What is single-cell RNA-Seq, and why is it useful?

SINGLE-CELL RNA-SEQ WITH BIOCONDUCTOR IN R

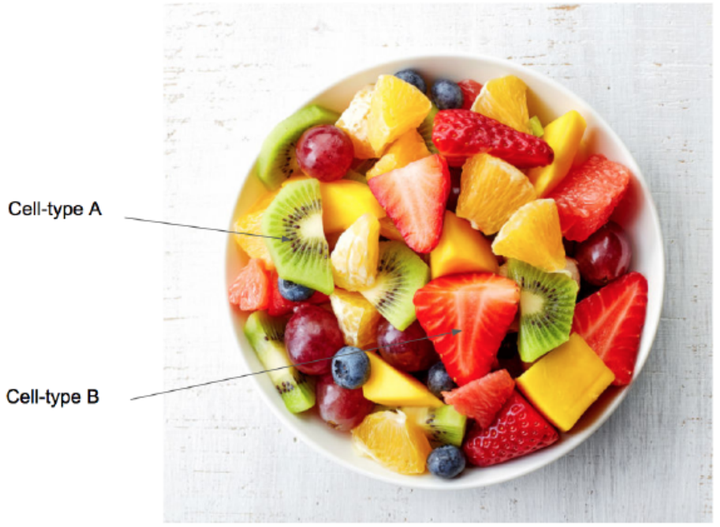
Fanny Perraudeau
Senior Data Scientist, Whole Biome

datacamp

Milkshake or fruit salad?.

