Funded by
the European Union

Bioinformatics Summer School

Long-reads Transcriptomics

Eamon McAndrew

Ipmc - Institute of Molecular
and Cellular Pharmacology

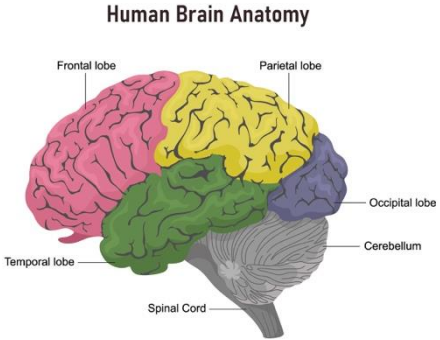
LongTREC - The Long-reads TRanscriptome European Consortium
Marie Skłodowska-Curie grant agreement No 101072892

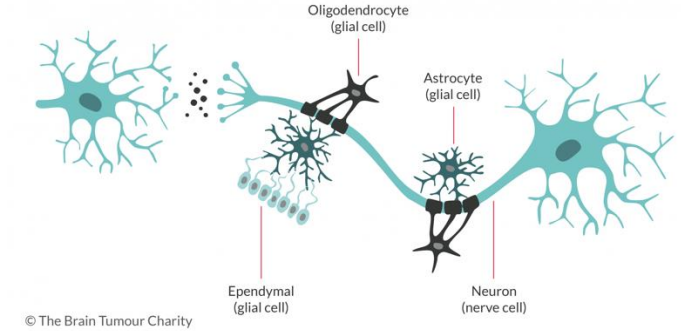
Section 4

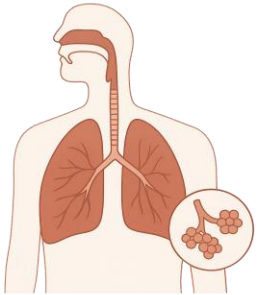
Long Read Single Cell Transcriptomics

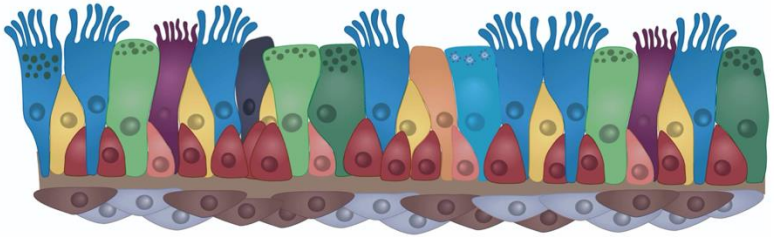
So after this week we know all about Long Read Bulk Transcriptomics:

- **But what about complex tissue?**



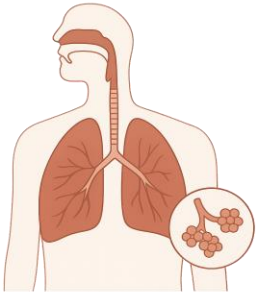


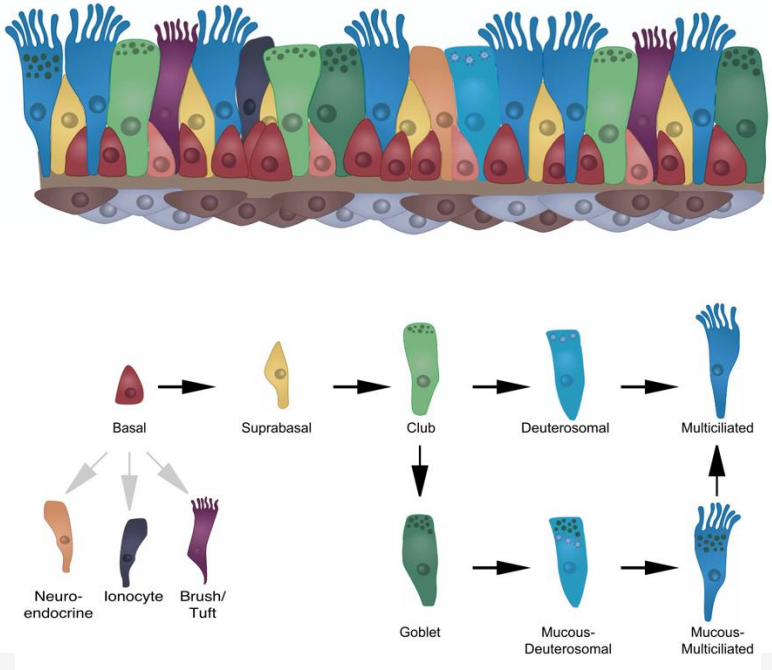




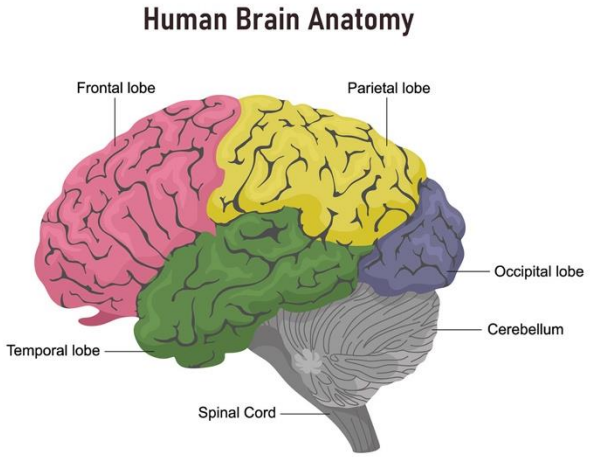
Why Single Cell?

In complex tissue the signal we generate with bulk assays is a linear combination derived from many cell types.



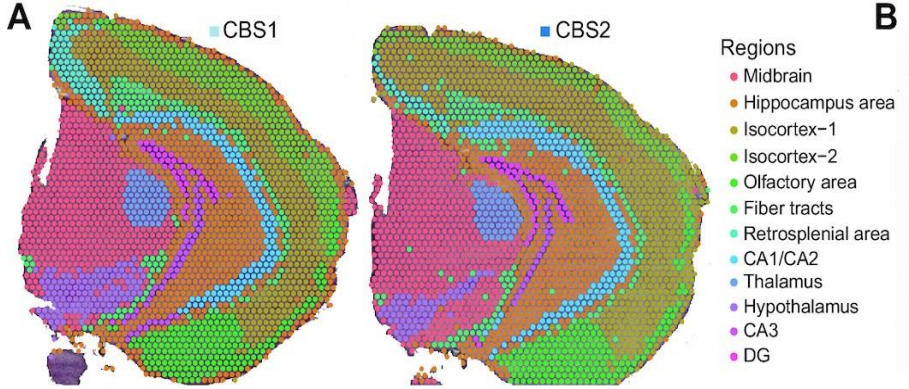


In complex tissue the signal we generate with bulk assays is a linear combination derived from many cell types (subtypes, states, diseased affected/healthy)

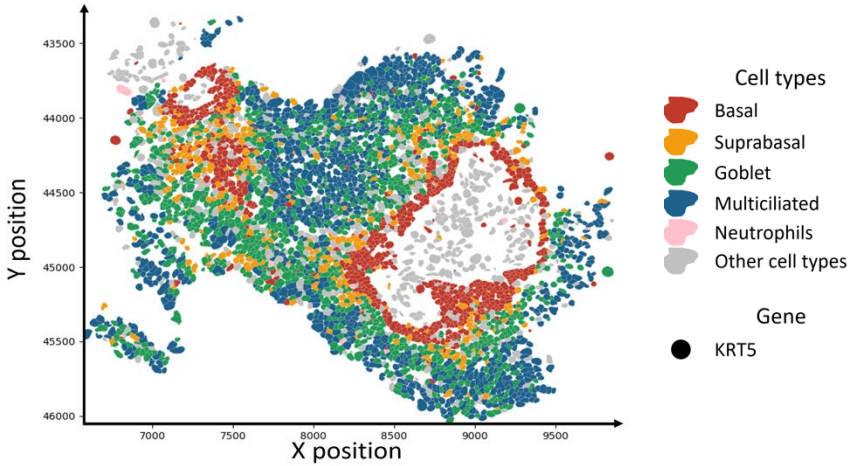


	Neuron	Astrocyte	Monocyte Macrophage	Brain microvascular endothelial cells	Oligodendrocyte
Healthy					
Parkinson's disease					
	• Cell death • Accumulation of α-synuclein • Susceptibility to oxidative stress • Aberrant neurite morphology • Mitochondrial dysfunction • Autophagosome/lysosome defects • Reduced inflammatory response • Epigenetic alterations: DNA hypermethylation	• Accumulation of α-synuclein • Autophagosome/lysosome defects • Extracellular matrix perturbation	• Altered inflammatory response: increased release of pro-inflammatory cytokines and reduced migration capacity • Accumulation of α-synuclein • Compromised phagocytosis	• Reduced P-glycoprotein function	• Endogenous expression of α-synuclein

Cells don’t exist or operate in Isolation.

