Analysis of single cell RNA-seq data

singlecellcourse.org/index.html

Alexander Predeus (apredeus)

Hugo Tavares (tavareshugo)

Simon Murray (simondmurray)

Vladimir Kiselev (wikiselev)

Tallulah Andrews (talandrews)

Jennifer Westoby (Jenni_Westoby)

Davis McCarthy (davisjmcc)

Maren Büttner (marenbuettner)

Jimmy Lee (THJimmyLee)

Krzysztof Polanski

Sebastian Y. Müller

Elo Madissoon

Stephane Ballereau

Maria Do Nascimento Lopes Primo

Rocio Martinez Nunez

Martin Hemberg (m_hemberg)

2022-08-08

1 About the course

Today it is possible to obtain genome-wide transcriptome data from single cells using high-throughput sequencing (scRNA-seq). The main advantage of scRNA-seq is that the cellular resolution and the genome wide scope makes it possible to address issues that are intractable using other methods, e.g. bulk RNA-seq or single-cell RT-qPCR. However, to analyze scRNA-seq data, novel methods are required and some of the underlying assumptions for the methods developed for bulk RNA-seq experiments are no longer valid.

In this course we will discuss some of the questions that can be addressed using scRNA-seq as well as the available computational and statistical methods available. The course is taught through the University of Cambridge Bioinformatics training unit, but the material found on these pages is meant to be used for anyone interested in learning about computational analysis of scRNA-seq data. The course is taught twice per year and the material here is updated prior to each event.

The number of computational tools is increasing rapidly and we are doing our best to keep up to date with what is available. One of the main constraints for this course is that we would like to use tools that are implemented in R and that run reasonably fast. Moreover, we will also confess to being somewhat biased towards methods that have been developed either by us or by our friends and colleagues.

1.1 Registration

Please follow this link and register for the **"Analysis of single cell RNA-seq data"** course: http://training.csx.cam.ac.uk/bioinformatics/search

1.2 GitHub

Updated repository: https://github.com/cellgeni/scRNA.seq.course

Original course (outdated): https://github.com/hemberg-lab/scRNA.seq.course

1.3 Course Docker image

The course can be reproduced without any package installation by running the course docker image which contains all the required packages.

1.3.1 Run the image

Make sure Docker is installed on your system. If not, please follow these instructions. To run the course docker image (use the latest version:

```
docker run -p 8888:8888 -e JUPYTER_TOKEN="jupyter" quay.io/cellgeni/scrna-seq-course:latest
```

Then follow the instructions provided, e.g.:

```
To access the notebook, open this file in a browser:
    file:///home/jovyan/.local/share/jupyter/runtime/nbserver-6-open.html
Or copy and paste one of these URLs:
    http://(a9ee1aad5398 or 127.0.0.1):8888/?
token=22debff49d9aae3c50e6b0b5241b47eefb1b8f883fcb7e6d
```

A Jupyter session will be open in a web browser (we recommend Chrome).

1.3.1.1 Windows users

On Windows operating system the IP address of the container can be different from `127.0.0.1` (`localhost`). To find the IP address please run:

```
docker-machine ip default
```

1.3.2 Download data/other files

Please click on `New -> Terminal` . In the new terminal window please run:

```
./poststart.sh
```

If you want to download data files outside of Docker image you can still use the same `poststart.sh` script but you will need to install AWS CLI on your computer.

Alternatively, you can browse and download the files in you web-browser by visiting this link.

1.3.3 RStudio

Now go back to Jupyter browser tab and change word `tree` in the url to `rstudio` . RStudio server will open with all of the course files, software and the data folder available.

1.3.4 Compute resources

Compute resources for the live course will be provided by Amazon.

1.4 Manual installation

If you are not using a docker image of the course, then to be able to run all code chunks of the course you need to clone or download the course GitHub repository and start an R session in the `course_files` folder. You will also need to install all required packages manually.

Alternatively, you can just install packages listed in a chapter of interest.

1.5 License

All of the course material is licensed under **GPL-3**. Anyone is welcome to go through the material in order to learn about analysis of scRNA-seq data. If you plan to use the material for your own teaching, we would appreciate if you tell us about it in addition to providing a suitable citation.

1.6 Prerequisites

The course is intended for those who have basic familiarity with Unix and the R scripting language.

We will also assume that you are familiar with mapping and analysing bulk RNA-seq data as well as with the commonly available computational tools.

We recommend attending the Introduction to RNA-seq and ChIP-seq data analysis or the Analysis of high-throughput sequencing data with Bioconductor before attending this course.

1.7 Contact

If you have any **comments**, **questions** or **suggestions** about the material, please contact Alexander Predeus, Hugo Tavares, or Vladimir Kiselev.