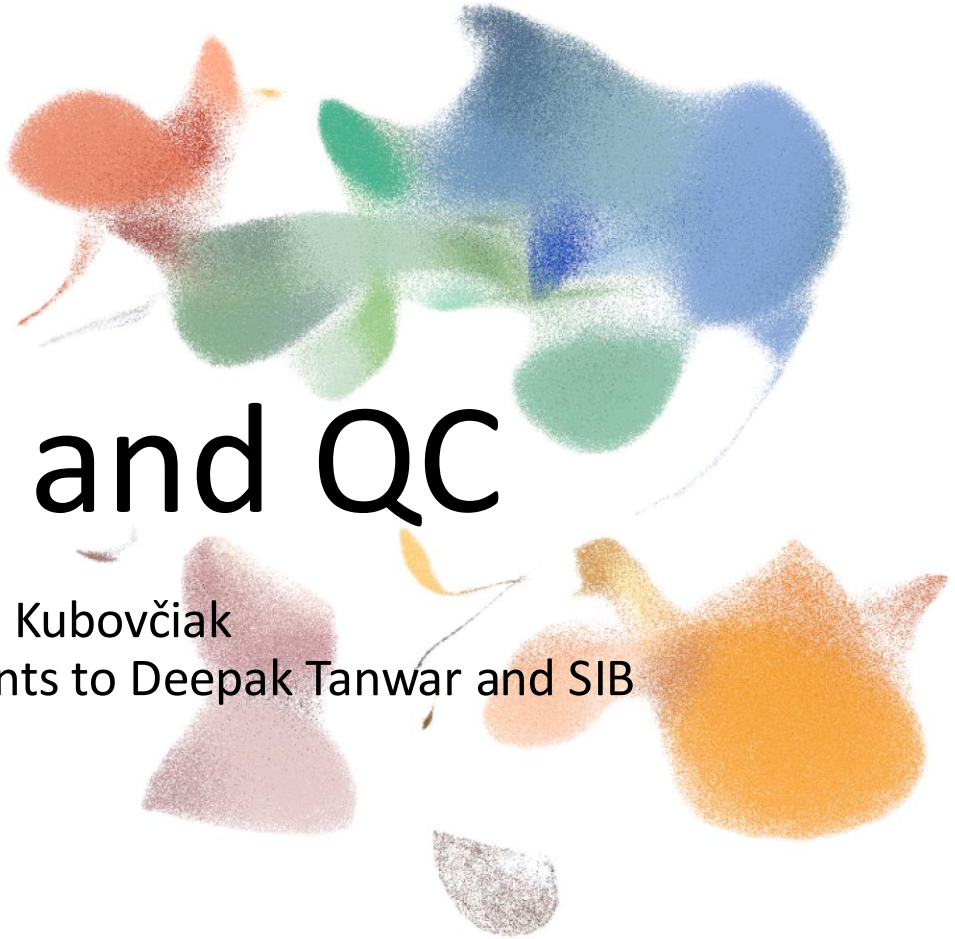


Tools and QC

Jan Kubovčiak
with acknowledgements to Deepak Tanwar and SIB

elixir CZECH REPUBLIC

IMG

cells ->

genes ->

	ATAC-1	CCG-1	GCGA-1
RPL22	5	13	3
PARK7	0	9	3
ENO1	1	35	0
PLA2G2A	0	0	0
CAMK2N1	0	6	0
CDC42	0	8	1
C1QA	0	0	25
C1QC	0	0	25
C1QB	0	0	29
ID3	0	35	0
RPL11	18	29	16
CLIC4	0	4	0

