

Introduction to STx analysis methods

Dr SIMON J COCKELL
ELEFTHERIOS (LEFTERIS) ZORMPAS

Biosciences Institute,
Faculty of Medical Sciences,
Newcastle University
23/07/2023





- Tool categories
- Mainstream tools in each category
- Promising tools in each category
- Walkthrough of a Bioconductor analysis pipeline



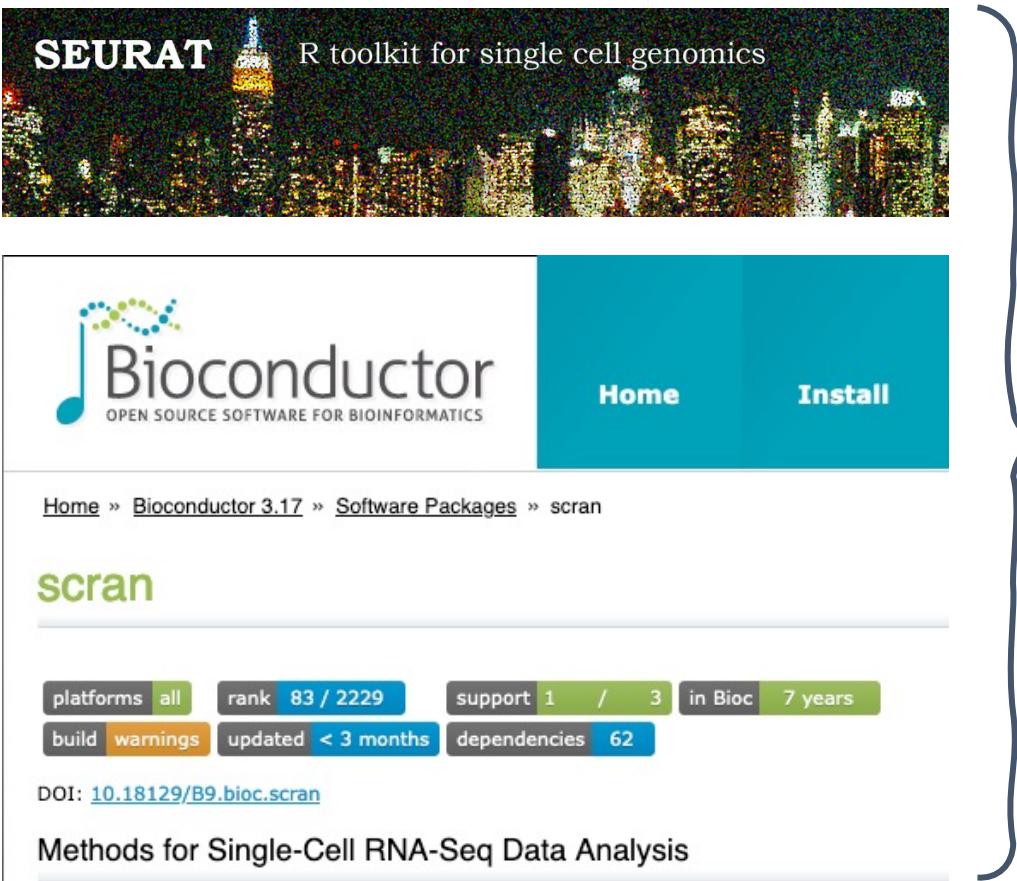
Part 1: Short overview of the STx data analysis tools landscape

CLUSTERING

Clustering of observations into statistically similar groups

Clustering approaches:

- k-means
- hierarchical
- Louvain → graph-based



- Originally developed for scRNA-seq data.
- SEURAT clustering:
 - Not Bioconductor.
 - graph-based building on PCA distances between top HVGs.
- scran clustering:
 - Bioconductor.
 - Mostly graph-based building on PCA distances.
 - igraph package-compatible for graph-based clustering algorithms.

DECONVOLUTION

Disentangling the cumulative gene expression profile and inferring the cell types present in the spot

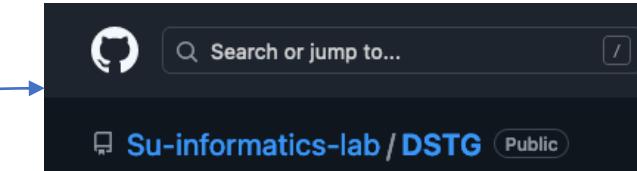
scRNA-seq + STx (*spot-based STx like 10X Visium*)

> 30 packages available

Approaches:

- deep learning and AI
- Bayesian models
- Statistical methods

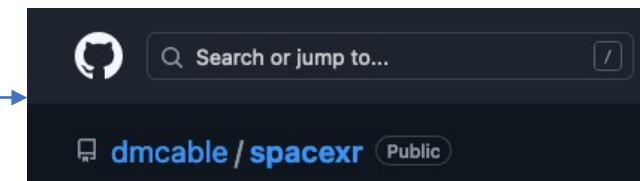
TANGRAM



BayesPrism



SpatialDecon

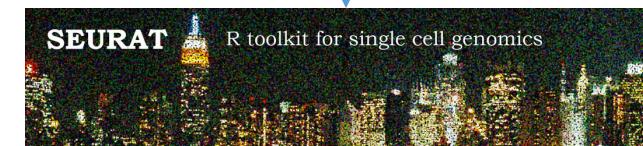


DATA INTEGRATION

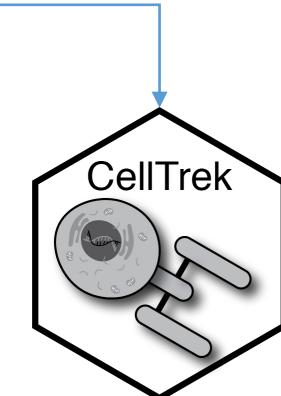
Combining STx with other related data types to enrich the interpretation



