CEITEC
Central European Institute of Technology
BRNO | CZECH REPUBLIC

UNIVERSITAS MASARYKIANA BRUNENSIS

Modern methods for genome analysis (PřF:Bi7420)

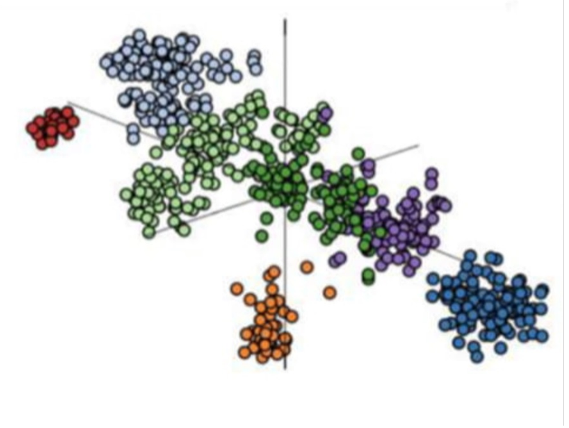
Lecture 7 : Single Cell RNA-seq analysis

Vojta Bystry
vojtech.bystry@ceitec.muni.cz

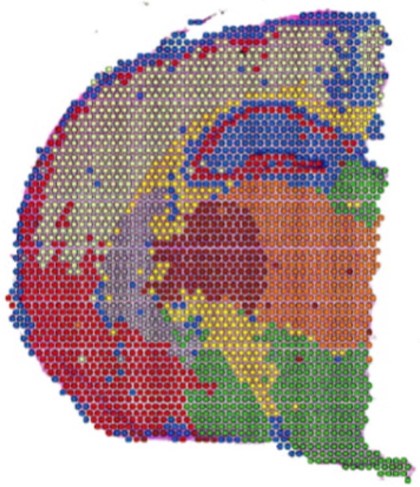
RNA-seq types

- Bulk RNA-seq

- SC-RNA-seq



- Spatially resolved RNA-seq



Bulk RNA-seq vs. SC-RNA-seq

- Bulk RNA-seq
    - Several samples
    - Difference in RNA levels between predetermined set of samples (conditions)

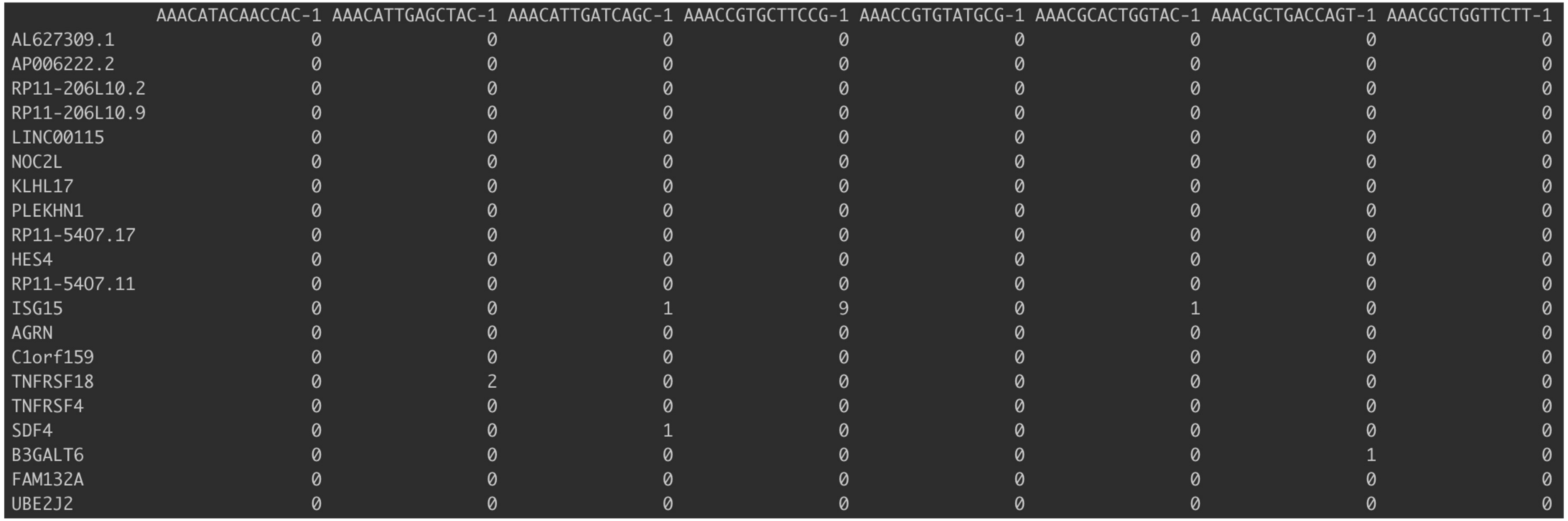
- SC-RNA-seq
    - Hundreds to thousand cells
    - Distinguish (cluster) cells based on the difference in RNA levels

SC-RNA-seq primary analysis

- In principle similar to the bulk RNA-seq
    - Map to genomic reference, demultiplex and count reads per gene
- For 10x Genomics tool from the company - Cell Ranger
- Cell Ranger - report example

SC-RNA-seq primary analysis results

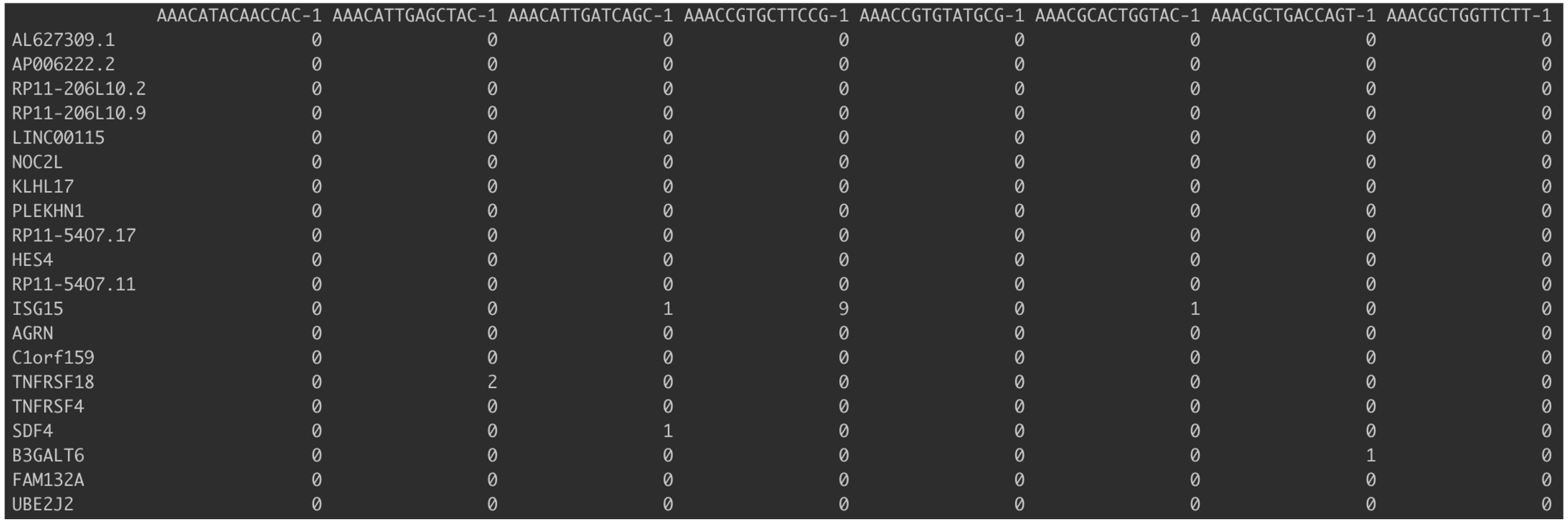
- per gene (feature) per cell read count



```
              AAACATACAACCAC-1 AAACATTGAGCTAC-1 AAACATTGATCAGC-1 AAACCGTGCTTCCG-1 AAACCGTGTATGCG-1 AAACGCACTGGTAC-1 AAACGCTGACCAGT-1 AAACGCTGGTTCTT-1
AL627309.1                   0                0                0                0                0                0                0                0
AP006222.2                   0                0                0                0                0                0                0                0
RP11-206L10.2                0                0                0                0                0                0                0                0
RP11-206L10.9                0                0                0                0                0                0                0                0
LINC00115                    0                0                0                0                0                0                0                0
NOC2L                        0                0                0                0                0                0                0                0
KLHL17                       0                0                0                0                0                0                0                0
PLEKHN1                      0                0                0                0                0                0                0                0
RP11-54O7.17                 0                0                0                0                0                0                0                0
HES4                         0                0                0                0                0                0                0                0
RP11-54O7.11                 0                0                0                0                0                0                0                0
ISG15                        0                0                1                9                0                1                0                0
AGRN                         0                0                0                0                0                0                0                0
C1orf159                     0                0                0                0                0                0                0                0
TNFRSF18                     0                2                0                0                0                0                0                0
TNFRSF4                      0                0                0                0                0                0                0                0
SDF4                         0                0                1                0                0                0                0                0
B3GALT6                      0                0                0                0                0                0                1                0
FAM132A                      0                0                0                0                0                0                0                0
UBE2J2                       0                0                0                0                0                0                0                0
```