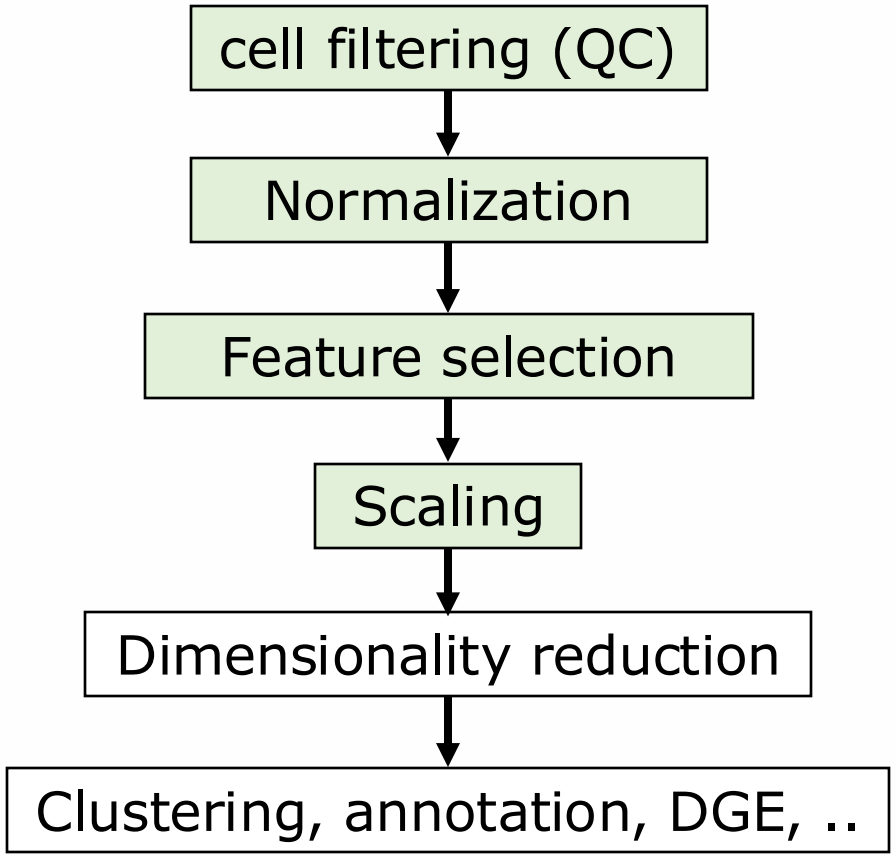
Frequently used analysis tools

- Major toolkits perform (at least) the following:
  - QC
  - normalization & scaling
  - dimensionality reduction
- `Seurat` (R, CRAN)
- `Bioconductor` (scater + scran etc, R)
- `scanpy` (python)

