Introduction to Workshop 5: scRNAseq

Genomics - BIO00087H

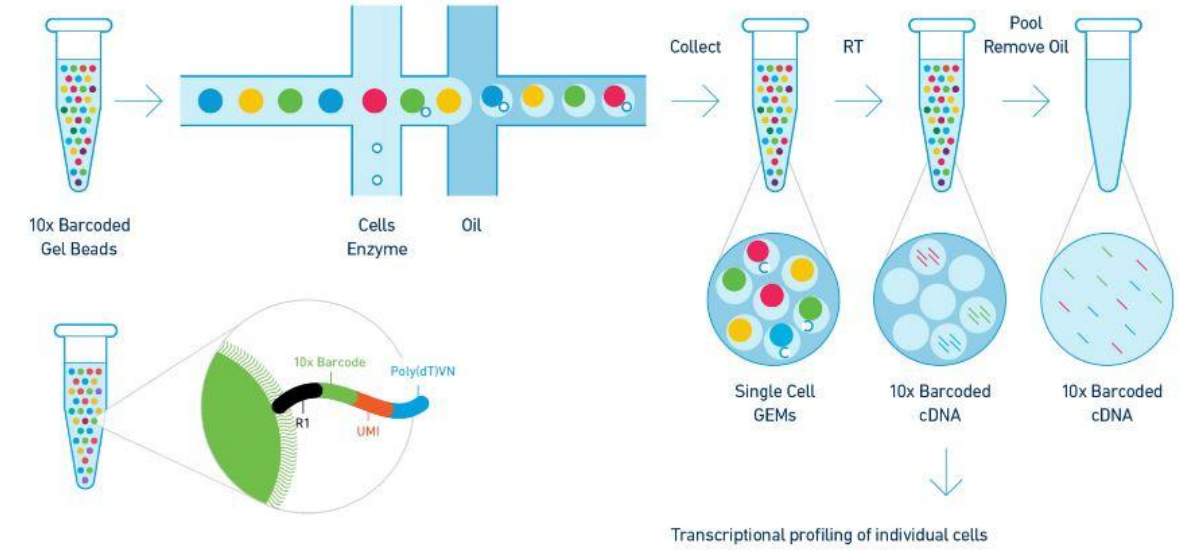
https://asmasonomics.github.io/courses/Genomics3_Workshop5_scRNAseq_Nov2025

scRNAseq – reminders from Lectures 5 & 9

Rather than taking an average of a cell population, we profile individual cells

- distinguish cells in a mixed population
- distinguish differential cell states in a mixed population

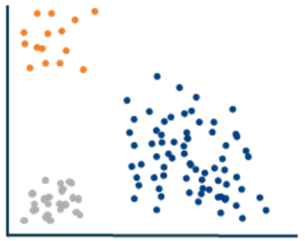
10x Genomics – oil droplet library preparation



1) Generate count matrix (more sparse)

	$Sample_1$	$Sample_2$	...	$Sample_n$
$Gene_1$	4200	3217	...	4231
$Gene_2$	0	2	...	1
...	...	...	...	...
$Gene_n$	178	142	...	205

2) Look for correlated features to reduce the data dimensions (PCA)



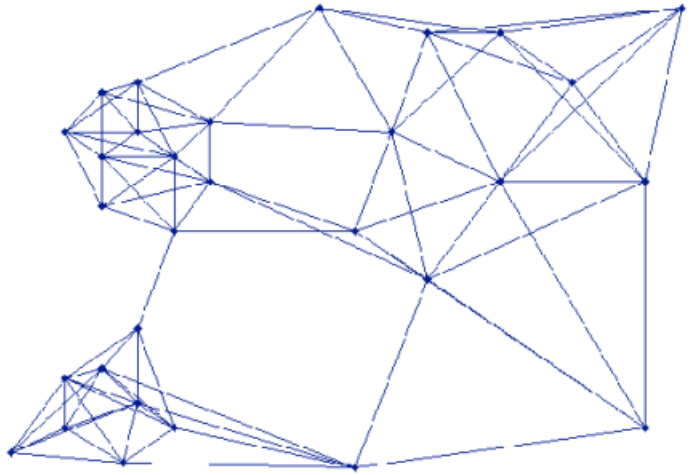
3) Identify 'clusters' and plot dimension-reduced data

