

Single cell transcriptomics

Analysis tools & QC

genes ->

cells ->

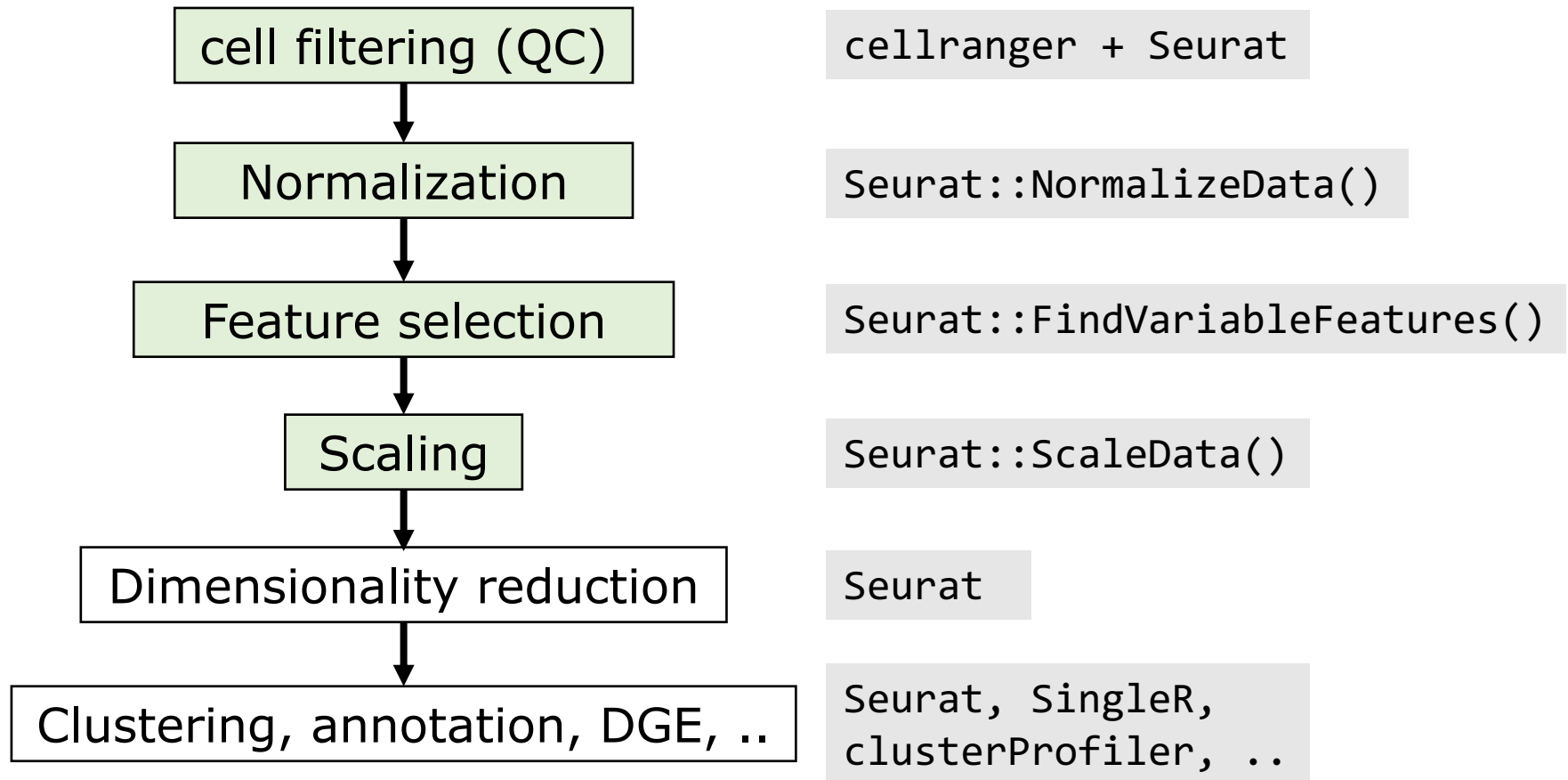
	ATAC-1	CCG-1	GCGA-1
RPL22	5	13	3
PARK7	0	9	3
ENO1	1	35	0
PLA2G2A	0	0	0
CAMK2N1	0	6	0
CDC42	0	8	1
C1QA	0	0	25
C1QC	0	0	25
C1QB	0	0	29
ID3	0	35	0
RPL11	18	29	16
CLIC4	0	4	0