https://bioconductor.org/books/3.22/OSCA

IMG

Analysis overview



`cellranger + Seurat`

`Seurat::NormalizeData()`

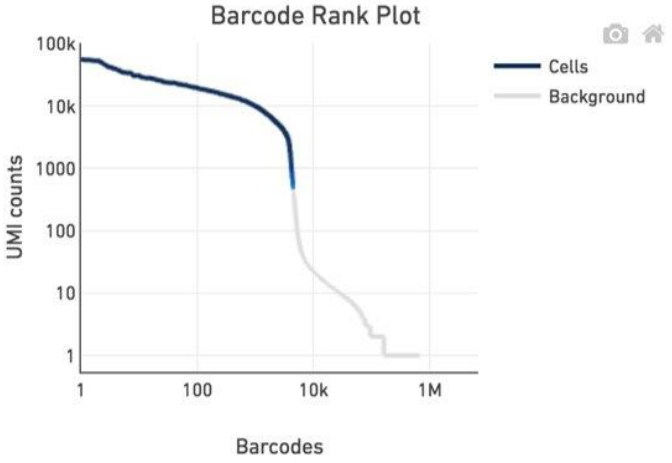
`Seurat::FindVariableFeatures()`

`Seurat::ScaleData()`

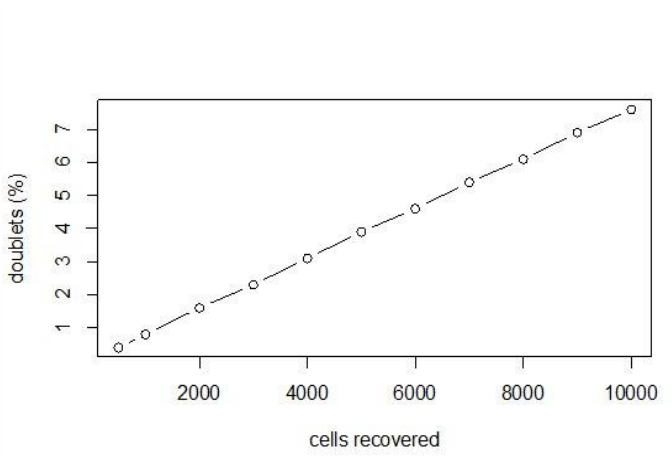
`Seurat`

`Seurat, SingleR, clusterProfiler, ..`

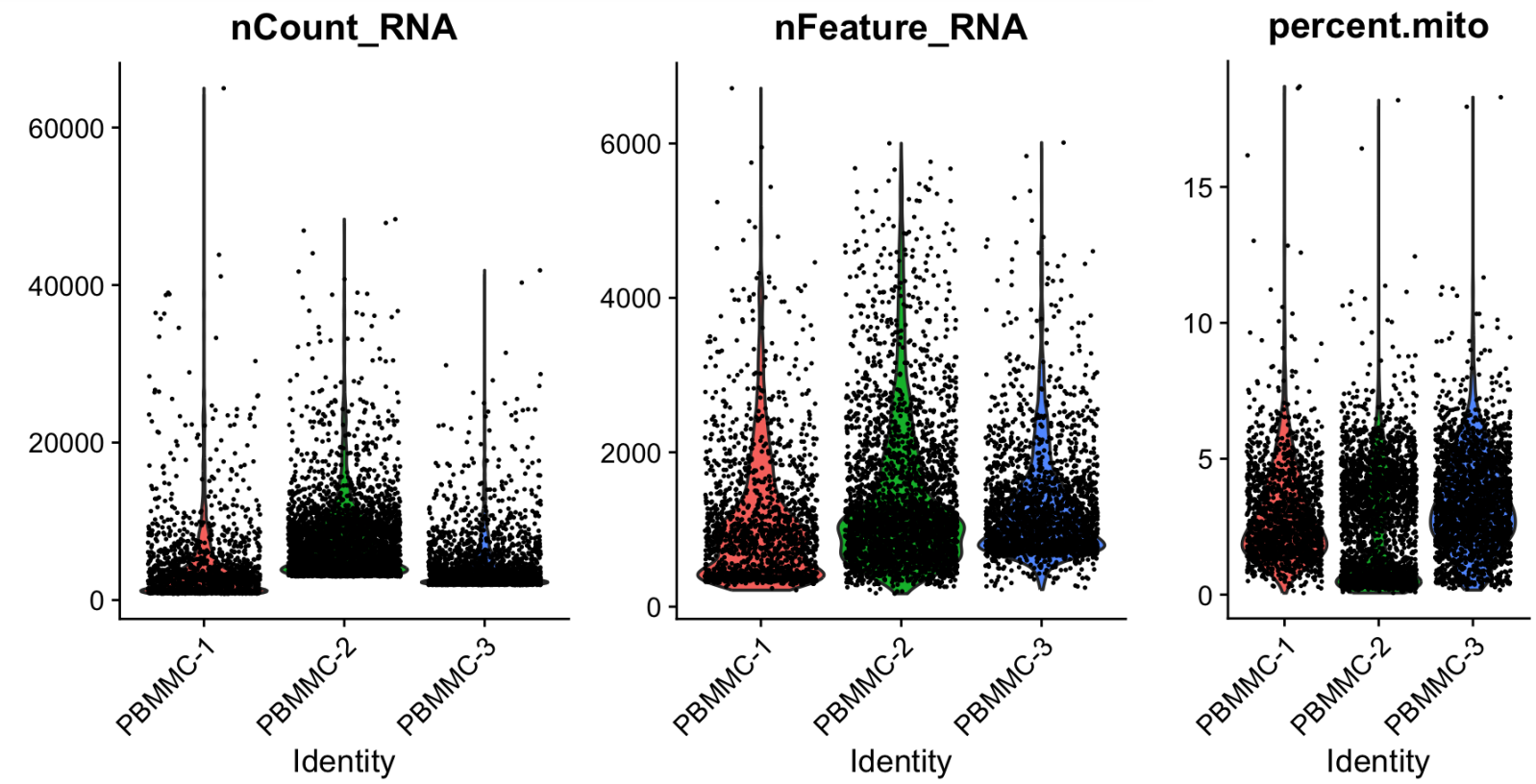
Cell filtering



- Cellranger:
  - cell calling (filter against low #UMI)
  - filtered vs raw matrix
- QC metrics to consider:
  - total expression
  - detected genes count
  - % mitochondrial UMI: dying cells
  - % globin or ribosomal UMI
  - Relationships between variables



