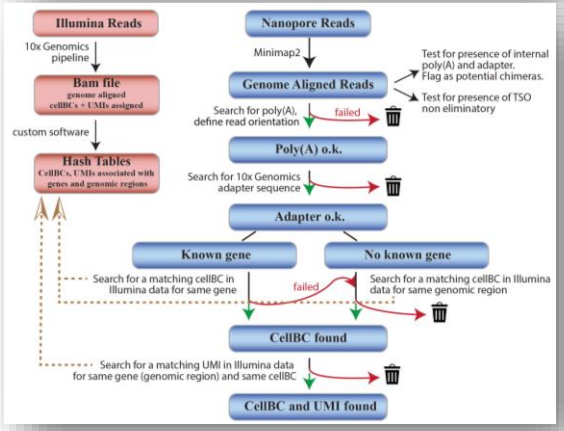
High throughput error corrected Nanopore single cell transcriptome sequencing

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

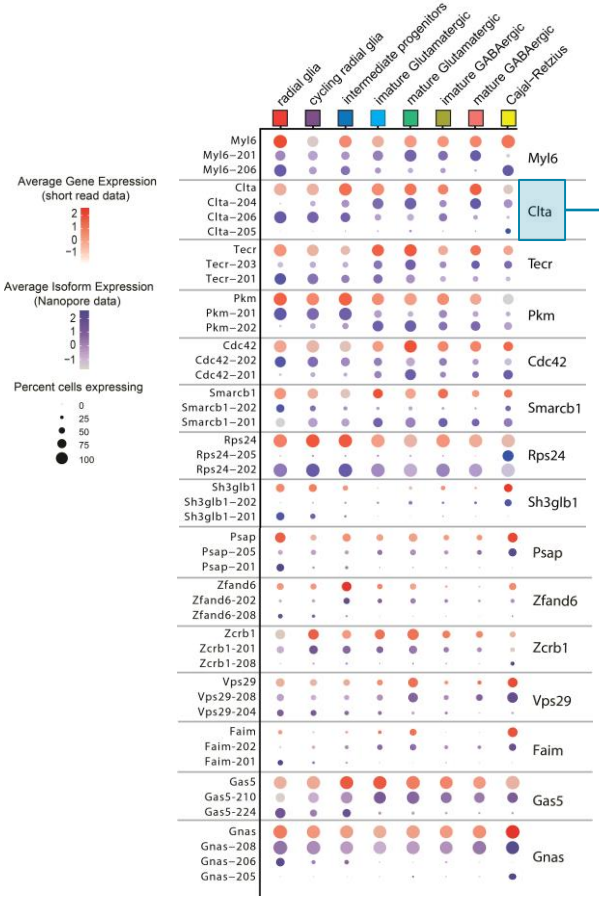
Rainer Waldmann

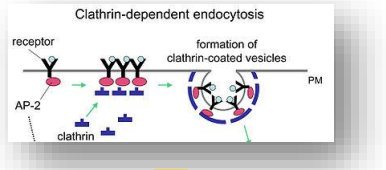


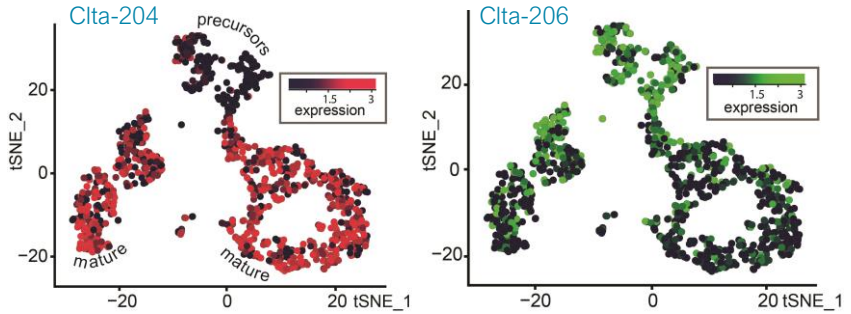
https://github.com/ucagenomix/sicelore

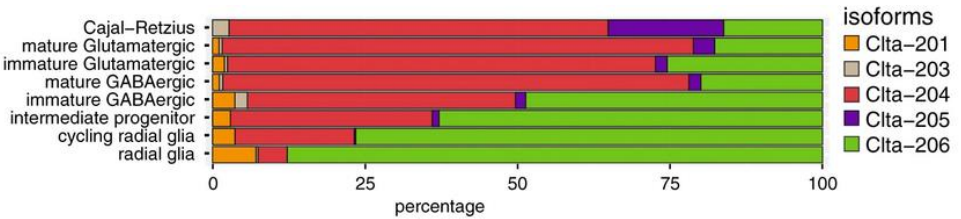
Single-cell long-read transcriptomics reveals diversity