Frequently used analysis tools

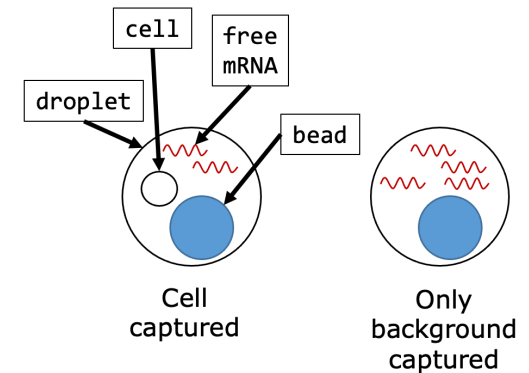
- Major tools perform (at least) the following:
 - QC
 - normalization & scaling
 - dimensionality reduction
- scanpy (python)
- scater + scan (R, Bioconductor)
- monocle3 (R, beta on github)
- Seurat (R, CRAN)



Analysis overview



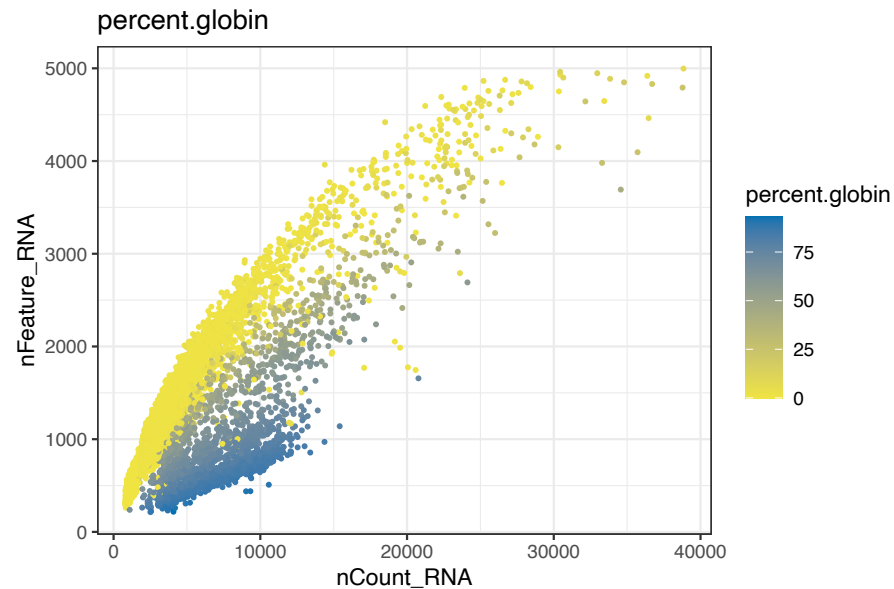
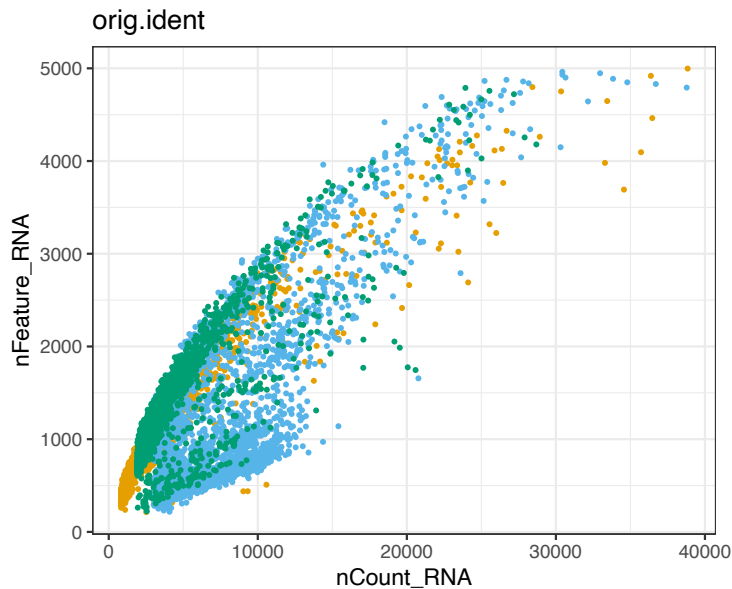
Cell filtering

