Single cell transcriptomics

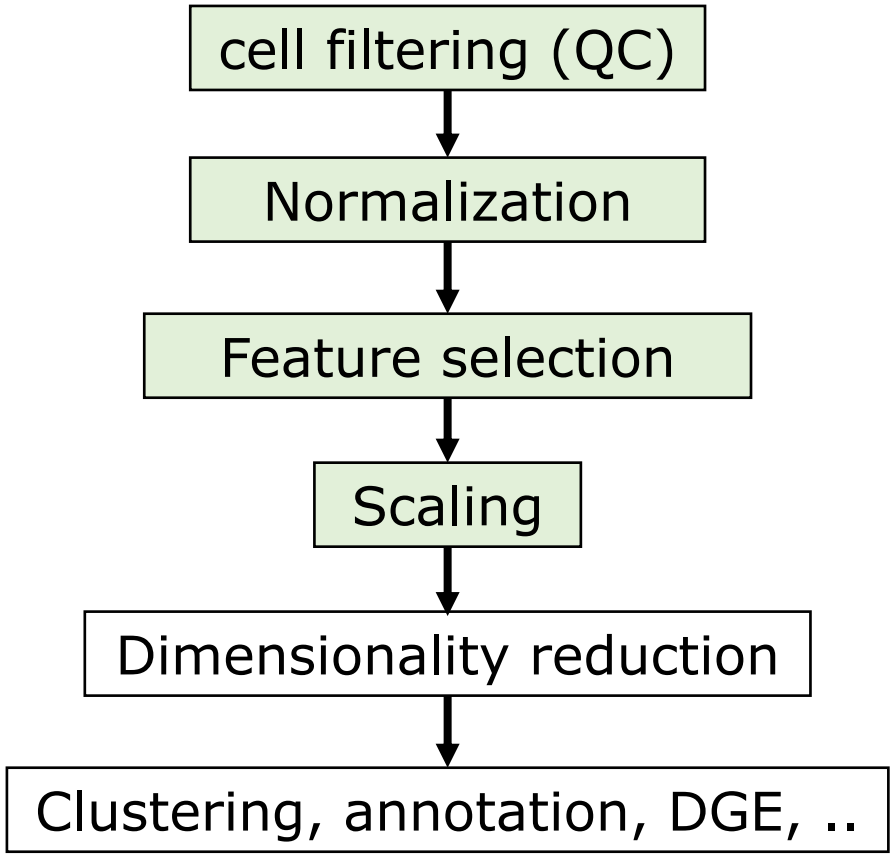
Quality control

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

Downstream analysis



`cellranger + Seurat`

`Seurat::NormalizeData()`

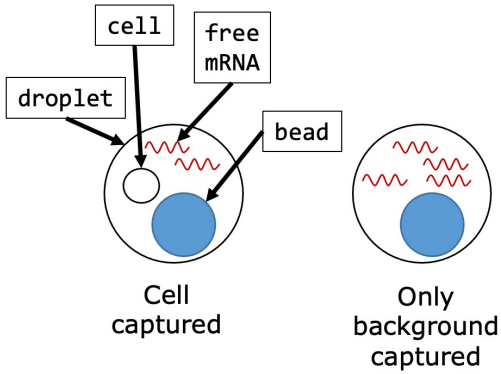
`Seurat::FindVariableFeatures()`

`Seurat::ScaleData()`

`Seurat`

`Seurat, SingleR, clusterProfiler, ..`

Cell filtering

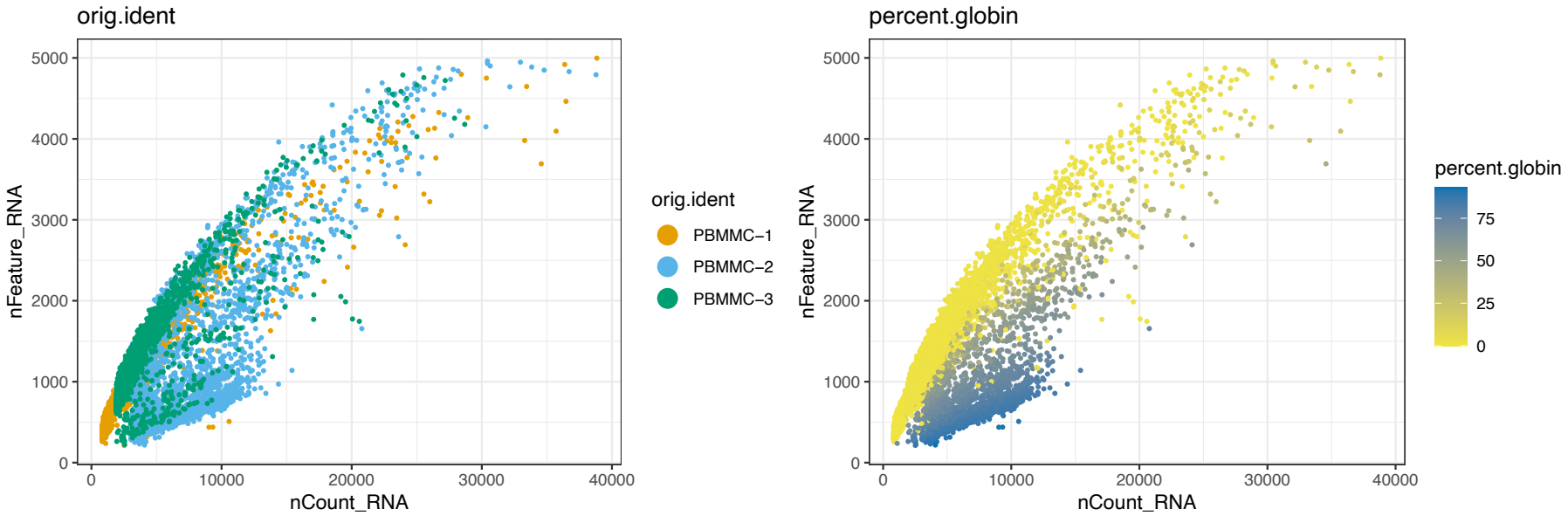


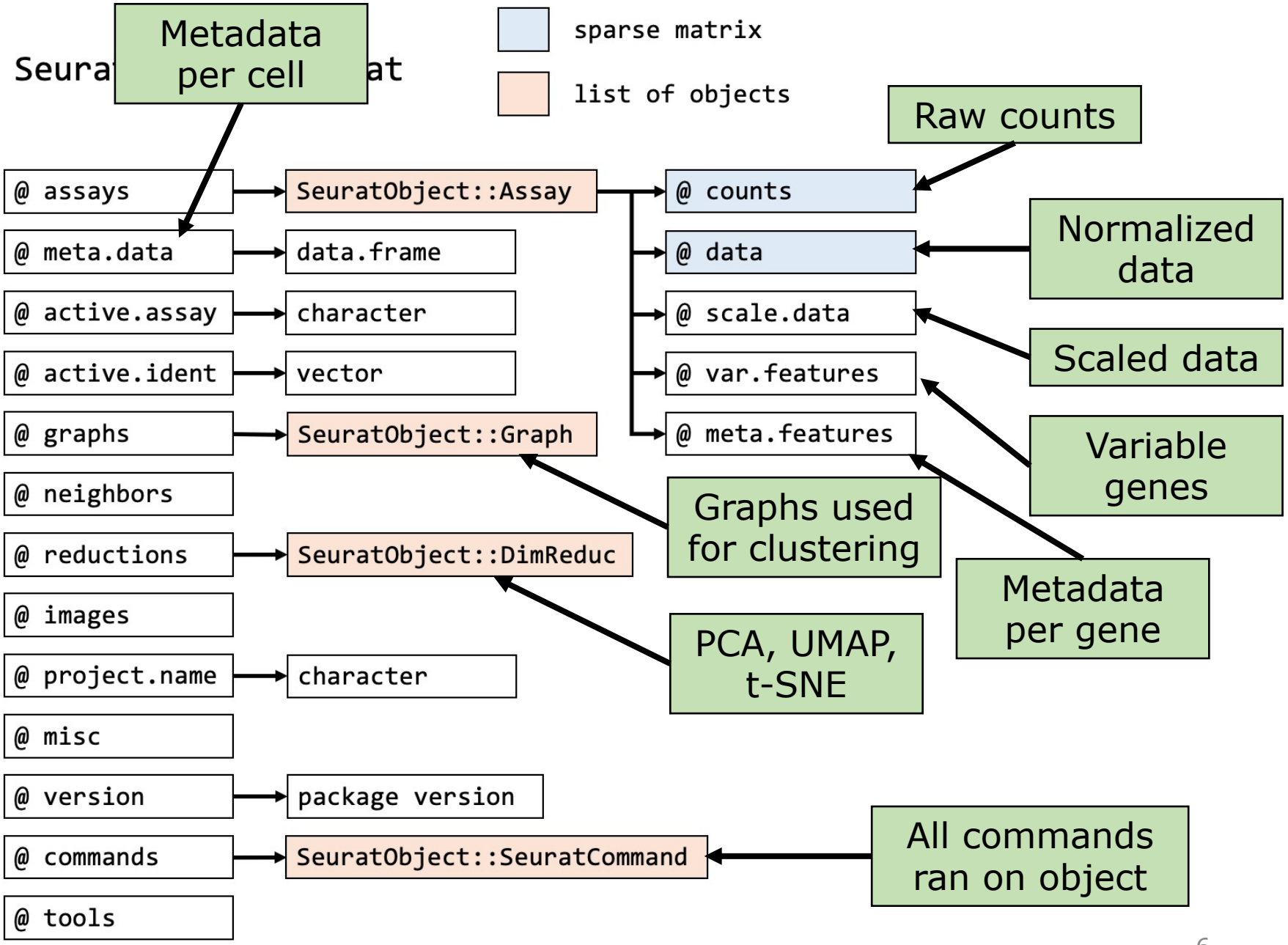
- Cellranger:
  - cell calling (filter against low #UMI)

- Manually (e.g. with Seurat):
  - #UMI: high -> possible doublet
  - #detected genes
  - % mitochondrial UMI: dying cells
  - % ribosomal UMI
  - % globin UMI
  - Relationships between variables

Cell filtering

Often it makes sense to look at relationships




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

Normalization & scaling

- **Normalization** (per cell): remove technical effects (i.e. library size)
- **Scaling** (per gene): standardize range, mean and variance

Both are mainly for the purpose of dimensionality reduction