Usually, STx with scRNA-seq to increase resolution of STx



TANGRAM

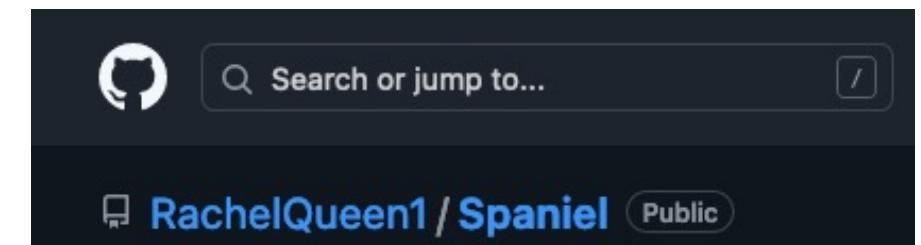
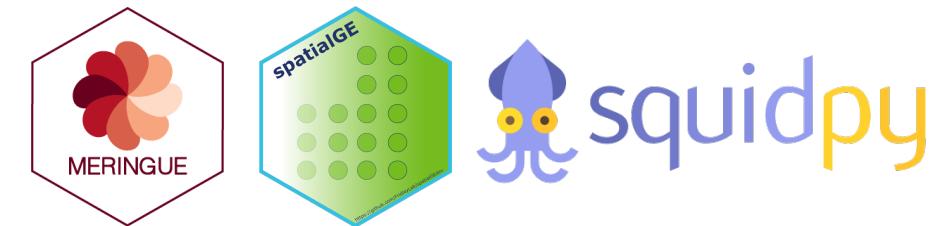
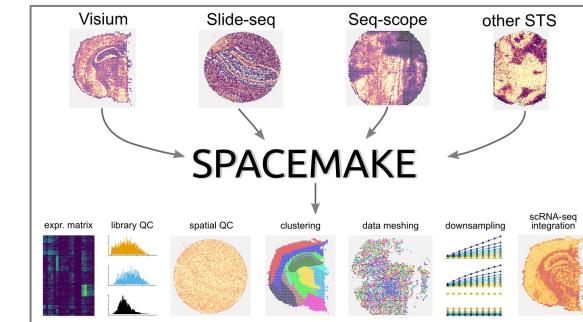
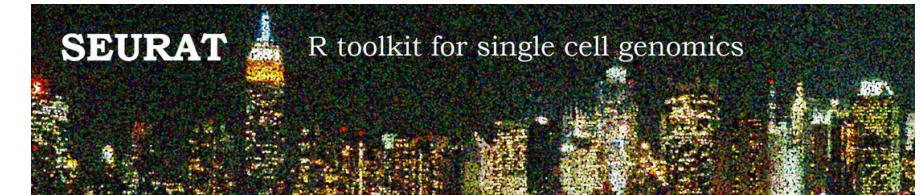


ST ANALYSIS TOOLKITS

Tools capable of undertaking more than one analysis task or are designed with the whole STx analysis pipeline in mind

Usually perform:

- Preprocessing
- Visualisation
- Integration
- Clustering
- Differential Gene Expression
- Any combination of the above with or without other specialised features



SPATIAL CLUSTERING

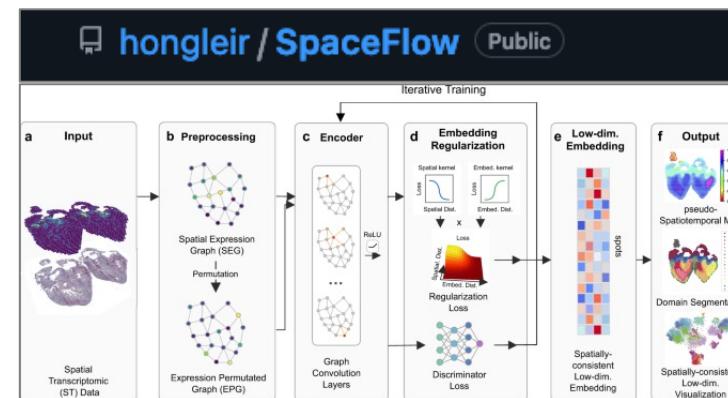
Spatial clustering seeks to define a geographically co-located group of similar observations

Geographically weighted clustering accounts for spatial data features (such as SA) when producing clusters.

Spatial Clustering



A screenshot of a GitHub search interface showing results for "Spatial Clustering". The top result is "maiziezhoulab / ADEPT" (Public), which has a dark theme and displays a 3D brain model. Below it is another result for "JinmiaoChenLab / GraphST" (Public).



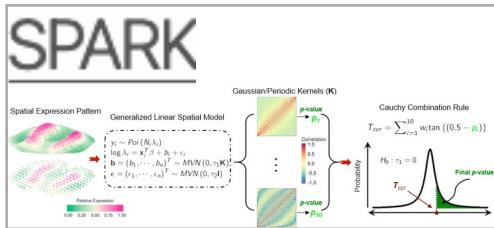
Geographically weighted Clustering



A screenshot of a GitHub search interface showing results for "Geographically weighted Clustering". The top result is "jianhuupenn / SpaGCN" (Public), which has a dark theme and displays a brain model.

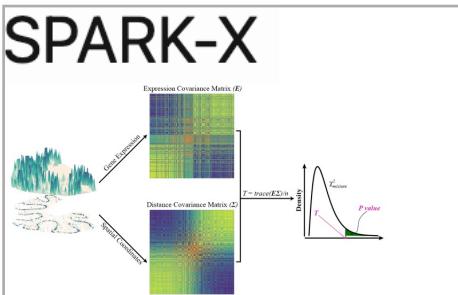
MAPPING SPATIALLY VARIABLE GENES (SVGs)

Generating activity maps of gene expression can reveal process-level insights into the functioning of genes



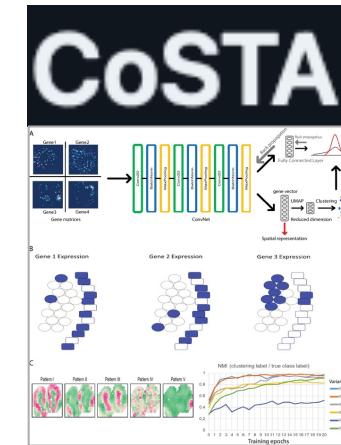
Sample size < 3000

Generalised
Linear Spatial
Model

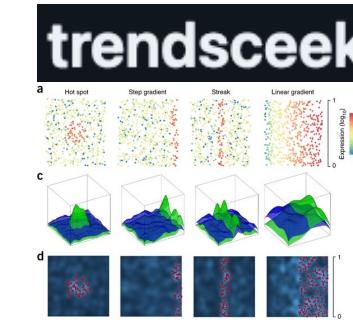


Sample size > 3000

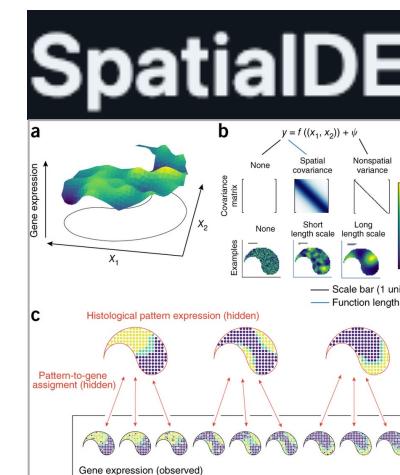
Covariance test
framework



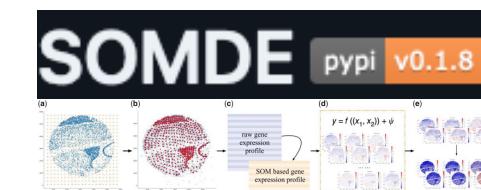
Convolutional
Neural
Networks



Statistical
method (marked
point process)