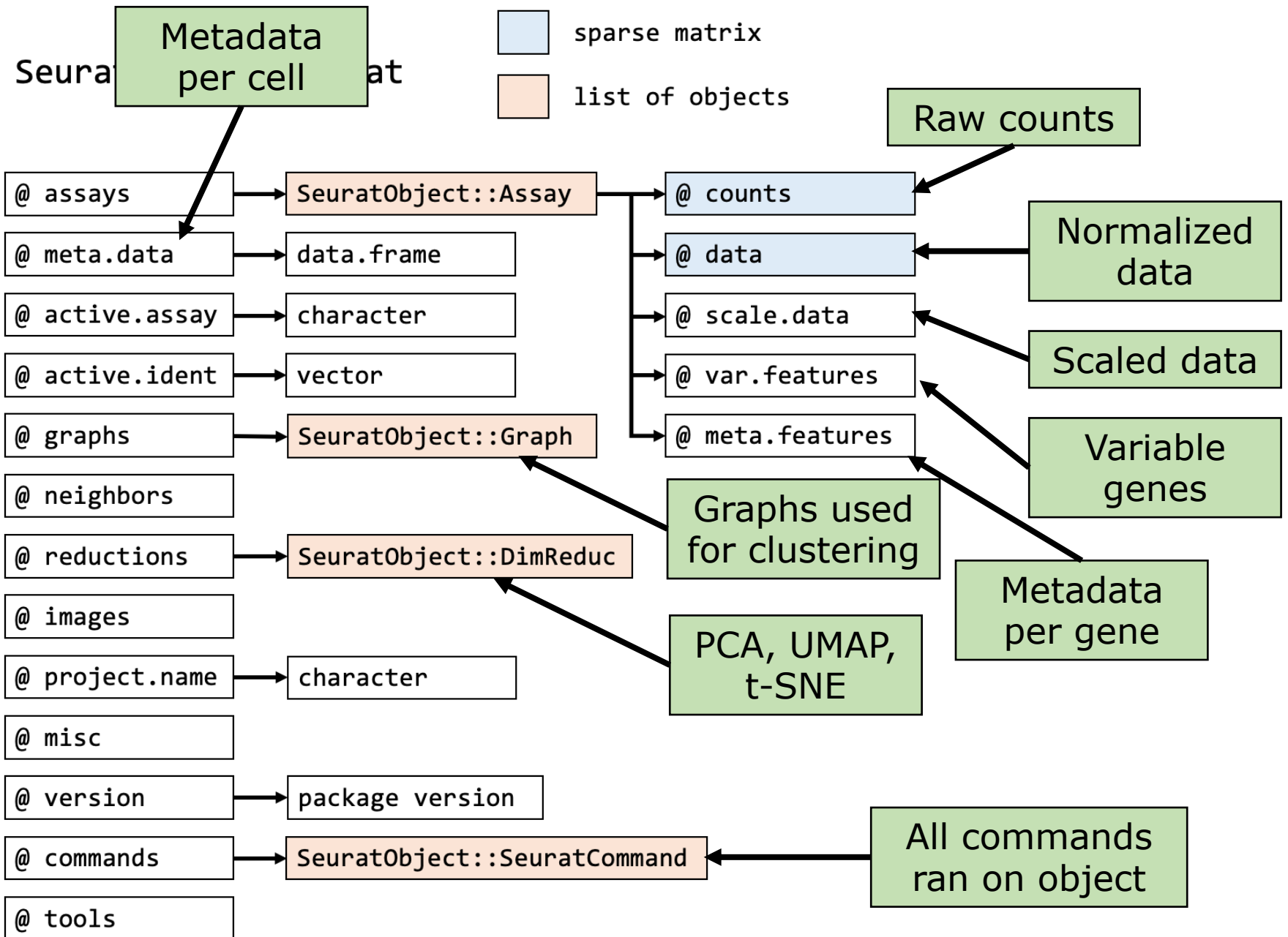


- Cellranger:
 - cell calling (filter against low #UMI)
- Manually (e.g. with Seurat):
 - #UMI: high -> possible doublet
 - #detected genes
 - % mitochondrial UMI: dying cells
 - % ribosomal UMI
 - % globin UMI
 - Relationships between variables

Cell filtering

Often it makes sense to look at relationships





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

Regress out variables

- Typically done after visualization with PCA/UMAP/t-SNE
- Embedding might be according to e.g. cell cycle phase -> but you want cell type

