



SPATIAL TRANSCRIPTOMICS DATA ANALYSIS: THEORY AND PRACTICE

PRACTICAL SESSION 2

DR SIMON J COCKELL

ELEFOTHERIOS (LEFTERIS) ZORMPAS

BIOSCIENCES INSTITUTE,
FACULTY OF MEDICAL SCIENCES,
NEWCASTLE UNIVERSITY

23/07/2023

Practical session 2

In this practical session, we will focus on some of the most common STx analysis tasks.

Practical session 2

Preprocessing and QC - the most important step of **any** Bioinformatics analysis!



Practical session 2

Preprocessing and QC

Quality Control (QC)



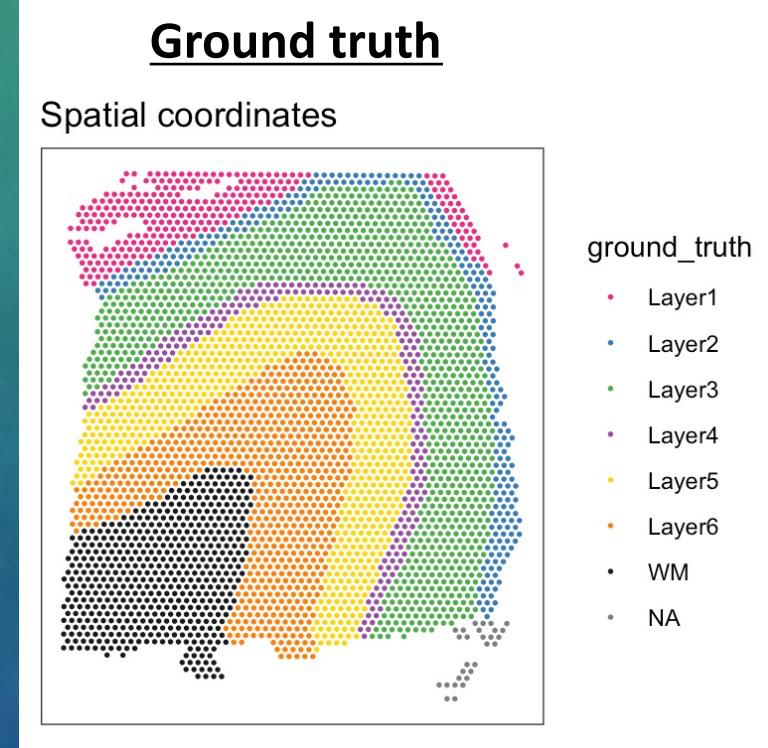
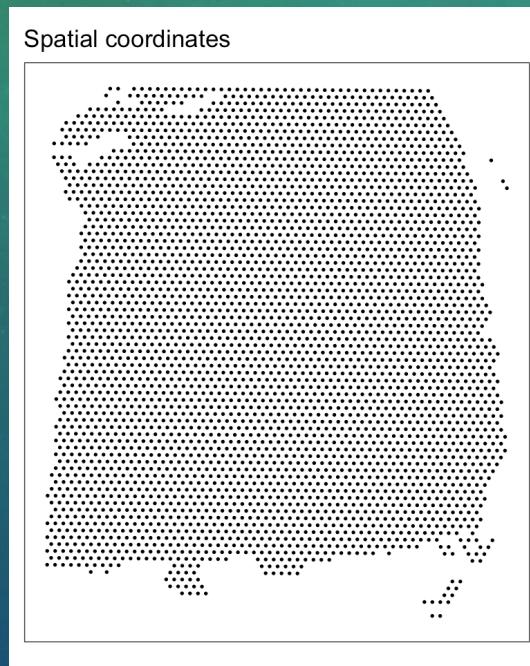
Practical session 2

In this tutorial session, we will also consider some global methods of STx analysis always inside the Bioconductor environment.