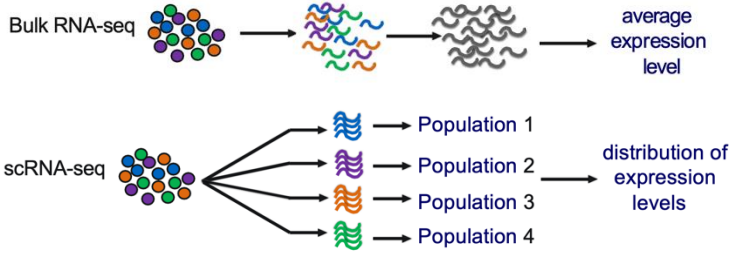


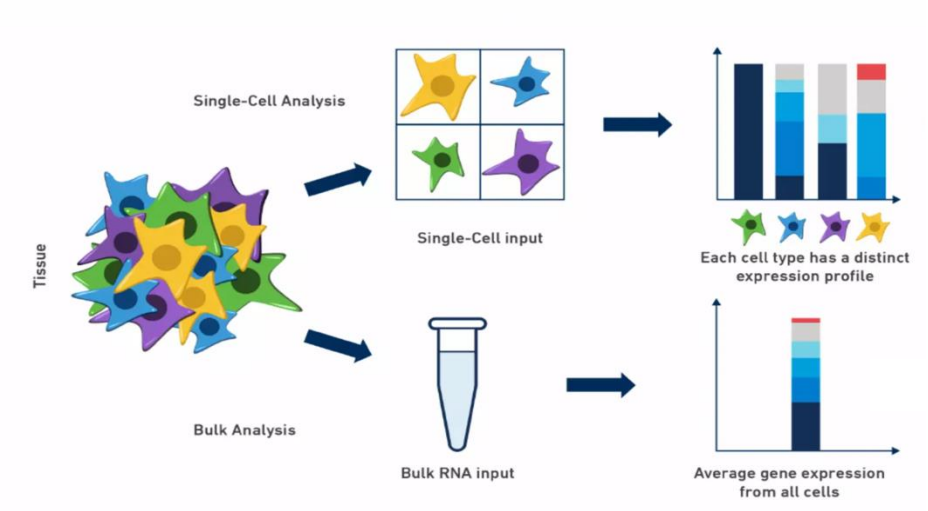
Mouse Coronal Brain Section



Bronchial COPD biopsy

Preserving cellular heterogeneity





Long Read single Cell Transcriptomics

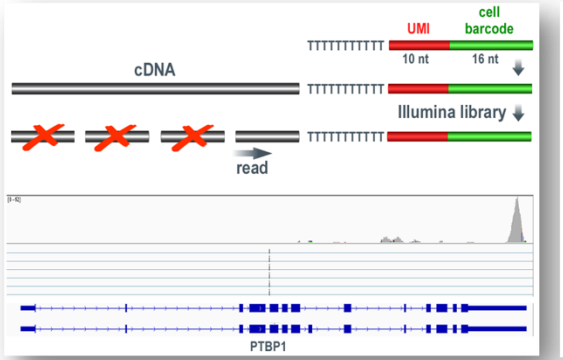
Standard short-read sequencing

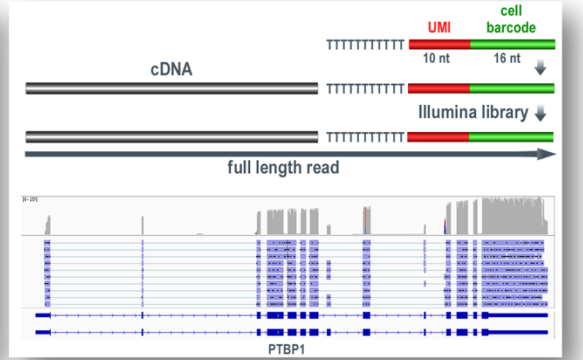
10x GENOMICS

Long-read full-length sequencing

illumina

Cell-X-Gene matrix



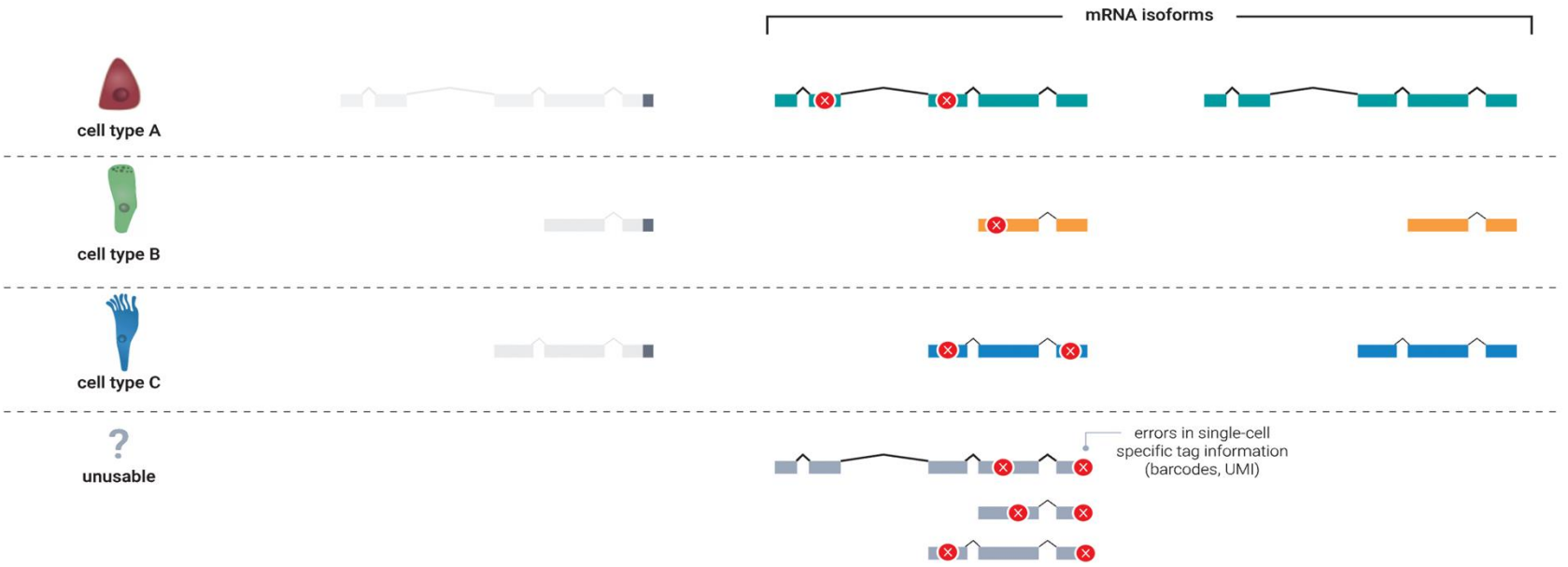


NANOPORE

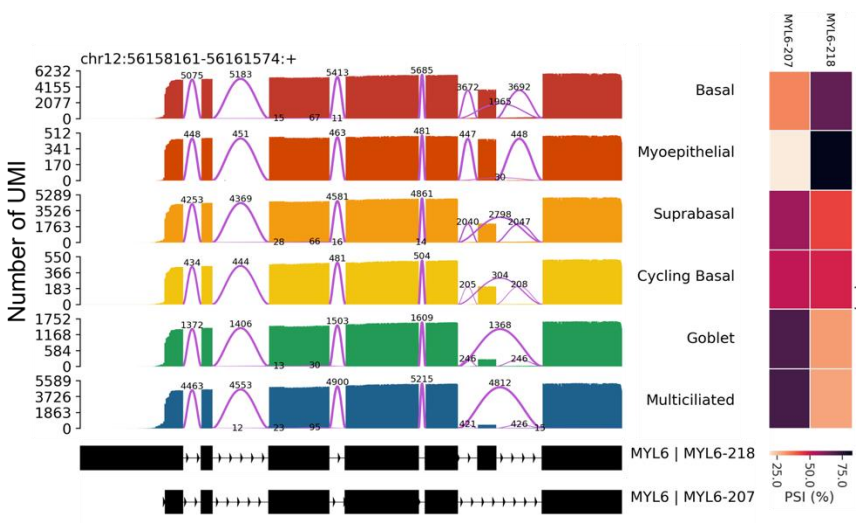
PacBio

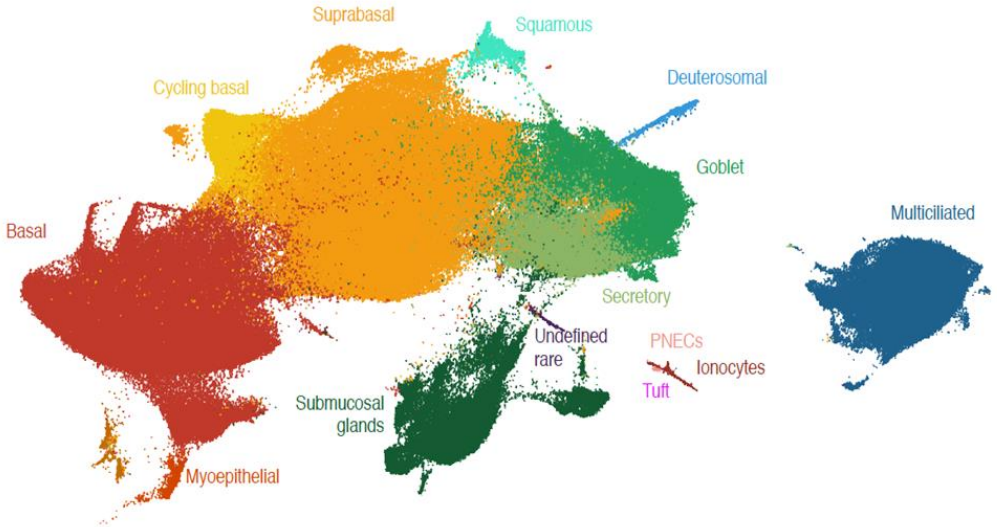
Cell-X-Isoform matrix

Individual reads assigned to individual cells



Why Single Cell?

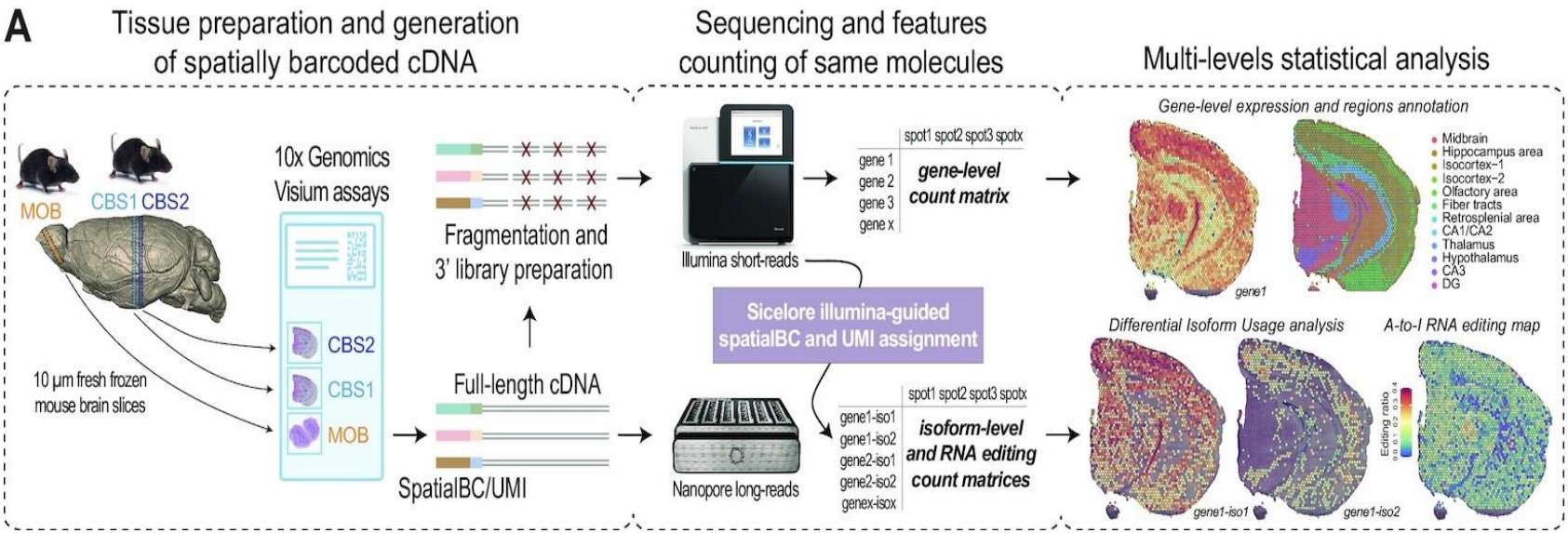




The challenges of Single Cell Long Reads

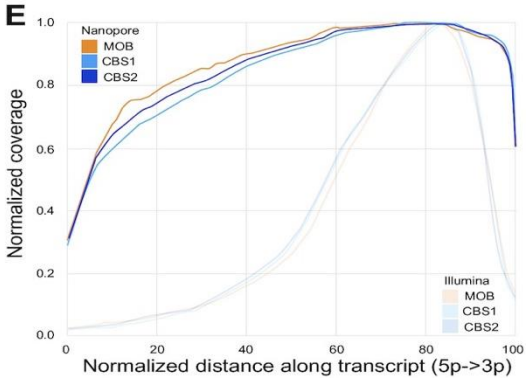
- High per-cell sequencing cost and lower throughput vs. short-read platforms.
- Higher raw read error rates complicate accurate isoform identification.
- Sparse coverage per cell limits detection of low-abundance transcripts -> SR Gene Level Single Cell is already sparse.
- Barcode/UMI assignment inefficiencies and cDNA length biases.
- Demanding compute + storage requirements for long-read datasets.
- Limited mature analysis tools.

What about spatial context?

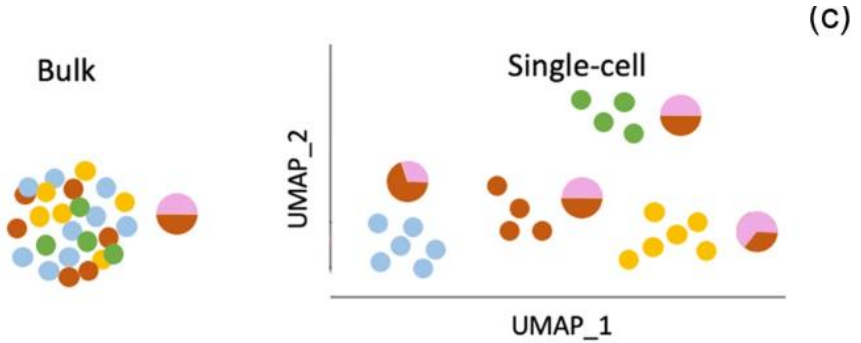


The promise of single cell long reads

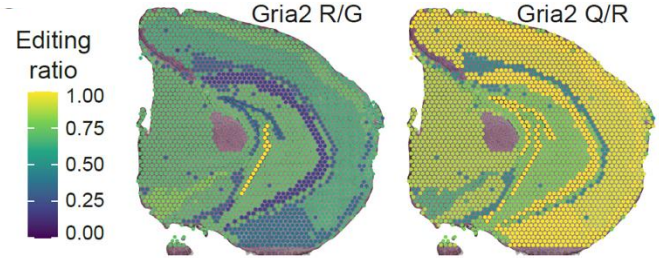
Full Gene Body Coverage



Allele Specific Expression

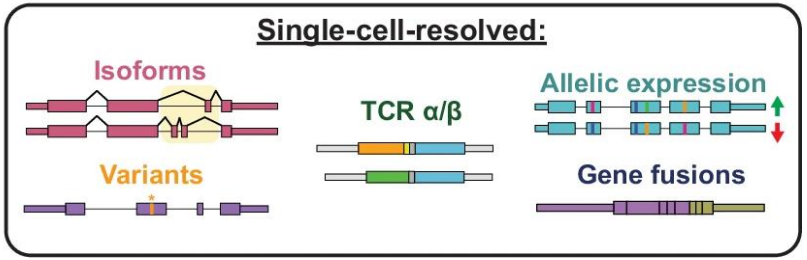


A -> I RNA-Editing



Individual A-to-I editing site editing ratio per region

And more – All at once* !



Question Time

- What does single cell offer over bulk transcriptomic approaches ?
- What are the challenges of long read single cell approaches?

Section 4

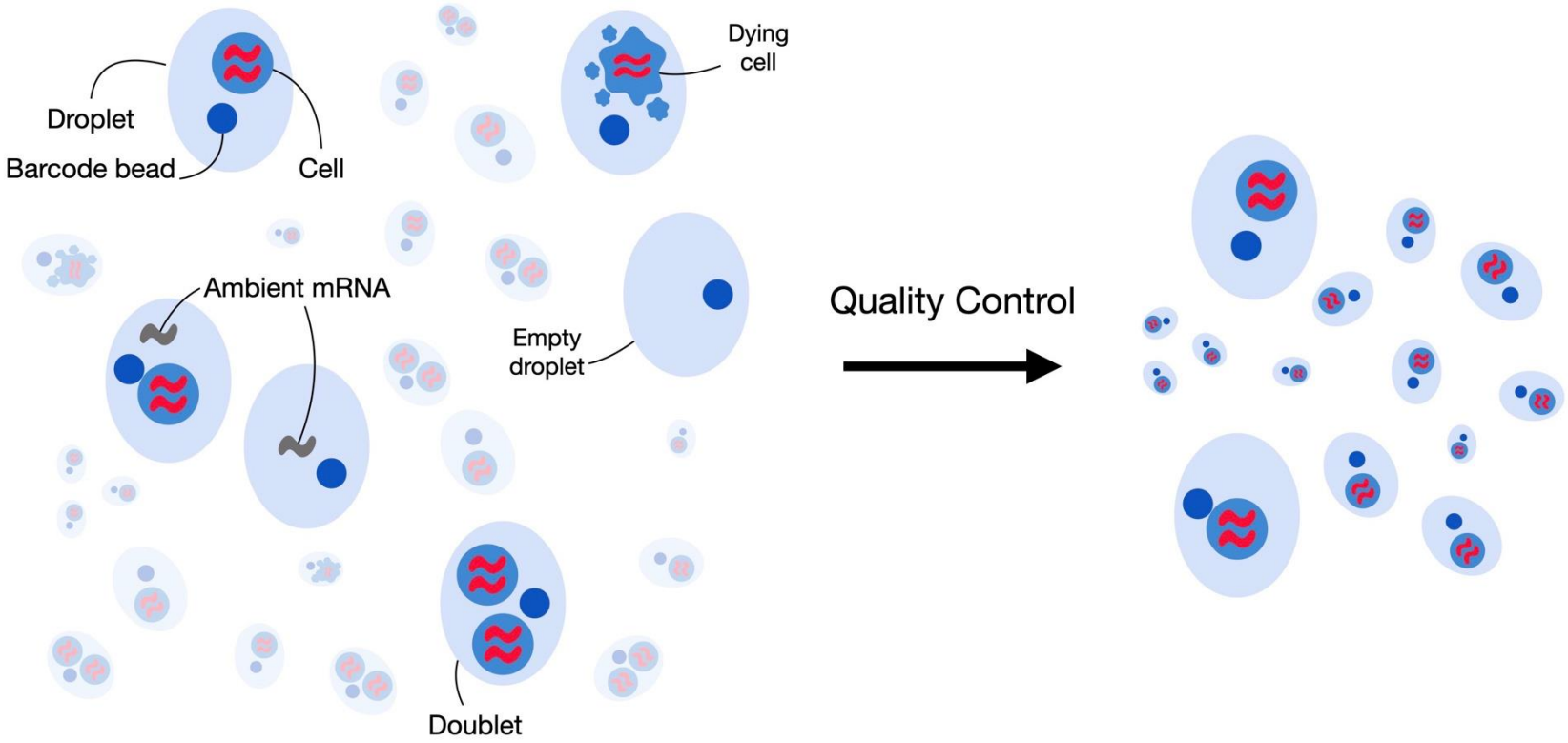
Tertiary Analysis

What do we need to begin tertiary analysis of Isoforms at Single Cell Resolution?