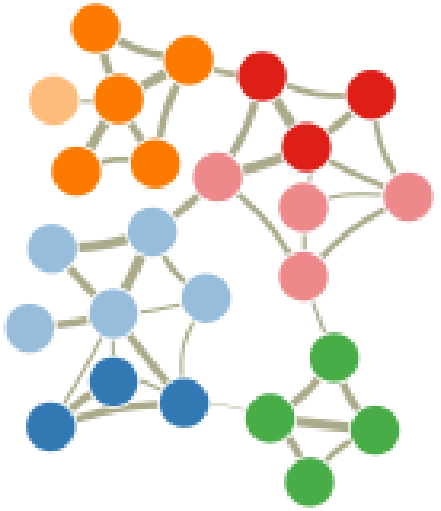
scRNAseq – reminders from Lectures 5 & 9

After QC, we look for shared features (i.e. expression patterns) in individual cells and use that to form cell ‘communities’ or ‘clusters’

The space between clusters is (largely) arbitrary – we space them out so they look nice

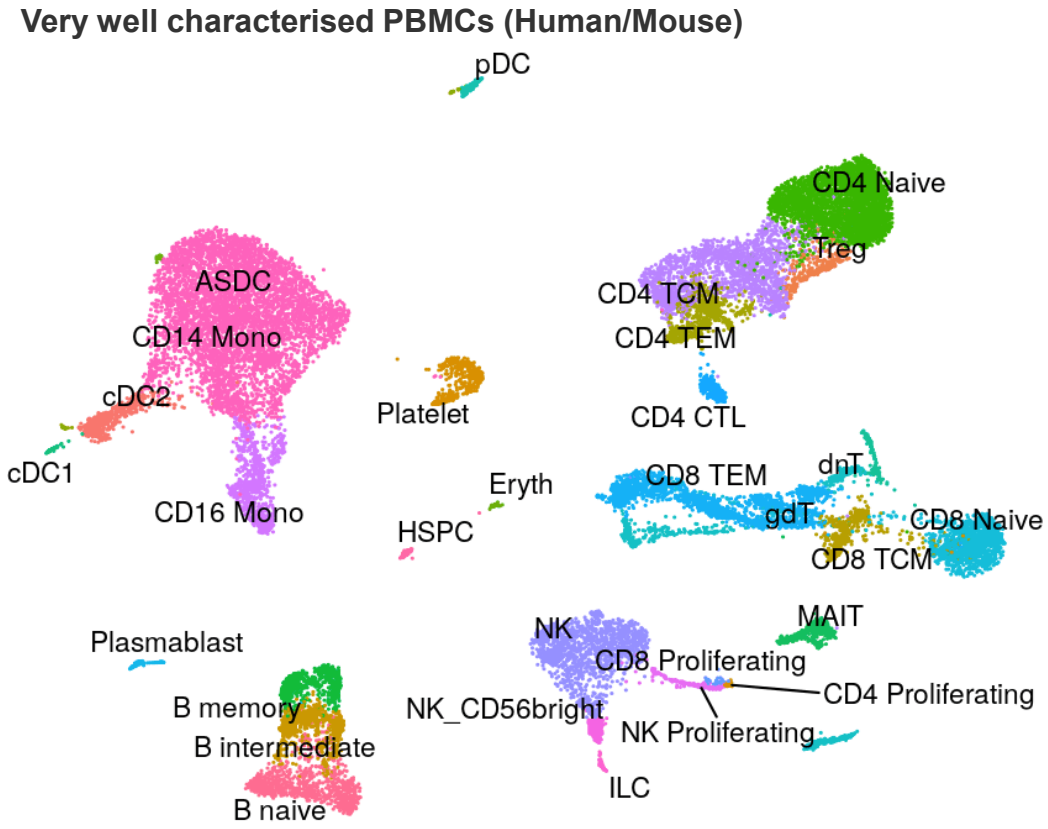
Detection algorithms require you to ~~know~~ guess how many clusters you have, and then you need to annotate them.





