RNA-seq & Single-cell-seq Workshop

An introductory course to RNA-seq and Single-cell seq

(Torino 9th-13th March 2020)

Further Information

For more information, please contact the course organizer:

…….

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 and 4SeqGUI applications, which are part of the Reproducible Bioinformatics Project.

Part of the Reproducible Bioinformatics Project is also SeqBox. SeqBox is a cheap, efficient and reproducible RNAseq-ChIPseq hardware/software solution based on NUC6I7KYK mini-PC. In SeqBox the analysis of RNAseq and ChIPseq data is supported by a friendly GUI. This allows access to fast and reproducible analysis also to scientists without scripting experience.
SeqBox will be available for test to participants, upon request (Please contact the organizer to reserve one for the course. The only requirement is an ethernet port/adapter and VNC available on participant PC/MAC)

Aims and Objectives

At the end of the course you will be able to:

- ✓ understand the importance of experimental design in order to ask sensible biological questions
- ✓ assess the quality of your data
- ✓ complete basic statistical tests on Next Generation Sequencing (NGS) data
- ✓ annotate and interpret your data and perform integration between gene-level expression and microRNA differential expression data
- ✓ understand some of the problems encountered when analyzing data
- ✓ acquire basic skills in R scripting

Audience

This course is suitable for biologists who are new to Next Generation Sequencing technology. Knowledge of statistics is not necessary prior to attending the course.

Course Description

Tools for RNA-seq data analysis

The course is based on the use of Bioconductor open-source software solutions. However, R coding skill is not required. Furthermore, RNAseq, miRNAseq and single-cell seq analyses will be performed using the tools available as part of the *Reproducible Bioinformatics Project* (Beccuti et al. Bioinformatics 2018, Kulkarni et al. BMC Bioinformatics 2018)

Experimental design

This section of the course discusses several criteria and principles of experiment design as well as related problems. Questions such as how many replicates one needs to detect differential gene/microRNA expression or alternative splicing events are addressed.

Quality control

This section will focus on RNA-seq quality controls. Approaches to check the quality of raw data will be presented as well as approaches to identify sequencing bias. Approaches to experimental replicates will also be considered. All approaches will be practically tested on real data provided during the practical training sessions.

Basic Statistics

This part will provide the biologist with a general overview on issues closely related to 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 gene/exon-level RNA-seq data, rather than offer a full treatment of the respective statistical notions and techniques. No previous statistical knowledge is assumed.

Selecting differentially regulated genes/microRNAs

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

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)

Marco beccuti is is Researcher at Dept. of Computer Science, University of Torino. He has a degree and a PhD in Computer Sciences. He is involved in NGS tools development.

Francesca Cordero is Researcher at Dept. of Computer Science, University of Torino. He has a degree and a PhD in Computer Sciences. He is involved in NGS tools development.

Vladimir Benes is Head of Genomics Core Facility at EMBL Heidelberg. He has a multi-years experience on transcriptomics, genomics data generation.

RC, MB and FC are the founders of the reproducible-bioinformatics.org, a community devoted to development of bioinformatic tools granting functional and computational reproducibility.

Selecting alternative splicing events

This portion presents approaches to identify alternative splicing events in a two groups experiment. The advantages and disadvantages of various methods are discussed in detail.

Single-cell RNAseq workflow

The field of single-cell genomics is advancing rapidly and is generating many new insights into complex biological systems, ranging from the diversity of microbial ecosystems to the genomics of human cancer. This module will briefly describe the differences existing between bulk RNAseq and single-cell sequencing. The module will provide the basic instrument for subpopulation discovery using rCASC workflow.

Gene Ontology enrichment and data visualization

This session will focus on the extraction of biological knowledge from a set of differentially expressed genes using on-line tools like as EnrichR. Furthermore, the use of PCA and hierarchical clustering will be described. As visualization tool will be used MeV, a Java application designed to allow the analysis of high-throughput data to identify patterns of gene expression.

R pills

Learning a scripting language might be frustrating for biologists. Thus for the first three days of the course we will have two session of R scripting training (60 mins theory/60 mins practice) one after breakfast and the other before dinner. The fourth day of course will be completely devoted to the use of R scripting in setting up a basic differential expression analysis using Bioconductor packages.

Practical sessions

The course is structured to provide practical analysis skills to the students. Datasets will be provided by B&Gu. Data provided by the organizers are based on cell lines experiments.

Dates Times and Locations

The RNA-seq workshop will last 5 days, in March 2020.

Day 1	9th March	9:30–18:00
Day 2	10th March	9:00–18:00
Day 3	11th March	9:00-18:00
Day 4	12ve March	9:00-18:00
		20:30 – 22:00 Social dinner sponsored by B&Gu
Day 5	13th March	9:00-18:00

Course Costs

The cost of the course is 500 Euros (**max 20 persons**)

A booklet with all presentations, coffee breaks, lunches and the social dinner are provided as part of the course.

Organized by:

Sponsored by: