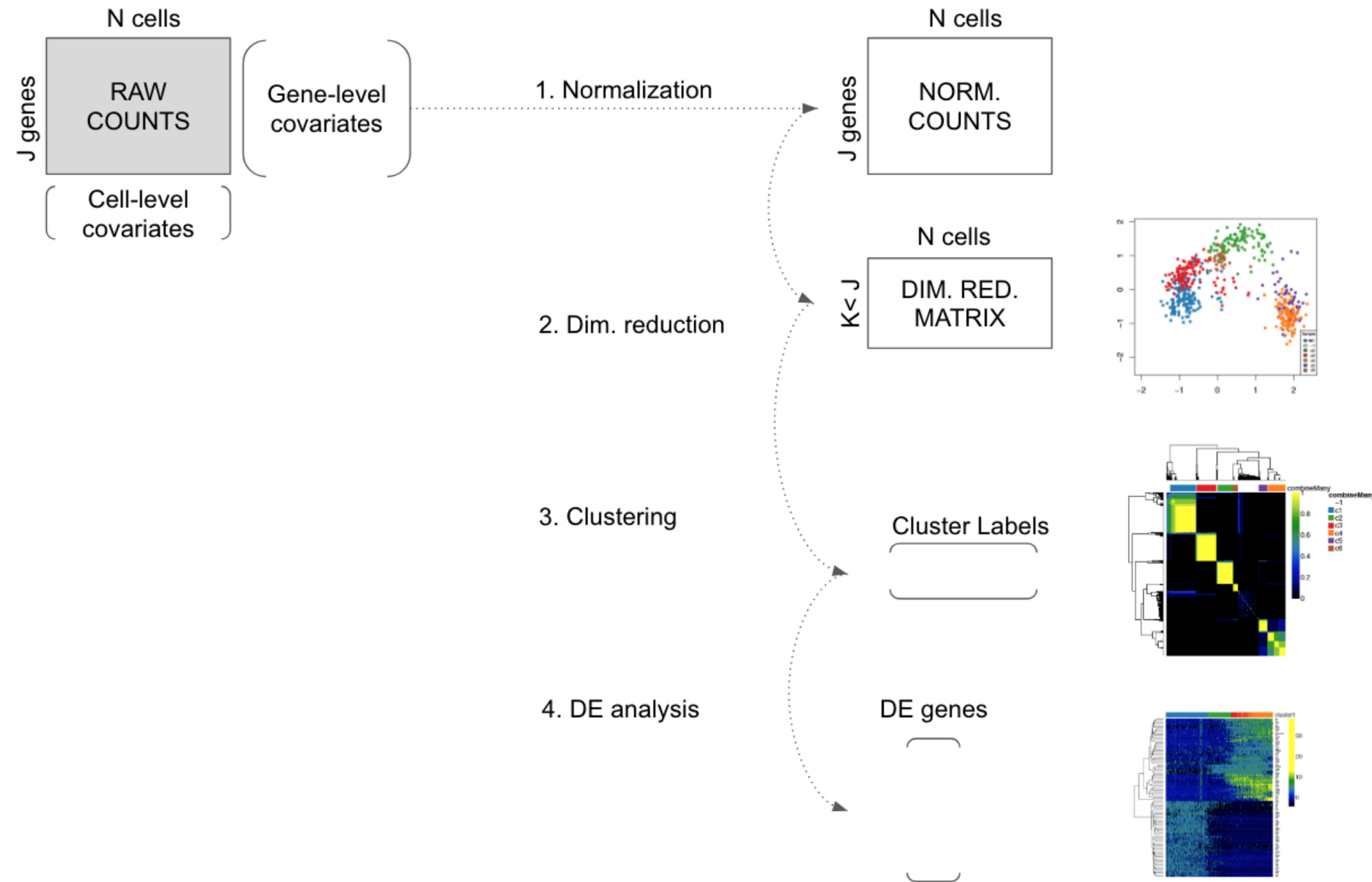# Mouse Epithelium Dataset

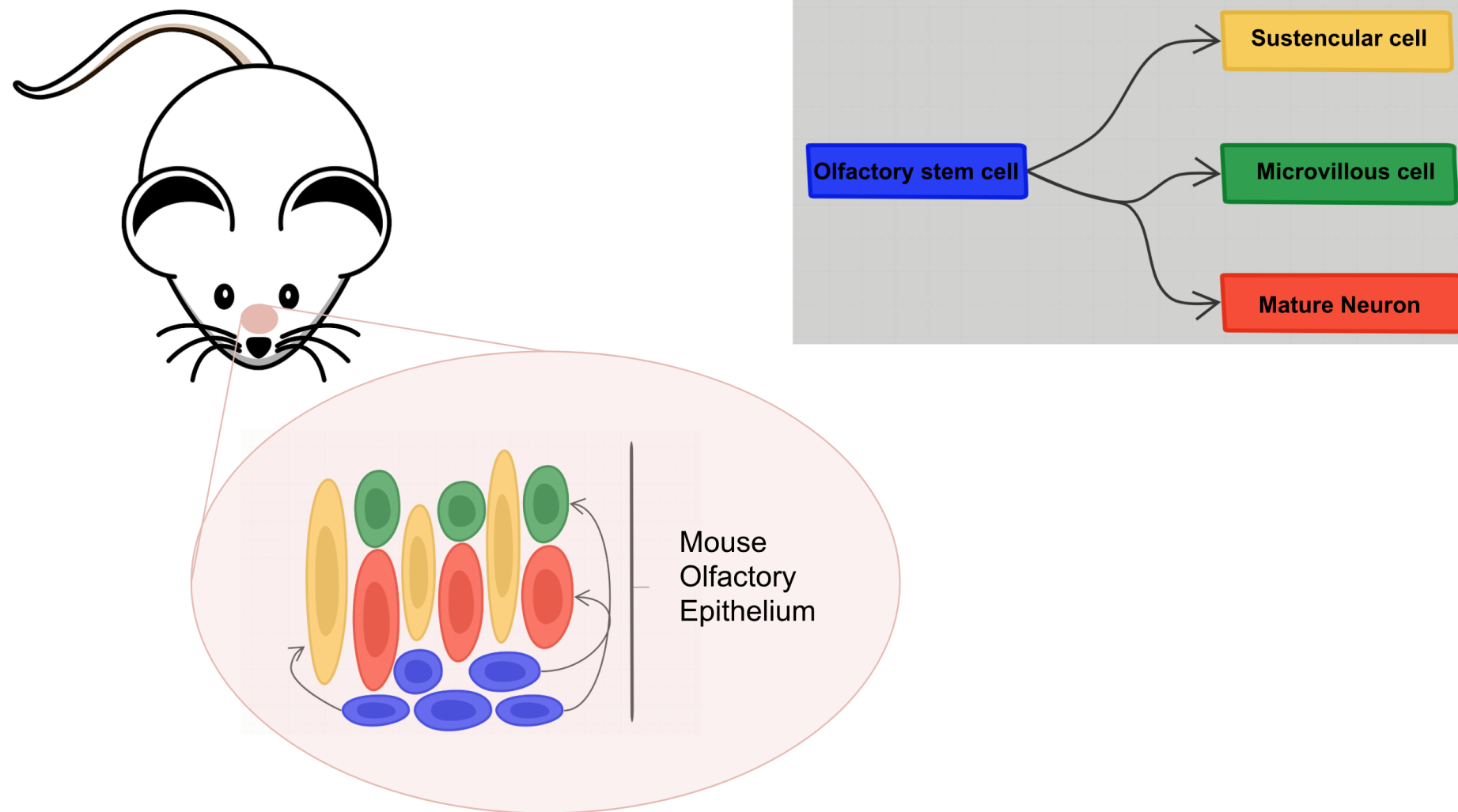## SINGLE-CELL RNA-SEQ WORKFLOWS IN R

Fanny Perraudeau
Senior Data Scientist, Whole Biome