Pre-processing workflow

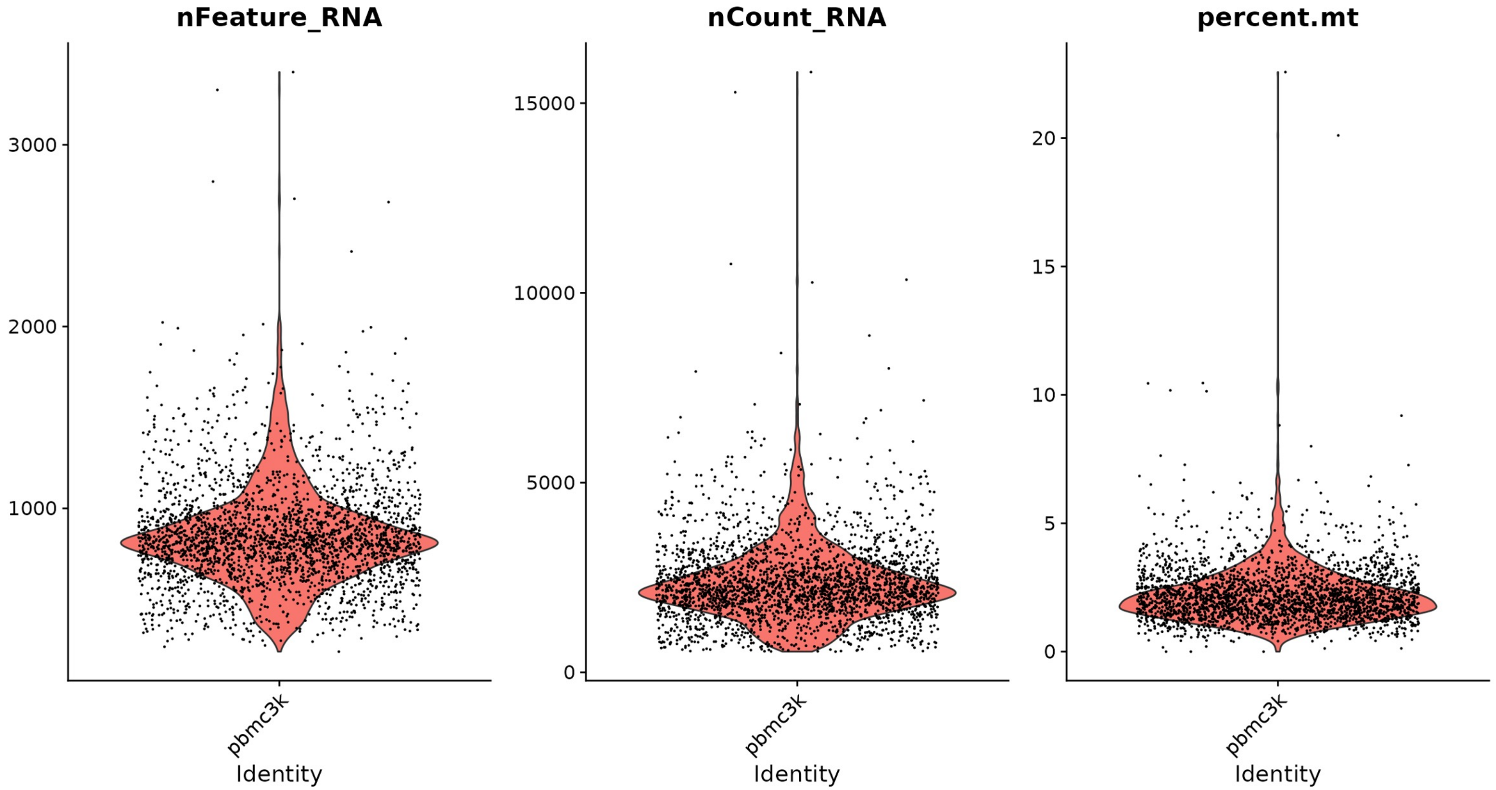
- Filtering of cells based on QC metrics
- Detection and filtering of highly variable features
- Data normalization and scaling



```
              AAACATACAACCAC-1 AAACATTGAGCTAC-1 AAACATTGATCAGC-1 AAACCGTGCTTCCG-1 AAACCGTGTATGCG-1 AAACGCACTGGTAC-1 AAACGCTGACCAGT-1 AAACGCTGGTTCTT-1
AL627309.1                   0                0                0                0                0                0                0                0
AP006222.2                   0                0                0                0                0                0                0                0
RP11-206L10.2                0                0                0                0                0                0                0                0
RP11-206L10.9                0                0                0                0                0                0                0                0
LINC00115                    0                0                0                0                0                0                0                0
NOC2L                        0                0                0                0                0                0                0                0
KLHL17                       0                0                0                0                0                0                0                0
PLEKHN1                      0                0                0                0                0                0                0                0
RP11-54O7.17                 0                0                0                0                0                0                0                0
HES4                         0                0                0                0                0                0                0                0
RP11-54O7.11                 0                0                0                0                0                0                0                0
ISG15                        0                0                1                9                0                1                0                0
AGRN                         0                0                0                0                0                0                0                0
C1orf159                     0                0                0                0                0                0                0                0
TNFRSF18                     0                2                0                0                0                0                0                0
TNFRSF4                      0                0                0                0                0                0                0                0
SDF4                         0                0                1                0                0                0                0                0
B3GALT6                      0                0                0                0                0                0                1                0
FAM132A                      0                0                0                0                0                0                0                0
UBE2J2                       0                0                0                0                0                0                0                0
```