➤ To Do: Re-load the data and work on the practical's code

2.1 Spot-level Quality Control

Plot the dataset



2.1 Spot-level Quality Control

```
## Dataset dimensions before the filtering  
dim(spe)
```

```
## [1] 33538 4992
```

```
## Subset to keep only on-tissue spots  
spe <- spe[, colData(spe)$in_tissue == 1]  
dim(spe)
```

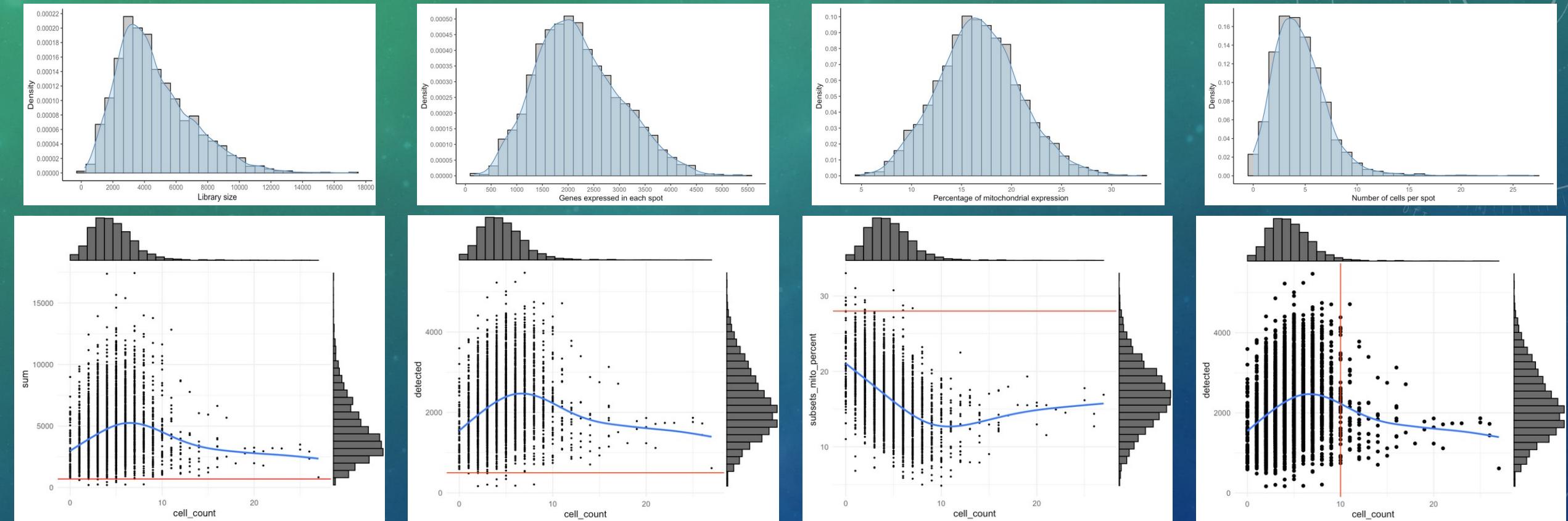
```
## [1] 33538 3639
```

```
## DataFrame with 6 rows and 13 columns  
##           barcode_id    sample_id in_tissue array_row array_col ground_truth cell_count  
sum detected      <character> <character> <integer> <integer> <character> <integer> <num  
eric> <numeric>  
## AAACAAGTATCTCCA-1 AAACAAGTATCTCCA-1 sample_151673      1     50     102   Layer3      6  
8458      3586  
## AAACAATCTACTAGCA-1 AAACAATCTACTAGCA-1 sample_151673      1      3     43   Layer1     16  
1667      1150  
## AACACCCAATAACTGC-1 AACACCCAATAACTGC-1 sample_151673      1     59     19      WM      5  
3769      1960  
## AACAGAGCGACTCCT-1 AACAGAGCGACTCCT-1 sample_151673      1     14     94   Layer3      2  
5433      2424  
## AACAGCTTCAGAAG-1 AACAGCTTCAGAAG-1 sample_151673      1     43      9   Layer5      4  
4278      2264  
## AACAGGGTCTATT-1 AACAGGGTCTATT-1 sample_151673      1     47     13   Layer6      6  
4004      2178  
##           subsets_mito_sum subsets_mito_detected subsets_mito_percent total  
##           <numeric> <numeric> <numeric> <numeric>  
## AAACAAGTATCTCCA-1        1407            13    16.6351    8458  
## AAACAATCTACTAGCA-1        204             11    12.2376   1667  
## AACACCCAATAACTGC-1        430             13    11.4089   3769  
## AACAGAGCGACTCCT-1       1316             13    24.2223   5433  
## AACAGCTTCAGAAG-1         651              12    15.2174   4278  
## AACAGGGTCTATT-1          621             13    15.5095   4004
```

