scRNA-seq hands-on course

(Torino 26th-29th September 2023)

B&Gu Turin

(Registration form: https://bit.ly/43mWN14)

Further Information

For more information, please contact the course organizer: raffaele.calogero@unito.it

Teaching Format

This course will include a series of theoretical sessions followed by practical exercises. This course will utilize open-source software. The course is mainly based on the use of, rCASC and 4SeqGUI applications, which are part of the Reproducible Bioinformatics Project.

Part of the Reproducible Bioinformatics Project is also SeqBox (Beccuti et al. Bioinformatics 2018)
The SeqBox is a cheap, efficient and reproducible RNAseq-ChIPseq hardware/software solution based on NUC Intel mini-PC. In SeqBox the analysis of RNAseq and scRNAseq data is supported by a friendly GUI. This allows access to fast and reproducible analysis also to scientists without scripting experience.
Computing infrastructure: All exercises will be performed using the B&Gu server located at HPC4AI infrastructure (https://hpc4ai.it/).

Aims and Objectives

At the end of the course, you will be able to:
✓ understand the importance of experimental design to ask sensible biological questions at single cell level.
✓ Assess the quality of your data.
✓ Understand limits and strength of data reduction and clustering in scRNA-seq.
✓ Identify genes driving cluster formation in scRNA-seq.
✓ Extract biological knowledge from cluster-specific biomarkers.

Audience

This course is suitable for biologists who are new to single cell gene expression technology. Knowledge of statistics as well as computing skills are not necessary prior to attending the course.
The course requires a very basic level of R, which can be reached following these videos and performing the associated exercises (https://bit.ly/44iW8PQ).

Course Description

Tools for scRNA-seq data analysis

The course is based on the use of open-source software solutions. scRNA-seq analyses will be performed using the tools available as part of the *Reproducible Bioinformatics Project* [1] (http://reproducible-bioinformatics.org/): rCASC [2] and its graphical interface.

Experimental design

This section of the course discusses several criteria and principles of experiment design as well as related problems. Questions such as
  i.   which are the minimal requirements for a single cell RNA-seq experiment
  ii. when a scRNA-seq is preferred to a bulk RNA-seq analysis
  iii. how to structure a successful scRNA-seq experiment
will be addressed.

Quality control

This section will focus on quality controls for single cells sequence outputs. Approaches to check the quality of raw data will be presented as well as approaches to identify sequencing bias. All approaches will be practically tested on real data provided during the practical training sessions.

Data analysis theoretical knowledge

This part will provide the biologist with a general overview on the theory behind the computing tools used in single cell RNA-seq data. The purpose is to give only as much information as needed to be able to make an informed choice during the subsequent data analysis. The aim of the training module is to put things in the perspective of someone who

analyzes single cell RNA-seq data, rather than offer a full treatment of the respective statistical/bioinformatics notions and techniques. No previous statistical knowledge is assumed.

Dimensionality reduction and clustering single cell RNA-seq data

This section presents several data reduction and clustering methods used to depict the cell sub-populations present in a single cell experiment. The advantages and disadvantages of all methods are discussed in detail.

Extracting biological knowledge integrating bulk and single cell RNA-seq data

This session will focus on the extraction of biological knowledge from cluster's data by mean of the integration of public available data.

Batch effect correction and experiments integration

This session will focus on the understanding of the limits of single cell data integration and on the available methods for batch effect correction.

Practical sessions

The course is structured to provide practical analysis skills to the students. Datasets will be provided by B&Gu.

Participants will have access to the computing infrastructure (https://hpc4ai.it/) for seven day after the end of the course. Special renting rate will be offered to participants to access to the scRNAseq computing infrastructure, for further info inquire to raffaele.calogero@unito.it.

Dates Times and Locations

The RNA-seq workshop will last 3 days and half, in September 2023 at MBC (Via Nizza 52, Torino).

Day 1   26th September   15:00-18:00
Day 2   27th September   9:00-17:00
Day 3   28th September   9:00-17:00
Day 4   29th September   9:00-17:00

Costs

The course is open to max 20 people. The registration to the course costs 300 euro and covers lunches, coffee breaks and social dinner.
Each participant should bring his/her own laptop, in case of need the organization can provide laptops.
For contact: raffaele.calogero@unito.it

Sponsors & Organisers

Registration form: https://bit.ly/43mWN14

Program

DAY 1	CONTENT
14:00-15:30	About single cell technologies (Dr. V. Benes Core lab Director EMBL)
15:30-16:00	Break
16:00-17:00	Tips and tricks single cell (Dr. V. Benes Core lab Director EMBL)

DAY 2	CONTENT
9:00-9:30	Course introduction, getting known each other
9:30-11:30	Experimental design
11:30-12:00	Introduction to the data analysis platform (rCASC/HPC4AI)
12:00-13:00	Lunch
13:00-14:00	QC for scRNA-seq, Cell ranger (10Xgenomics) output, sparse matrices, and imputation
14:00-14:30	Exercises
14:30-15:30	Data reduction for scRNA-seq
15:30-15:45	Break
15:45-16:30	Exercises
16:30-17:00	Q&A

DAY 2	CONTENT
9:00-10:30	Sub-population discovery in scRNA-seq: clustering
10:30-10:45	Break
10:45-12:00	Exercises
12:00-13:00	Lunch
13:00-14:30	Differential expression analysis
14:30-15:30	Exercises
15:30-15:45	Break
15:45-16:30	Batch effect correction and experiments integration
16:30-17:00	Exercises
20:00-22:00	Social dinner

DAY 3	CONTENT
9:00-10:00	Cells annotation
10:00-11:00	Exercises
11:00-12:00	Mining cells sub-populations (sparsely-connected autoencoders)
12:00-13:00	Lunch
13:00-13:30	Presentation of the final exercises: an exercise will be provided with some questions to be answered.
13:30-16:00	Participants will perform the requested tasks and prepare a brief report explaining what their findings were.
16:00-16:30	Discussion on the revision exercise
16:30-17:00	Filling the feedback form

References

1.   Kulkarni, N., et al., *Reproducible bioinformatics project: a community for reproducible bioinformatics analysis pipelines.* BMC Bioinformatics, 2018. **19**(Suppl 10): p. 349.
2.   Alessandri, L., et al., *rCASC: reproducible classification analysis of single-cell sequencing data.* Gigascience, 2019. **8**(9).

Additional information

All software used during the course is open-source and can be retrieved from https://github.com/kendomaniac/rCASC. Minimal hardware requirements are: Linux multi core workstation, 2 TB SSD, 32 GB RAM.

In case course participants would be interested to work on their data using the fully configured infrastructure used during the course, they can rent an account at HPC4AI with the following fee:

1 month 147,94€ + 40€ account activation

3 months 443,82€ + 30€ account activation

12 months 1775,28€ + 30€ account activation