Introduction to Transcriptomics Analysis

Class 00 - Introduction to the Course.

INSTRUCTOR:
Aureliano Bombarely
Department of Bioscience
Universita degli Studi di Milano
aureliano.bombarely@unimi.it

The workshop will be organised in three days 10 hours each including breaks:

- June 23th, 2020
- June 25th, 2020
- June 30th, 2020

- June 23th, 2020 (Tuesday)
  - 09:00 – Presentation of the Course.
  - 09:30 – Lecture: First Steps using a CLI in Linux.
  - 10:30 – Lecture: Next Generation Sequencing applied to RNA-Seq.
  - 11:30 – Practice: Basic File Navigation with Linux (Unsupervised).
  - 12:30 – Lunch break.
  - 13:30 – Lecture: RNA-Seq Experimental Design and Analysis.
  - 14:30 – Practice: Manipulations of Sequence Files in Linux I.
  - 15:00 – Lecture: Read Mapping for RNA-Seq data.
  - 16:00 – Break.
  - 16:45 – Practice: Manipulations of Sequence Files in Linux II.
  - 17:15 – Practice: Read Mapping.
  - 17:45 – 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.

Some notes:

- Please use the tag "I2TA2020" for all the communications by email. It will help me to find your emails.
- The assistance to the workshop is mandatory in order to get the credits according the PhD program coordinator, so I'll ask you to write your name + present + timestamp four times per day (e.g. Aureliano Bombarely - Present - June 19th, 2020, 11:08). If you can't attend some of the classes, just let me know in advance.
- First day I will ask everyone to connect the video and introduce yourself.
- I will upload the presentations before each class in the "File" section. I will also share the presentation on the chat.
- I will record the classes on video. I will share the videos on request (they are going to be heavy).

Some notes:

- You will be connecting to the Asimina Linux server to make the exercises. In order to do that, you will need a terminal or a terminal simulator (e.g. Terminal for MacOS or MobaXTerm for Windows). Please install them before tomorrow. You will need also an address, a username and a password. The address is 159.149.160.43. I will send you the username and password in two other different emails with the subjects "I2TA2020-AccessInfo1" and "I2TA2020-AccessInfo2". Once you have all the information, you can follow the instructions of the attached presentation.

- For the Thursday and Tuesday next wee, you will run some exercises in your computer for which you will need:

- FilleZilla to transfer the files from the server into your computer (see the Instructions for more details).

- R and R Studio with a different libraries to run the analysis. Specifically you will need: Ballgown, RColorBrewer, genefilter, dplyr, devtools, knitr, topGO, VennDiagram, pheatmap and ggplot2. In my case, I am running R version v3.5.1. Other versions may work too, I did not check.

Some notes:

- For the exercise, we will be using some experimental data from:

OPEN

ARTICLE

Citation: Cell Discovery (2016) **2**, 16027; doi:10.1038/celldisc.2016.27
www.nature.com/celldisc

The chromatin remodeler DDM1 promotes hybrid vigor by regulating salicylic acid metabolism

Qingzhu Zhang[1], Yanqiang Li[1], Tao Xu[2], Ashish Kumar Srivastava[1], Dong Wang[1], Liang Zeng[1], Lan Yang[1], Li He[1], Heng Zhang[1], Zhimin Zheng[1], Dong-Lei Yang[1], Cheng Zhao[1], Juan Dong[3], Zhizhong Gong[4], Renyi Liu[1], Jian-Kang Zhu[1,5]

[1]Shanghai Centre for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; [2]Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai, China; [3]Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ, USA; [4]State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing, China; [5]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN, USA

Some notes:

- For the exercise, we will be using some experimental data from:

Plant growth and phenotyping of different genotypes

The *Arabidopsis* seeds from WT Columbia-0 and C24 background were available in the lab. The epigenetic mutants in Col were procured from *Arabidopsis* Biological Resource Centre (http://www.biosci.ohio-state.edu/pcmb). The mutants in C24 background were from previous studies [24, 25]. Seeds were surface sterilized and then plated on Murashige and Skoog medium [51]. To ensure homogenous germination, seeds were kept at 4 °C. After 7 days of stratification, seeds were transferred to growth room and then allowed to grow in continuous, cool fluorescent white light (100 $\mu E\, m^{-2}\, s^{-1}$) at 22 °C under long day conditions. The 7 day-old seedlings were transplanted on soil pots and at 25 day after sowed, phenotyping was performed in terms of plant width and leaf width. For root phenotyping, 12 day-old seedlings were transplanted on vertical plates and after 7 days increase in root length was measured.

RNAseq and data analysis

All the different genotypes were grown on plates in triplicates. At 14 day after sowed, entire seedling was collected, RNA was isolated and RNA sequencing (RNAseq) was performed on Illumina HiSeq 2000 platform (San Diego, CA, USA).