total → Library size

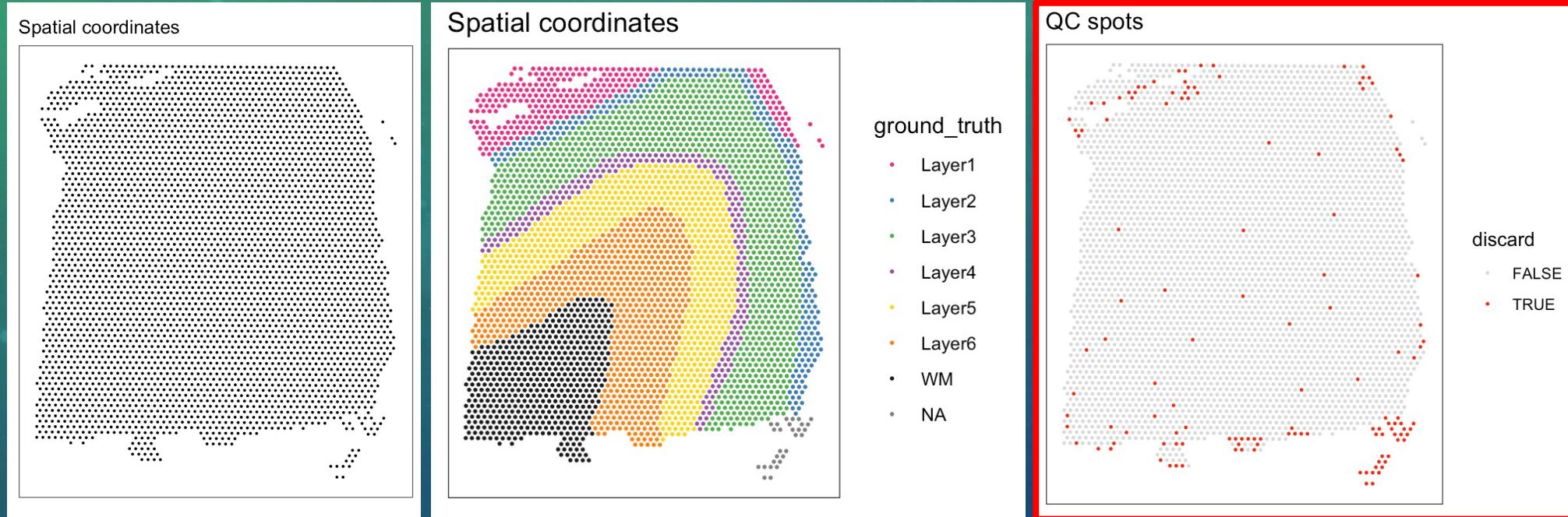
2.1 Spot-level Quality Control

Selecting thresholds



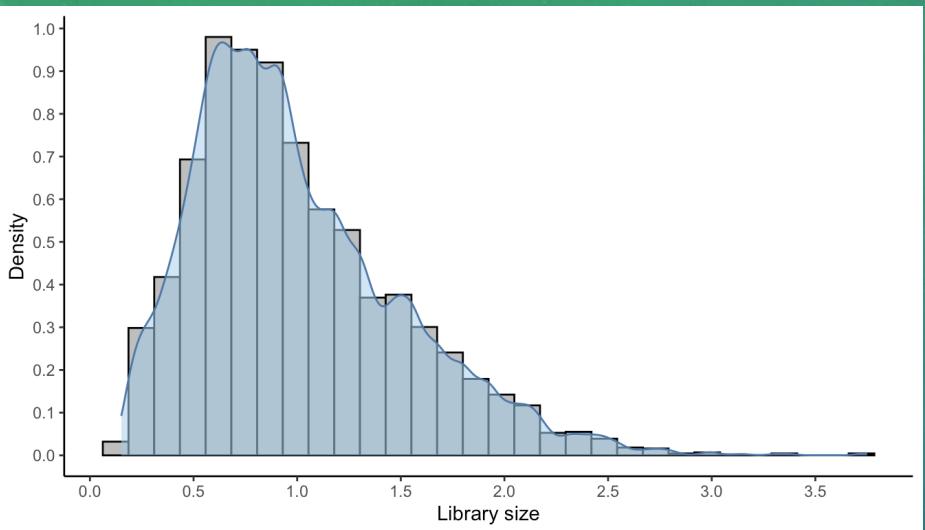
2.1 Spot-level Quality Control

Applying thresholds



Our QC does **not** remove any biology – so we can assume it is correct and move on.

2.2 Normalisation of counts



Library size factors are used for the normalisation

This is arguably the simplest approach for STx data.