scRNA-seq is capturing gene expression at the cellular level

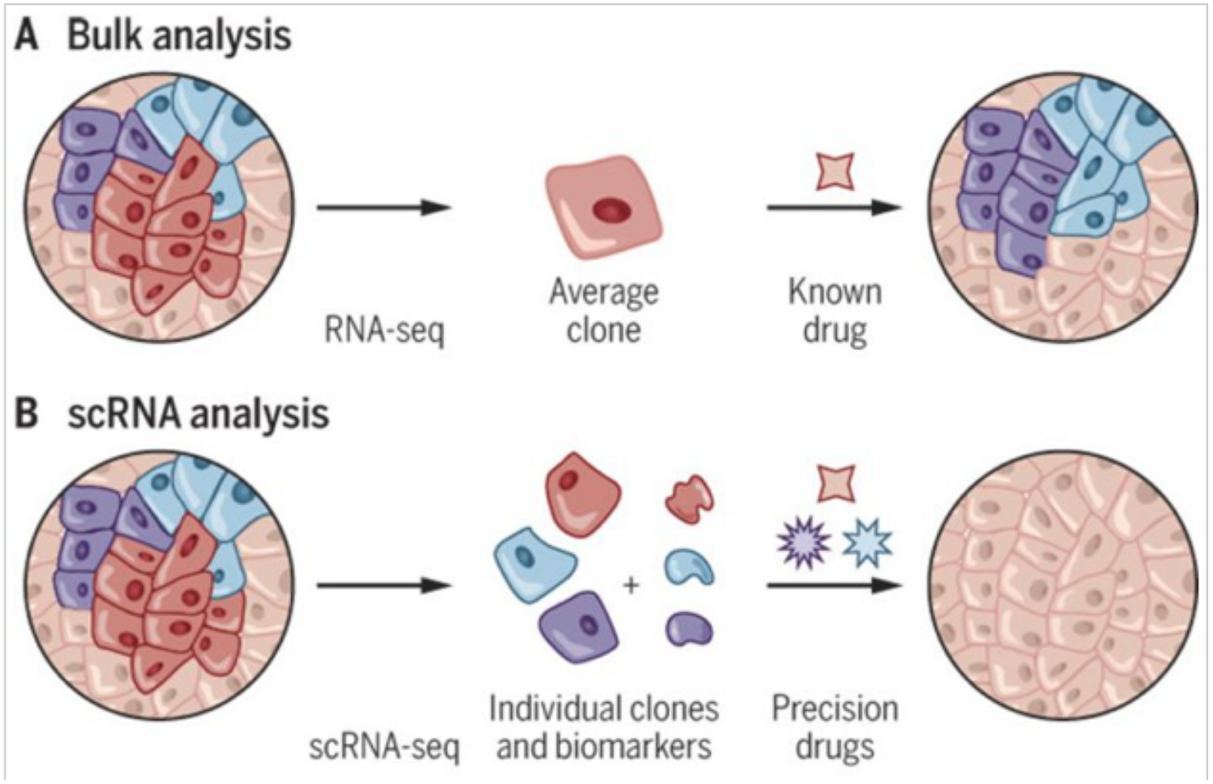
Single-cell



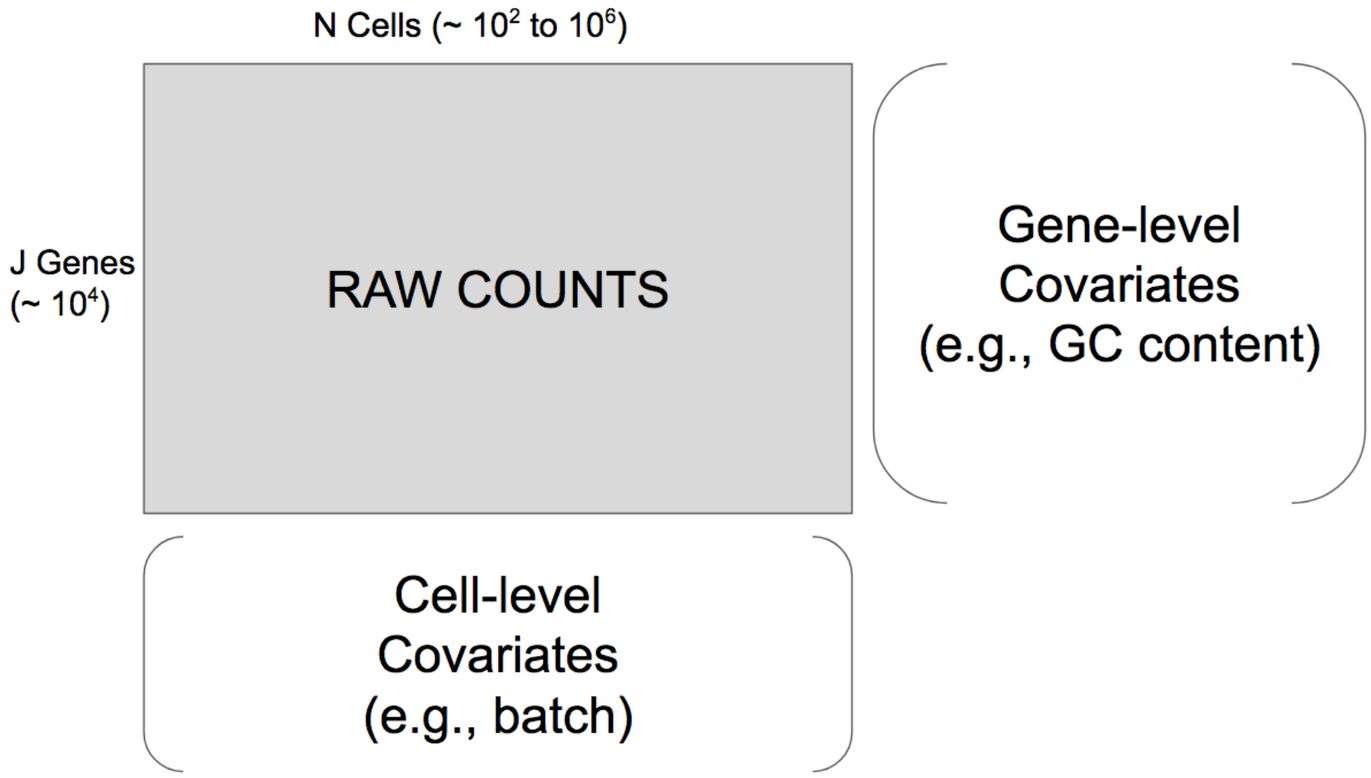
Bulk

Shalek and Regev (2016)