76 isoform-switching genes along neuronal maturation





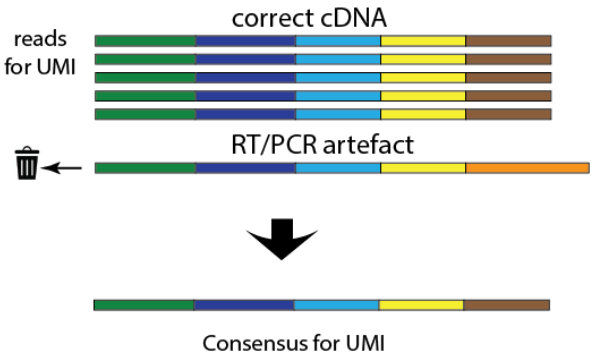




Single-cell long-read transcriptomics reveals sequence heterogeneity

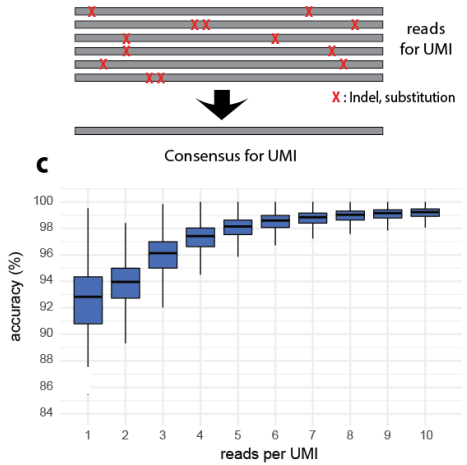
Consensus sequence computation per UMI

UMIs enable elimination of PCR artifacts



Crucial for accurate novel isoform discovery

UMIs enable correction of sequencing errors



High accuracy for Single Nucleotide Variation call

02

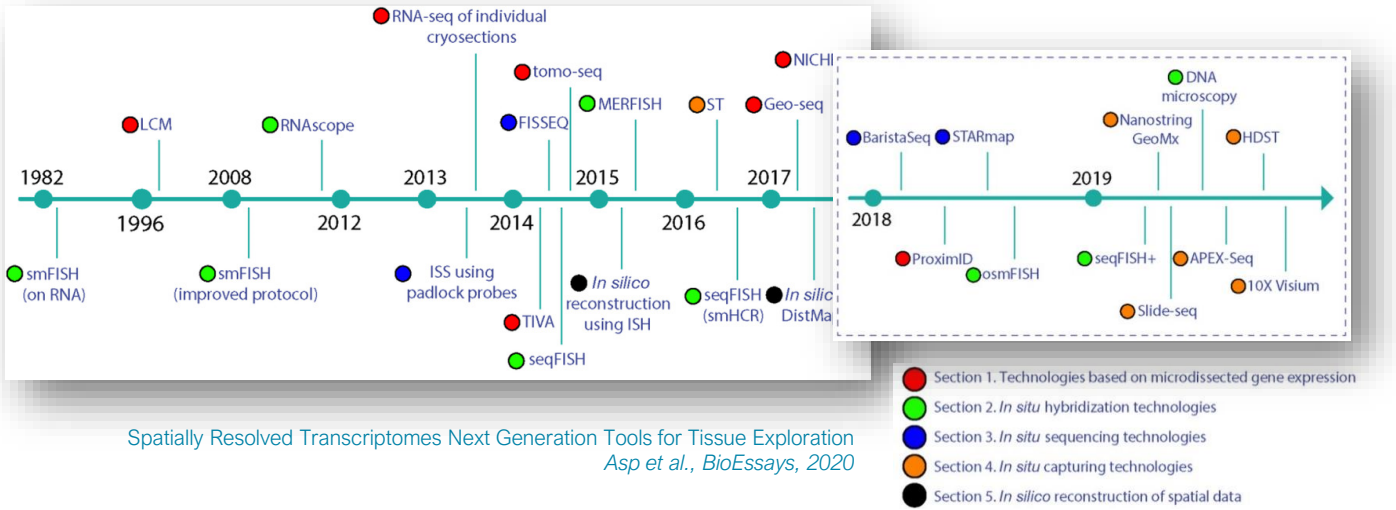
Spatial Transcriptomics (NGS)

Nature Method Of the Year 2020

Spatial transcriptomics approaches

Timeline

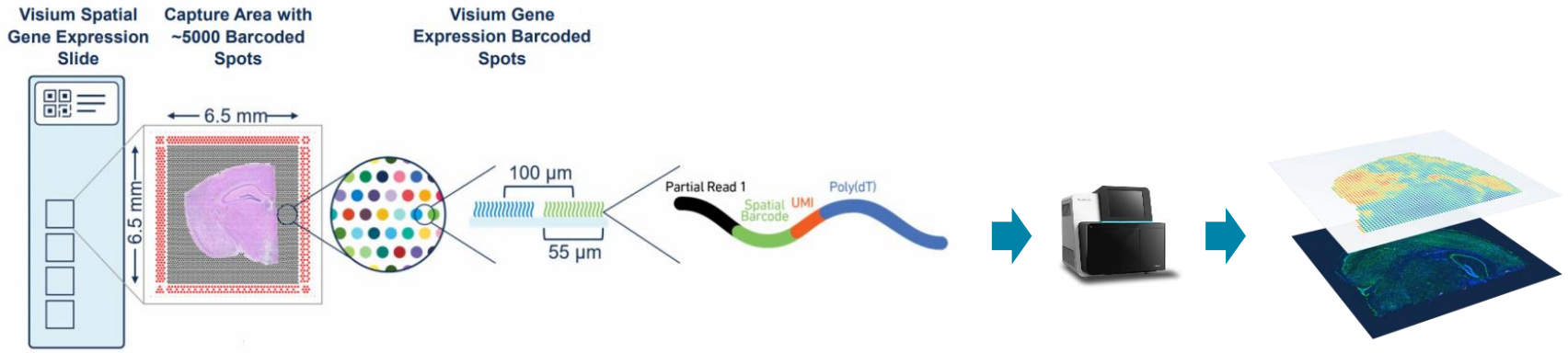
- Spatial transcriptomics aims at directly visualize gene expression in their original environment,
- It tackles the main limitation of single cell experiment missing the spatial organization,
- A lot of developments in the last years thanks to recent advances in different fields,

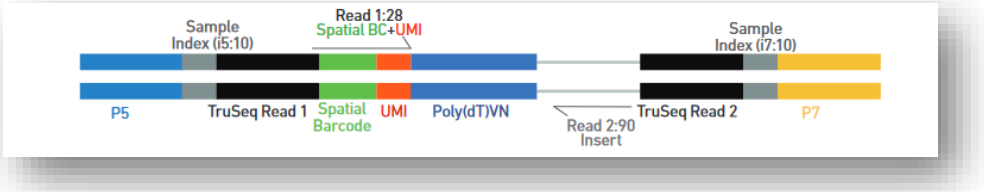


Spatially Resolved Transcriptomes Next Generation Tools for Tissue Exploration
Asp et al., BioEssays, 2020

In-situ capture spatial transcriptomics

10x Genomics Visium (2019)





→ Spatial barcode / UMI assignment strategy identical to single cell transcriptomics

Spatial isoform Transcriptomics (SiT)

The spatial landscape of gene expression isoforms in tissue sections

The spatial landscape of gene expression isoforms in tissue sections

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

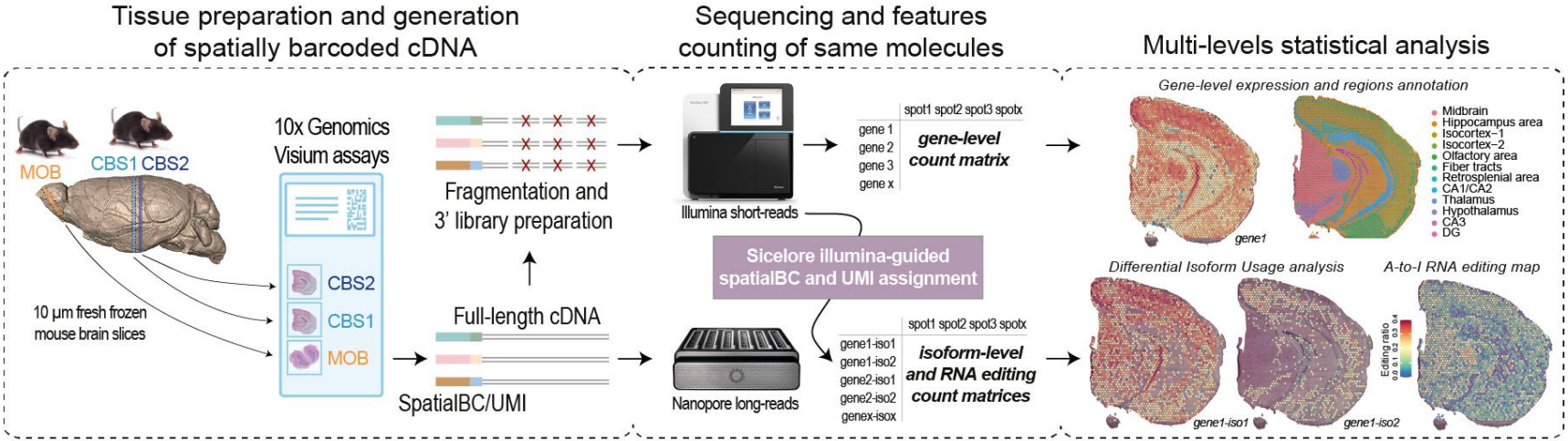
doi: https://doi.org/10.1101/2020.08.24.252296