Other scRNA-seq methods can be applied but:

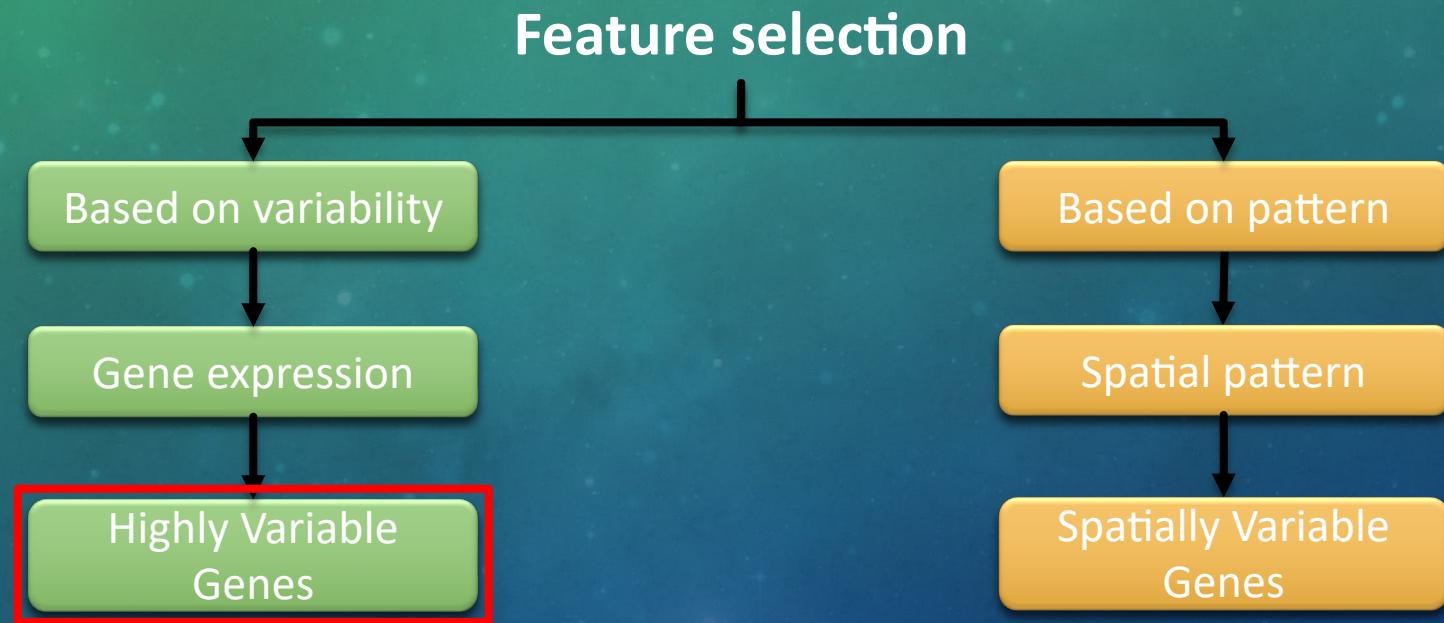
1. Spots can contain multiple cells of different cell types.
2. Datasets can include multiple tissue samples which will lead to different clusterings.

```
## Calculate logcounts and store in the spe object
spe <- logNormCounts(spe)

## Check that a new assay has been added
assayNames(spe)

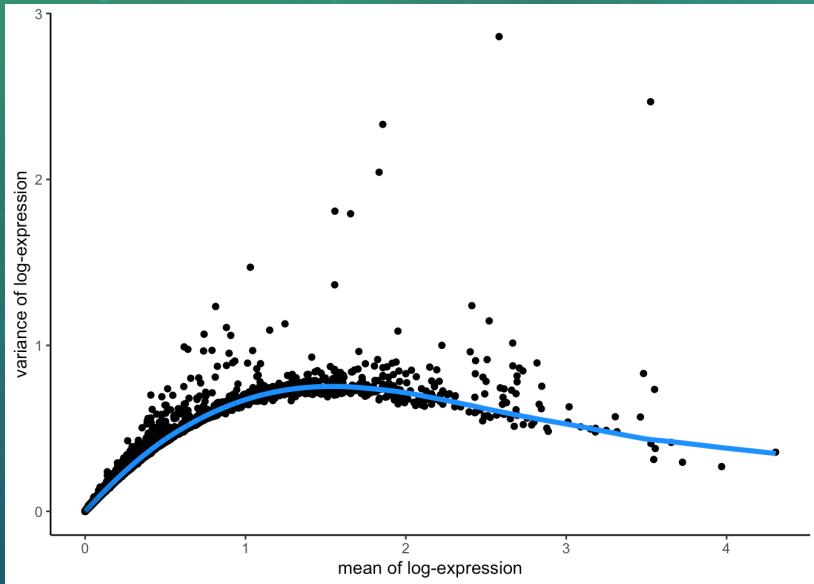
## [1] "counts"    "logcounts"
```