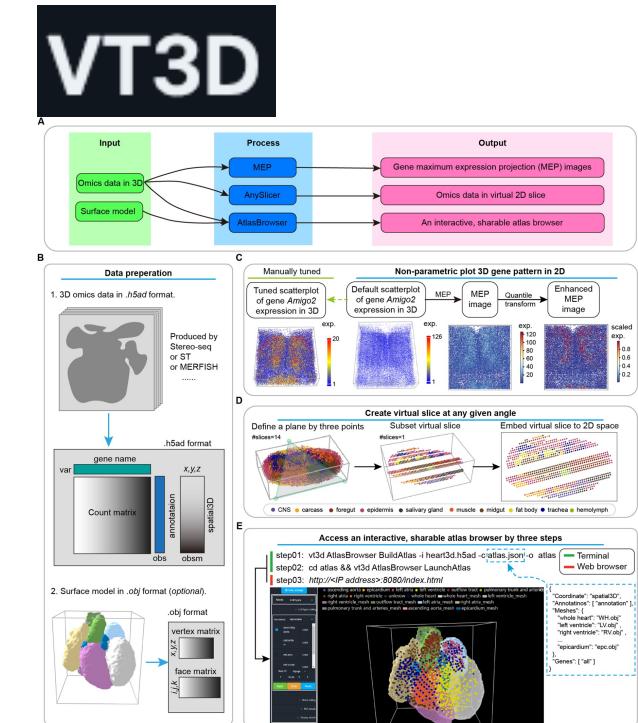
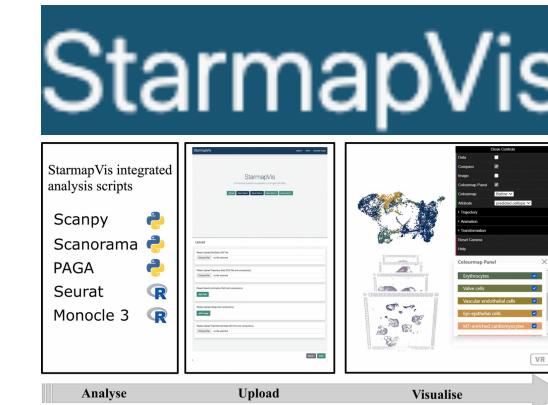
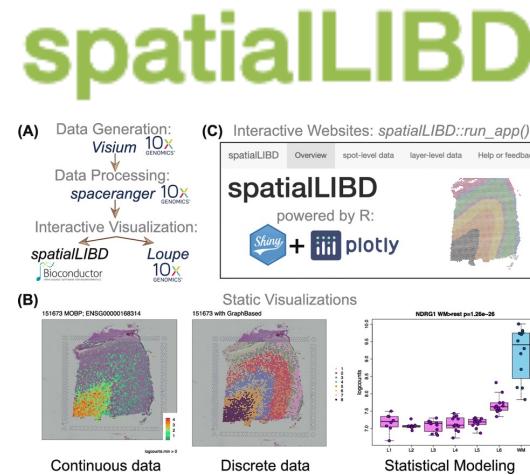
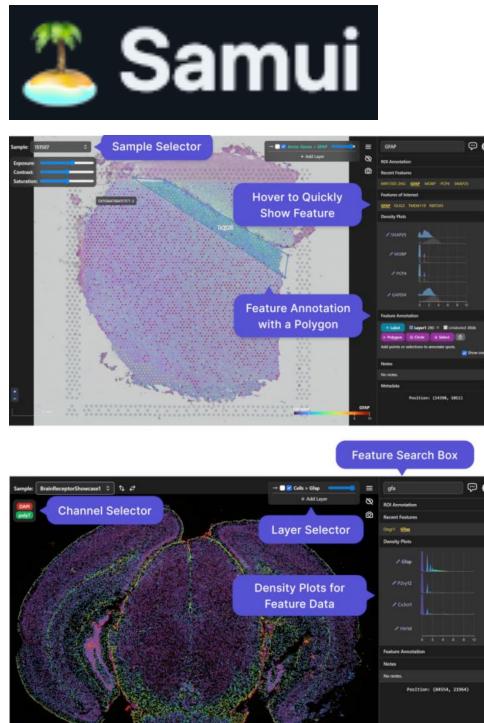
Statistical
method
(regression/
Bayes
model)

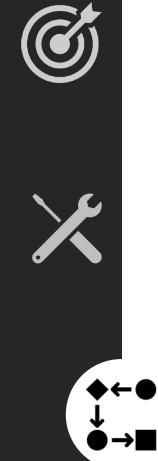


Neural
Networks

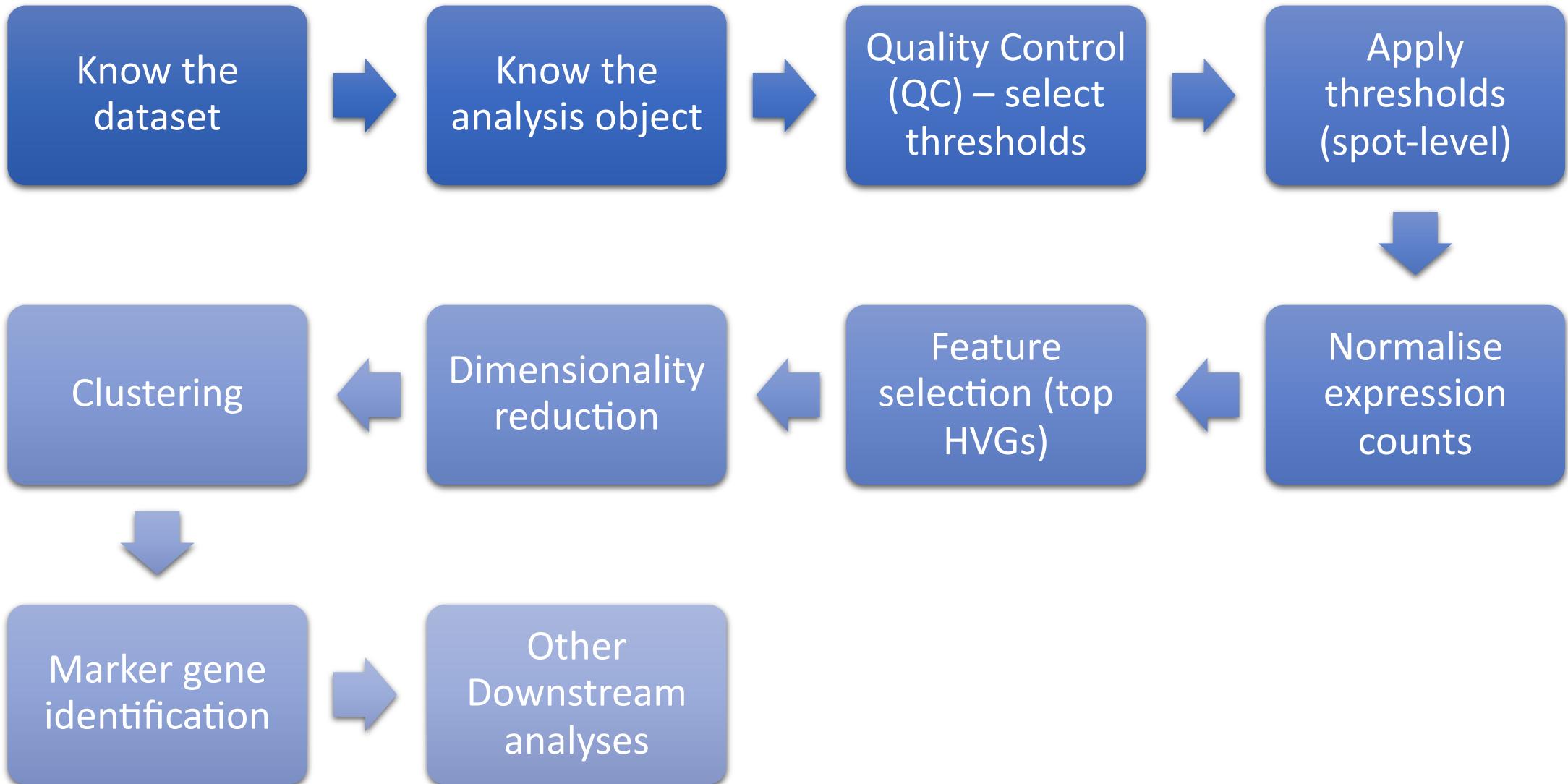
VISUALISATION

Visualisation toolboxes to allow users explore their data





Part 2: Walkthrough of a Bioconductor STx analysis pipeline



1

The dataset

2

3

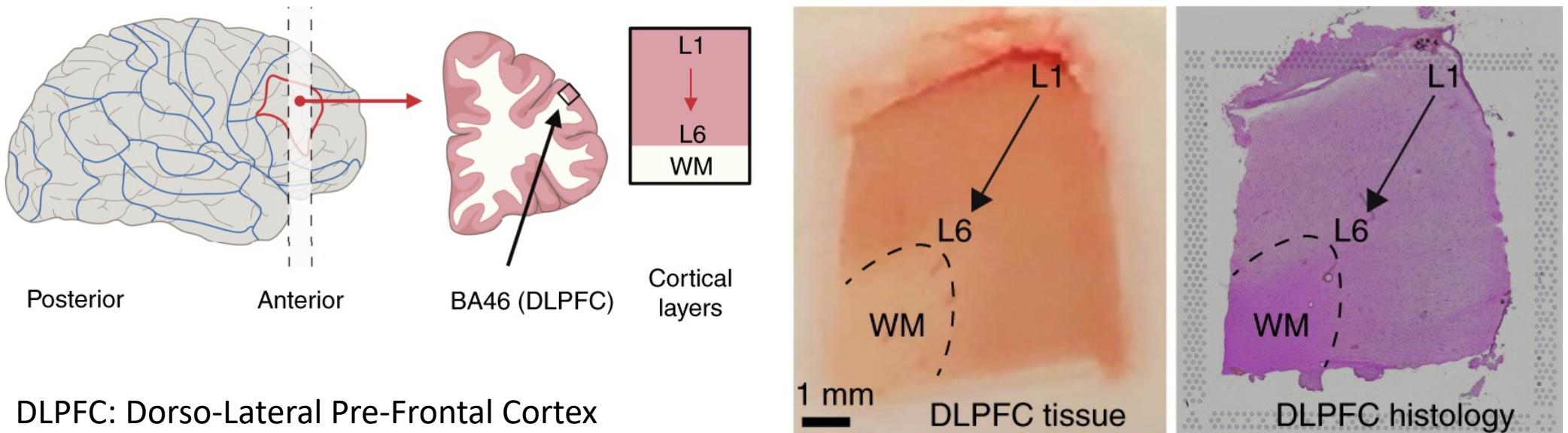
4

5

6

7

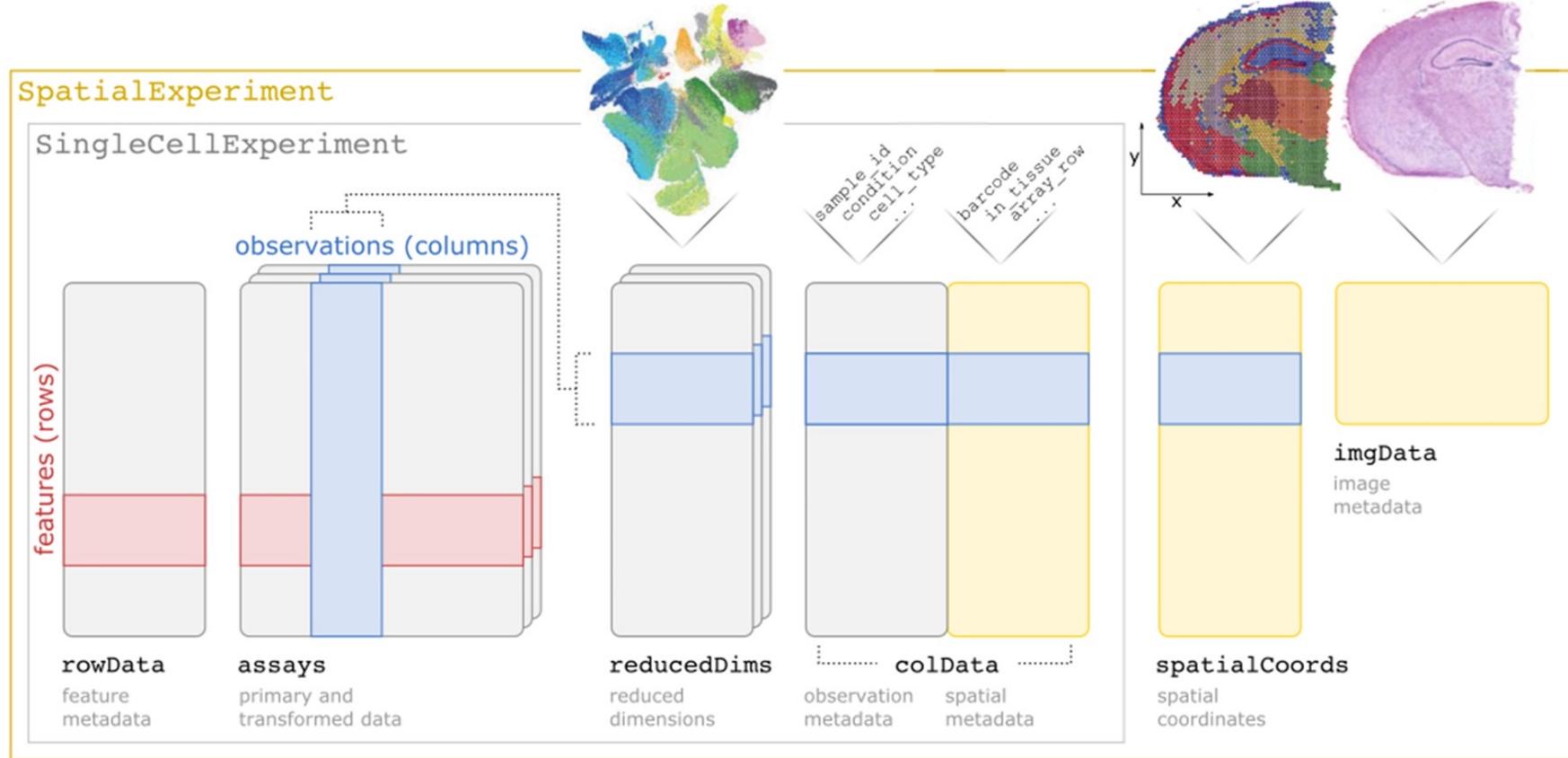
8



For the purpose of this tutorial, we will be using 1 of the 12 samples: sample 151673

1

The SpatialExperiment class



1. **assays**: gene expression counts
2. **rowData**: information about features, usually genes
3. **colData**: information on spots (non-spatial and spatial metadata)
4. **spatialCoords**: spatial coordinates
5. **imgData**: image data

PRACTICAL SESSION 2

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

2



3

4

5

6

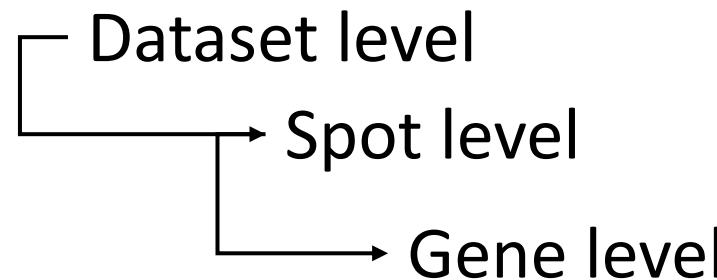
7

8

1
Preprocessing and QC

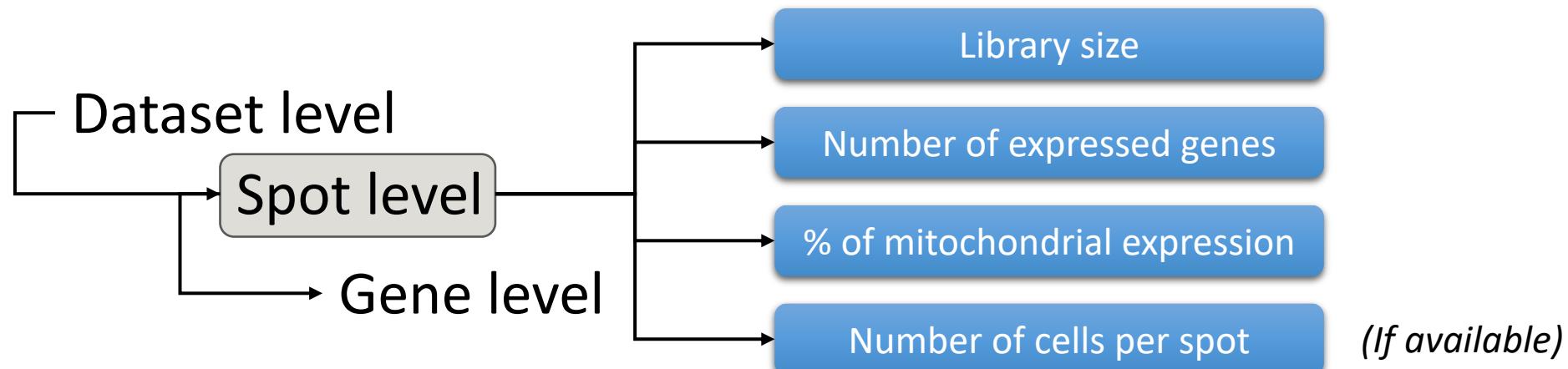
2

Quality Control (QC)



1
2
3
4
5
6
7
8
Preprocessing and QC

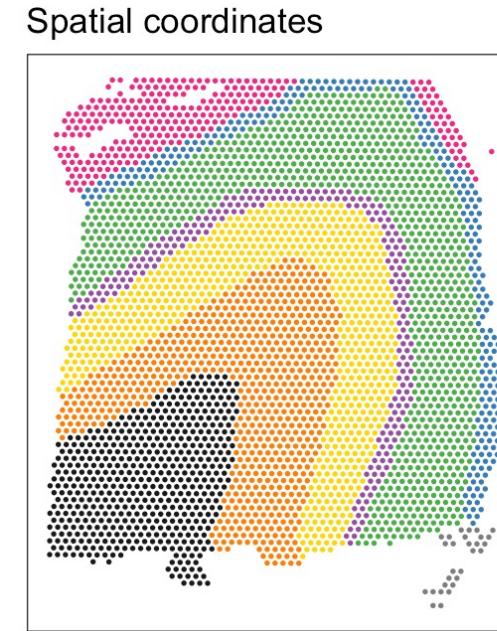
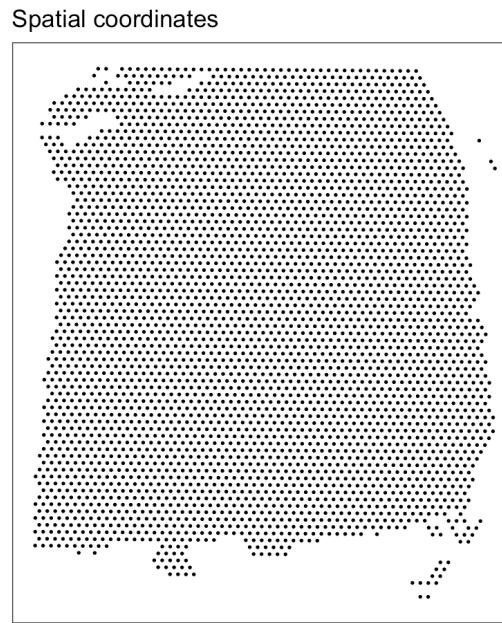
Quality Control (QC)