bioRχiv THE PREPRINT SERVER FOR BIOLOGY

August 24th, 2020
January 5th, 2022

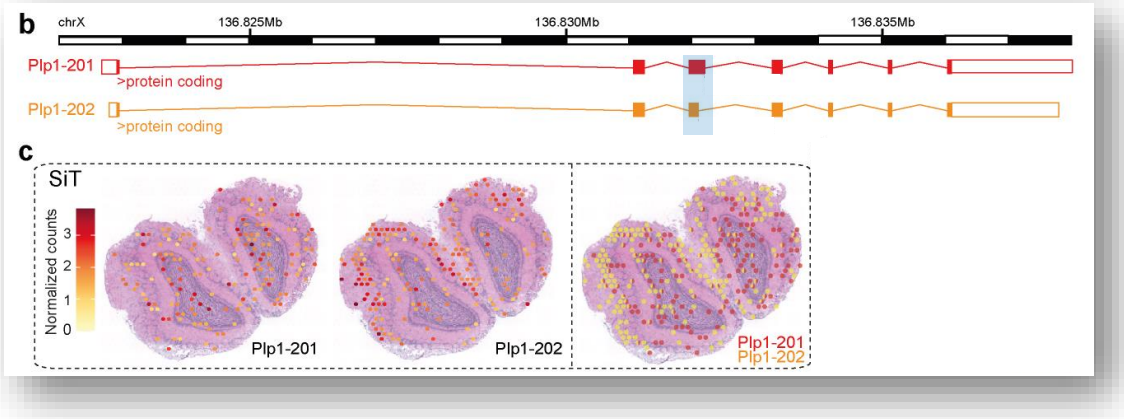
Nucleic Acids Research

in press



SiT reveals specific splicing pattern across MOB regions

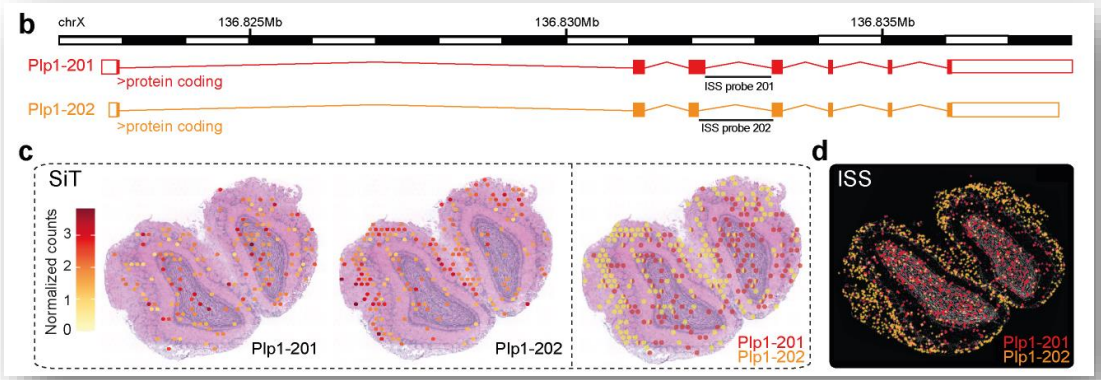
Plp1 Differential Transcript Usage (DTU)



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)



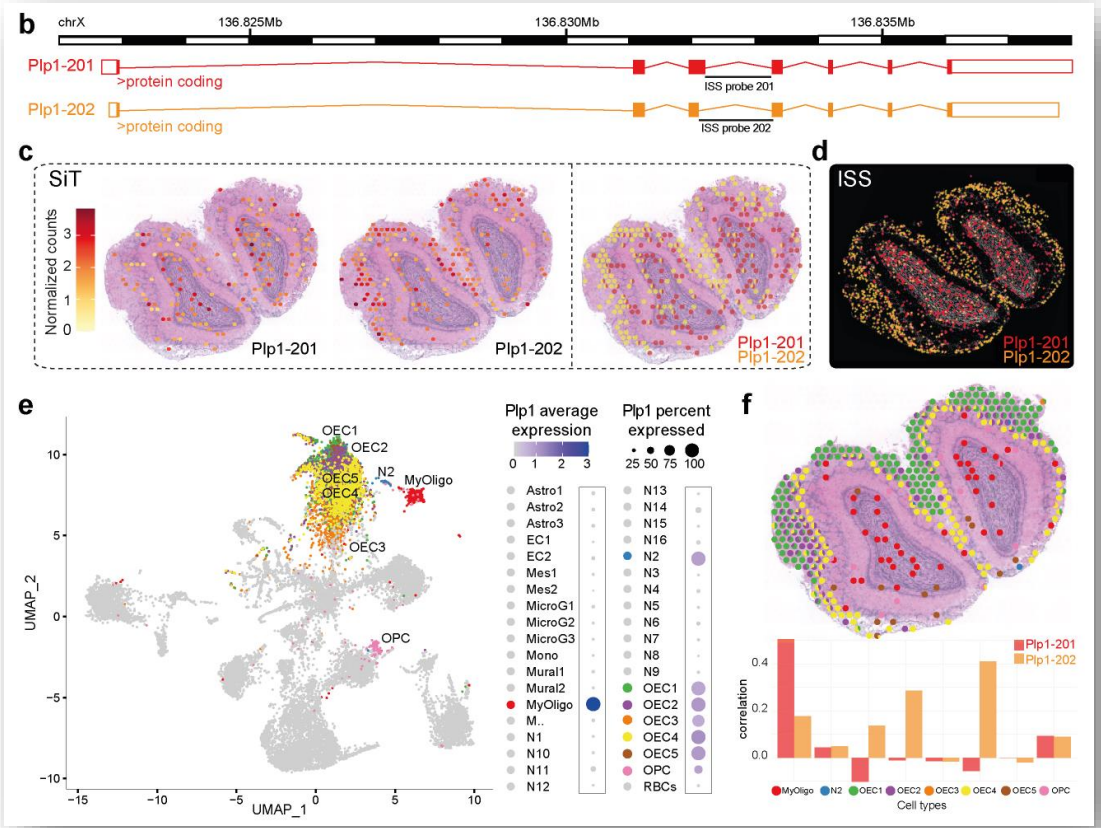
Proteolipid Protein 1 (Plp1) is a gene involved in severe pathologies associated with CNS dysmyelination