2.3 Selecting genes



2.3 Selecting genes

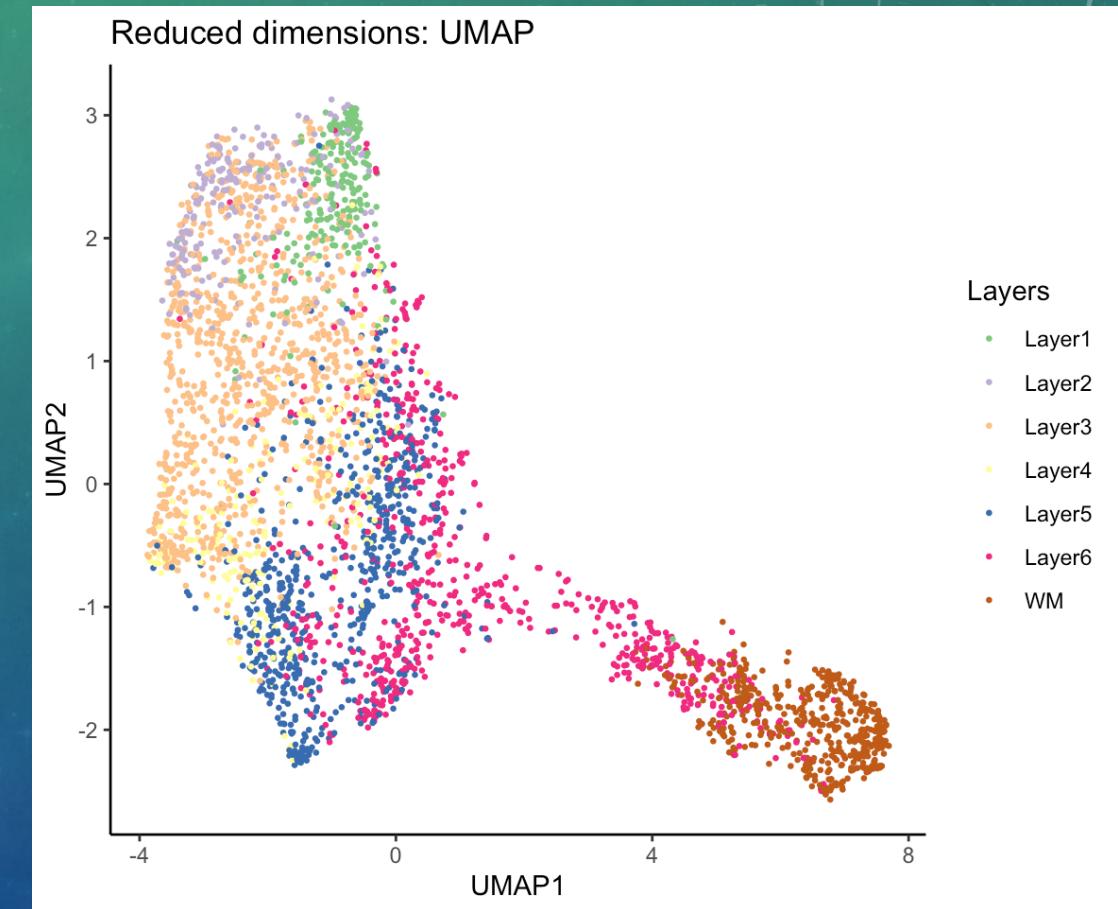
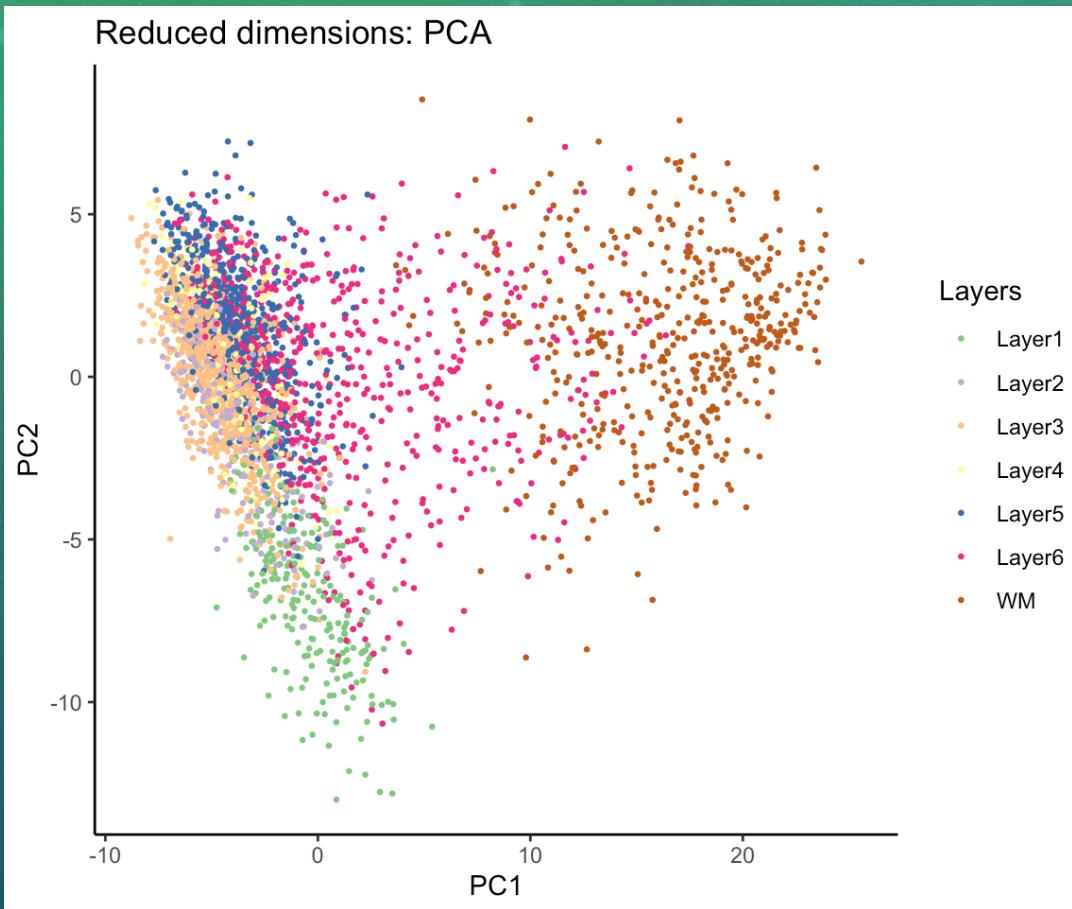
Highly Variable Genes (HVGs)



```
## Select top HVGs  
top_hvgs <- getTopHVGs(dec, prop = 0.1)  
  
## How many are the HVGs?  
length(top_hvgs)  
  
## [1] 1429
```