scRNAseq – reminders from Lectures 5 & 9

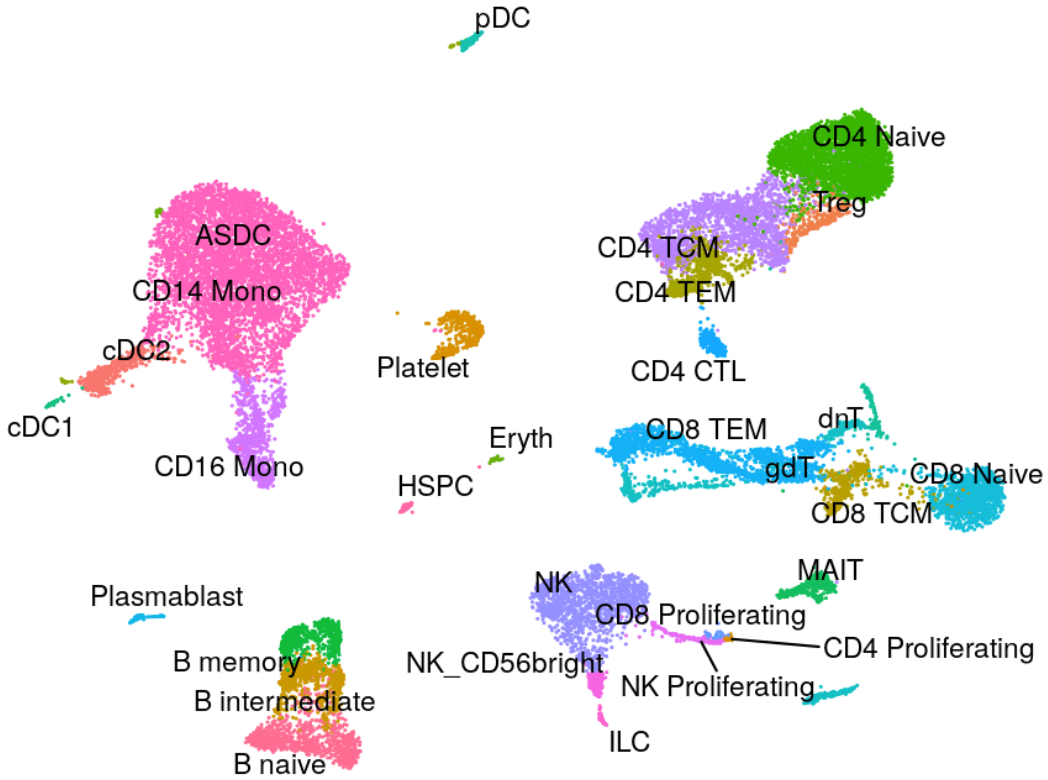
Annotation requires a lot of domain-specific knowledge, and very careful curation



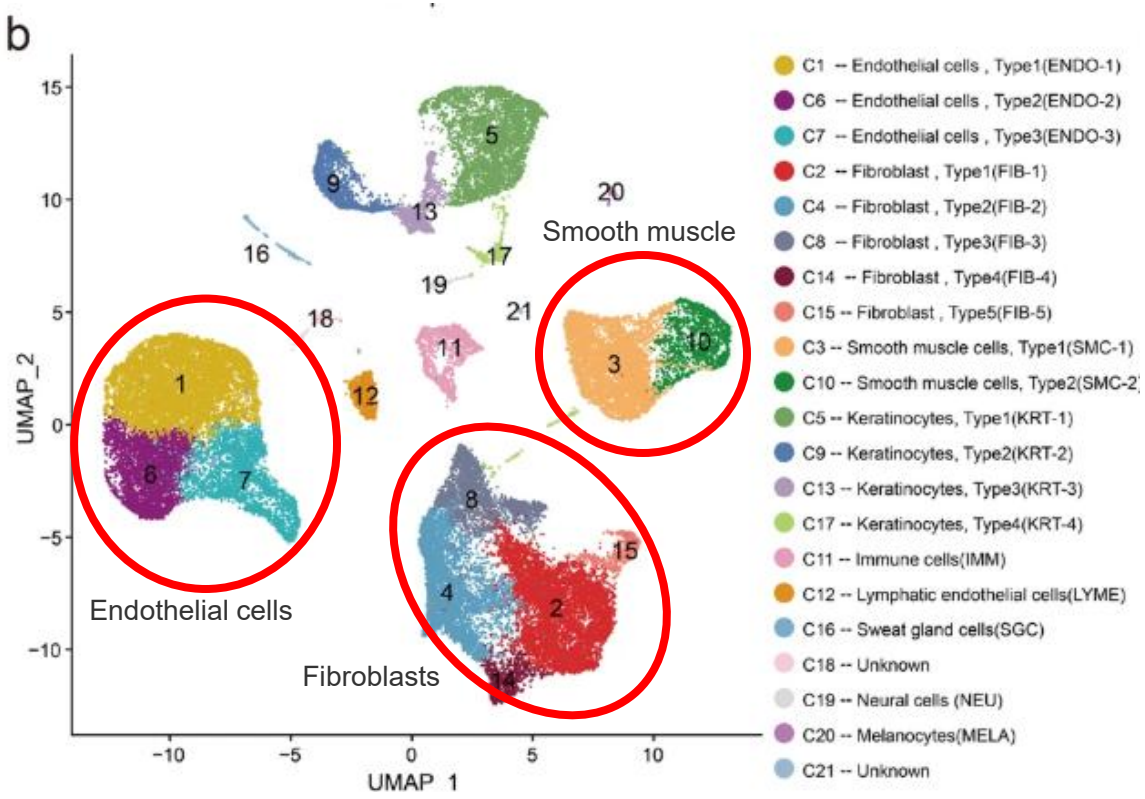
scRNAseq – reminders from Lectures 5 & 9

