

Introduction to Transcriptomic Analysis

Instructor: Dr. Aureliano Bombarely (aureliano.bombarely@unimi.it).

Credits: 6.

Hours: 30.

Format: Online using the Microsoft Teams platform.

Dates: 3 days (Tu, Th and Tu), specifically June 23rd, June 25th and June 30th 2020.

Language: English

Attendance Limit: 15 people.

Registration Deadline: June 18th, 2020.

Registration Process: Follow the official instructions for the registration of graduate studies at the University of Milan. Additionally, please send an email to the instructor of the course

General Goal:

Learn the basics about a gene expression transcriptomic analysis as well as develop the bioinformatic skills to perform such analysis.

Specific Goals:

- Develop the basic skills to navigate through a Linux environment and run commands using a command line interface.
- Learn about the basics of Next Generation Sequencing specially applied to the transcriptomic analysis.
- Learn about the design of transcriptomic experiments as well as assess possible drawbacks.
- Develop the basic skills to process NGS data and evaluate its quality.
- Learn about the basic transcriptomic analysis reference based.
- Develop the basic skills for RNA-Seq data mapping to reference as well as evaluate the quality of the read mapping.
- Learn about the expression quantification methods associated with the RNA-Seq analysis as well as the differential expression analysis.
- Develop the skills for expression quantification associated to a RNA-Seq experiment and testing of the differential expression.
- Develop the skills to assess the quality of the experiment based in the expression distribution and the sample distances.
- Learn about the different downstream analysis such as gene expression profile clustering as well as analysis associated to the gene annotation such as Gene Set Enrichment Analysis (GSEA).
- Develop the skills to perform basic gene expression clustering and GSEA.
- Develop the skills to visualize the gene expression data using R.
- Learn about the guidelines to present a RNA-Seq analysis in a scientific publication.

Requirements:

- Basic knowledge about genetics and molecular biology.
- Basic knowledge about R and basic statistics.
- Hardware: A computer with a connection to the Internet, 20 Gb of free disk space and 8 Gb of RAM memory.

- Software:
 - R and R Studio with the following packages installed: Ballgown, RColorBrewer, genefilter, dplyr, devtools, knitr, topGO, VennDiagram, pheatmap and ggplot2.
 - Filezilla.
 - Terminal (for Linux or MacOS) or MobaXTerm (Windows).

Evaluation:

The course evaluation will be done in three parts:

1. Attendance and participation (20%).
2. Homework after the lessons (50%).
3. Final quiz (30%).

Timetable:

- June 23th, 2020 (Tuesday)
 - 09:00 – Presentation of the Course.
 - 09:30 – Lecture: First Steps using a CLI in Linux.
 - 10:30 – Practice: Basic File Navigation with Linux.
 - 11:00 – Break.
 - 11:30 – Lecture: Next Generation Sequencing applied to RNA-Seq.
 - 12:30 – Practice: Manipulations of Sequence Files in Linux I.
 - 13:00 – Lunch break.
 - 14:00 – Lecture: RNA-Seq Experimental Design and Analysis.
 - 15:00 – Practice: Manipulations of Sequence Files in Linux II.
 - 15:30 – Break.
 - 16:00 – Lecture: Read Mapping for RNA-Seq data.
 - 17:00 – Practice: Read Mapping.
 - 18:00 – Lecture: Basic Notions of R.
 - 19:00 – End of Day 1 Session.
- June 25th, 2020 (Thursday)
 - 09:00 – Questions from last session.
 - 09:30 – Lecture: Methods to Quantify Gene Expression.
 - 10:30 – Practice: Gene Expression Quantification.
 - 11:30 – Break.
 - 12:00 – Lecture: Statistical Methods to Detect Differential Expression.
 - 13:00 – Lunch Break.
 - 14:00 – Practice: Detection of the Differential Expression.
 - 15:00 – Lecture: Downstream Analysis I, Clustering.
 - 16:00 – Break
 - 16:30 – Practice: Gene Expression Profile Clustering.
 - 17:30 – Consulting.
 - 19:00 – End of Day 2 Session.
- June 30th, 2020 (Tuesday)
 - 09:00 – Questions from last session.
 - 09:30 – Lecture: Downstream Analysis II, Gene Set Enrichment Analysis.
 - 10:30 – Practice: Applying GSEA to the Differently Expressed Genes.
 - 11:30 – Lecture: Visualization of Gene Expression, Heatmaps.

- 12:00 – Practice: Creating Heatmaps for Gene Expression.
- 13:00 – Lunch break.
- 14:00 – Lecture: Visualization of Results associated to Gene Ontology Terms.
- 14:30 – Practice: GO Terms and Visualization with R.
- 15:00 – Lecture: Guidelines for RNA-Seq Scientific Publishing, Do and Don't.
- 16:00 – Practice: Analysis of RNA-Seq Publications.
- 17:00 – Lecture: Special Cases and Pitfalls for RNA-Seq.
- 17:30 – Consulting.
- 19:00 – End of Day 3 Session.

Notes:

1. Exercises and examples will be focused in plant related examples. The exercises will be performed using an *Arabidopsis thaliana* public dataset.
2. The results of the evaluation will be delivered 2 weeks after the submission deadline.