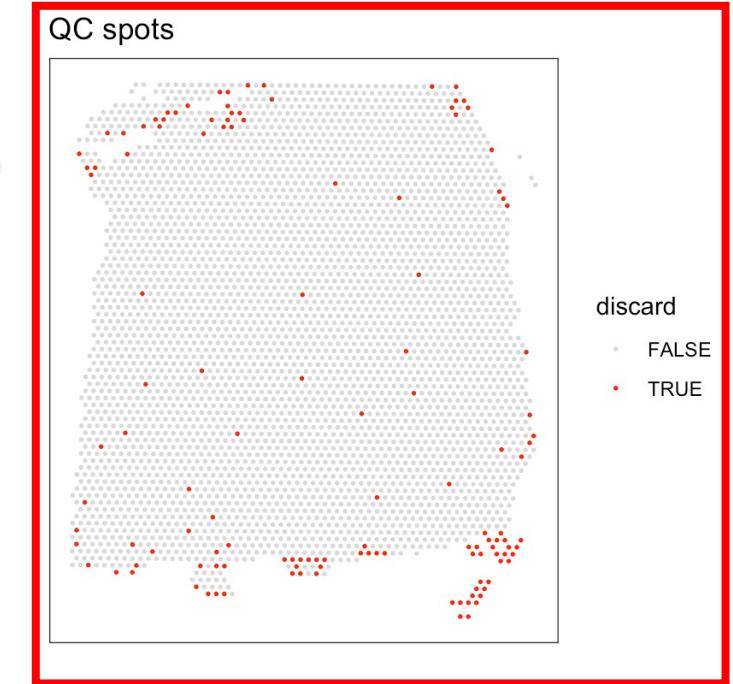
PRACTICAL SESSION 2

Applying thresholds

3



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

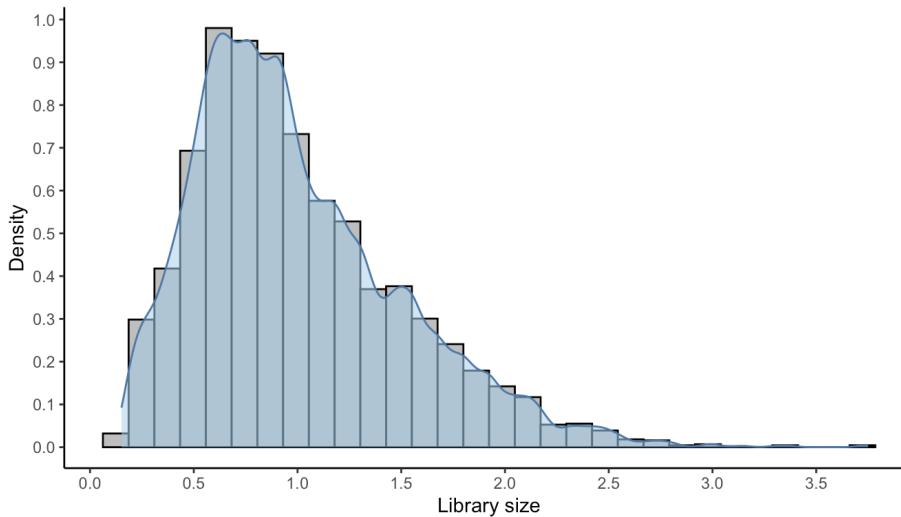


Our QC does not remove any biology – so we can assume it is correct and move on.

8

PRACTICAL SESSION 2

Counts normalisation



Library size factors are used for the normalisation

4

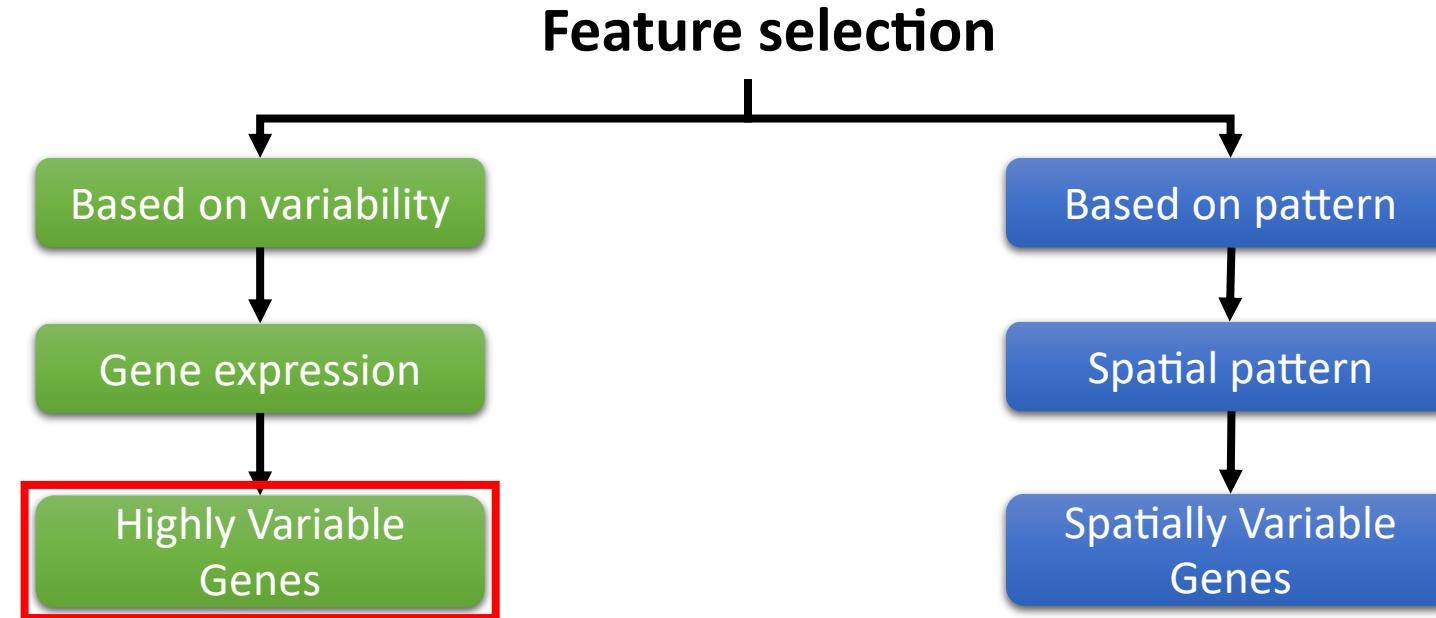
```
## Calculate logcounts and store in the spe object  
spe <- logNormCounts(spe)
```