Annotation requires a lot of domain-specific knowledge, and very careful curation

Very well characterised PBMCs (Human/Mouse)



Defensible/reproducible sub-clusters??



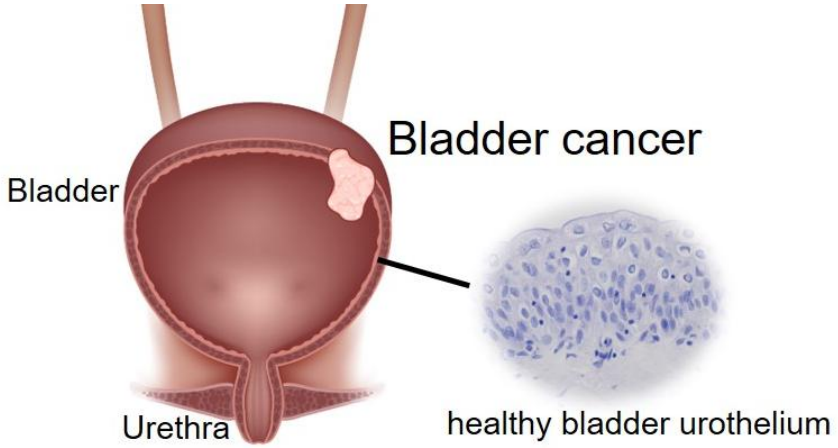
Think about narrative and over-interpretation!