Typical workflow

Stem Cell Differentiation in the Mouse Olfactory Epithelium

[1] Cell Stem Cell, Fletcher et al, Deconstructing Olfactory Stem Cell Trajectories at Single [2] Cell Resolution (2017)
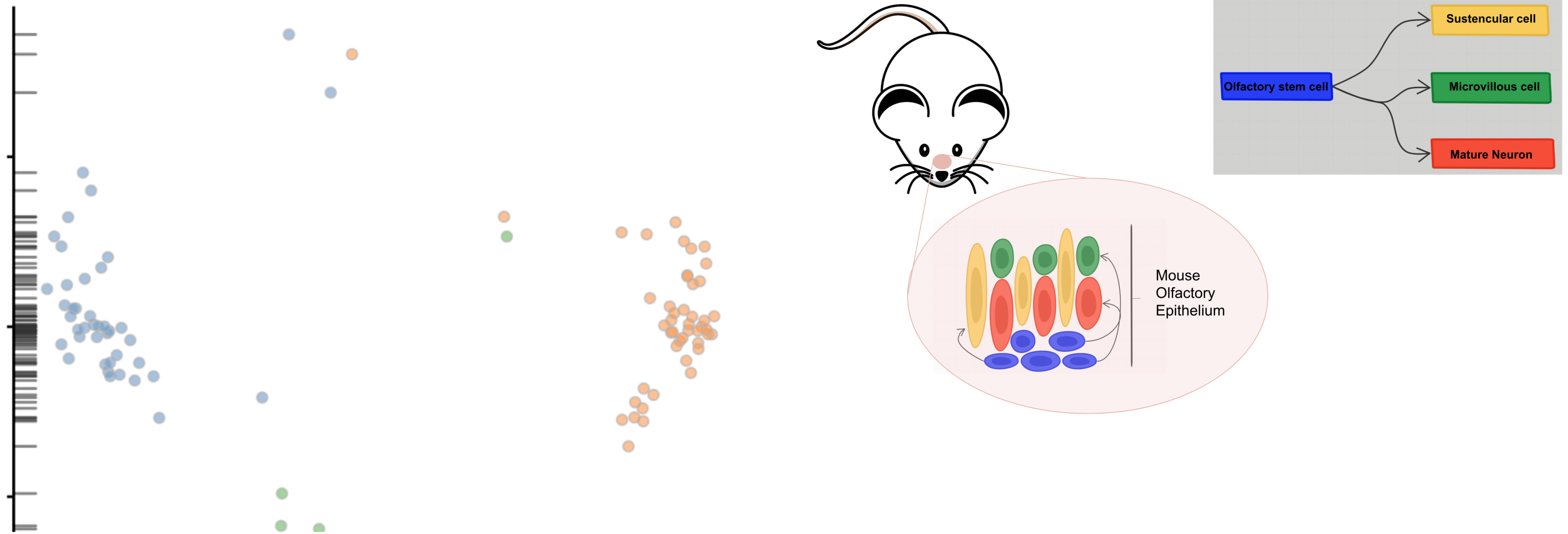
Stem Cell Differentiation in the Mouse Olfactory Epithelium

[1] Cell Stem Cell, Fletcher et al, Deconstructing Olfactory Stem Cell Trajectories at Single [2] Cell Resolution (2017)