We select the top 10% of genes based on their variability.

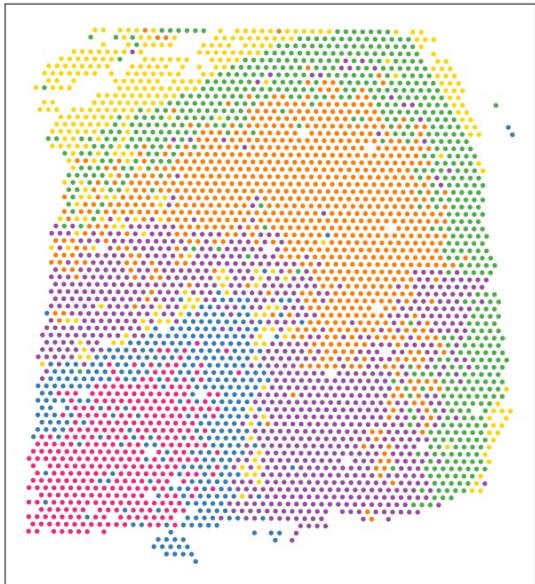
2.4 Dimensionality reduction



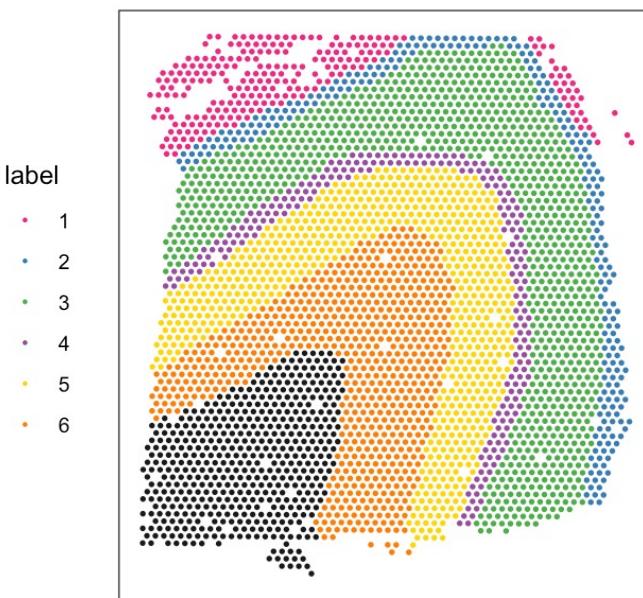
2.5 Clustering

```
## clus  
##  1   2   3   4   5   6  
## 350 354 661 895 366 885
```