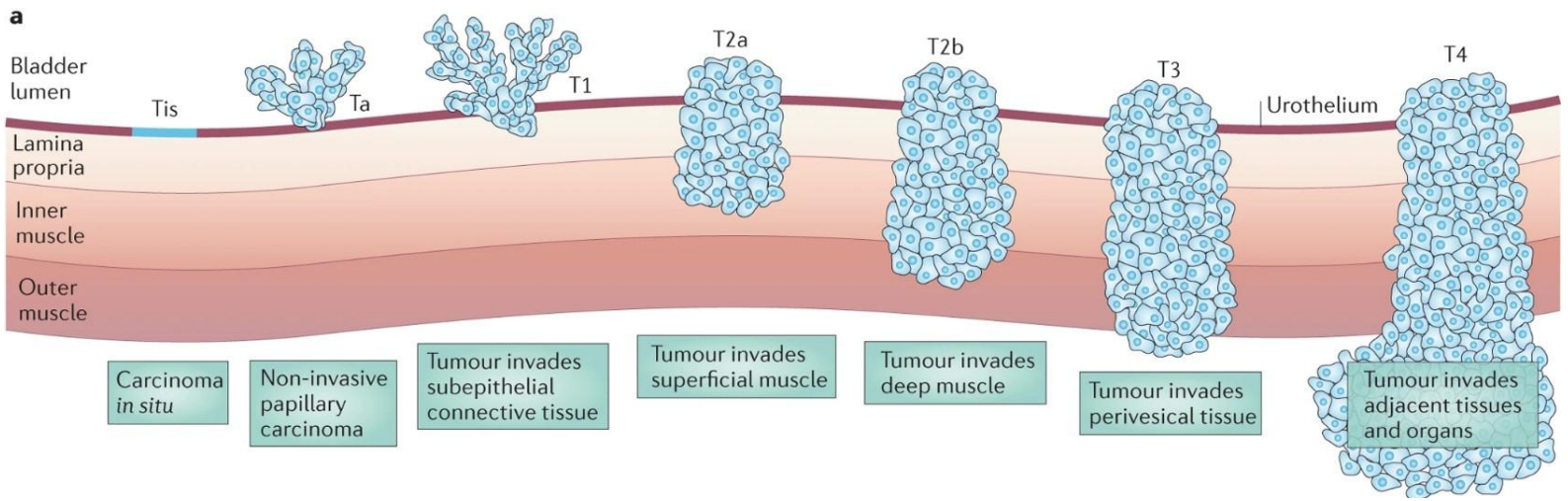
Workshop 5 – the plan

1) Prepare your Linux directory structure, but then launch RStudio
2) Load, quality check and filter your dataset
3) Integrate the two datasets together, find informative genes
4) Find and annotate cell communities
5) Understand cell cycle state of all cells
6) Compare the epithelial cancer cell transcriptomes between samples