Let's practice!
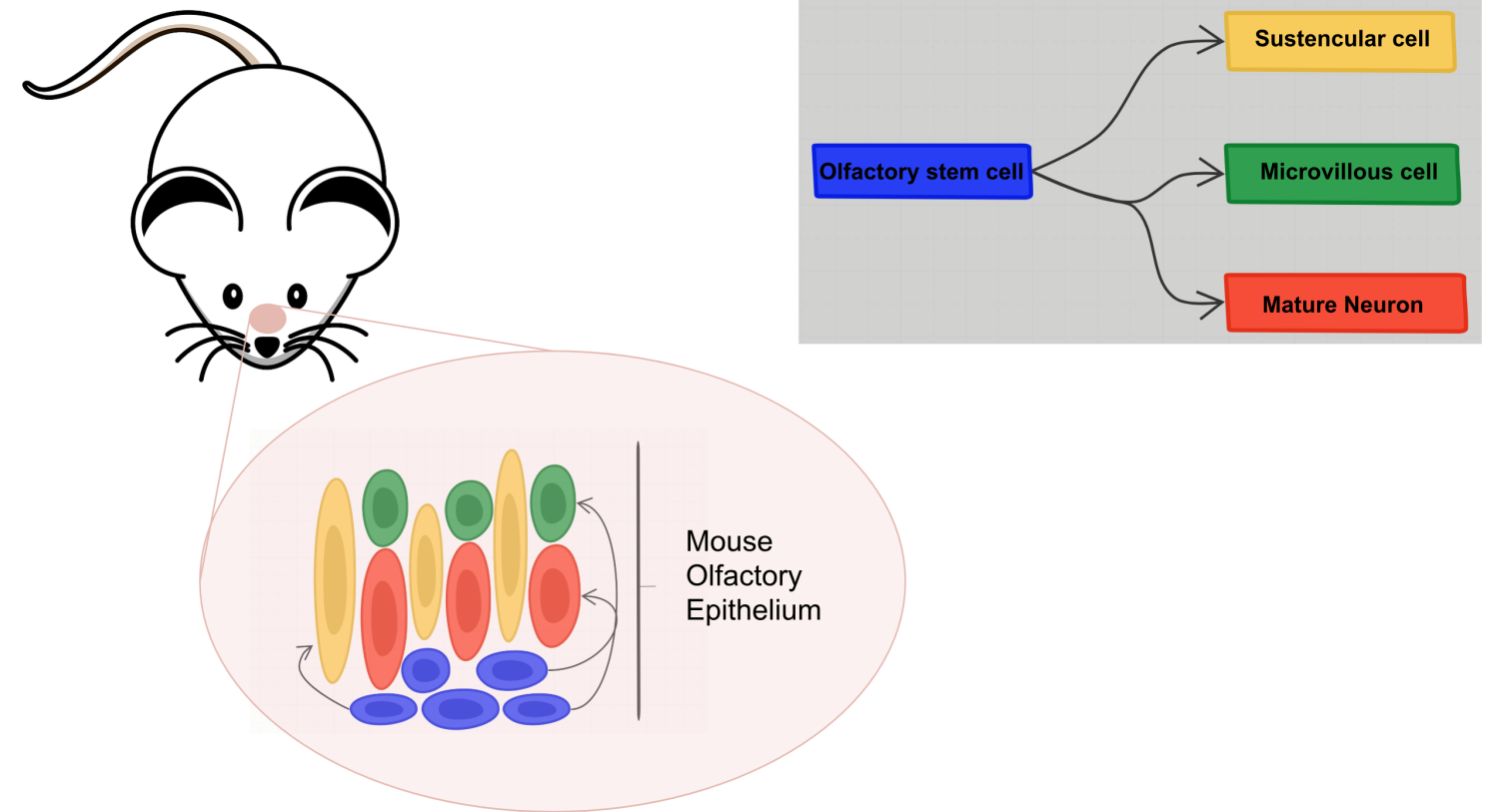