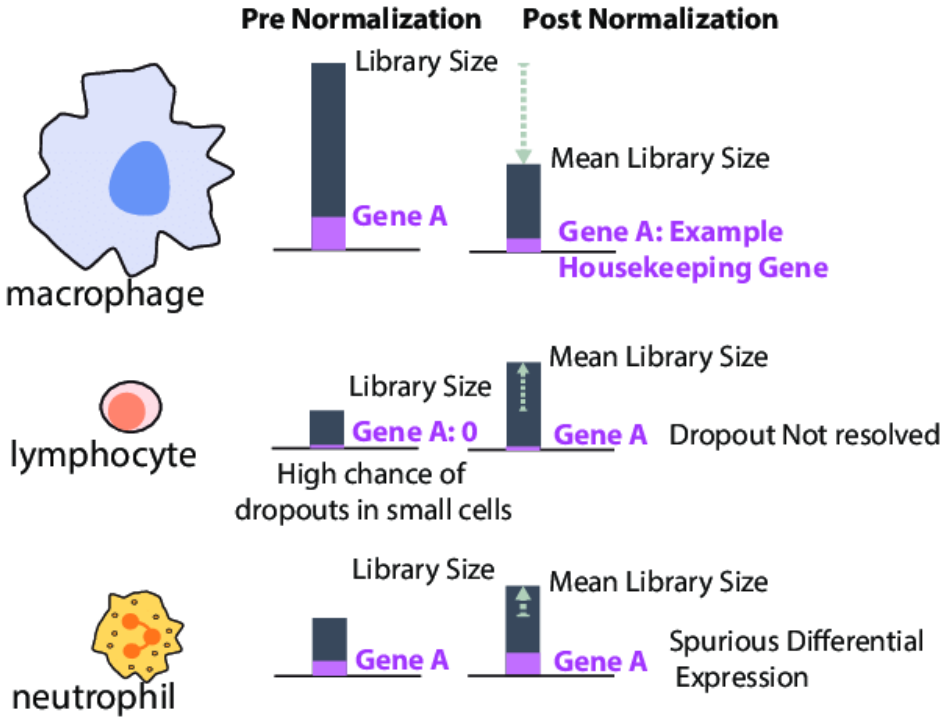
Cell X Isoform Matrix

Annotation

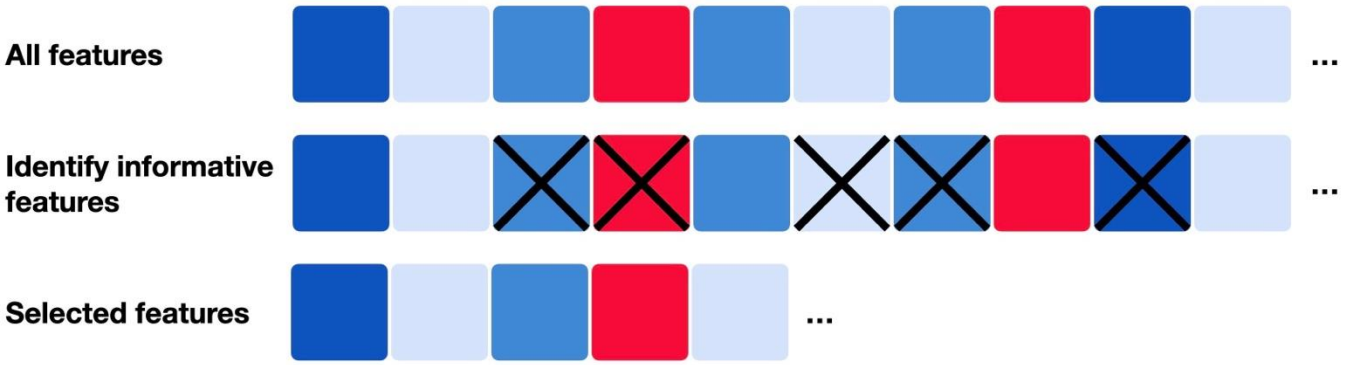
Removing Low Quality Cells If necessary



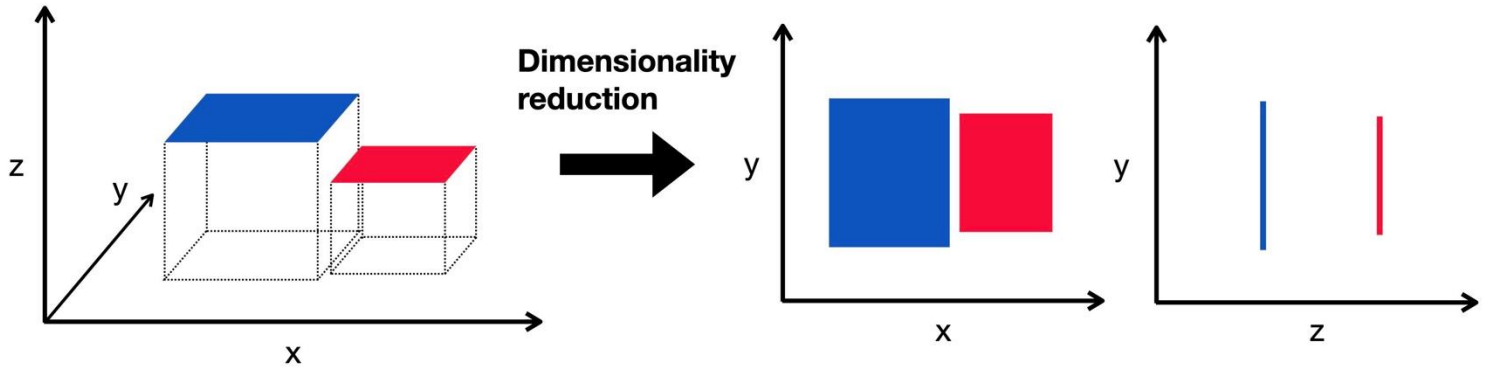
Normalization and why it matters

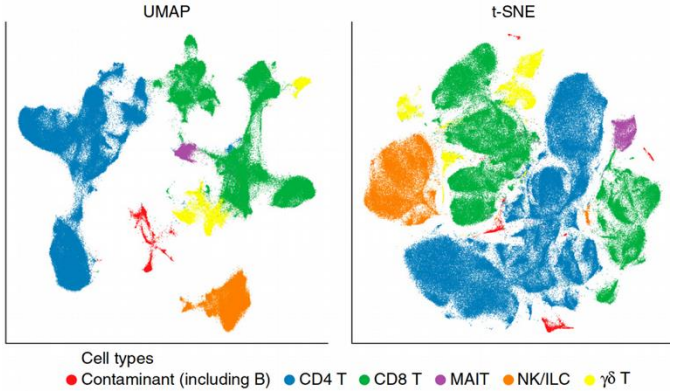


Selecting the most Informative Features

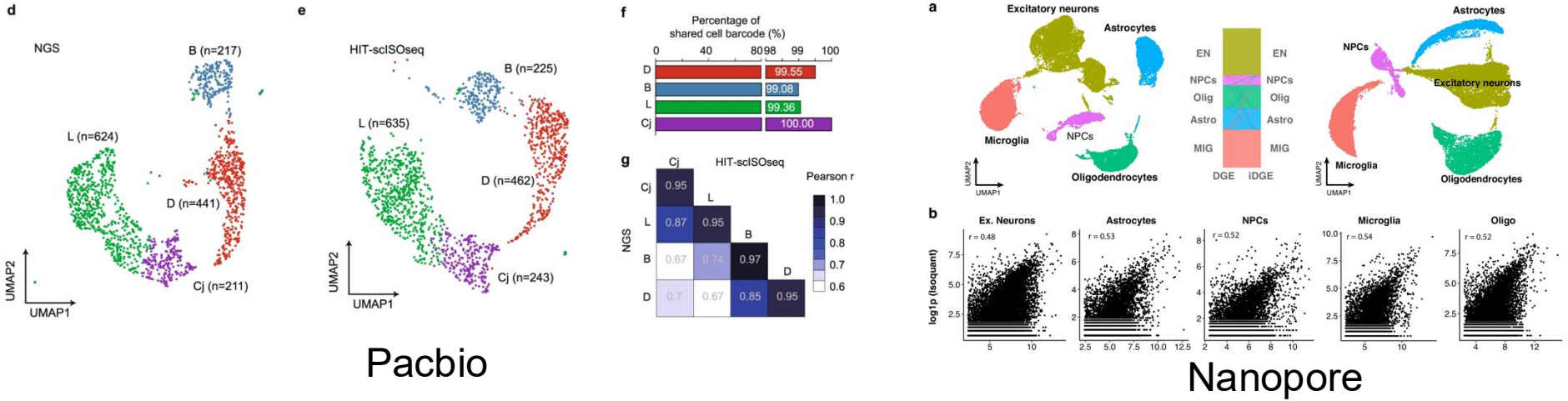


Reducing Dimensionality





Do I define my clusters on Gene or Isoform Features?