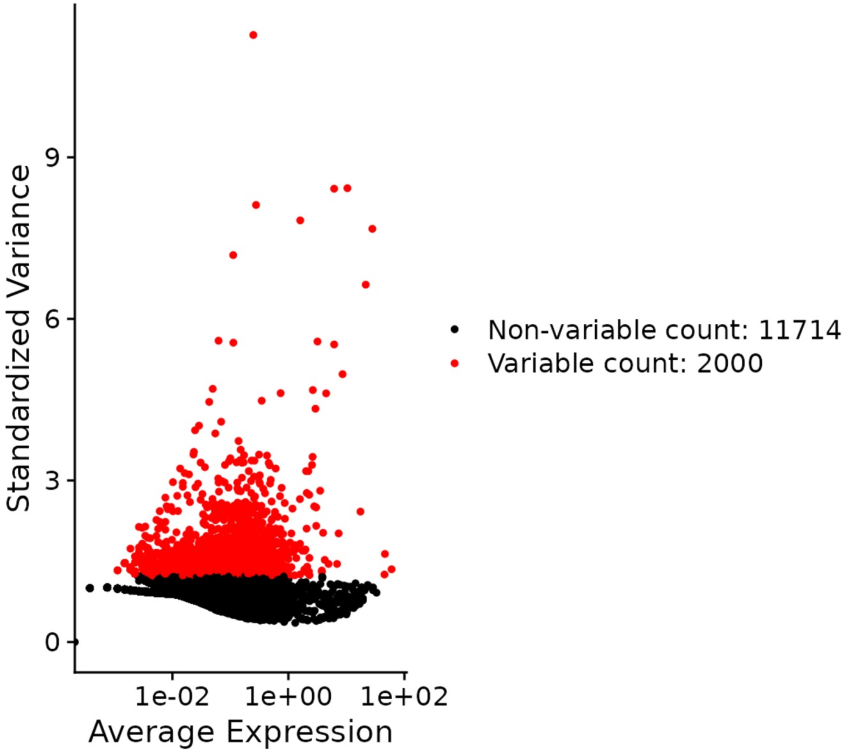
Filtering of cells QC

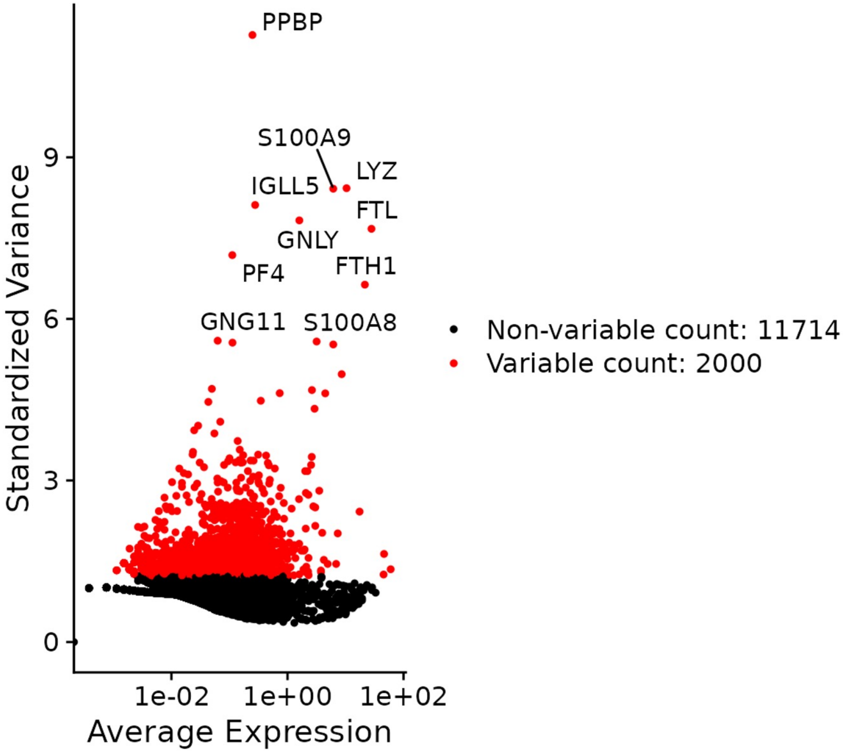
- Number of genes
- Number of reads
- % of mitochondrial RNA
    - Sign of dying cells



- Example filtering:
    - filter cells that have unique feature counts over 2,500 or less than 200
    - filter cells that have >5% mitochondrial counts

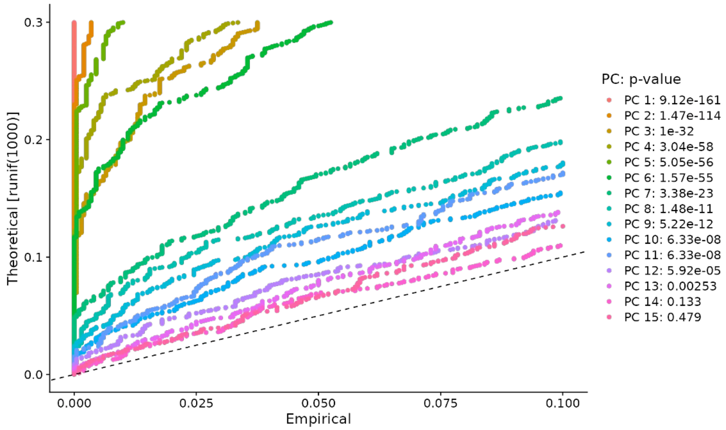
Identification of highly variable features

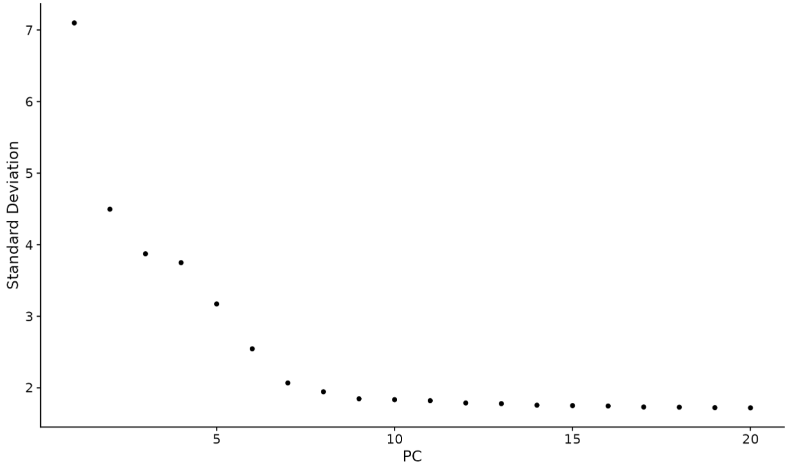


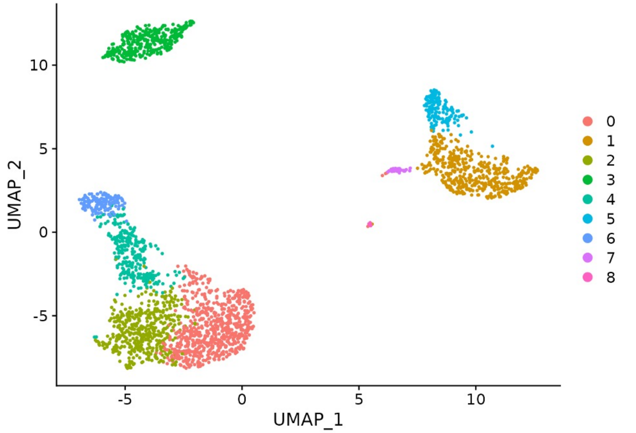


Determine the ‘dimensionality’ of the dataset

- How many clusters (cell types) is in the experiment (we want to explore)
- Analysis of PCA results