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

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

8

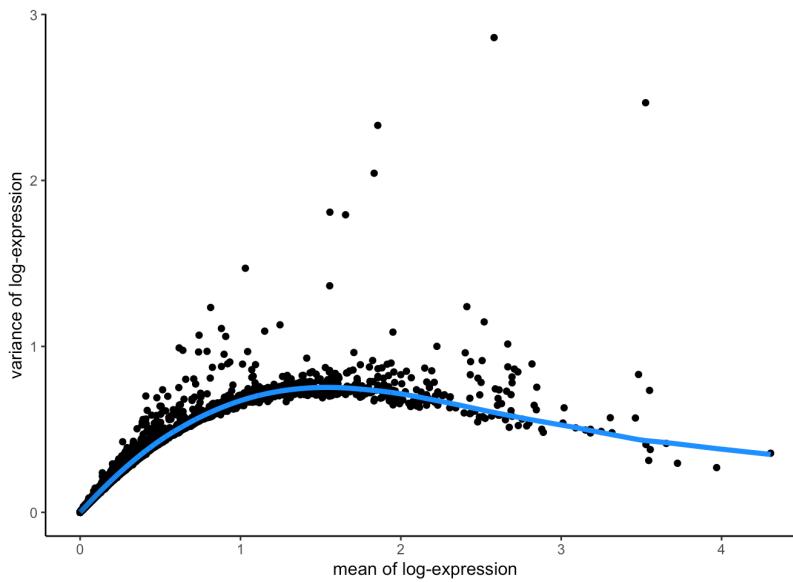
Selecting genes



PRACTICAL SESSION 2

Selecting genes

Highly Variable Genes (HVGs)

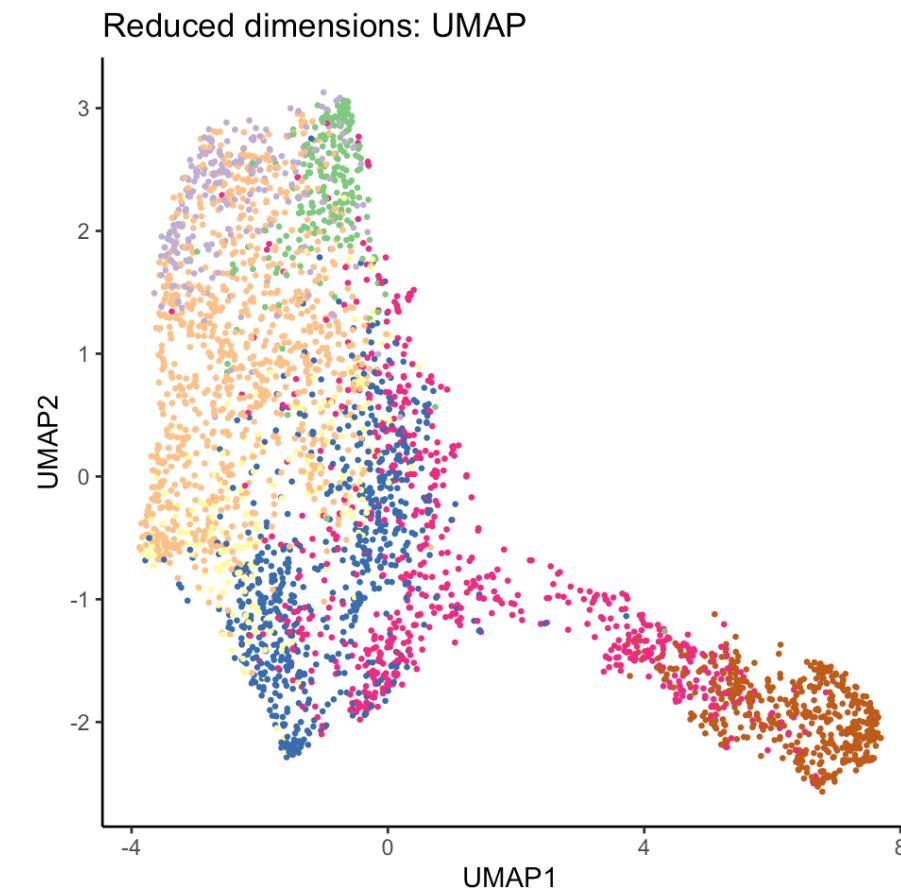
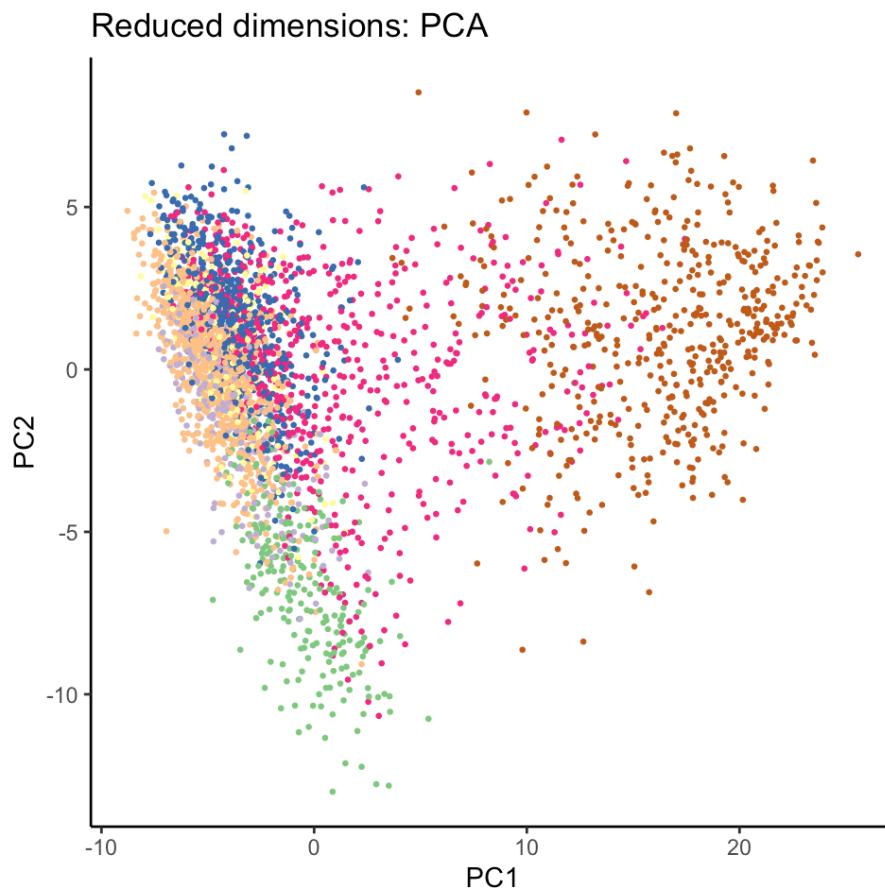


```
## Select top HVGs  
top_hvgs <- getTopHVGs(dec, prop = 0.1)  
  
## How many are the HVGs?  
length(top_hvgs)  
  
## [1] 1429
```