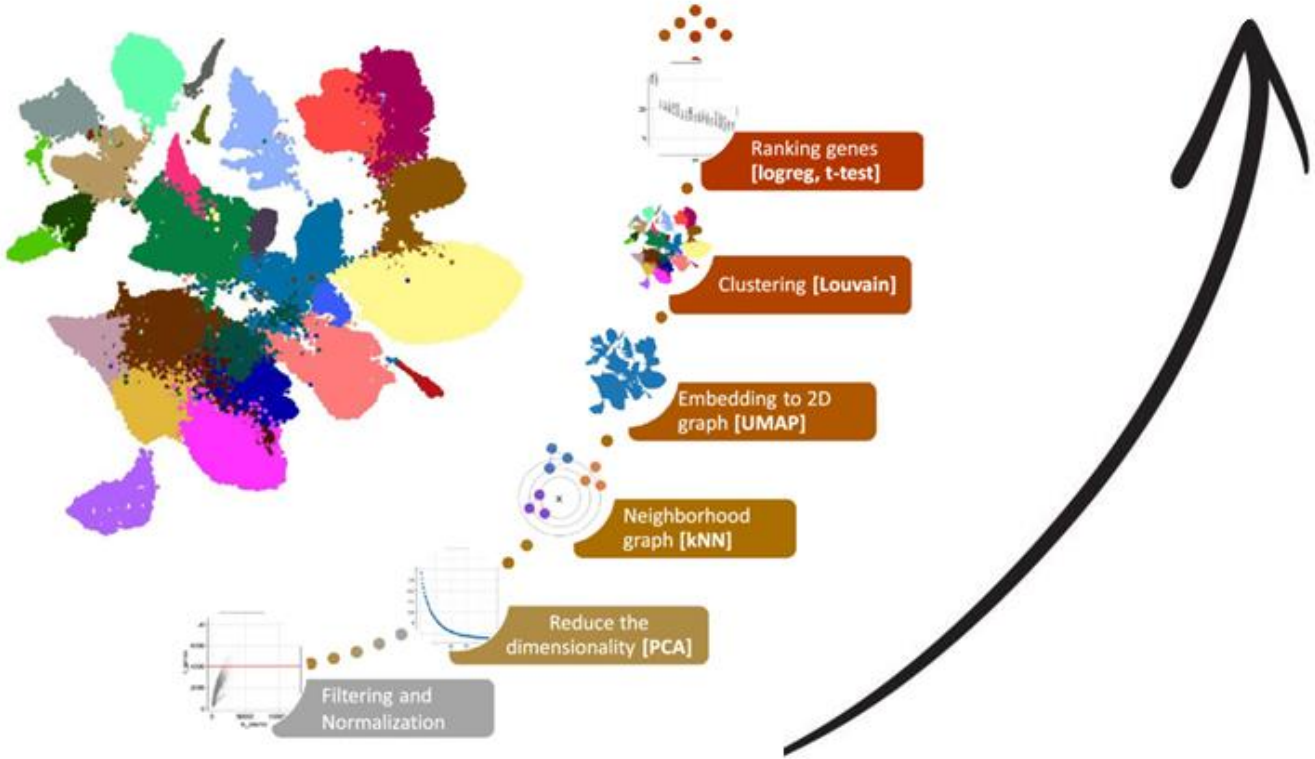
Pacbio

Nanopore

Same cells, two feature spaces.

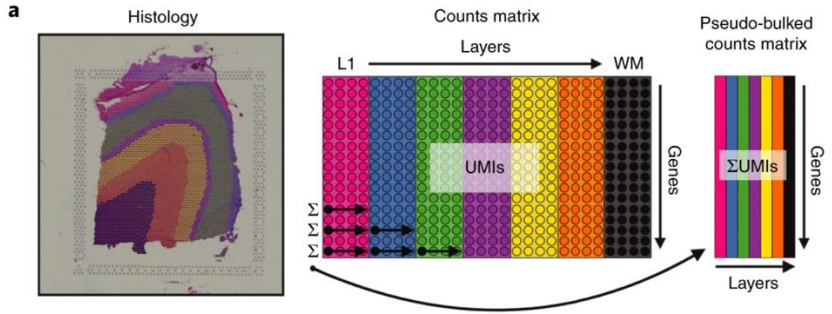
(Shi et al., 2023)

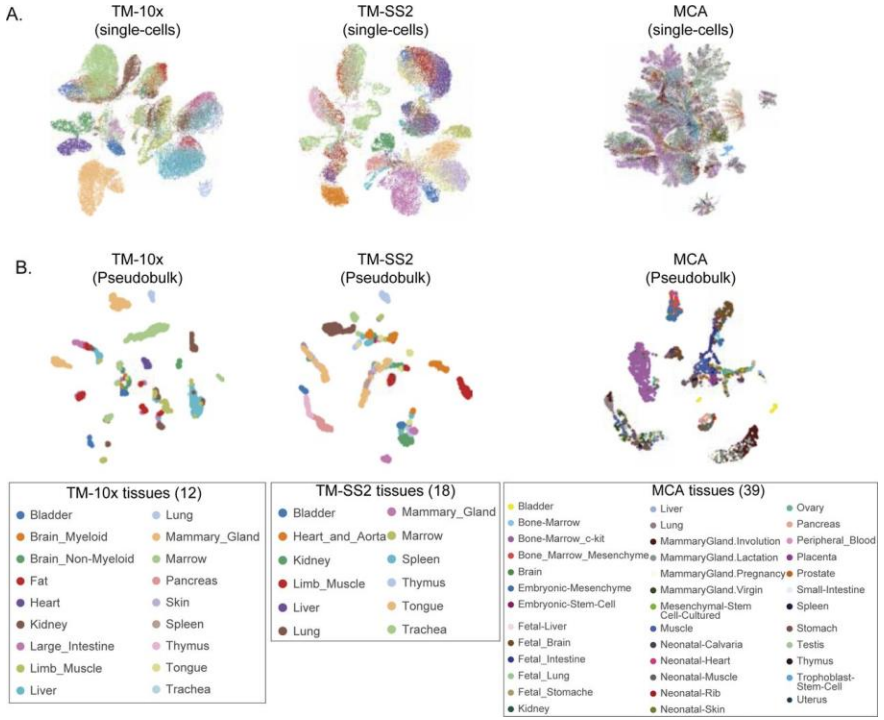
Putting it all together to Annotate Cell Types



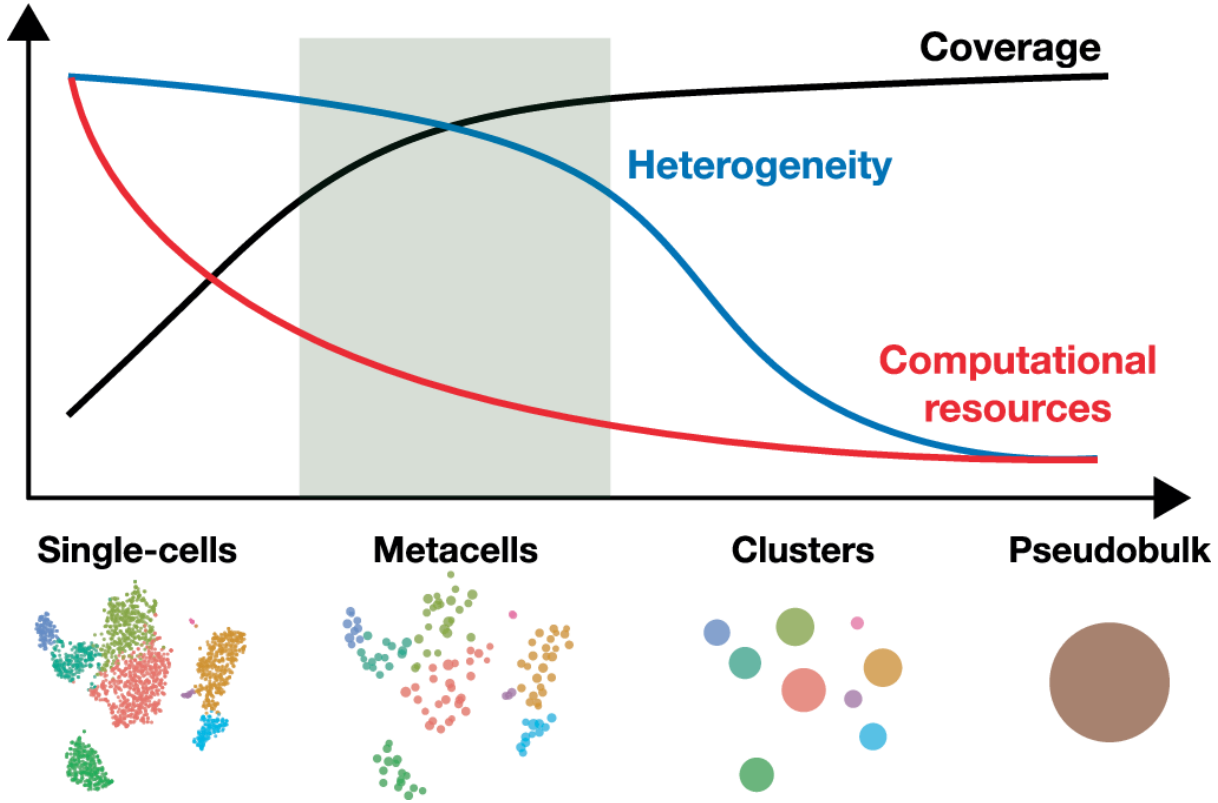
What is Pseudobulk Analysis?