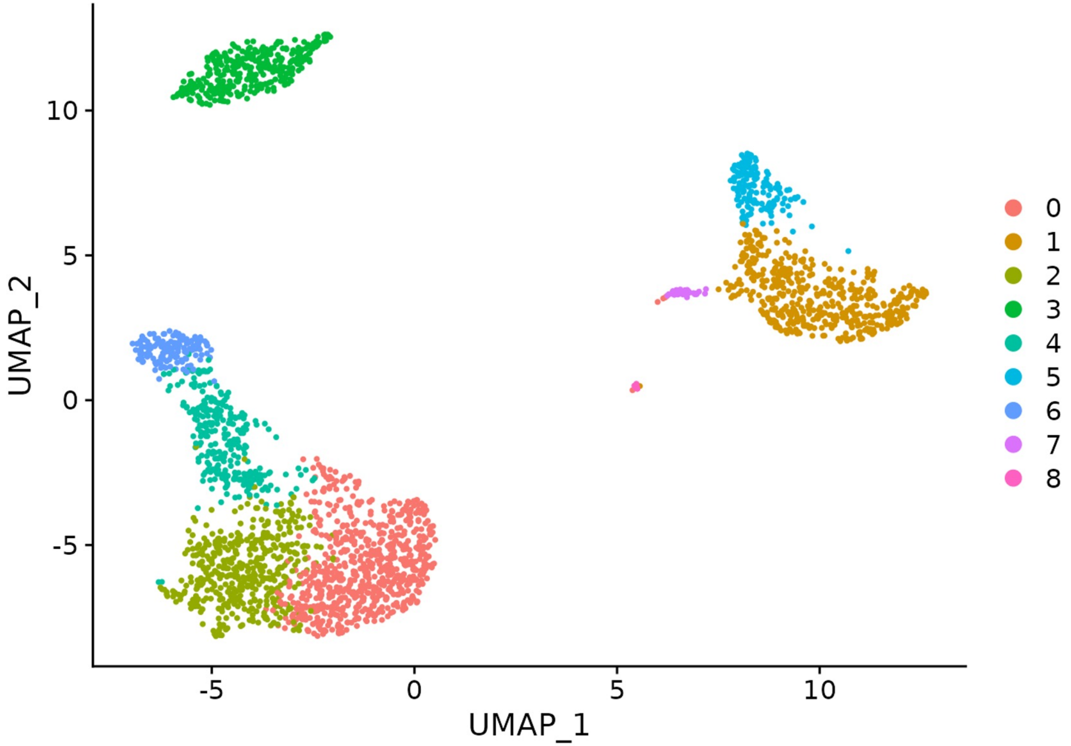




- Better to choose higher value then smaller

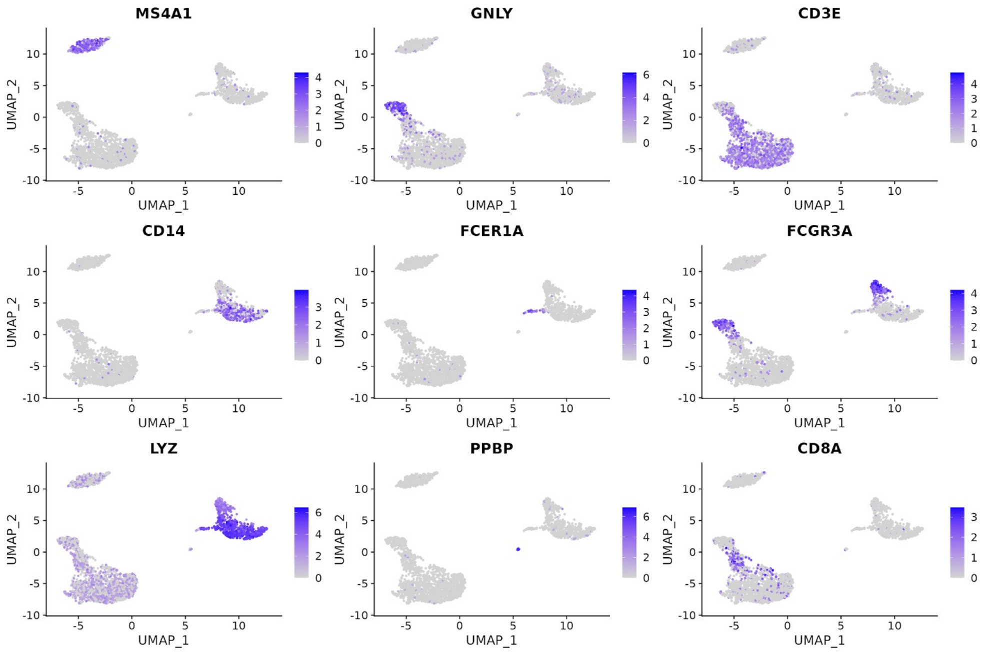
Cluster the cells

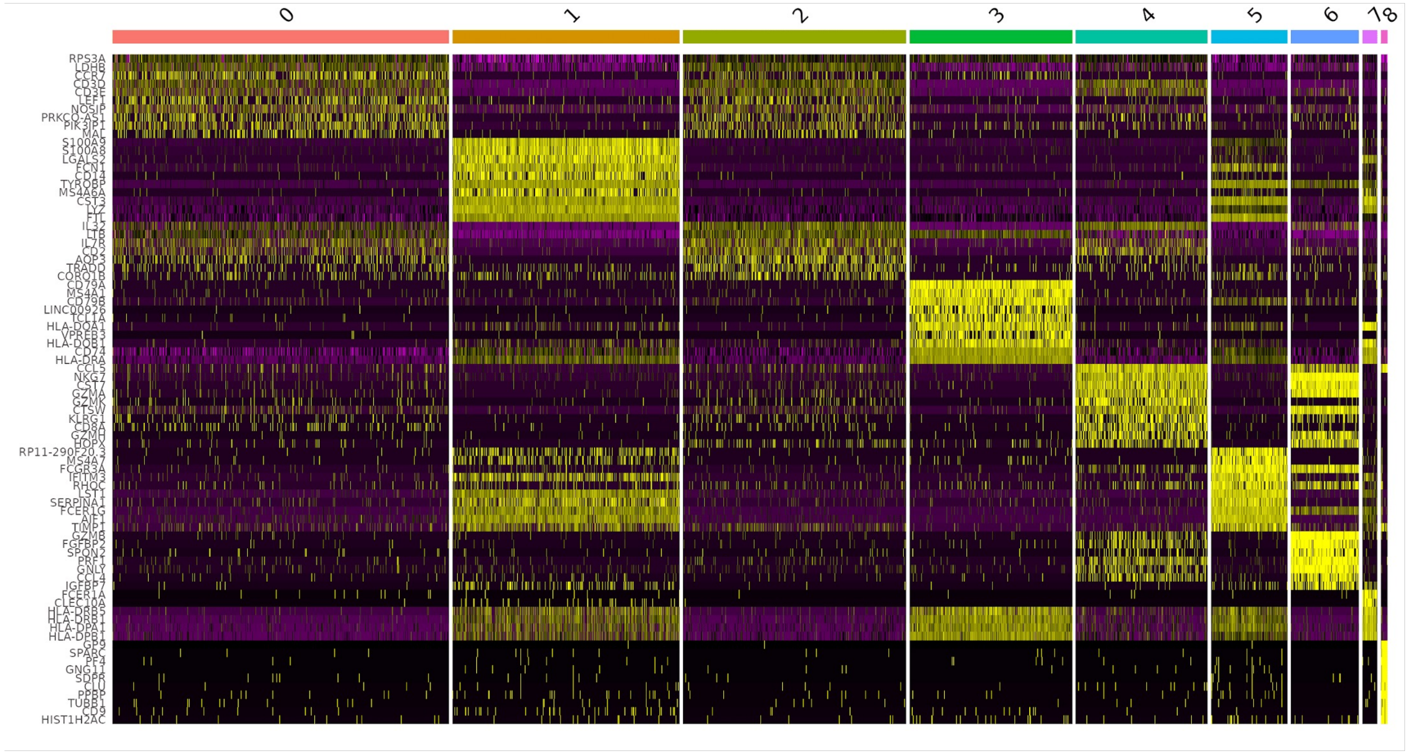
- graph-based algorithms
- UMAP/tSNE visualization