Cell filtering

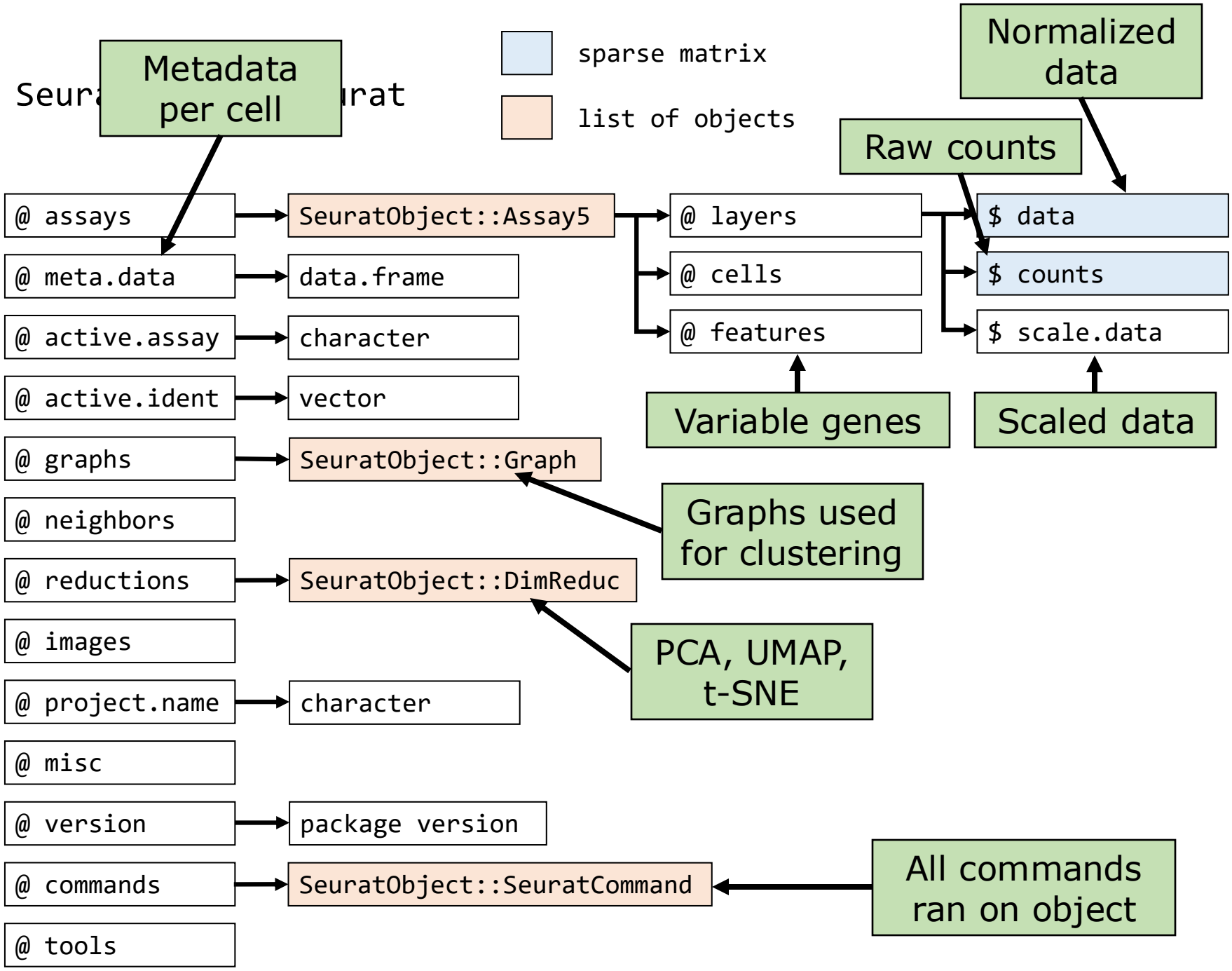


Think about thresholds for each experiment

Cell filtering

- Dataset-sensitive filtering
  - scuttle::perCellQCFilters() - MAD approach
  - assumes high quality data
  - observe the results

Gene filtering?


Seura rat
Metadata per cell
sparse matrix
list of objects
Normalized data
Raw counts
@ assays
SeuratObject::Assay5
@ layers
$ data
@ meta.data
data.frame
@ cells
$ counts
@ active.assay
character
@ features
$ scale.data
@ active.ident
vector
Variable genes
Scaled data
@ graphs
SeuratObject::Graph
@ neighbors
Graphs used for clustering
@ reductions
SeuratObject::DimReduc
@ images
PCA, UMAP, t-SNE
@ project.name
character
@ misc
@ version
package version
@ commands
SeuratObject::SeuratCommand
All commands ran on object
@ tools

Thanks for your attention!