Introduction to Workshop 4:
RNAseq

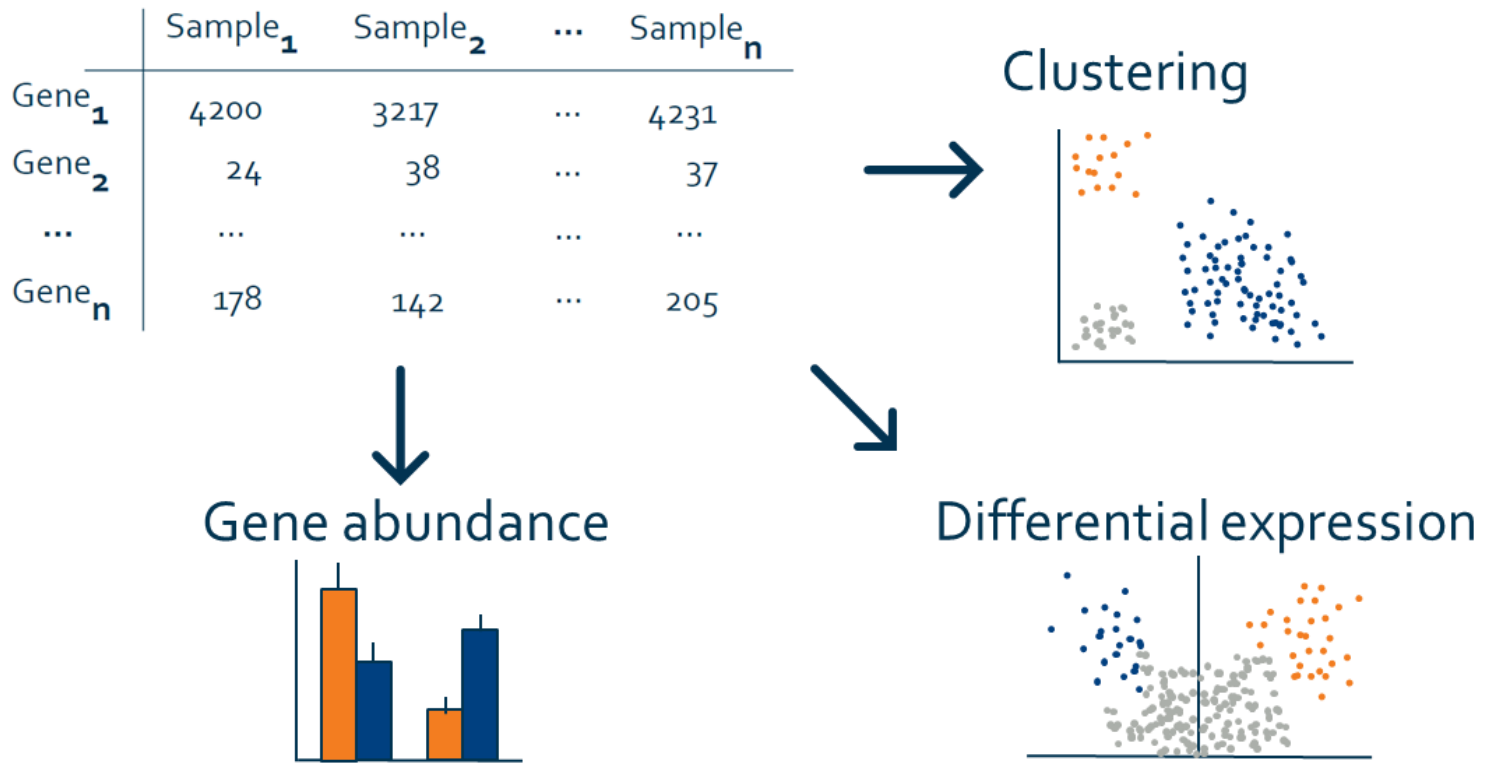
Genomics - BIO00087H

https://asmasonomics.github.io/courses/Genomics3_Workshop4_RNAseq_Nov2025

Transcriptomics – a brief L5 reminder

Bulk mRNAseq

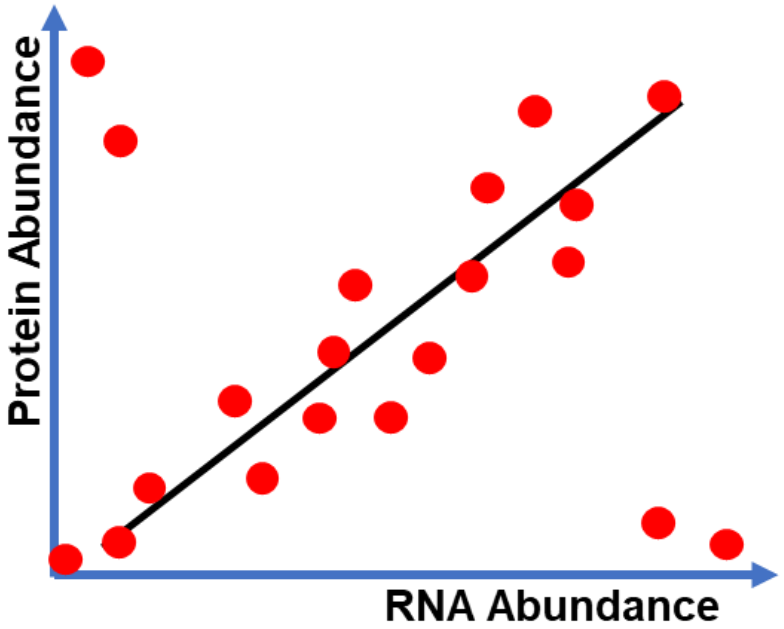
- What most people think of when they say RNAseq



Transcriptomics – a brief L5 reminder

Bulk mRNAseq

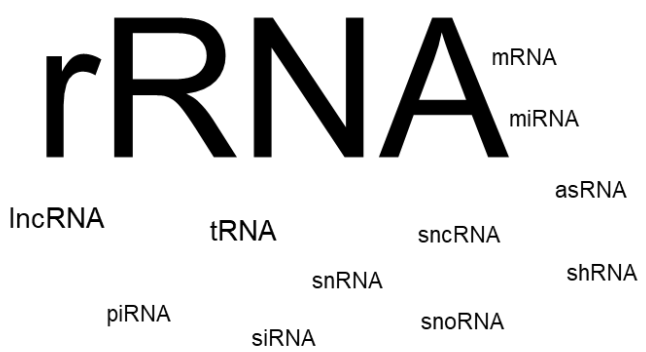
- What most people think of when they say RNAseq
- Functional molecules in cells are hard to measure directly
- RNA is easy and cheap to sequence and mostly indicative of phenotype

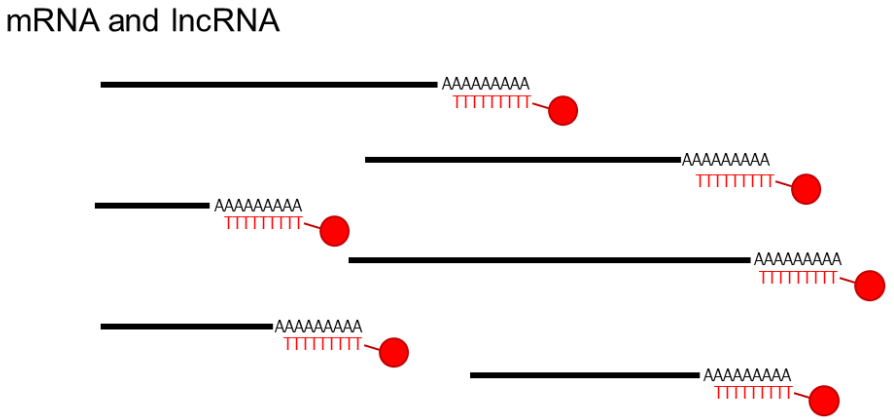


Transcriptomics – a brief L5 reminder

Bulk mRNAseq

- What most people think of when they say RNAseq
- Functional molecules in cells are hard to measure directly
- RNA is easy and is mostly indicative of cellular phenotype
- Combination of mRNA and lncRNA (both polyA)
  - Other RNAs still left in your library are technical noise
  - If you want to study other RNAs you need a different library prep (i.e. ribodep, sRNA)





Workshop 4 – the plan

1) Quality check our sequencing data
2) (Pseudo)align our reads to a reference human transcriptome
3) Merge individual transcript expression to the gene-level
4) Differential expression analysis (DEA) of control and treatment groups
5) Gene set enrichment analysis (GSEA) to look at changing pathways