Analysing clusters

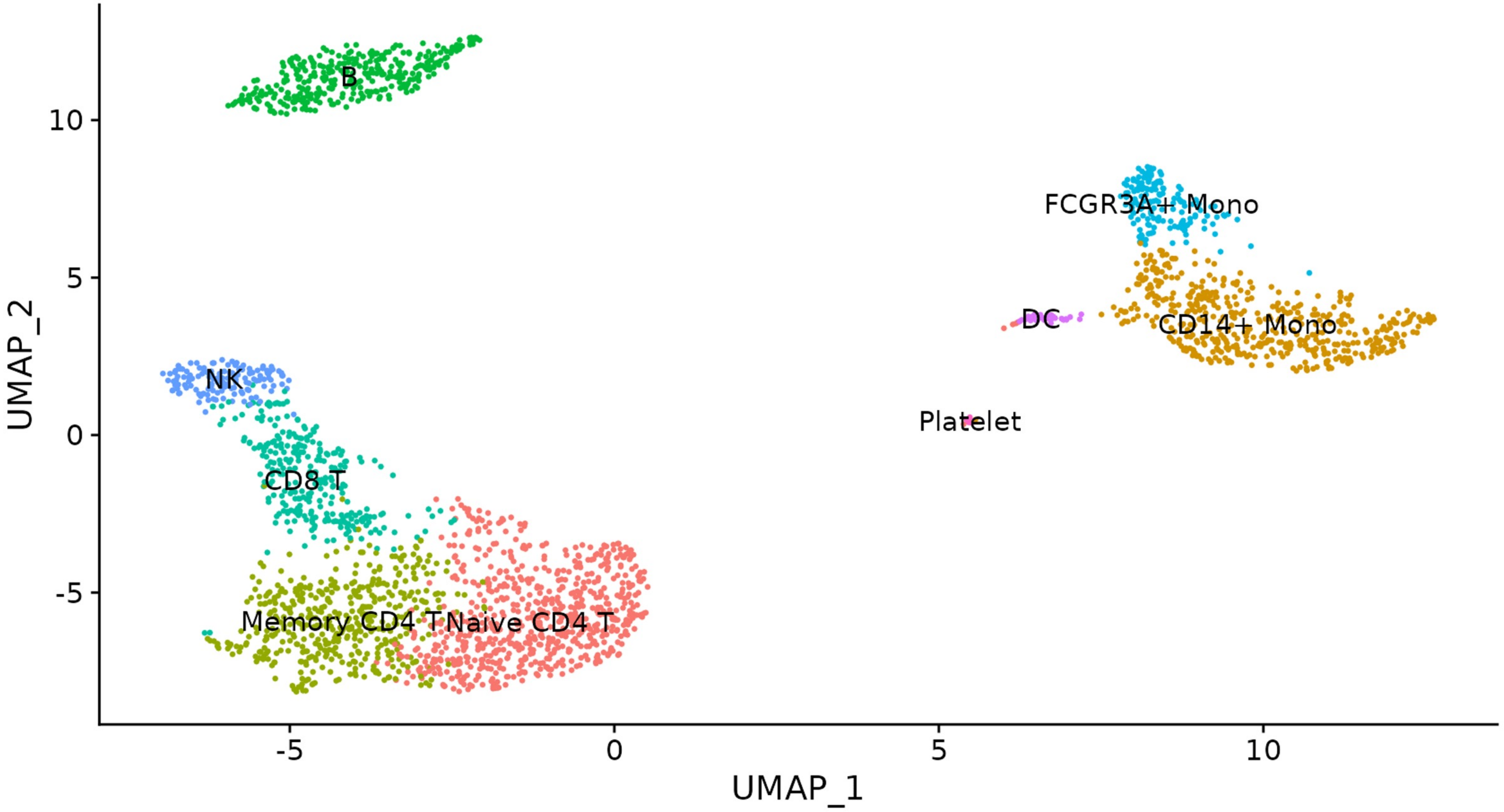
- Finding differentially expressed features (cluster biomarkers)





Assigning cell type identity to clusters

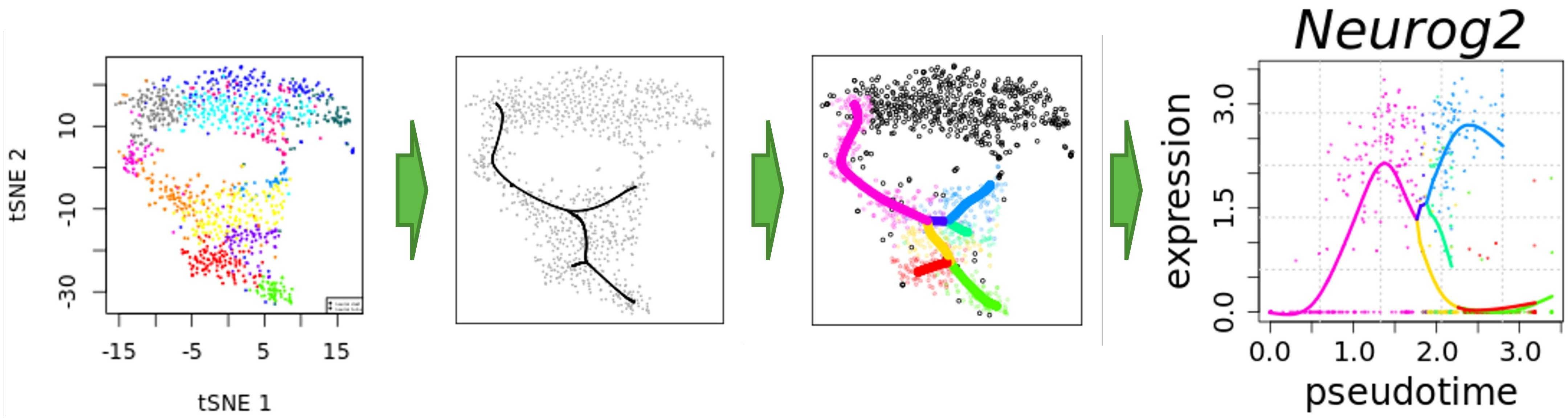
Cluster ID	Markers	Cell Type
0	IL7R, CCR7	Naive CD4+ T
1	CD14, LYZ	CD14+ Mono
2	IL7R, S100A4	Memory CD4+
3	MS4A1	B
4	CD8A	CD8+ T
5	FCGR3A, MS4A7	FCGR3A+ Mono
6	GNLY, NKG7	NK
7	FCER1A, CST3	DC
8	PPBP	Platelet



