Project 1

Project 1 is about a single cell sequencing project of zebrafish retina. Photoreceptors were damaged with MNU and the response was investigated with help of transgenic fish that contained contstruct with a non-coding element (careg) regulating attached to a EGFP transcript, that can be used as regenerative activation marker. For single-cell transcriptomics analysis, cell suspensions were created from retinal cells, and processed with the 10x 3' kit.

Available data

Data has been downloaded and prepared for you from GEO GSE202212. The count matrices are created with `cellranger`. To create the count tables, the EGFP sequence was added to the reference genome. The gene name of EGFP is `EGFP`.

In order to download the data, run:

Code starts here:

```
wget https://single-cell-transcriptomics.s3.eu-central-1.amazonaws.com/projects/data/project1.tar.gz
tar -xzcvf project1.tar.gz
```

Code ends here

After extracting, a directory `project1` appears with the following content:

```
.
├── data
│   ├── 10dp1
│   │   ├── filtered_feature_bc_matrix
│   │   │   ├── barcodes.tsv.gz
│   │   │   ├── features.tsv.gz
│   │   │   └── matrix.mtx.gz
│   │   └── web_summary.html
│   ├── 10dp2
│   │   ├── filtered_feature_bc_matrix
│   │   │   ├── barcodes.tsv.gz
│   │   │   ├── features.tsv.gz
│   │   │   └── matrix.mtx.gz
│   │   └── web_summary.html
│   ├── 3dp1
│   │   ├── filtered_feature_bc_matrix
│   │   │   ├── barcodes.tsv.gz
│   │   │   ├── features.tsv.gz
│   │   │   └── matrix.mtx.gz
│   │   └── web_summary.html
│   ├── 3dp2
```

```
│   │   ├── filtered_feature_bc_matrix
│   │   │   ├── barcodes.tsv.gz
│   │   │   ├── features.tsv.gz
│   │   │   └── matrix.mtx.gz
│   │   └── web_summary.html
│   ├── 7dp1
│   │   ├── filtered_feature_bc_matrix
│   │   │   ├── barcodes.tsv.gz
│   │   │   ├── features.tsv.gz
│   │   │   └── matrix.mtx.gz
│   │   └── web_summary.html
│   ├── 7dp2
│   │   ├── filtered_feature_bc_matrix
│   │   │   ├── barcodes.tsv.gz
│   │   │   ├── features.tsv.gz
│   │   │   └── matrix.mtx.gz
│   │   └── web_summary.html
│   ├── ctrl1
│   │   ├── filtered_feature_bc_matrix
│   │   │   ├── barcodes.tsv.gz
│   │   │   ├── features.tsv.gz
│   │   │   └── matrix.mtx.gz
│   │   └── web_summary.html
│   └── ctrl2
│       ├── filtered_feature_bc_matrix
│       │   ├── barcodes.tsv.gz
│       │   ├── features.tsv.gz
│       │   └── matrix.mtx.gz
│       └── web_summary.html
└── paper.pdf

17 directories, 33 files
```

Showing us that we have two replicates per treatment, and four treatments:

- ctrl: controls
- 3dp: 3 days post injury
- 7dp: 7 days post injury
- 10dp: 10 days post injury

Now create a new project in the `project1` directory (Project (None) > New Project ...), and create Seurat object from the count matrices:

Code starts here:

```r
library(Seurat)

# vector of paths to all sample directories
datadirs <- list.files(path = "data", full.names = TRUE)
```

```r
# get the sample names
# replace underscores with hyphen to correctly extract sample names later on
samples <- basename(datadirs) |> gsub("_", "-", x = _)

# files are in filter_feature_bc_matrix
datadirs <- paste(datadirs, "filtered_feature_bc_matrix", sep = "/")

names(datadirs) <- samples

# create a large sparse matrix from all count data
sparse_matrix <- Seurat::Read10X(data.dir = datadirs)

# create a seurat object from sparse matrix
seu <- Seurat::CreateSeuratObject(counts = sparse_matrix,
                                  project = "Zebrafish")
```

Code ends here

> **Project exercise**
>
> with this dataset, go through the steps we have performed during the course, and try to reproduce the results provided in the paper. Pay specific attention to quality control, clustering and annotation.

Tips

- For mitochondrial genes, ribosomol genes and hemoglobin genes you can use the following patterns: `"^mt-"`, `"^rp[sl]"` and `"^hb[^(p)]"`.
- Work iterative; meaning that based on results of an analyis, adjust the previous analysis. For example, if clustering is not according to cell types, try to adjust the number of components or the resolution.