CARTANA
Part of 10x Genomics

In Situ Sequencing Data

SiT reveals specific splicing pattern across MOB regions

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



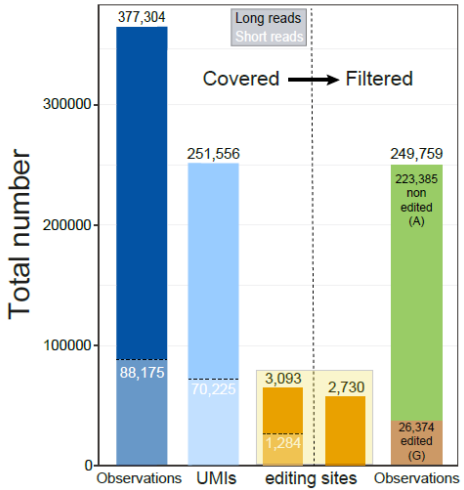
Proteolipid Protein 1 (Plp1) is a gene involved in severe pathologies associated with CNS dysmyelination

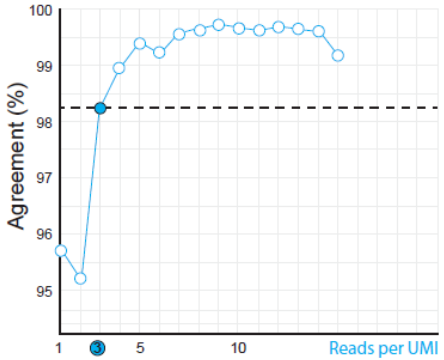
Spatial spot deconvolution of prominent *Plp1* expresser cell types. Correlation Deconvolution score / *Plp1* isoforms expression correlation shows that *Plp1* is predominantly expressed as Plp1-202 by olfactory ensheathing cells (OEC) in the ONL and as Plp1-201 isoform by myelinating-oligodendrocytes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

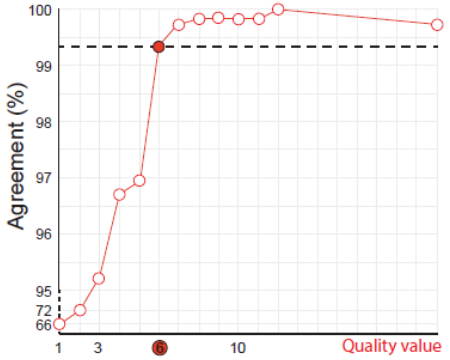
SiT reveals full-length sequence heterogeneity

High confidence SNV call (>99%) calibration using long- vs short-read agreement

- Exploration of 5,817 A-to-I RNA editing sites described in the literature (Ramaswami et al., 2013 (RADAR), Licht et al., 2019)
- Calibration by looking at agreement between long and short read base calls for 88,175 shared UMI / Editing site observations
  - number of reads per UMI >= 3
  - consensus Phred score QV >= 6



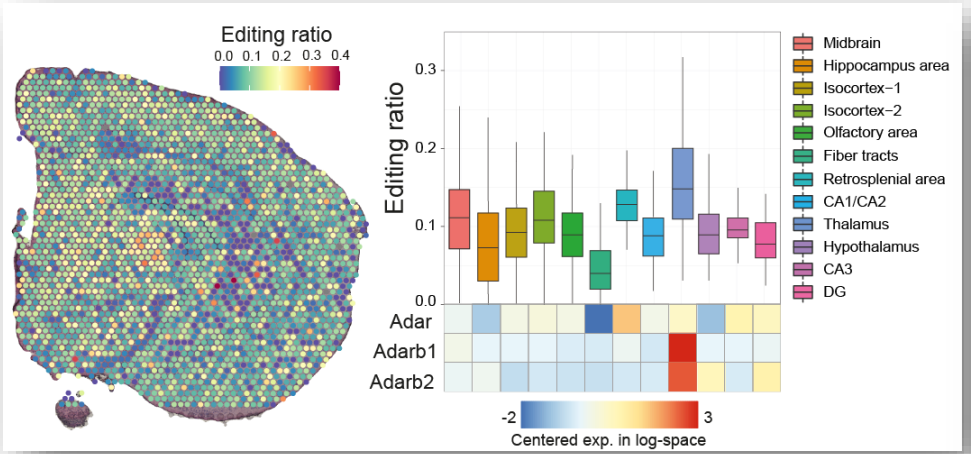


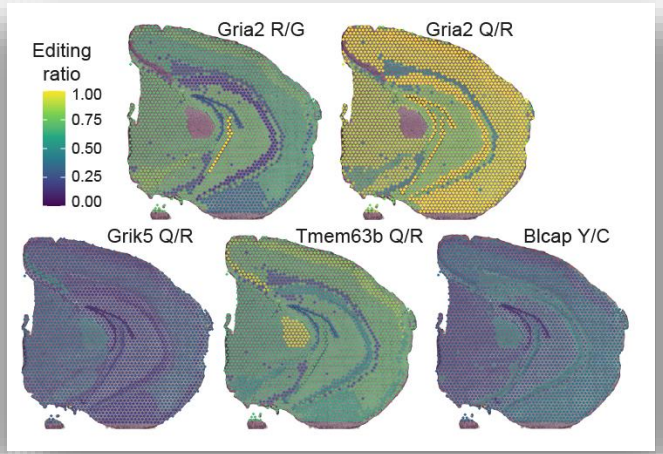


SiT reveals full-length sequence heterogeneity

Mouse brain map of A-to-I RNA editing activity

- Exploration of 5,817 A-to-I RNA editing sites described in the literature (Ramaswami et al., 2013 (RADAR), Licht et al., 2019)





Single cell and Spatial isoform transcriptomics

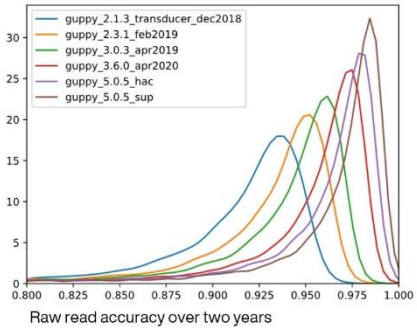
Summary

- 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

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/sicelore-2.1

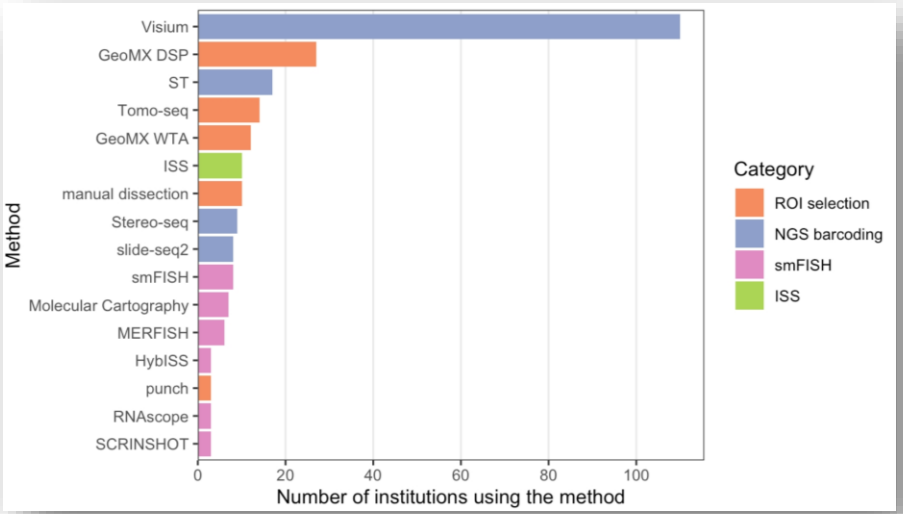
- Visium and single-cell 3' and 5' libraries
- Illumina-free profiling available