Advanced analysis

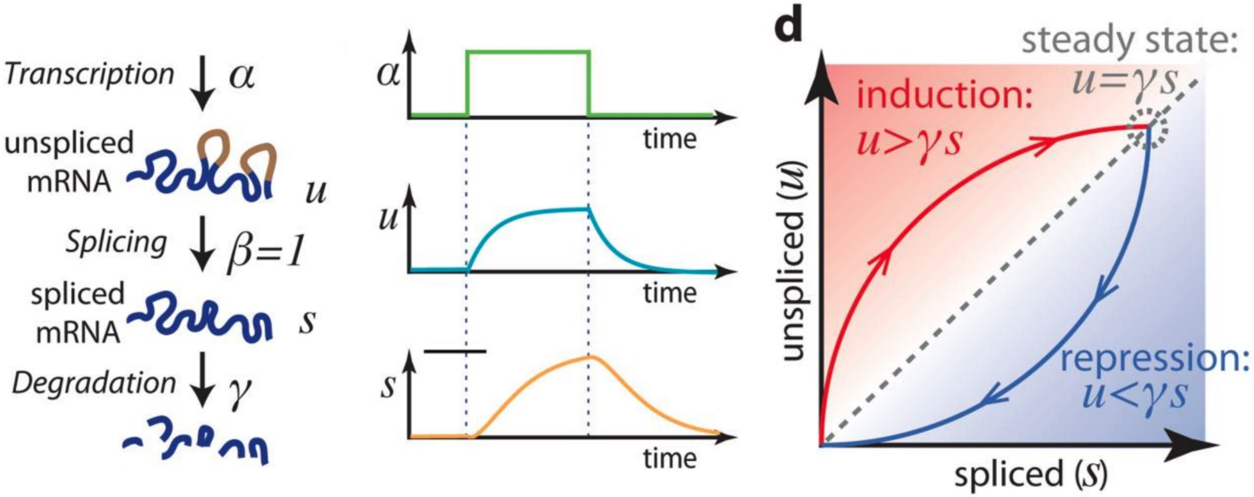
- Cell state/development directions
    - Pseudo-time
    - Velocity

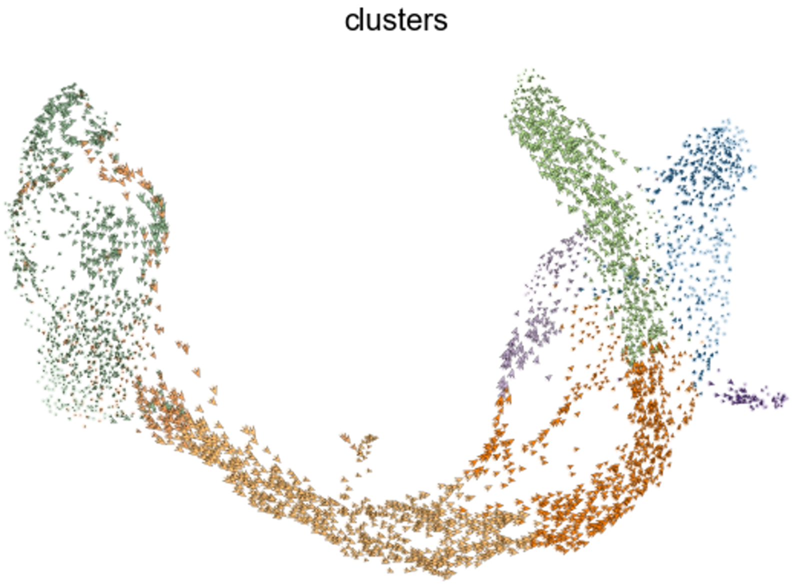
Pseudotime

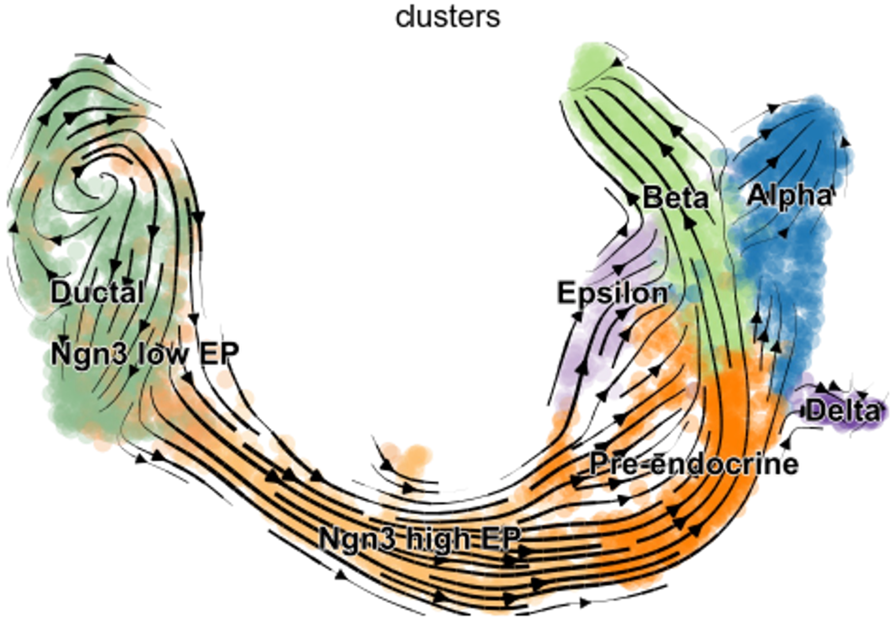


- does not provide information about directionality of dynamics

Velocity


Transcription
α
unspliced mRNA
u
Splicing
β=1
spliced mRNA
s
Degradation
γ
α
time
u
time
s
time
d
steady state:
u=γs
induction:
u>γs
repression:
u<γs
unspliced (u)
spliced (s)