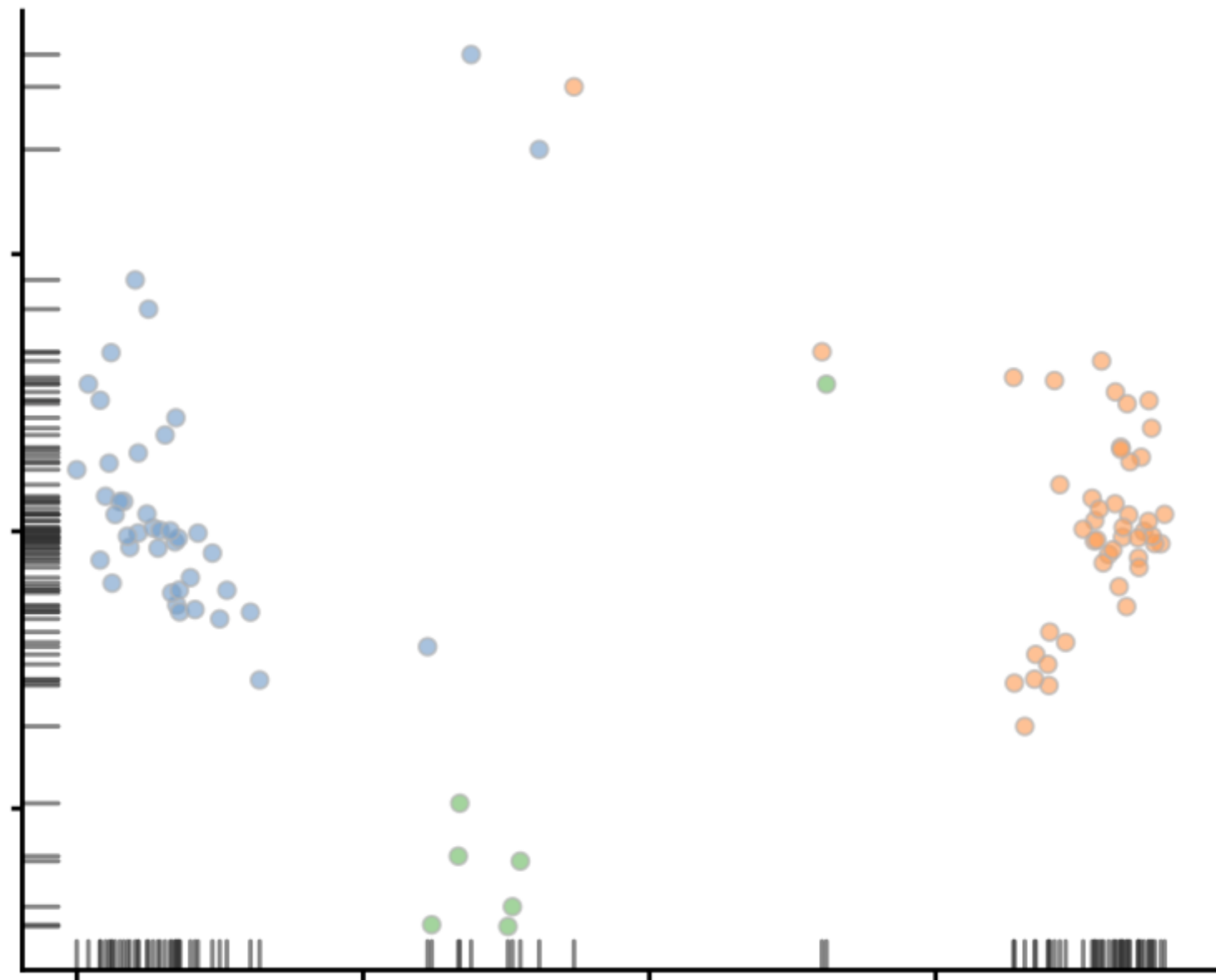
SINGLE-CELL RNA-SEQ WORKFLOWS IN R

