RNA-seq & scRNA-seq Workshop

An introduction to bulk and single cell RNA-seq data analysis.

(DUKE-NUS 25th-29th July 2022)

Further Information

For more information, please contact the course organizer: Elisabeth.tan@duke-nus.edu.sg

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 docker4seq , 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.

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 both at bulk and single cell level.
- ✓ Assess the quality of your data.
- ✓ Master the analysis tools required to detect differentially expressed genes from bulk RNA-seq.
- ✓ Understand limits and strength of clustering in scRNA-seq.
- ✓ Decript the biological information from differential analysis results (buk RNA-seq).
- ✓ Identify genes driving cluster formation in scRNA-seq.

Audience

This course is suitable for biologists who are new to bulk and 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 RNA-seq and scRNA-seq data analysis

The course is based on the use of open-source software solutions. RNA-seq, and scRNA-seq analyses will be performed using the tools available as part of the *Reproducible Bioinformatics Project* (http://reproducible-bioinformatics.org/): docker4seq, rCASC and their 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 successful bulk or single cell RNA-seq experiment
ii. when a scRNA-seq is preferred to a bulk RNA-seq analysis
iii. how to structure a successful bulk/scRNA-seq experiment

will be addressed.

Quality control

This section will focus on quality controls for bulk and 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 bulk and 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

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 RNA-seq/scRNAseq 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). He is the co-founder of the Reproducible Bioinformatics Community, which is an open community devoted to the development of bioinformatics workflows granting computational and functional reproducibility.

bulk and 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.

Selecting differentially regulated genes

This portion presents several methods used to select differentially regulated genes in comparative experiments. The advantages and disadvantages of all methods are discussed in detail.

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 a set of differentially expressed genes (bulk RNA-seq) or cluster's marker genes (scRNA-seq) using on-line tools like as EnrichR and omicsnet. 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.

One-to-one discussion

On the last part of each day, revision exercises to be provided. During revision exercises, over the 5 days of the course, Prof. Calogero will spend at least 15 minutes discussing with each of the participants about their project. Participants, if interested to one-t-one discussion, will be requested to provide to Prof. Calogero (**before the beginning of the course**) a brief description of their experimental design and desired aim.

Dates Times and Locations

The RNA-seq workshop will last 5 days, in July 2022.

Day 1 25th July 9:00–17:30
Day 2 26th July 9:00–17:30
Day 3 27th July 9:00-17:30
Day 4 28th July 9:00-17:30
Day 5 29th July 9:00-16:00