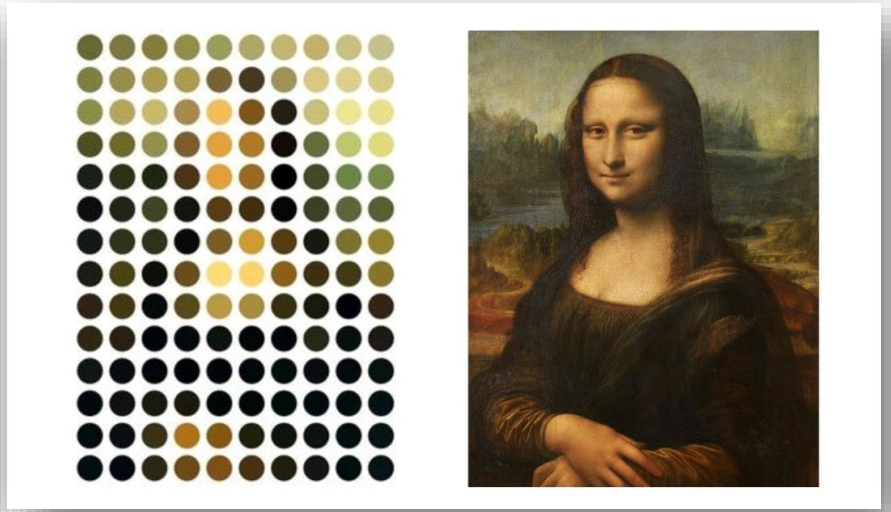
03

Spatial single-cell transcriptomics (imaging)

Spatial transcriptomics technologies (2020-2022)

Visium is widely adopted by academics





But is not the ideal readout for spatial biology

(Akoya credit rough caricature)

Spatial imaging technologies (2023)

No more sequencing for direct single-cell resolution

Nanostring CosMx

ISH-based

- 960 targets
- Sensitivity : < 30-80%
- Resolution: 200 nm
- Imaging area: 16 mm2

Vizgen Merscope

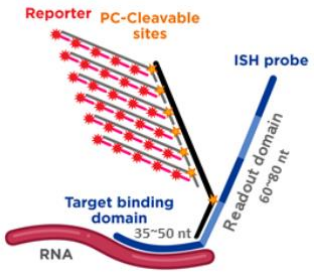
Merfish

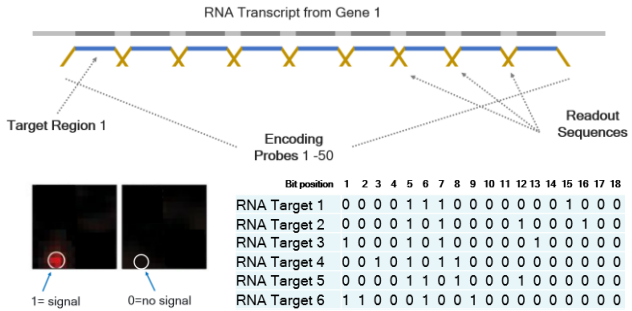
- 500 targets (1,000 soon)
- Sensitivity : 30-80%
- Resolution: 100 nm
- Imaging area: 100 mm2

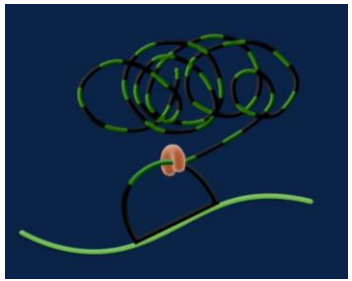
10xGenomics Xenium

Cartana ISS, padlock probes / RCA

- 400 targets
- Sensitivity : 5-30%
- Resolution: 200 nm
- Imaging area: 25 mm2







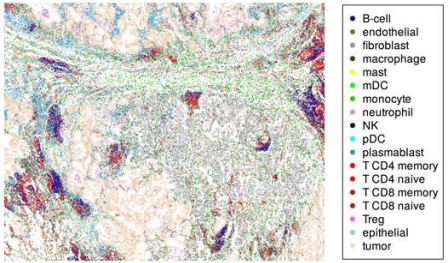
Spatial imaging technologies (2023)

Compare available datasets

Vizgen Merscope

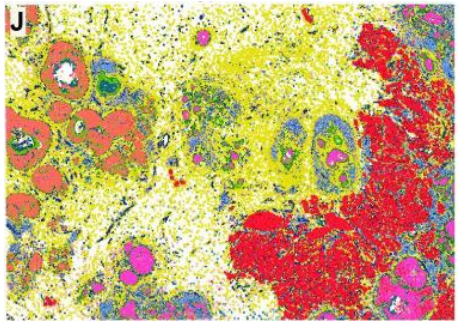
- Xiaowei Zhuang's lab merfish publications
  - Chen et al., Science (2015)
  - Moffitt et al., PNAS (2016), Science (2018)
  - Emanuel G et al., Nature Methods (2017)
  - Xia C. et al., PNAS (2019, Scientific Reports (2019)
  - Zhang M. et al., Nature (2021)
  - …
- Internal data release program
  - Human Immuno-oncology (**breast**, colon, **lung**, liver, skin, prostate, uterine and ovarian) 500 genes, >4 billion transcripts, 9 million cells
  - Mouse Liver Map (347genes)
  - Mouse brain Receptor Map (483 genes)
- External labs publications
  - Dixon E. et al., Kidney Int. (2022): Kidney
  - Wang et al., Nat. Neuro. (2022): Mouse olfactory Glomerular map
  - Stogsdill et al., Nature (2022): Neocortex microglia
  - …

Nanostring CosMx



- Release date: 11/2021
- FFPE Human NSCLC (**Lung**)
- 960 gene targets
- 8 sections for 800k cells
- Imaging area: 8 x 16 mm^2
- 259,604,214 transcripts
- Mean transcripts/cell: 265