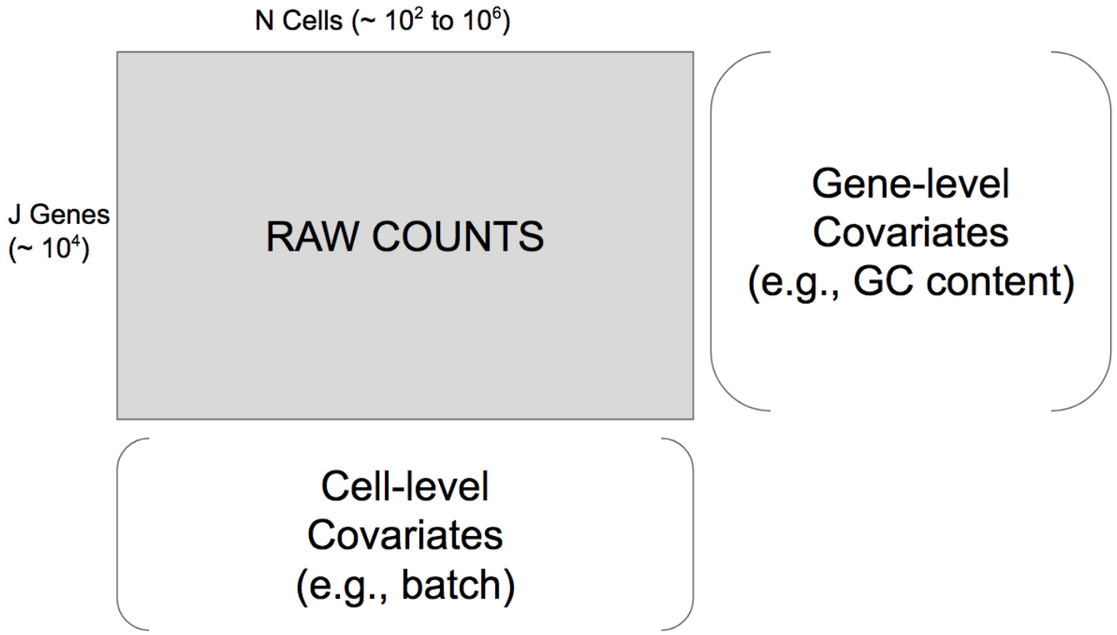
scRNA-Seq could revolutionize personalized medicine in cancer



Data structure



Zero inflation in single-cell transcriptome sequencing



- Biological zeros (e.g., cell cycle genes).
- Technical (false) zeros: dropouts.

Let's practice!

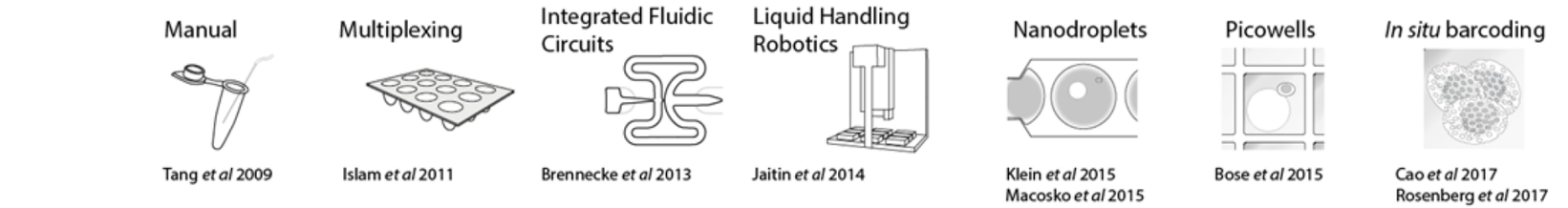
SINGLE-CELL RNA-SEQ WITH BIOCONDUCTOR IN R