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

PRACTICAL SESSION 2

Dimensionality reduction

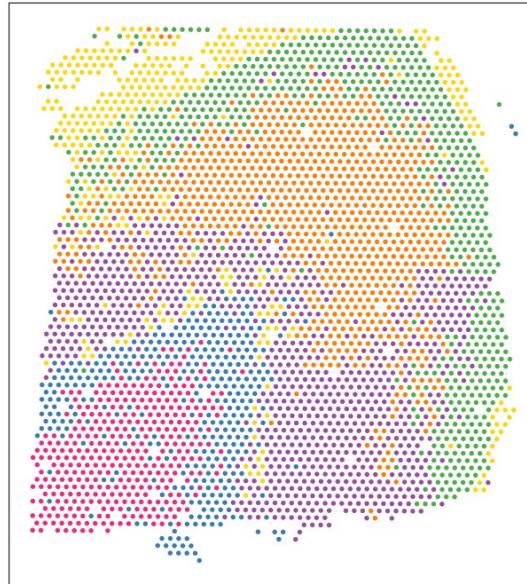


PRACTICAL SESSION 2

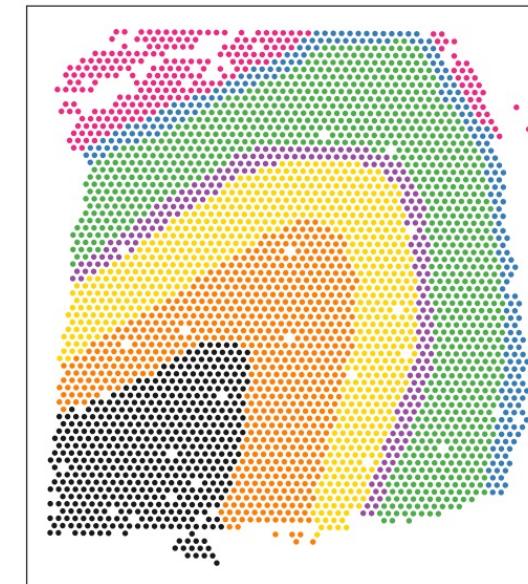
Clustering

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

Spatial coordinates



Spatial coordinates



ground_truth

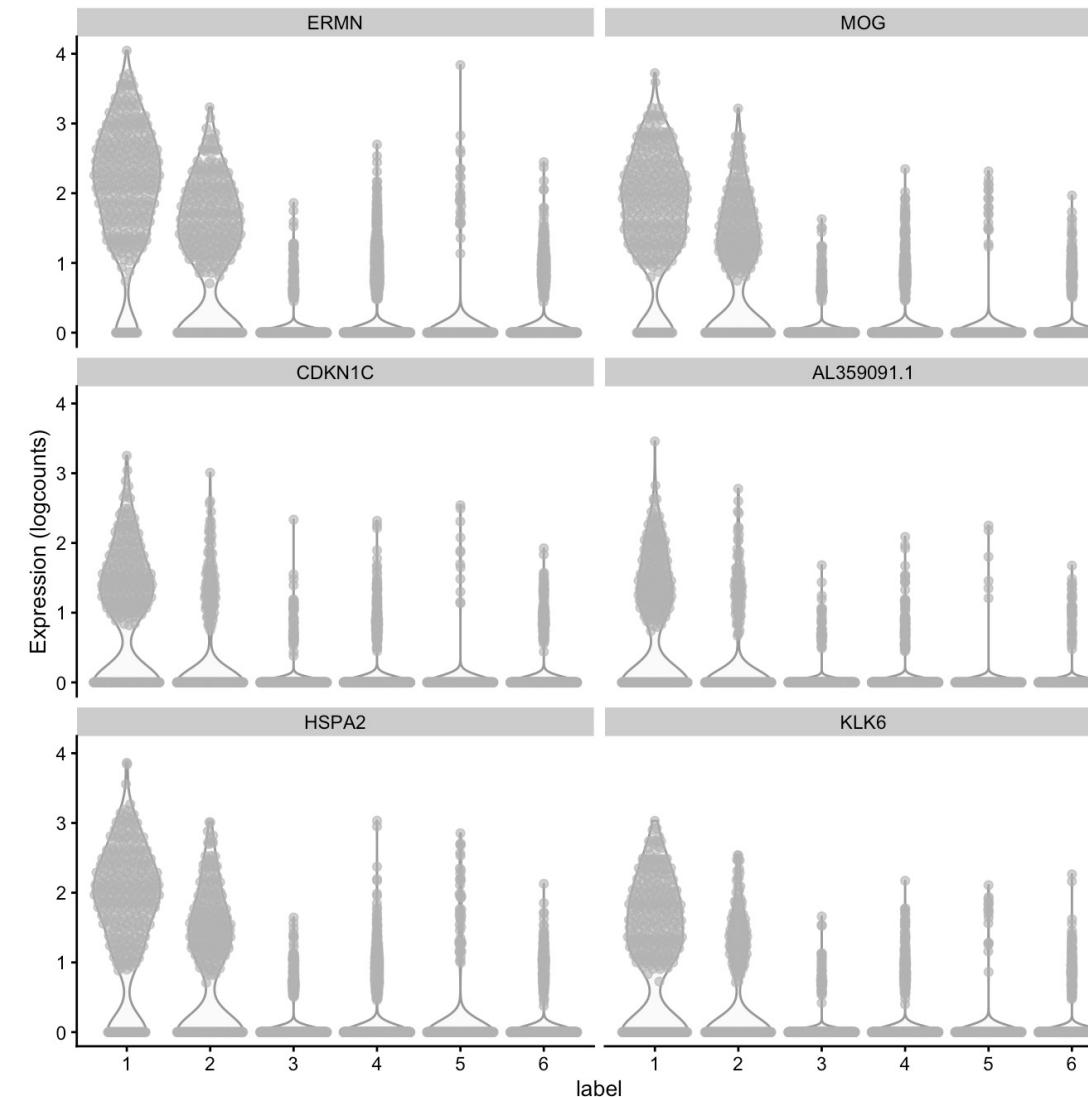
- Layer1
- Layer2
- Layer3
- Layer4
- Layer5
- Layer6
- WM

7

8

PRACTICAL SESSION 2

Differentially expressed genes (DGEs)



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
Partnership²⁵