10xGenomics Xenium

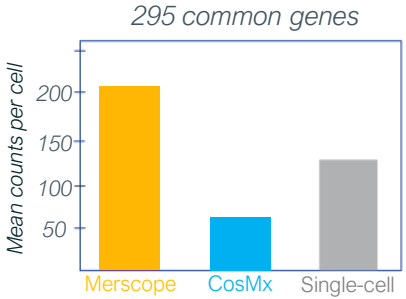


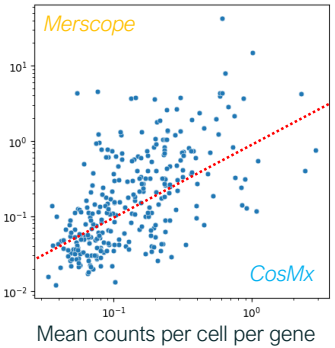
- Release date: 10/2022
- FFPE Human **Breast** cancer
- 313 gene targets
- 167,885 cells,
- 36,944,521 transcripts
- Imaging area: 40 mm^2
- Mean transcripts/cell: 193

Spatial imaging technologies comparison

Compare available datasets: Lung and Breast cancer samples

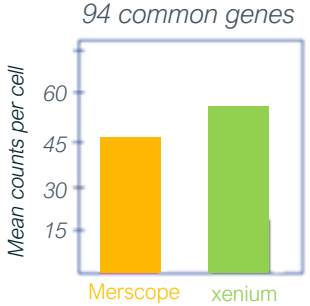
FFPE Human Lung Cancer	Merscope	CosMx
Total cells	353 k (x4)	92 k
Detected transcripts	107 M (x4)	26 M
Gene targets	500	960 (x2)
Total RPKM	9,204	61,680 (x6)
Mean transcripts/cell	302	284

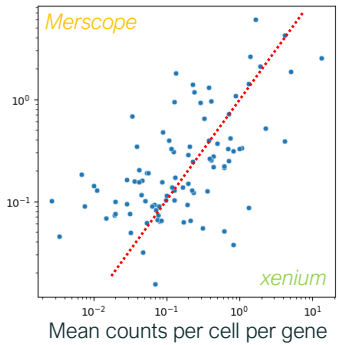




https://vizgen.com/wp-content/uploads/2022/12/Vizgen-Spatial-Genomics-Data-Quality-eBook-1.pdf

FFPE Human Breast Cancer	Merscope	Xenium
Total cells	713 k (x4)	168 k
Detected transcripts	353 M (x10)	32 M
Gene targets	500	313
Total RPKM	9,909	7,912
Mean transcripts/cell	495	193

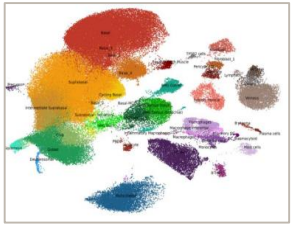




MERSCOPE @ UCAGenomiX (Nice-Sophia-Antipolis)

October 2022

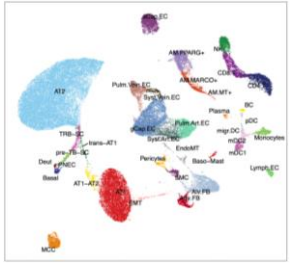
- Human Lung Cell Atlas (CZI) discovAIR
  Discovering the Cellular Landscape of the Airways and Lung Tissue



  - 12 control / 2 IPF / 10 COPD patients
  - 415k cells (117 samples)
  - 48 cell types

- DSRemodeling (mouse model brain, Massimo Mantegazza, IPMC)
  Molecular and functional remodeling in the developmental and epileptic encephalopathy Dravet Syndrome (DS)

- Pulmonary Arterial Hypertension (Christophe Guignabert, Paris-Saclay)



  - 7 control / 7 PAH patients
  - 69,949 cells
  - 39 cell types

Acknowledgments

Institut de Pharmacologie Moléculaire et Cellulaire

Pascal Barbry Lab (IPMC, CNRS, France)
- Rainer Waldmann

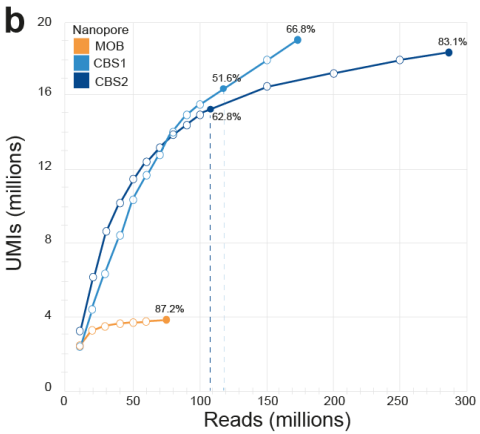
Joakim Lundeberg Lab (KTH Royal Institute of Technology, Sweden)
- Joseph Bergenstråhle
- Kim Thrane

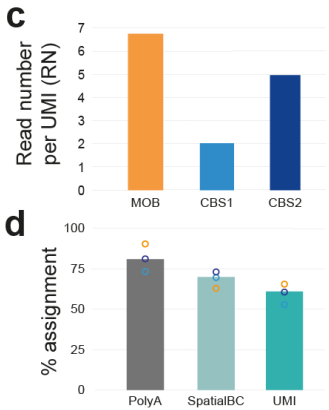
UCAGenomiX platform (IPMC, CNRS, France)
- Marie-Jeanne Arguel

Spatial isoform Transcriptomics (SiT)

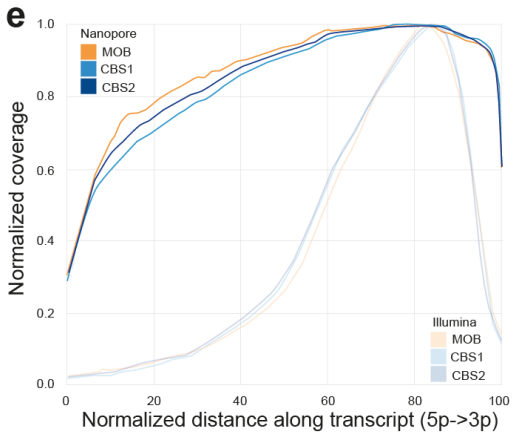
Nanopore long-read sequencing

Sequencing saturation curves per samples





Transcripts full-length coverage

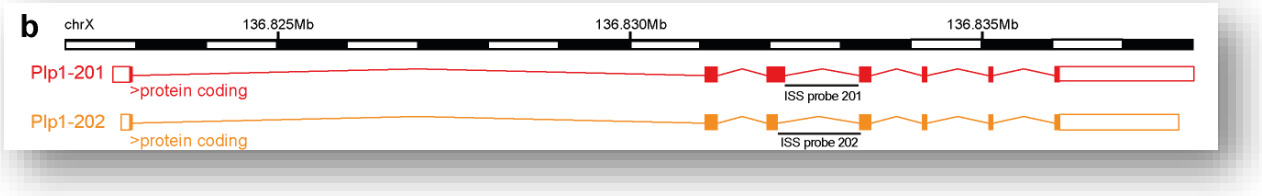


Reads	MOB		CBS1			CBS2											
Date	18 feb. 20	20 mar. 20	18 feb. 20	20 mar. 20	24 feb. 21	12 may 20	13 may 20	19 may 20	25 may 20	25 may 20	26 may 20	27 may 20	09 feb. 21	Total			
Flow cells	PAE06474	PAE59649	PAE01745	PAE59645	PAG52067	PAE59606	PAE59231	PAE32756	PAE32753	PAE31188	PAE21339	PAD99555	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 or 2 Promethion flow cells per slice - R9.4, LSK-110 chemistry (2021)