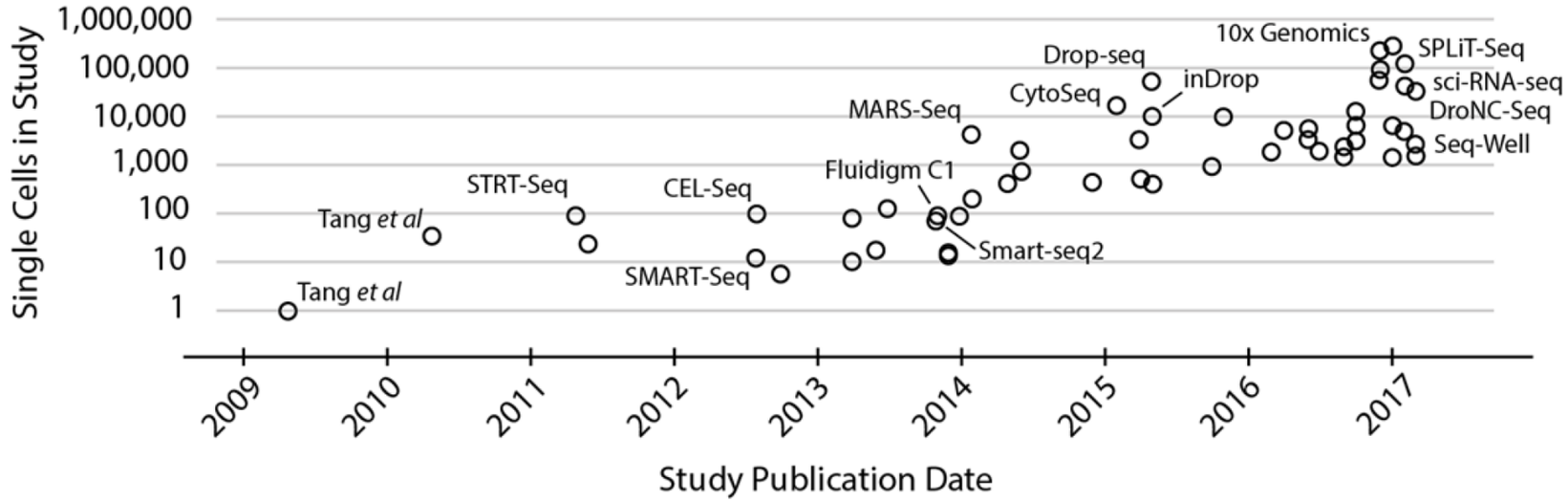
Typical workflow

SINGLE-CELL RNA-SEQ WITH BIOCONDUCTOR IN R

Fanny Perraudeau
Senior Data Scientist, Whole Biome

Exponential scaling of scRNA-Seq in the last decade





[1] "Exponential scaling of single cell RNAseq in the last decade". Valentine Svensson, Roser Vento [2] Tormo, Sarah A Teichmann

Aspects of scRNE-Seq methods

- **Quantification**: determines types of analyses
  - Full-length protocols -- uniform coverage of RNA seq
  - Tag-based protocols -- one of the ends of each RNA
- **Capture**: determines throughput
  - microwell-based
  - microfluidic-based
  - droplet-based

