scRNA-seq Workshop

An introduction to single cell RNA-seq data analysis.

(Torino 28th-30th September 2022)

(Registration form: https://bit.ly/3qIzKNP)

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.

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

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/).

Instructor Credentials

Raffaele Calogero is Associate Professor at Turin University and the P.I. of the Bioinformatics and Genomics unit. The Bioinformatics and Genomics unit (B&Gu) is a core facility to support researchers in multiplatform microarray/RNA-seq experimental design, analysis and mining. Since 2002 he has led theoretical/practical training courses on microarray data analysis. Since 2010 he is part of the training team of the EMBL Whole transcriptome data analysis course (Heidelberg,DE)

Francesca Cordero is associate professor at Dept. of Computer Sciences, University of Torino. works on the development of new algorithms and pipelines to analyze deep sequencing data, i.e. genomic, different type of transcriptomic data (mRNA, short-noncodingRNs, long-noncodingRNA, circ RNA, methylation data) or single cell experiments.

Marco Beccuti is currently associate professor at Dept. of Computer Sciences, University of Torino. Part of his research work is focused on the design of bioinformatics algorithms and workflows to analyze deep sequencing data (i.e. genomic, transcriptomic, and single-cell data), emphasizing reproducibility aspects.

respective statistical/bioinformatics notions and techniques. No previous statistical knowledge is assumed.

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 from bulk and single cell RNA-seq data

This session will focus on the extraction of biological knowledge from cluster's data using tools like Sparsely-connected-autoencoders [3,4] and omicsnet [5]. Hierarchical clustering will be also used as tools to understand samples heterogeneity given a specific gene set signature.

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, in September 2022.

Day 1 28th September 9:00-17:00
Day 2 29th September 9:00-17:00
Day 3 30th September 9:00-17:00

Costs

The course is open to max 20 people. The registration to the course costs 220 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/3qIzKNP

Program

DAY 1	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
14:00-14:30	Exercises
14:30-15:30	Cell ranger (10Xgenomics) output, sparse matrices, and imputation
15:30-15:45	Break
15:45-16:30	Exercises
16:30-17:00	Q&A

DAY 2	CONTENT
9:00-10:30	Data reduction for scRNA-seq
10:30-10:45	Break
10:45-12:00	Exercises
12:00-13:00	Lunch
13:00-14:30	Sub-population discovery in scRNA-seq: clustering
14:30-15:30	Exercises
15:30-15:45	Break
15:45-16:30	Exercises
16:30-17:00	Q&A
20:00-22:00	Social dinner

DAY 3	CONTENT
9:00-10:00	Depicting cluster specific biomarkers
10:00-11:00	Exercises
11:00-12:00	Mining cells sub-populations (sparsely-connected autoencoders and omninet)
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., Alessandri, L., Panero, R., Arigoni, M., Olivero, M., Ferrero, G., Cordero, F., Beccuti, M., Calogero, R.A.: Reproducible bioinformatics project: a community for reproducible bioinformatics analysis pipelines. BMC Bioinformatics **19**(Suppl 10), 349 (2018). doi:10.1186/s12859-018-2296-x
2. Alessandri, L., Cordero, F., Beccuti, M., Arigoni, M., Olivero, M., Romano, G., Rabellino, S., Licheri, N., De Libero, G., Pace, L., Calogero, R.A.: rCASC: reproducible classification analysis of single-cell sequencing data. Gigascience **8**(9) (2019). doi:10.1093/gigascience/giz105
3. Alessandri, L., Cordero, F., Beccuti, M., Licheri, N., Arigoni, M., Olivero, M., Di Renzo, M.F., Sapino, A., Calogero, R.: Sparsely-connected autoencoder (SCA) for single cell RNAseq data mining. NPJ Syst Biol Appl **7**(1), 1 (2021). doi:10.1038/s41540-020-00162-6
4. Alessandri, L., Ratto, M.L., Contaldo, S.G., Beccuti, M., Cordero, F., Arigoni, M., Calogero, R.A.: Sparsely Connected Autoencoders: A Multi-Purpose Tool for Single Cell omics Analysis. Int J Mol Sci **22**(23) (2021). doi:10.3390/ijms222312755
5. Zhou, G., Xia, J.: OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. Nucleic Acids Res **46**(W1), W514-W522 (2018). doi:10.1093/nar/gky510

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