Spatial imaging technologies

No more sequencing not compatible with isoform transcriptomics



- Validating one exon-exon junction is not really the same as validating a full-length isoform (complete exons layout required)
- Current limitations in # target genes is even more an issue when dealing with the huge amount of exon-exon junctions, Human Ensembl GRCh38.108:
  - 20,036 protein coding genes,
  - 170,185 described isoforms,
  - 626,750 unique exons,
  - 616,988 unique exon-exon junctions
- Even in case of no more limitation in # targets, we would not be able to link together individual exon-exon signal to reconstitute unambiguously each gene isoforms expression

From Gene-level to Isoform-level long-read profiles

Enhanced description of single-cell transcriptome

Gene-level matrix

cells

Genes

Clta	5	4	6	7	1	4	3	9	5	5	.	1	4	3	6	4	.	7	5	2	4
Myl6	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
...																					

Isoform-level matrix

cells

Isoforms

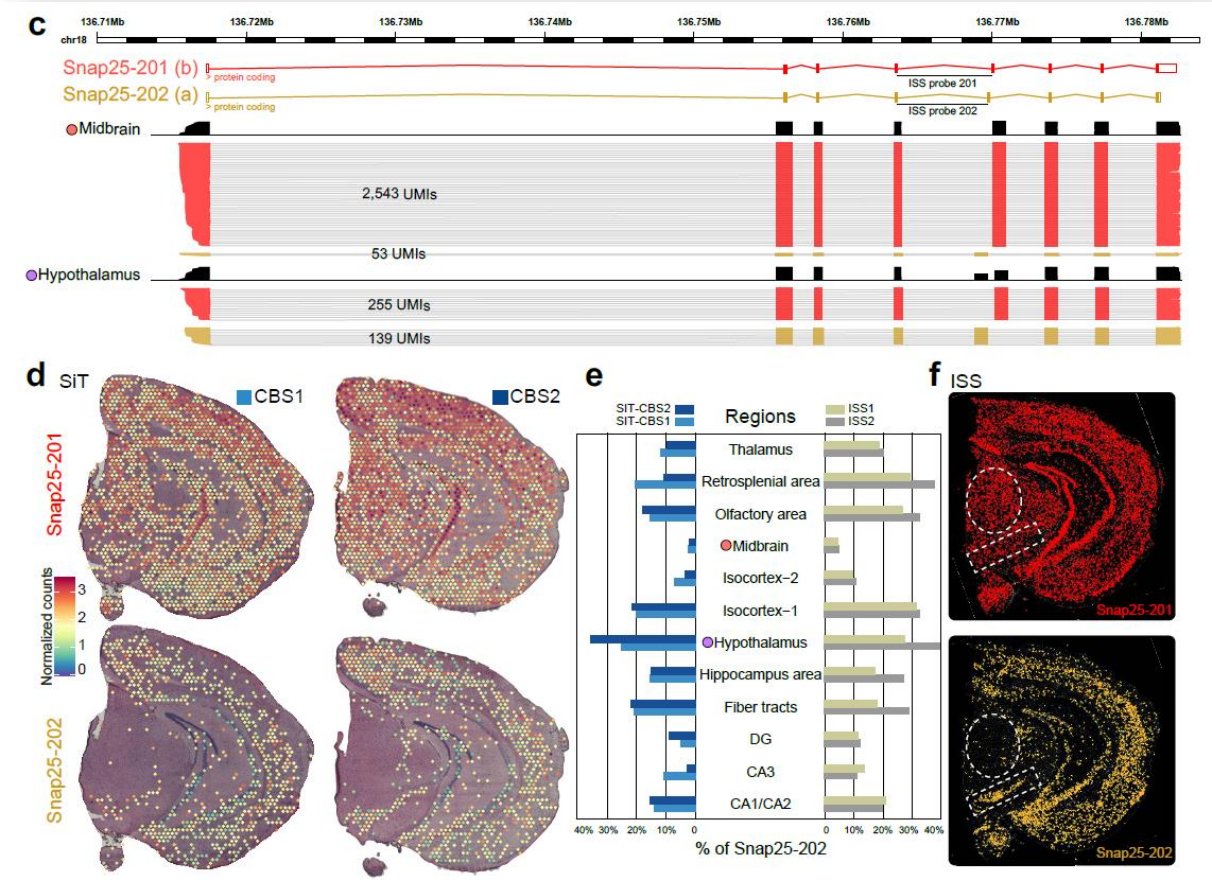
Clta..ENSMUST00000107851.9	1	.	.	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
Clta..ENSMUST00000107845.3	.	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.
Clta..ENSMUST00000170241.7	3	2	.	.	.	1	2	2	.	1	.	.	.	.	.	.	.	1	1	.	.
Clta..ENSMUST00000107849.9	1	.	3	5	.	3	.	5	3	2	.	1	3	3	4	4	.	3	4	2	3
Clta..ENSMUST00000107847.9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
Myl6..ENSMUST00000217733.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Myl6..ENSMUST00000218127.1	1	.	1	.	.	1	8	.	.	.	.	.	.	1	1	.	.	1	.	.	.
Myl6..ENSMUST00000217776.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Myl6..ENSMUST00000219236.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Myl6..ENSMUST00000164181.1	4	1	2	1	1	2	1	1	3	2	3	2	.	.	5	3	2	1	2	1	.
Myl6..ENSMUST00000219554.1	.	.	.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.	.	.
Myl6..ENSMUST00000218170.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Myl6..ENSMUST00000220307.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Myl6..ENSMUST00000217969.1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
Pkm..ENSMUST00000214571.1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	.	.	.	.
Pkm..ENSMUST00000163694.3	2	2	.	.	.	3	.	7	2	6	2	.	.	2	9	1	1	1	3	1	.
Pkm..ENSMUST00000034834.15	1	.	1	.	.	3	3	.	.	1	.	.	.	.	1	.	.	1	2	.	3
Meis2..ENSMUST00000028639.12	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
Meis2..ENSMUST00000189640.6	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
Meis2..ENSMUST00000110908.8	.	.	.	.	2	.	.	1	.	1	.	.	.	.	1	.	.	.	.	.	.
Meis2..ENSMUST00000110907.7	.	.	.	.	1	.	.	1	.	.	1	.	.	.	1	.	.	.	.	.	.
Meis2..ENSMUST00000110906.8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Meis2..ENSMUST00000074285.7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Meis2..ENSMUST00000102538.10	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.
...																					