[1] https://hemberg [2] lab.github.io/scRNA.seq.course/introduction [3] to [4] single [5] cell [6] rna [7] seq.html

scRNE-seq workflow

1. Quality control
2. Normalization
3. Dimensionality reduction
4. Clustering
5. Differential expression analysis

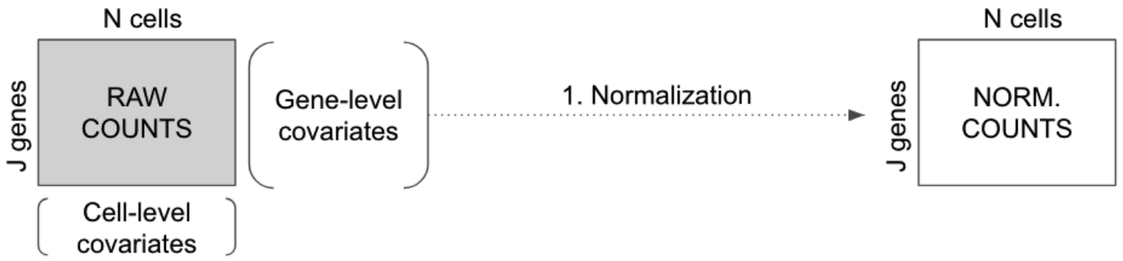
First step: quality control

Filter out low-quality cells:

- by library size: total number of reads aligned to each cell (a library refers to a cell)
- by cell coverage: average number of expressed genes in each cell

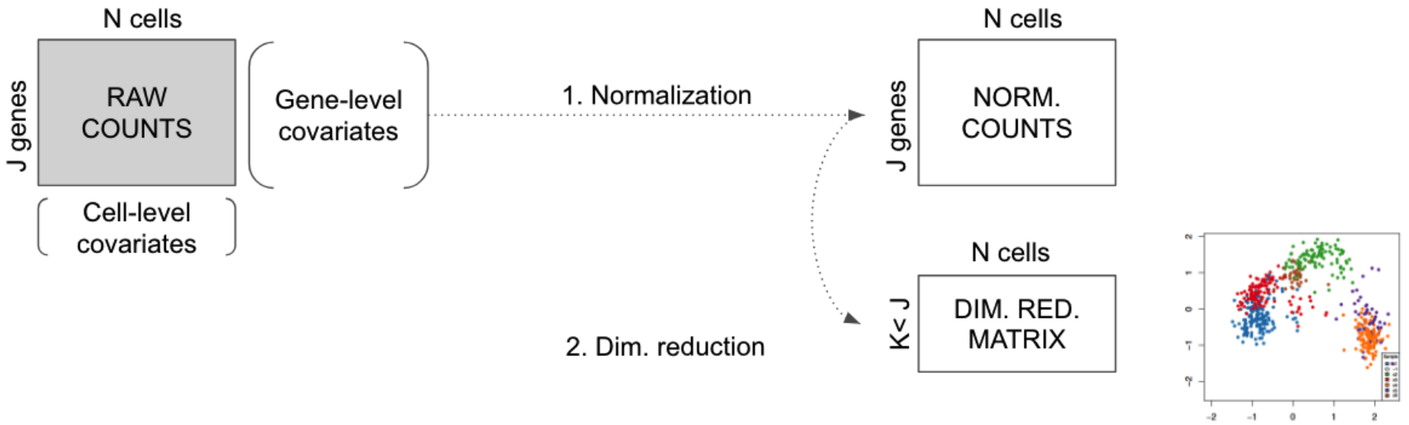
[1] "A step [2] by [3] step workflow for low [4] level analysis of single [5] cell RNA [6] seq data". Lun ATL, McCarthy DJ and Marioni JC