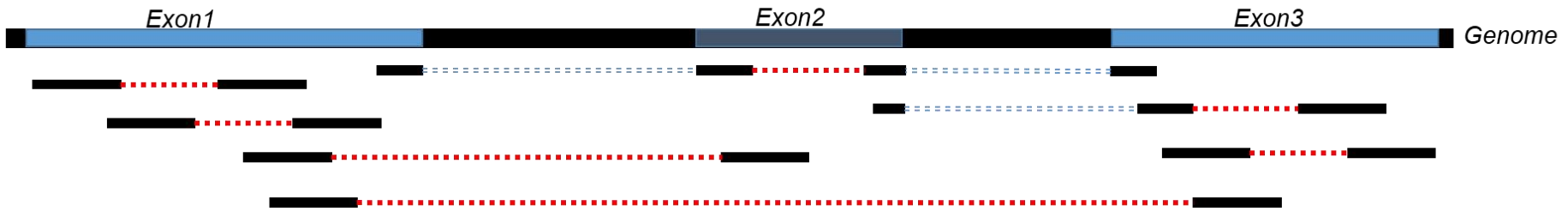
Workshop 4 – the plan

1) Quality check our sequencing data
2) (Pseudo)align our reads to a reference human transcriptome
3) Merge individual transcript expression to the gene-level
4) Differential expression analysis (DEA) of control and treatment groups
5) Gene set enrichment analysis (GSEA) to look at changing pathways

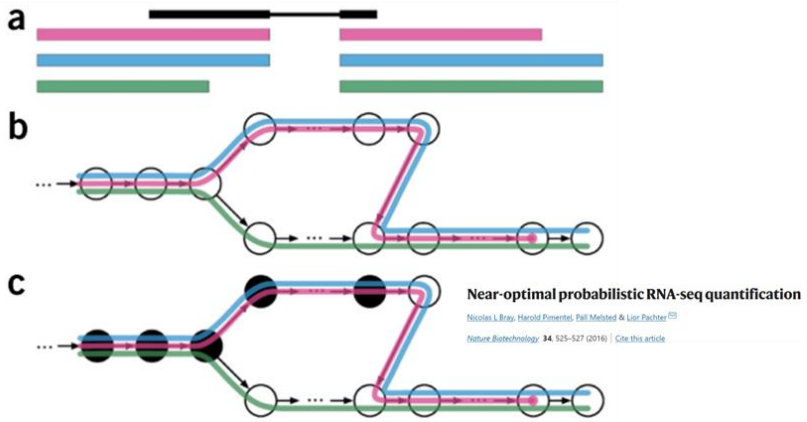
Pseudoalignment and transcript-level data

Very rapid and accurate mapping strategy → requires very well annotated genome

1) Gapped alignment to the genome



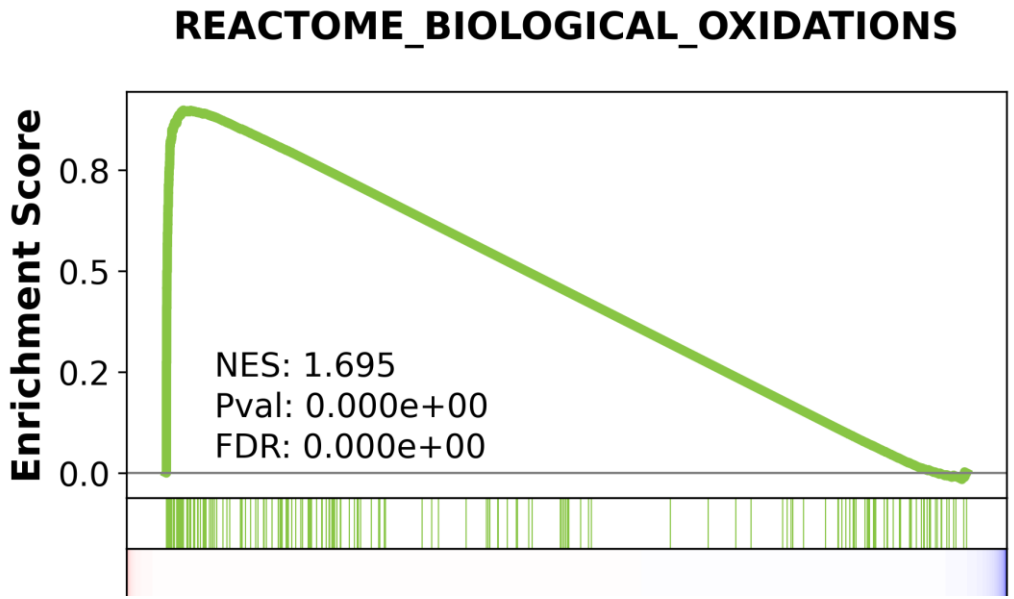
2) Pseudoalignment to the transcriptome
(well annotated genomes only)