- Aggregates single-cell or spatial counts into group-level profiles.
- Treats groups as “bulk samples” → improves statistical power.
- Allows plug and play with existing bulk methods.
- Reduces false positives compared to cell-level models (at least a gene-level*)

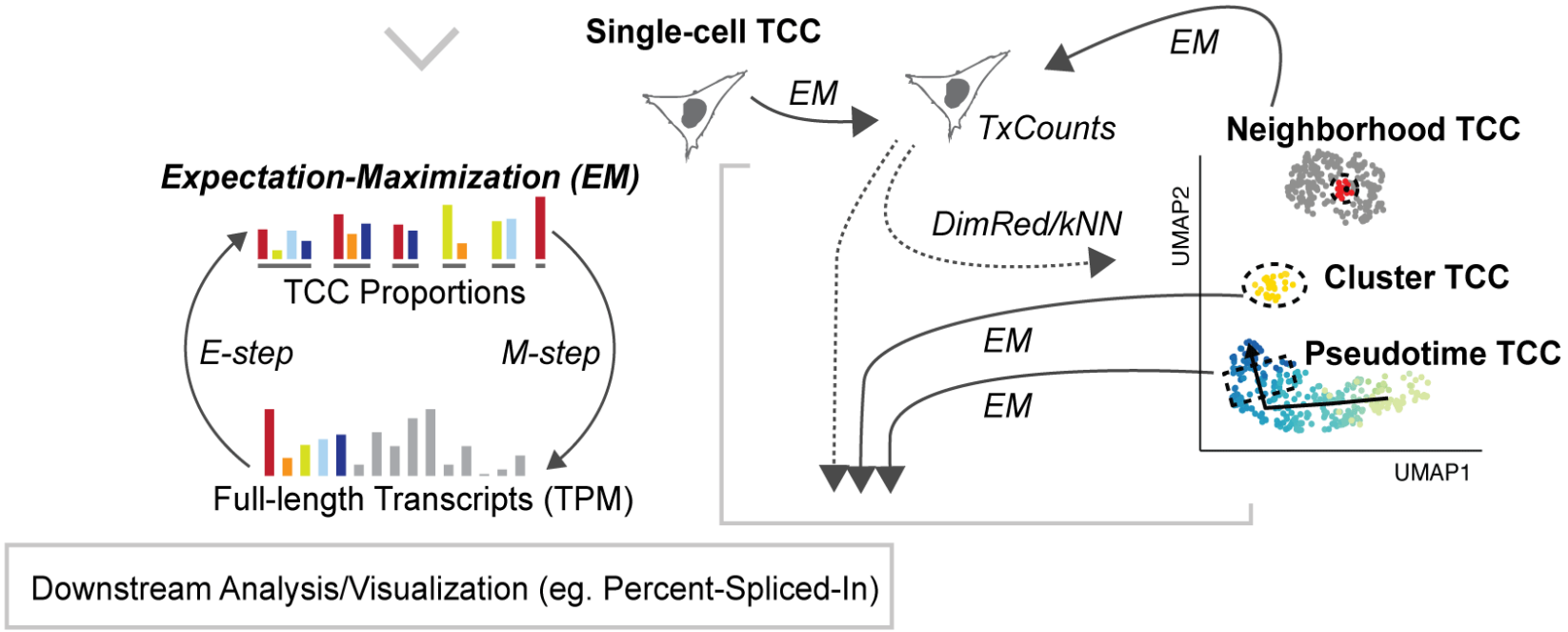




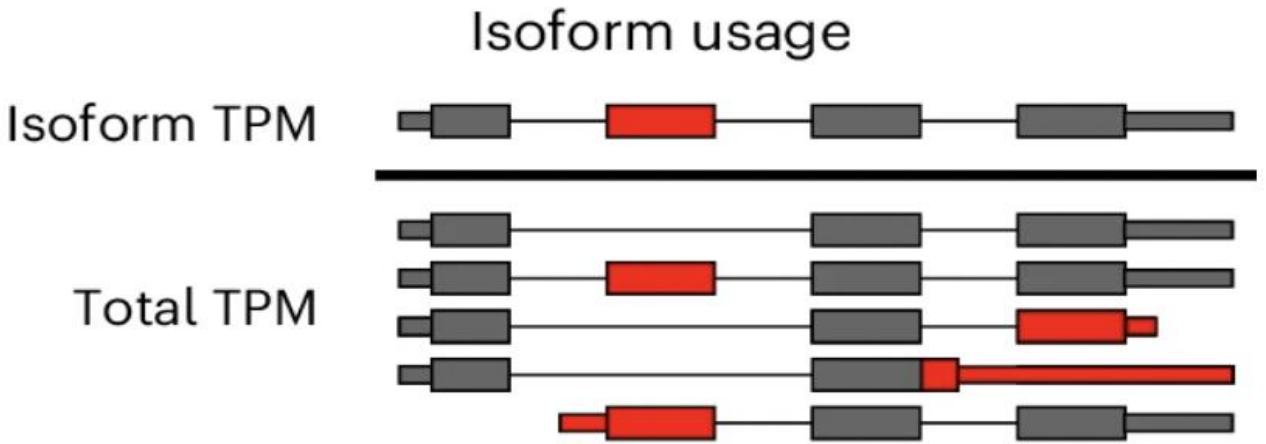
Coverage/Heterogeneity Trade-off



Finding the optimal Resolution

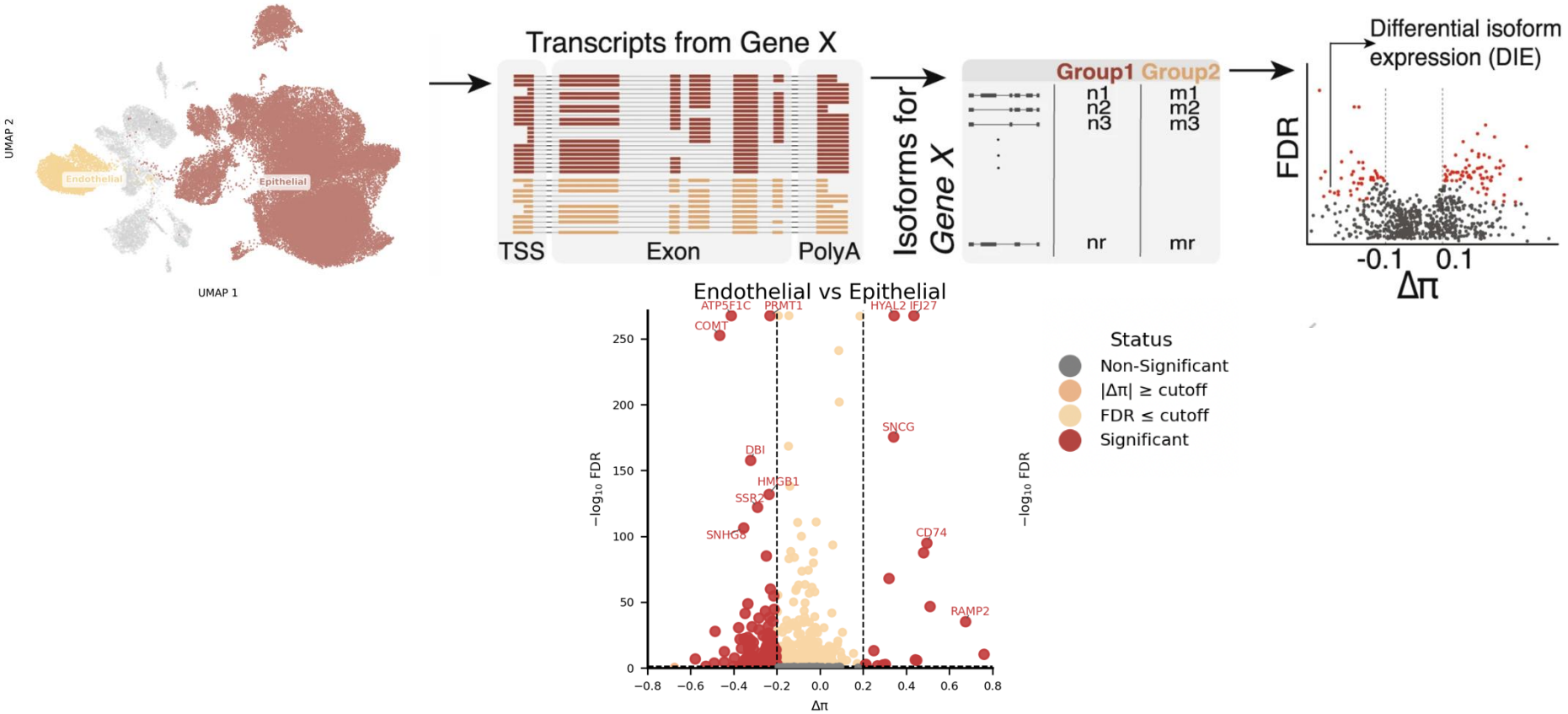


Percent spliced in as a metric for Isoform changes

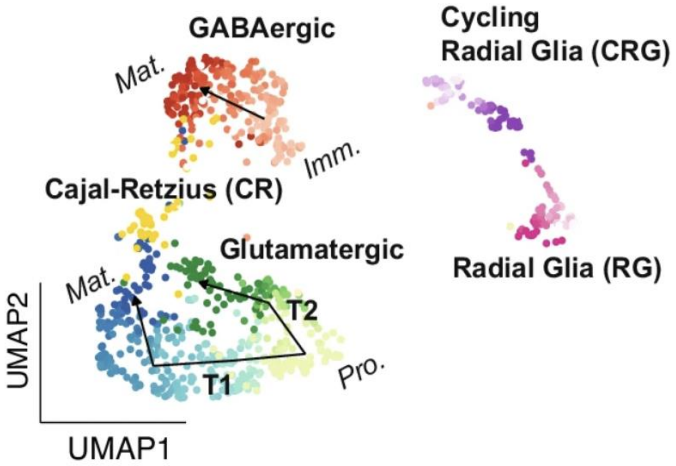


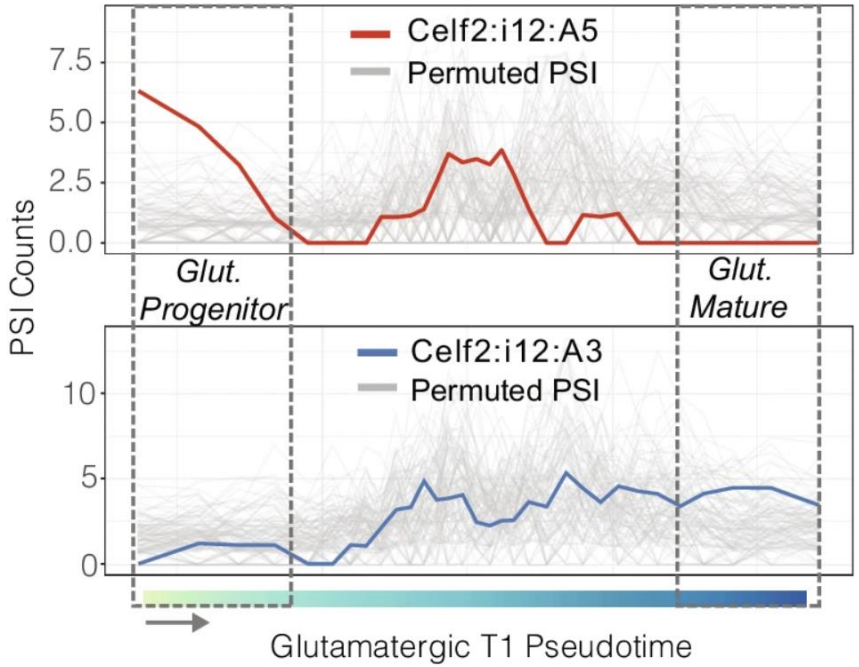
$$Isoform\ Usage/\ Percent\ Spliced\ in = \frac{Individual\ \ Count}{Total\ Gene\ Count}$$

Detecting Differential Isoform Usage across Cell Types



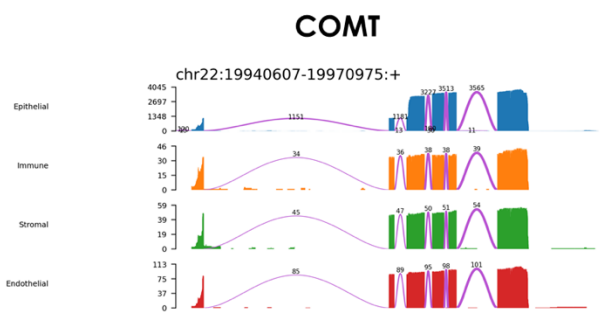
Isoform switching Across a differentiation trajectory

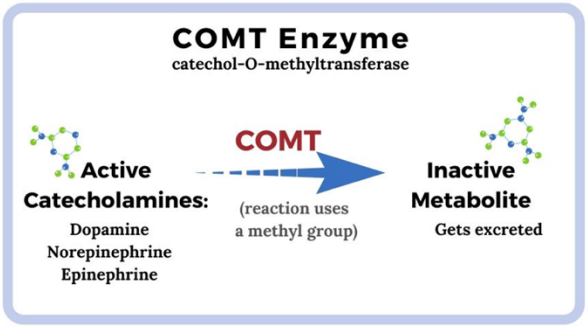


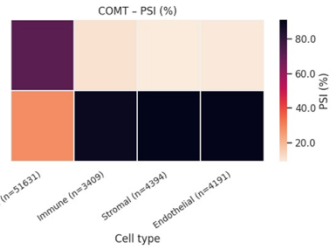


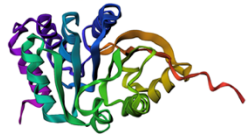
Cells are then grouped into sliding windows (here 30 cells per window, 15-cell step) so that PSI and gene counts are smoothed along pseudo time.