Typical workflow



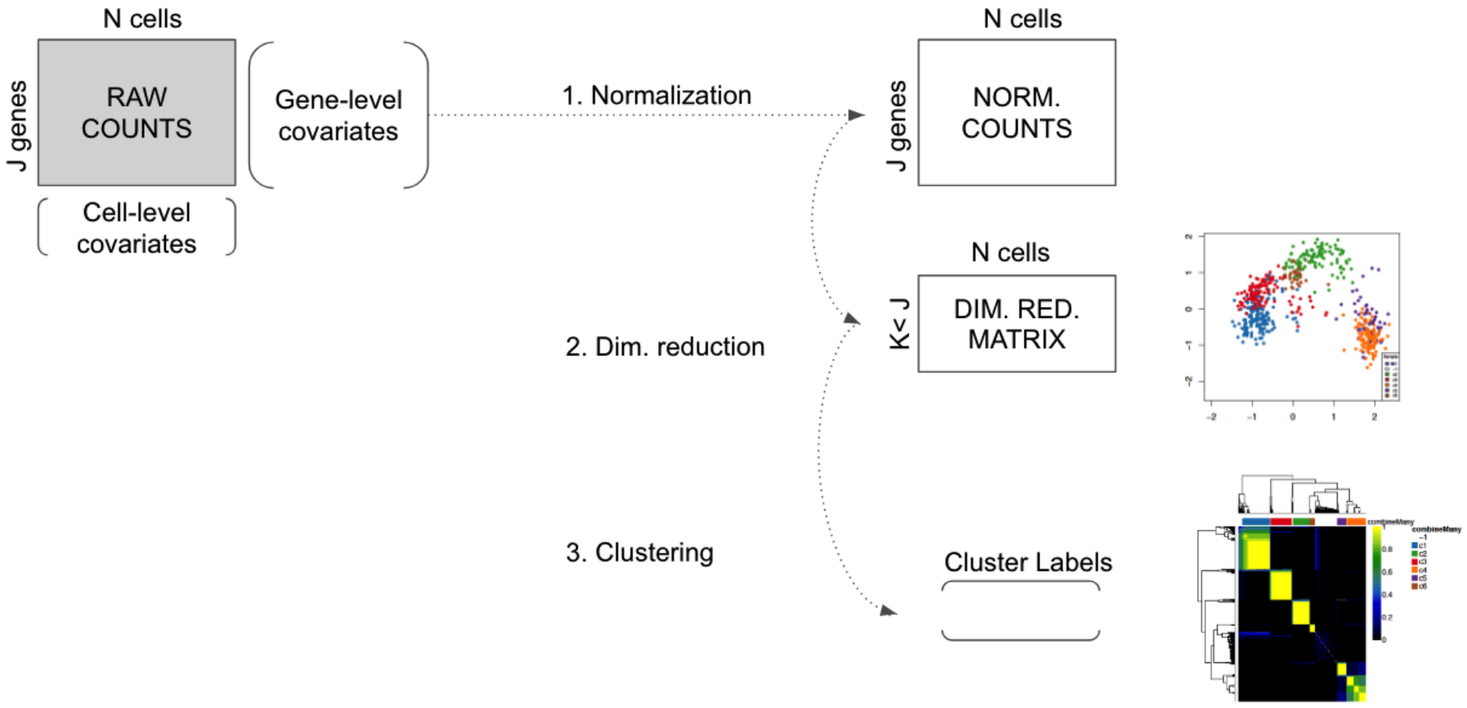
[1] "Bioconductor workflow for single [2] cell RNA sequencing". Perraudeau F, Risso D, Street K et al