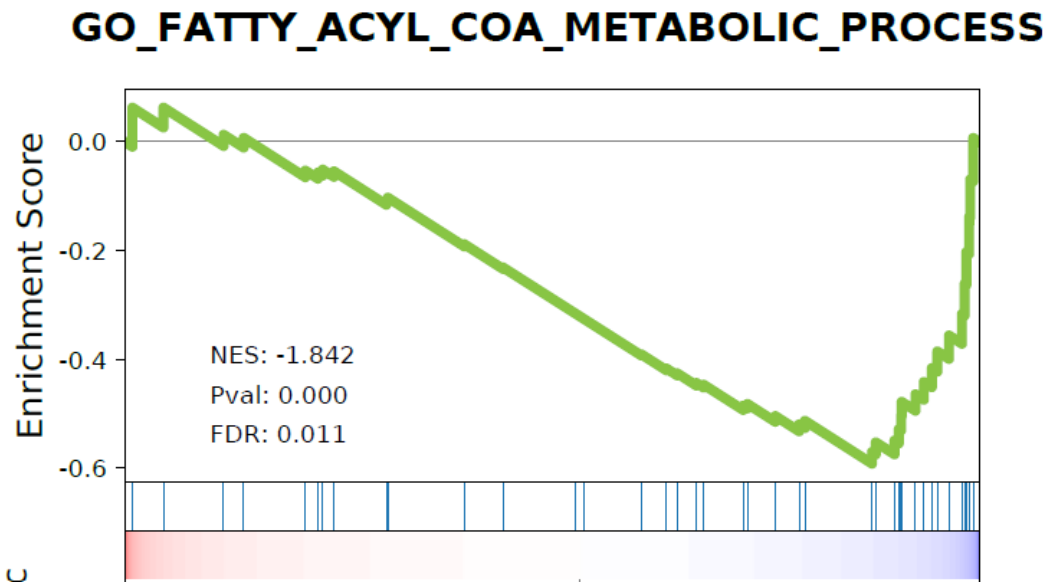


Gene Set Enrichment Analysis (GSEA)

Rather than googling every gene which is different between your groups, GSEA looks for “functional” groups of genes which are changing significantly

- This saves a lot of time (but can be misleading)
- The plots can be “a bit” gross





Workshop 4 – the question

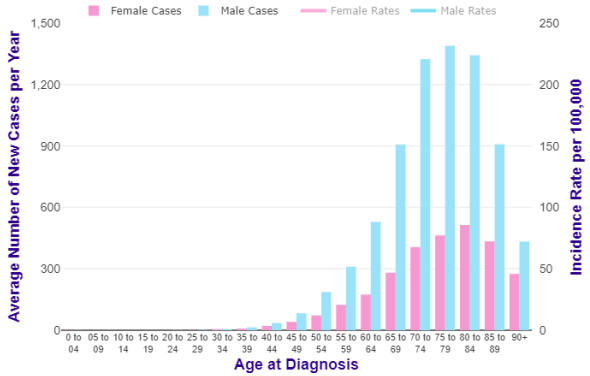
Question – what causes bladder cancer?

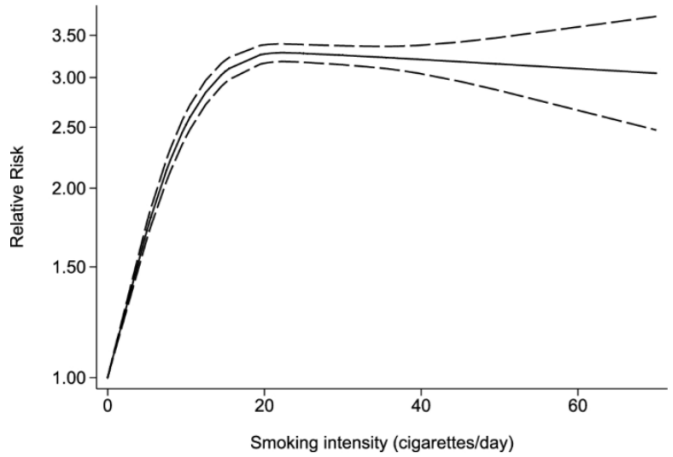
Zhao *et al* (2022) meta-analysis in bladder cancer PMID:35332429