Dimensionality reduction

[1] Cell Stem Cell, Fletcher et al, Deconstructing Olfactory Stem Cell Trajectories at Single [2] Cell Resolution (2017)

Dimensionality reduction methods

- Principal component analysis (PCA)

- t-Distributed Stochastic Neighbor Embedding (tSNE)

- Zero inflated factor analysis (ZIFA) [1]

- Zero-inflated negative binomial wanted variation extraction (ZINB-WaVE) [2]

[1] ZIFA, Dimensionality reduction for zero [2] inflated single [3] cell gene expression analysis (Emma Pierson and Christopher Yau) Genome Biology [4] A general and flexible method for signal extraction from single [5] cell RNA [6] seq

Plot PCA

```
plotPCA(sce, exprs_values = "logcounts", shape_by = "Batch",
    colour_by = "publishedClusters")
```
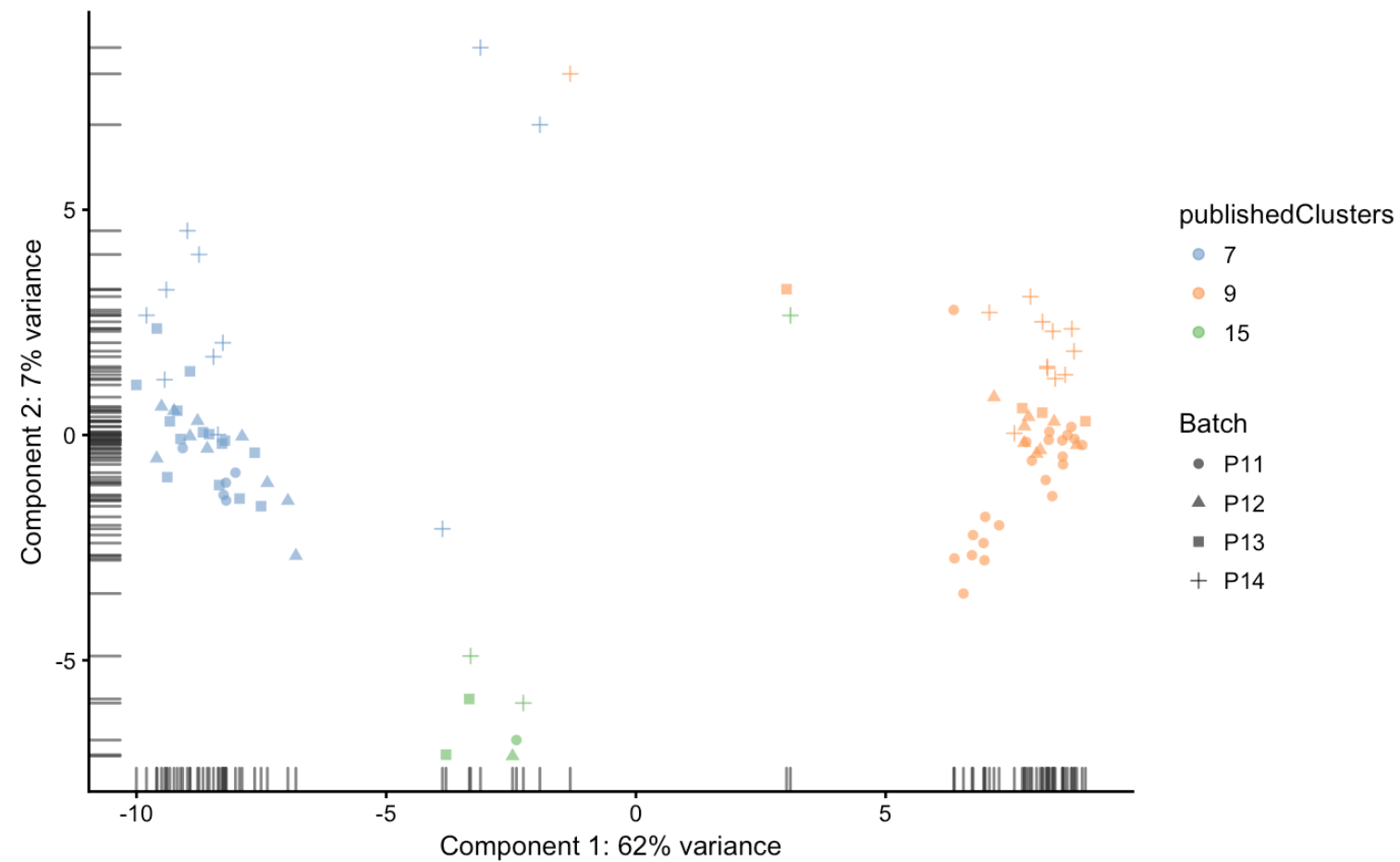
Plot t-SNE

```
plotTSNE(sce, exprs_values = "logcounts", shape_by = "Batch",
    colour_by = "publishedClusters", perplexity = 5)
```

Let's practice!

SINGLE-CELL RNA-SEQ WORKFLOWS IN R

DataCamp

Dimensionality Reduction

SINGLE-CELL RNA-SEQ WORKFLOWS IN R

Fanny Perraudeau
Senior Data Scientist, Whole Biome

DataCamp

Most variable genes

```r
library(magrittr)
vars <- assay(sce) %>% log1p %>% rowVars
names(vars) <- rownames(sce)
vars <- sort(vars, decreasing = TRUE)
head(vars)
```

```
   Cyp2g1   Sec14l3      Rgs5      Sdc4      Cbr2   Cyp2f2
 15.40474  14.72372  14.08343  13.25418  13.10627  12.68986
```

Subset sce

```r
sce_sub <- sce[names(vars[1:50]),]
sce_sub
```

```
class: SingleCellExperiment
dim: 50 94
metadata(0):
assays(1): counts
rownames(50): Cyp2g1 Sec14l3 ... Calb2 Tmprss13
rowData names(0):
colnames(94): OEL19_N724_S503 OEL19_N719_S502 ...
  OEL23_N701_S511 OEL23_N703_S502
colData names(20): Experiment Batch ... ERCC_reads
  colPublishedClusters
reducedDimNames(0):
spikeNames(0):
```