Typical workflow



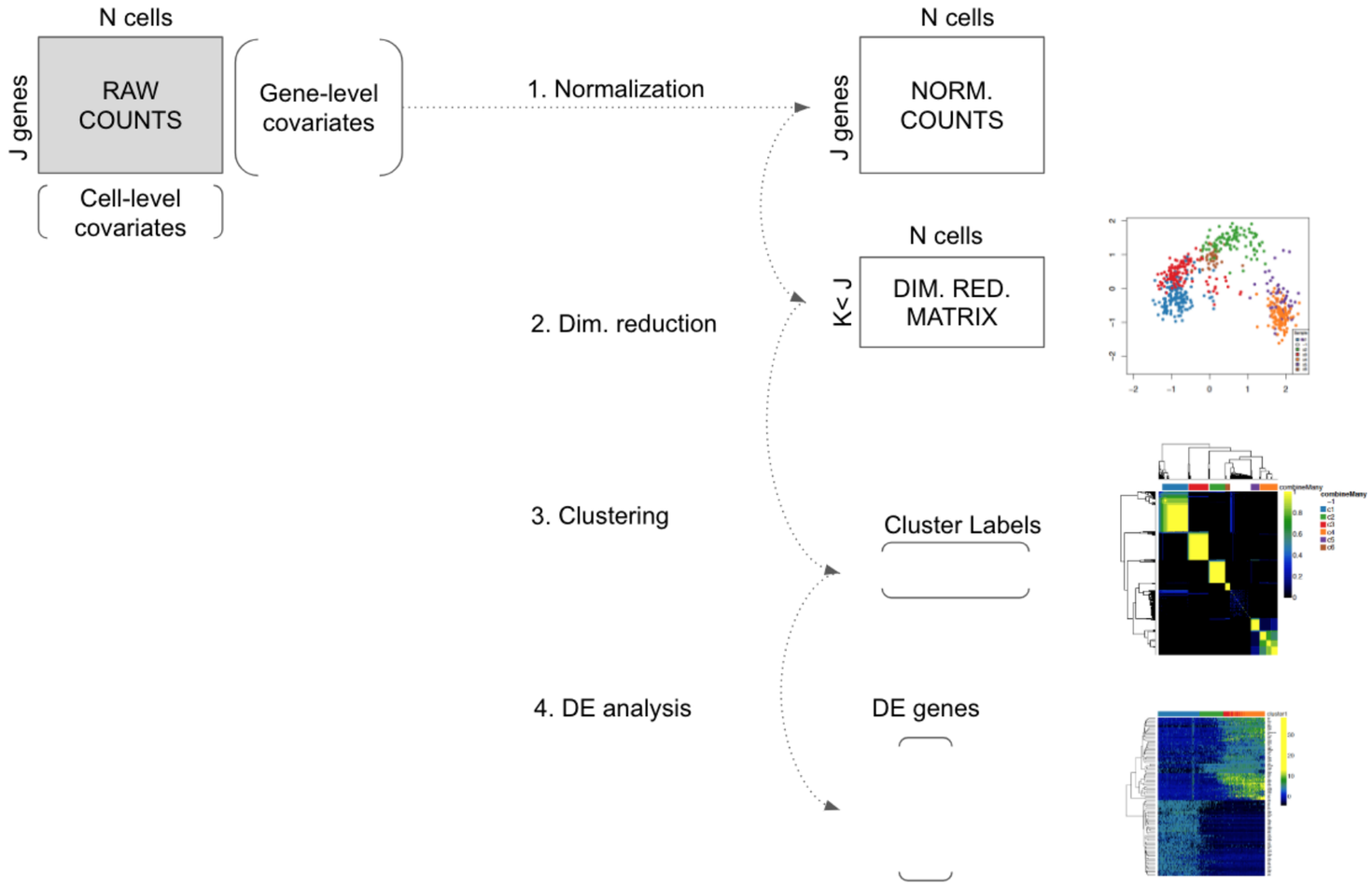
[1] "Bioconductor workflow for single [2] cell RNA sequencing". Perraudeau F, Risso D, Street K et al

Typical workflow



[1] "Bioconductor workflow for single [2] cell RNA sequencing". Perraudeau F, Risso D, Street K et al

Typical workflow



[1] "Bioconductor workflow for single [2] cell RNA sequencing". Perraudeau F, Risso D, Street K et al

Let's practice!

SINGLE-CELL RNA-SEQ WITH BIOCONDUCTOR IN R

Load, create, and access single-cell datasets in R

SINGLE-CELL RNA-SEQ WITH BIOCONDUCTOR IN R

Fanny Perraudeau
Senior Data Scientist, Whole Biome

datacamp

SingleCellExperiment class

`SingleCellExperiment` (SCE) is a S4 class for storing data from single-cell experiments.