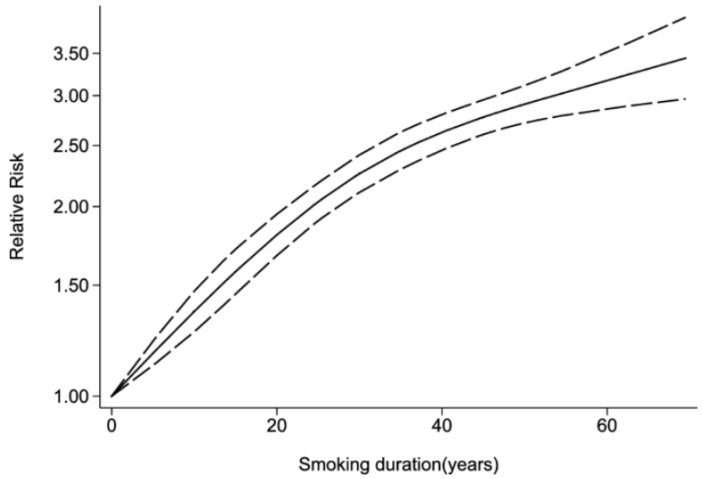
Didkowska *et al* (2011; PMID:20553096) from data in Simonata *et al* (2001; PMID:11275995)

CANCER RESEARCH UK

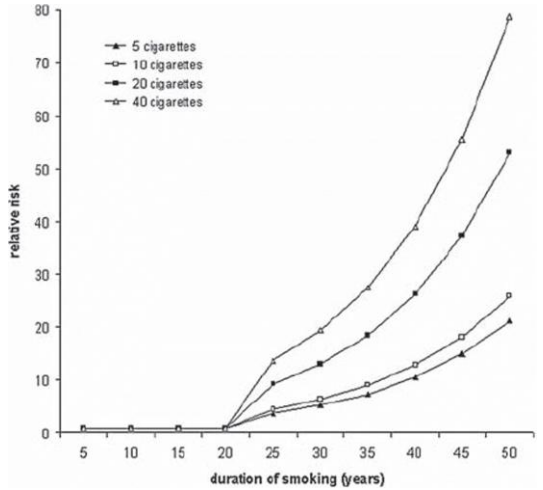
Older age is the main risk factor for cancer. This largely reflects cell DNA damage accumulating over time. Damage can result from biological processes or from exposure to risk factors.







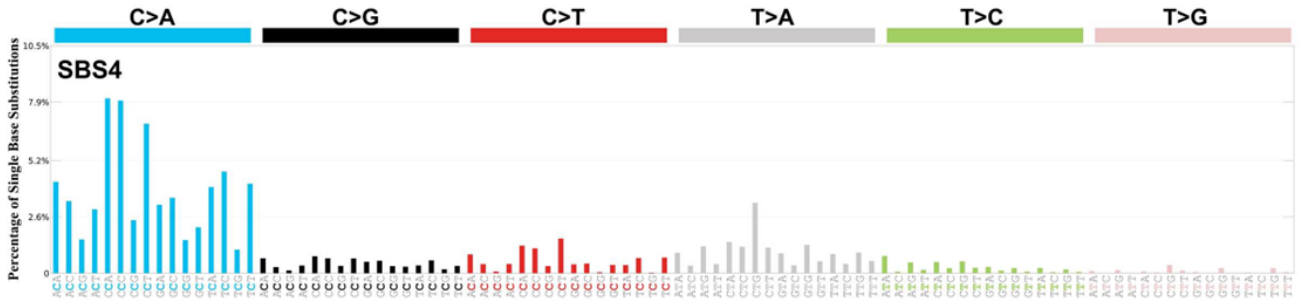
Smoking relationship different in lung cancer