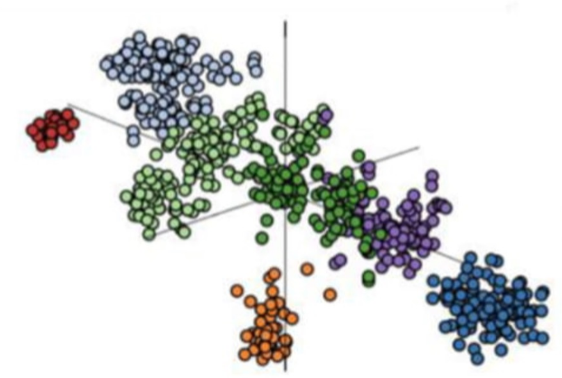

clusters


clusters
Ductal
Ngn3 low EP
Ngn3 high EP
Pre-endocrine
Epsilon
Beta
Alpha
Delta

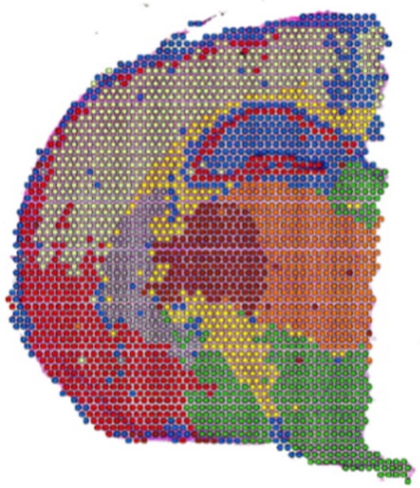
RNA-seq types

- Bulk RNA-seq

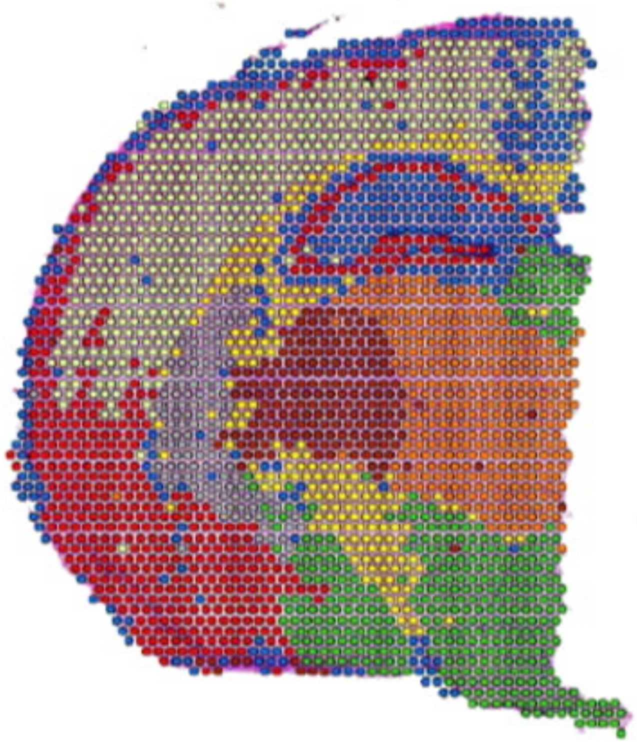
- SC-RNA-seq



- Spatially resolved RNA-seq



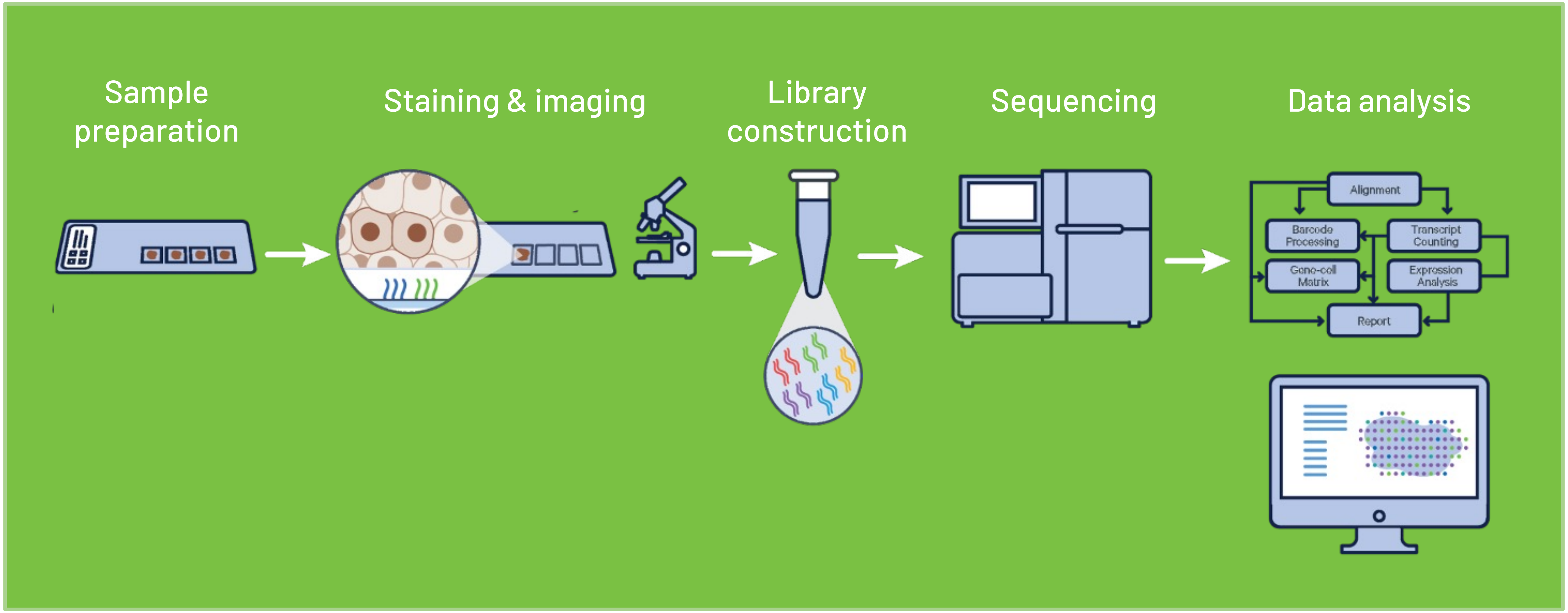
Spatial Transcriptomics



Spatially Resolved Gene Expression

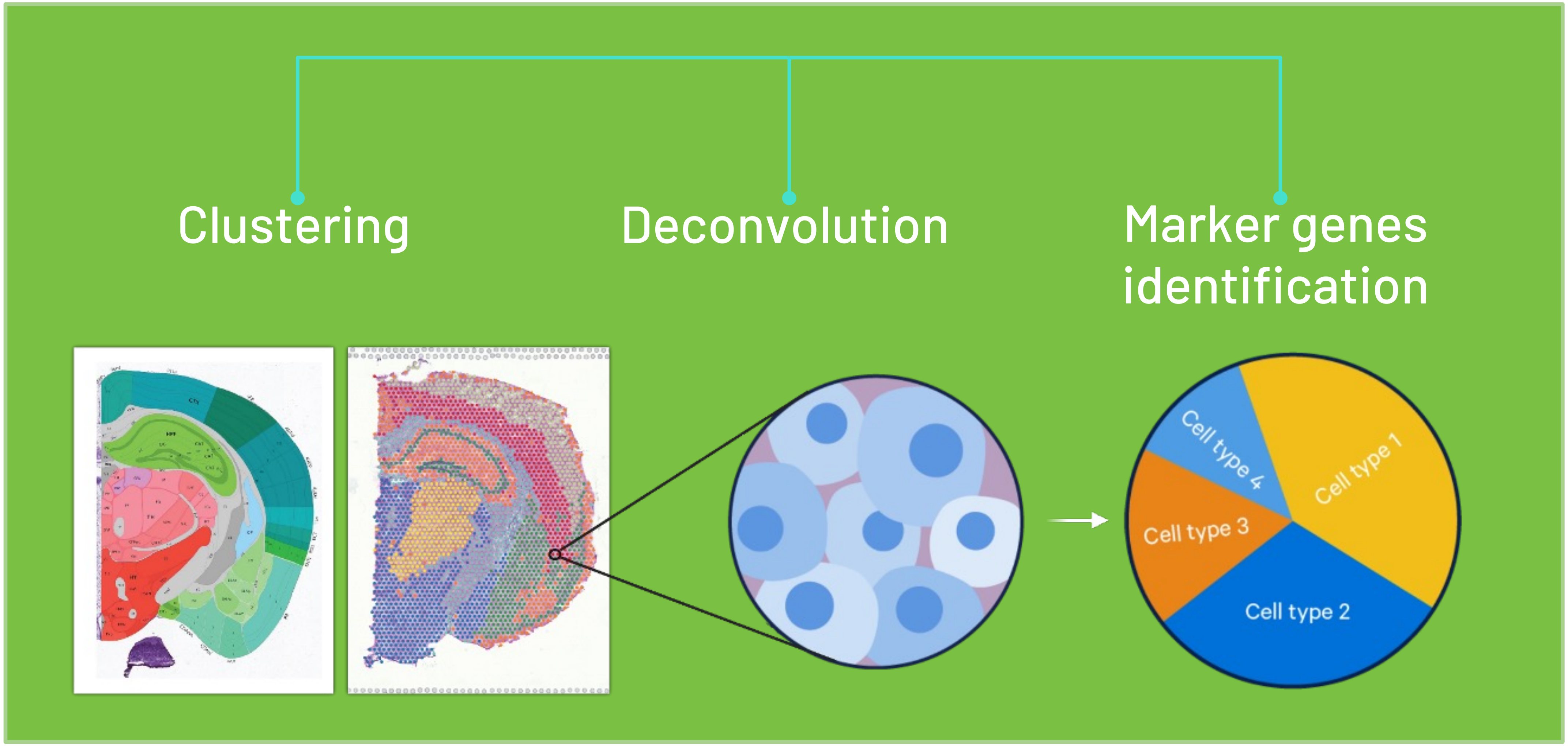
Tissue Section

Visium sequencing library



Sample preparation → Staining & imaging → Library construction → Sequencing → Data analysis

Visium spatial data analyses

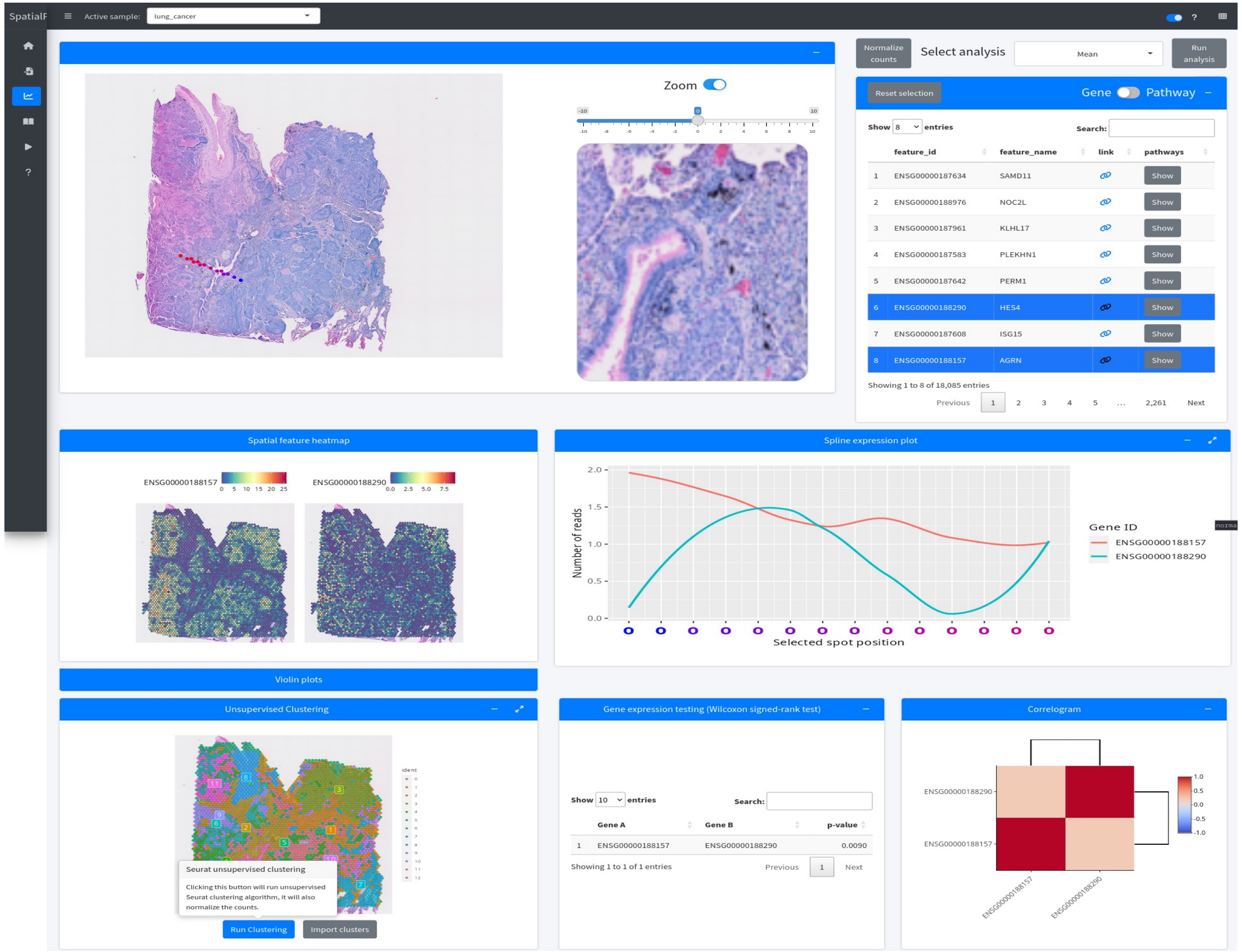


Clustering

Deconvolution

Marker genes identification

Spatial Transcriptomics – Spatial Path



Active sample: lung_cancer

Select analysis: Mean

Show 8 entries

Search:

	feature_id	feature_name	link	pathways
1	ENSG00000187634	SAMD11	🔗	Show
2	ENSG00000188976	NOC2L	🔗	Show
3	ENSG00000187961	KLHL17	🔗	Show
4	ENSG00000187583	PLEKHN1	🔗	Show
5	ENSG00000187642	PERM1	🔗	Show
6	ENSG00000188290	HES4	🔗	Show
7	ENSG00000187608	ISG15	🔗	Show
8	ENSG00000188157	AGRN	🔗	Show

Showing 1 to 8 of 18,085 entries

Previous 1 2 3 4 5 … 2,261 Next

Seurat unsupervised clustering

Clicking this button will run unsupervised Seurat clustering algorithm, it will also normalize the counts.

Show 10 entries

Search:

	Gene A	Gene B	p-value
1	ENSG00000188157	ENSG00000188290	0.0090

Showing 1 to 1 of 1 entries

Previous 1 Next

Vojtech Macala
http://cfb.ceitec.muni.cz/spatialpath/

CEITEC

@CEITEC_Brno

Thank you for your attention!

Vojta Bystry
vojtech.bystry@ceitec.muni.cz

CEITEC

www.ceitec.eu