Can store and retrieve:

- matrix of counts
- cell and gene information
  - spike-in information,
  - dimensionality reduction coordinates,
  - size factors for each cell,
  - usual metadata for genes and cells.

in a single R object!

[1] https://bioconductor.org/packages/3.9/bioc/html/SingleCellExperiment.html (by Aaron Lun and Davide Risso)

Load and install

Install `SingleCellExperiment` package

```
source("https://bioconductor.org/biocLite.R")
biocLite("SingleCellExperiment")
```

Load `SingleCellExperiment` package

```
library(SingleCellExperiment)
```

SCE object from a counts matrix

```
# create a counts matrix from Poisson distribution
counts <- matrix(rpois(8, lambda = 10), ncol = 2, nrow = 4)
# assign row and column names of counts matrix
rownames(counts) <- c("Lamp5", "Fam19a1", "Cnr1", "Rorb") #genes
colnames(counts) <- c("SRR2140028", "SRR2140022") #cells
# print the counts matrix
counts
```

```
        SRR2140028 SRR2140022
Lamp5           13          3
Fam19a1          9         10
Cnr1             8         10
Rorb             5          7
```

```
# create a SingleCellExperiment object
sce <- SingleCellExperiment(assays = list(counts = counts),
                            rowData = data.frame(gene = rownames(counts)),
                            colData = data.frame(cell = colnames(counts)))
# print the SCE object
sce
```

```
class: SingleCellExperiment
dim: 4 2
metadata(0):
assays(1): counts
rownames(4): Lamp5 Fam19a1 Cnr1 Rorb
rowData names(1): gene
colnames(2): SRR2140028 SRR2140022
colData names(1): cell
reducedDimNames(0):
spikeNames(0):
```

SCE object from SummarizedExperiment

```
# create a SummarizedExperiment object from the counts matrix
se <- SummarizedExperiment(assays = list(counts = counts))
# convert to SingleCellExperiment
sce <- as(se, "SingleCellExperiment")
sce
```

```
class: SingleCellExperiment
dim: 4 2
metadata(0):
assays(1): counts
rownames(4): Lamp5 Fam19a1 Cnr1 Rorb
rowData names(0):
colnames(2): SRR2140028 SRR2140022
colData names(0):
reducedDimNames(0):
spikeNames(0):
```

[1] SummarizedExperiment package: https://bioconductor.org/packages/3.9/bioc/html/SummarizedExperiment.html

Real single-cell dataset

```
# load the allen dataset from scRNAseq
library(scRNAseq)
data(allen)
# print allen
allen
```

```
class: SummarizedExperiment
dim: 20908 379
metadata(2): SuppInfo which_qc
assays(4): tophat_counts cufflinks_fpkm rsem_counts rsem_tpm
rownames(20908): 0610007P14Rik 0610009B22Rik ... Zzef1 Zzz3
rowData names(0):
colnames(379): SRR2140028 SRR2140022 ... SRR2139341 SRR2139336
colData names(22): NREADS NALIGNED ... Animal.ID passes_qc_checks_s
```

[1] Tasic et al "Adult mouse cortical cell taxonomy revealed by single cell transcriptomics"

```
# covert to a SingleCellExperiment
sce <- as(allen, "SingleCellExperiment")

#print the sce object
sce
```

```
class: SingleCellExperiment
dim: 20908 379
metadata(2): SuppInfo which_qc
assays(4): tophat_counts cufflinks_fpkm
  rsem_counts rsem_tpm
rownames(20908): 0610007P14Rik 0610009B22Rik
  ... Zzef1 Zzz3
rowData names(0):
colnames(379): SRR2140028 SRR2140022 ...
  SRR2139341 SRR2139336
colData names(22): NREADS NALIGNED ...
  Animal.ID passes_qc_checks_s
reducedDimNames(0):
spikeNames(0):
```

Let's practice!

SINGLE-CELL RNA-SEQ WITH BIOCONDUCTOR IN R