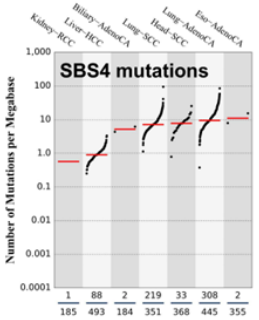


Workshop 4 – the question

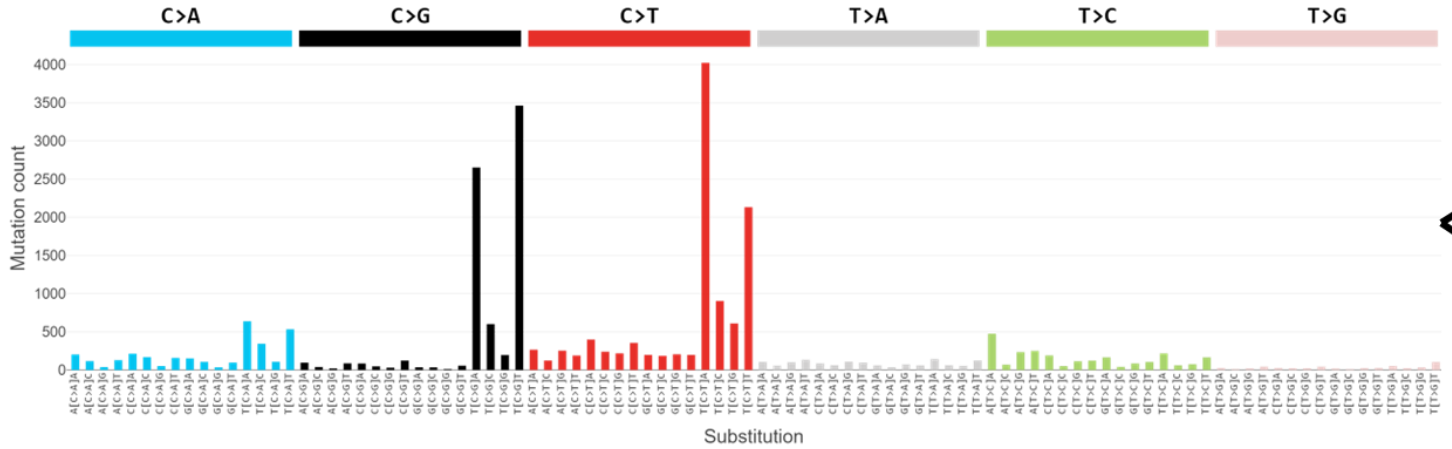
Question – what causes bladder cancer?

Smoking signature characteristic in lung cancer

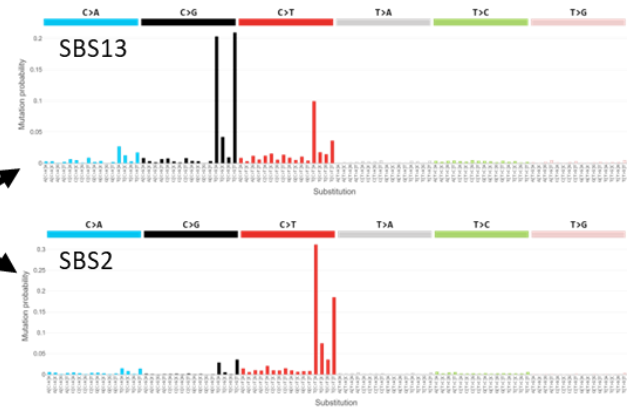




Typical bladder cancer mutational signature



APOBEC mutagenesis



Workshop 4 – the question

Baker lab (York) – hypothesized a causative virus could be BK Polyomavirus (BKPyV)

A key issue here is that bladder cancers do not contain virus

- HPV driven cancers do have bits of virus in them

BKPyV can infect human urothelial cells (bladder lining – originating cells in bladder cancer). BKPyV will kill the cells, but treating with IFNɣ can protect the cells (i.e. an innate immune response). BKPyV can be cleared from cells.

But what has BKPyV done to the remaining cells? Are they now pre-cancerous?

We did RNAseq to investigate this → what does infection do to the cells?

This is what you will investigate in today's workshop.

Choosing this workshop for your report

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

Do not analyse the BKPyV dataset for your report!

I have provided a dataset from healthy human urothelial cells grown in hypoxic (low oxygen) or normoxic conditions.

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 project. You *could* choose anything, but **stick with human data** – we have provided the necessary reference datasets for human, but not anything else.

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_Workshop4_RNAseq_Nov2025