Perform PCA

```r
logcounts <- log1p(assay(sce_sub))
pca <- prcomp(t(logcounts))


reducedDims(sce_sub) <- SimpleList(PCA = pca$x)
sce_sub
```

```
class: SingleCellExperiment
dim: 50 94
metadata(0):
assays(1): counts
rownames(50): Cyp2g1 Sec14l3 ... Calb2 Tmprss13
rowData names(0):
colnames(94): OEL19_N724_S503 OEL19_N719_S502 ...
  OEL23_N701_S511 OEL23_N703_S502
colData names(20): Experiment Batch ... ERCC_reads
  colPublishedClusters
reducedDimNames(1): PCA
spikeNames(0):
```
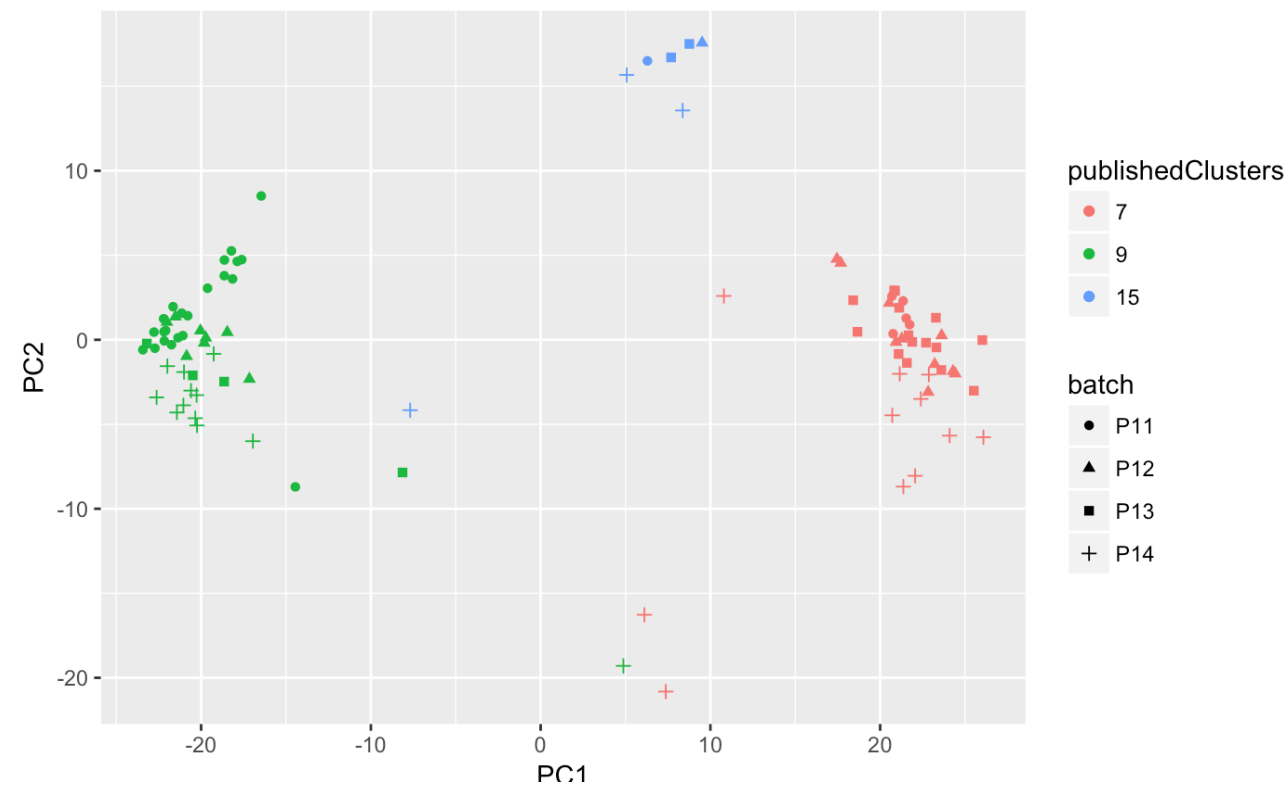
Perform PCA

```r
head(reducedDim(sce_sub, "PCA")[, 1:2])
```

```
                      PC1        PC2
OEL19_N724_S503 -18.63651  3.7905674
OEL19_N719_S502  21.53071  1.2738851
OEL21_N712_S507  20.93405 -0.1382121
OEL19_N723_S506 -17.60803  4.7438350
OEL19_N720_S507  20.69562  2.5815635
OEL19_N721_S510 -18.62802  4.7205441
```

Plot PCA

```
pca <- reducedDim(sce_sub, "PCA")[, 1:2]
col <- colData(sce)[, c("publishedClusters", "batch")]
df <- cbind(pca, col)
ggplot(df, aes(x = PC1, y = PC2, col = publishedClusters, shape = batch)) +
    geom_point()
```

Let's practice!

SINGLE-CELL RNA-SEQ WORKFLOWS IN R