Why cell type context matters?



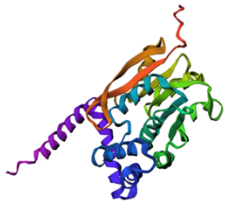






s-COMT (Soluble)

~216 aa (~24 kDa)



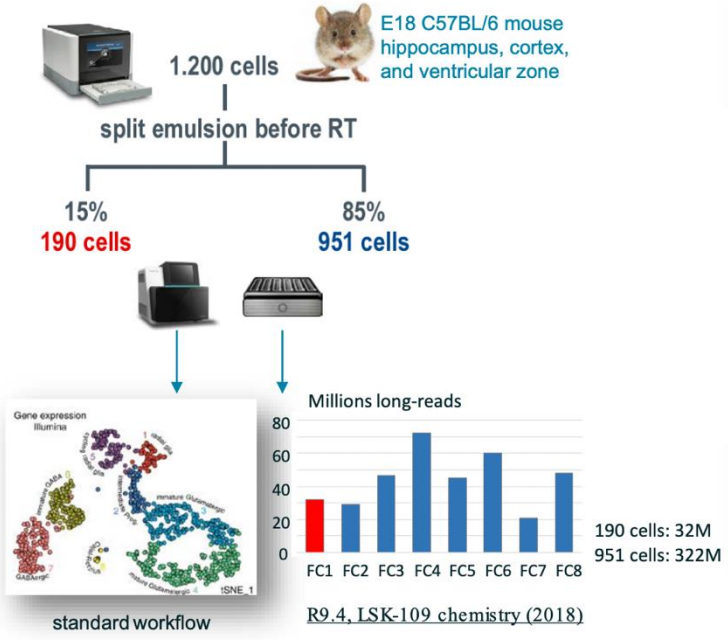
m-COMT (Membrane-bound)

~271 aa (~30 kDa)

Question Time

- Why is normalisation a crucial step of single cell analysis ?
- Should I do dimensionality reduction and clustering on gene or Isoform features?
- When aggregating cells (psuedobulk, metacells etc.) What do we trade-off as we increase the number of cells aggregated into a single unit?

Practical session : Exploring Differential Isoform Usage in the Mouse Brain



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

44k Accesses | **182** Citations | **64** Altmetric | Metrics

Practical session : Exploring Differential Isoform Usage in the Mouse Brain

1) Connect to VM
2) download new notebook from git
3) conda activate Single_cell
4) cd longTREC/day4
5) jupyter lab/ Jupyter Notebook
6) Open Notebook

Thank You!

For more information about the LongTREC Summer School:

https://longtrec.eu