Workshop 5 – the question

Location and stages of bladder cancer

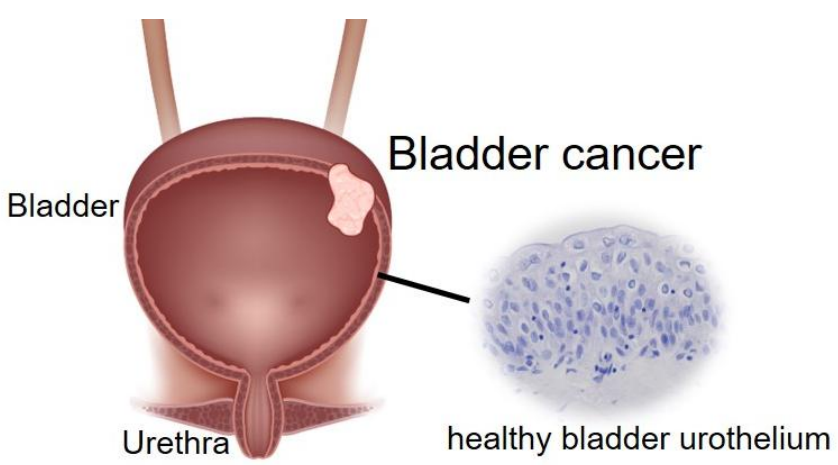


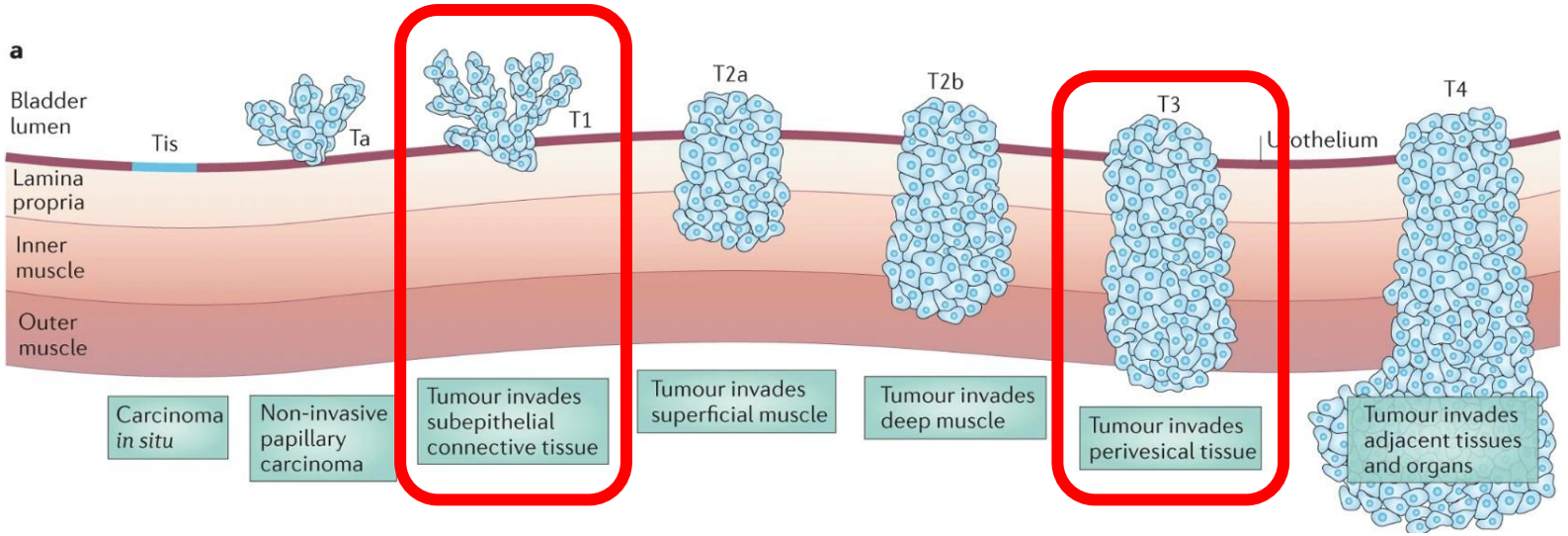


- 80% BLCA are Ta/T1 at initial presentation
- >50% recur and ~10% progress to T2+

Workshop 5 – the question

Location and stages of bladder cancer





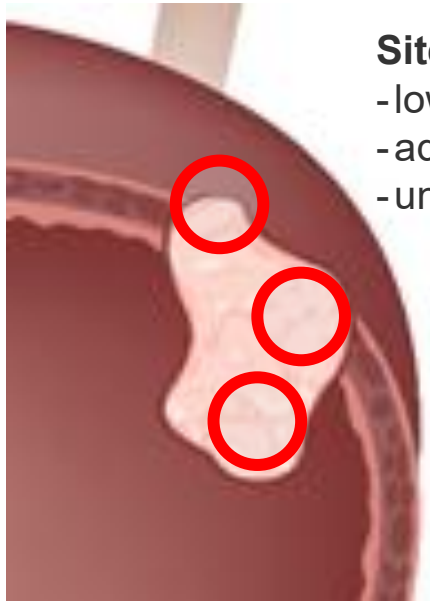
- 80% BLCA are Ta/T1 at initial presentation
- >50% recur and ~10% progress to T2+
- Today we will compare T1 and T3 tumours – trigger to muscle-invasive disease

(*remember this is n=1…)*

Workshop 5 – the question

Some (biological) thoughts/considerations:

- Extrapolating from 1 sample to a whole disease state
- Tissue sampling bias vs biological differences
  - What other information might we need?
- Limits of detection
  - If a gene is not detectable, is it absent?
- Tissue/disease heterogeneity



Site 1
- low tumour purity
- adjacent tissue included
- unclear origin of non-epi cells

Site 2
- high tumour purity
- infiltrating cells are really infiltrating

Site 3
- different subclone contributions to site 2?

Sometimes you simply cannot address these issues in your analysis. That's ok! (if you don't over-interpret in your results & discussion)

Choosing this workshop for your report

There are some helpful tips at the bottom of the workshop webpage.

Do not analyse the workshop BC2 vs BC5 dataset for your report!

I have provided four datasets, each able to address different questions in bladder cancer. You can analyse multiple sets, and you can compare with today's workshop results, but analyse them separately and then compare – **don't try to combine within Seurat**.

You can find your own dataset if you want to.

Make sure you have permission. You shouldn't use the dataset from your BSc capstone or MBiol group project. You *could* choose anything, be careful about gene names (e.g. PPARG vs Pparg in human vs mouse), and availability of reference datasets (GSEA etc.)

How do I get the excellence criteria?

Extend your analysis beyond what we've given you. Think critically. Think biologically. Do something interesting!

Get started!

1. Reboot into Linux if you haven't already
2. Follow the guide carefully (check where you are and check for typos)
3. Think about the biology and ask us questions!

https://asmasonomics.github.io/courses/Genomics3_Workshop5_scRNAseq_Nov2025