Spatial coordinates



Spatial coordinates

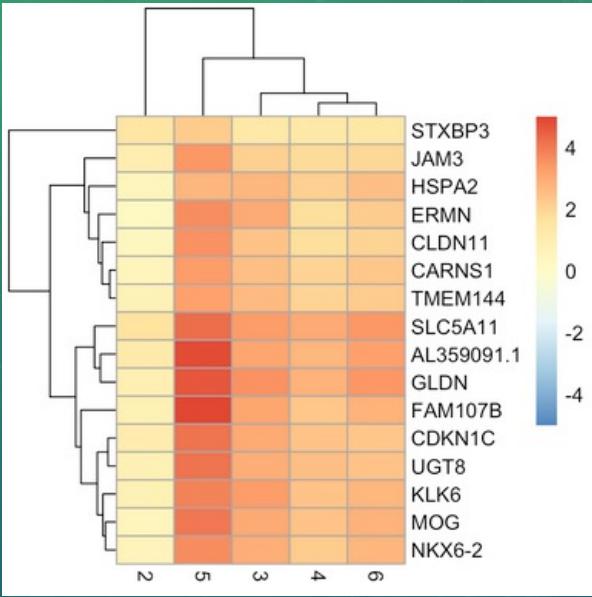


ground_truth

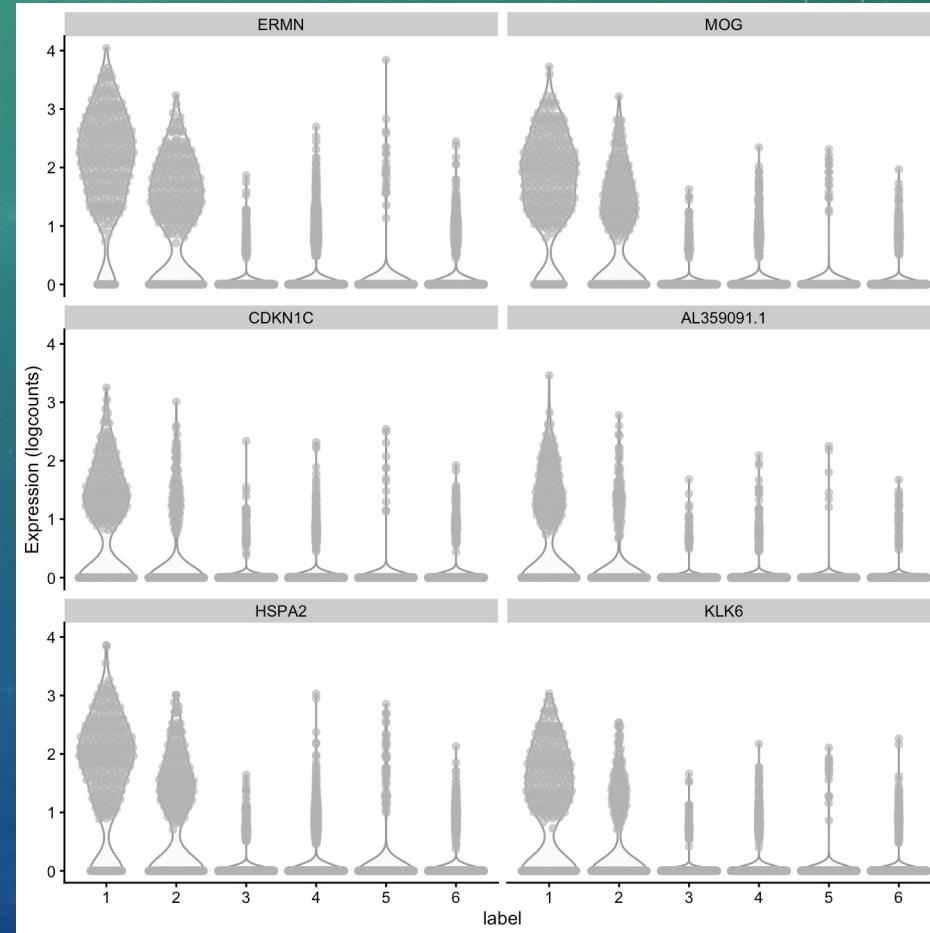
- Layer1
- Layer2
- Layer3
- Layer4
- Layer5
- Layer6
- WM

2.6 Inter-cluster differentially expressed genes (DGEs)

LogFCs for cluster 1 against all other clusters



Log counts of a selection of genes in all clusters



AKNOWLEDGEMENTS

Eleftherios Zormpas



Dr Simon J Cockell



Dr Rachel Queen



Prof. Alex Comber



iSMB feedback form:



© ICBAM research group, Newcastle University, UK



Medical
Research
Council



**MRC DiMeN
Doctoral Training
Partnership¹⁶**