Isoform-allelic-level matrix

Actually in development

Snap25 DTU across CBS regions

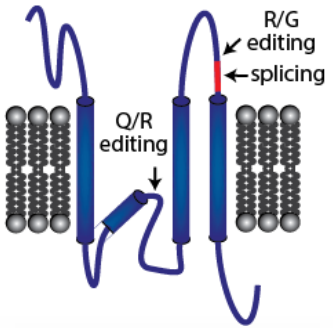
Presynaptic plasma membrane protein involved in the regulation of neurotransmitter release

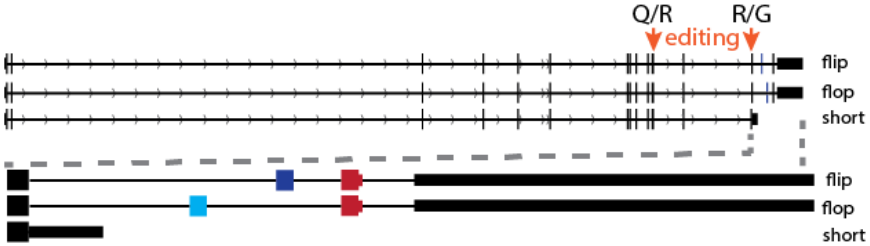


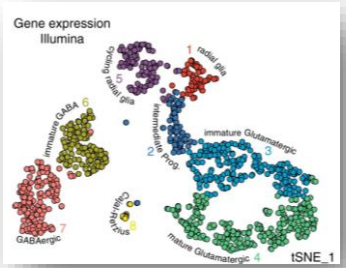
Single-cell long-read transcriptomics reveals sequence heterogeneity

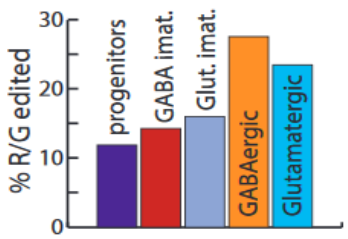
RNA editing of the AMPA receptor Gria2

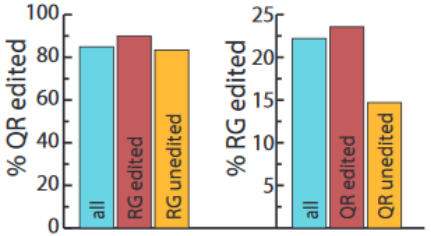
Q/R site regulates AMPA receptor Ca^{2+}-permeability
R/G site is involved in desensitization and recovery of the receptor









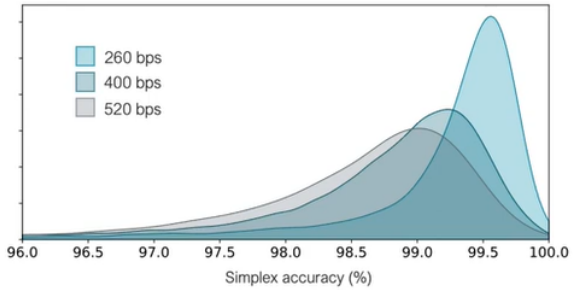


Nanopore Chemistry

Tuneable run settings

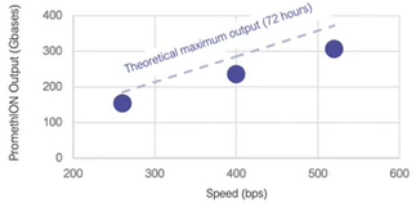
Same chemistry, different conditions

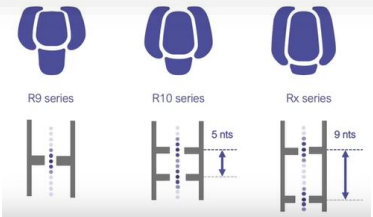
- Depending on the application requirement R10.4 / Kit14 can be run at different temperatures to tune the accuracy vs output
- Appropriate models will be chosen automatically or selected by user
- Example speeds and outputs
  - 260, 400, 520 bps
- Different accuracies
  - Simplex: 99.6%, 99.2%, 99.0%
  - Duplex: Q31, Q30, Q29



Choice of run condition

- Current focus is the 400 bps default condition
  - Demonstrated 236 Gbases from PromethION flow cell
  - Modal accuracy 99.2 % (Q21)
- Choice of run condition coming soon
  - "Accuracy" 260 bps, "Default" 400 bps, "Output" 520 bps
  - 99.6% (Q24) @ 260 bps, 307 Gbases @ 520 bps





In your hands from $1,000

Scalability and accessibility

LC 2022 LONDON CALLING

Flongle	MinION	MinION Mk 1c	GridION	PromethION P2 *solo* (in EA)	PromethION P2 (in dev)	PromethION 24 With A100 compute	PromethION 48 With A100 compute
$1,640	$1,000	$4,900	$49,955	$10,455	$59,995	$225,000	$310,000
• 12 flow cells	• 1 flow cell	• 6 flow cells	• 60 flow cells	• 8 flow cells	• 48 flow cells	• 192 flow cells	• 288 flow cells
• Flongle adapter	• 1 kit	• 1 kit	• 10 kits	• 2 kits	• 8 kits	• 42 kits	• 48 kits
Use with GridION or MinION		• 12M SW licence	• 12M SW licence	• 12M SW licence	• 12M SW licence	• 12M SW licence	• 12M SW licence

Same sample prep – same sequencing chemistry – same data analysis

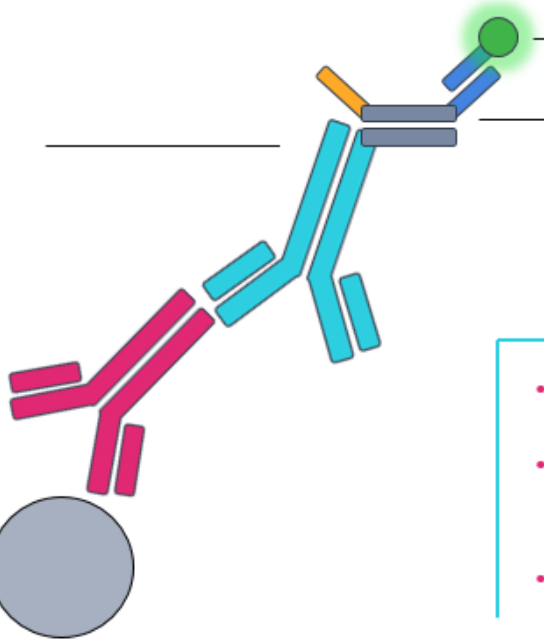
Technical Concept

MERSCOPE Protein Detection

Auxiliary Detection Probe - Binds to encoding probe and conjugated fluor is detected

Anti-Species Oligo-Conjugated Secondary Antibody - Binds to primary antibody and is oligo conjugated (pre-conjugated)

Auxiliary Encoding Probe - Binds to oligo on Secondary Antibody



Primary Antibody Supplied by Customer - Binds to protein target

Protein Target

- Signal detected - equated with protein presence
- Auxiliary rounds allow single fluor detection without barcode scheme - restricts sequential gene use
- Allows simultaneous protein detection with RNA transcript detection