MSc in Bioinformatics for Health Sciences
IEO. Information Extraction from "omics" Technologies

Syllabus Information

Academic Course: 2019/20

Academic Center: 804 - Official Postgraduate Programme in Biomedicine

Study: 8045 – Bioinformatics for Health Sciences - MSc

Subject: 31032 – IEO. Information Extraction from "Omics" Technologies

Credits: 5.0

Course: 1st

Teaching languages: English

Teachers: Robert Castelo

Teaching Period: 3rd term

Presentation

This course provides an introduction to the computational analysis of data obtained from high-throughput molecular profiling assays, such as DNA-seq or RNA-seq, using R and Bioconductor. The topics covered span a full data analysis cycle, including data acquisition, quality assessment, inference of molecular changes and functional analysis.

Associated skills

General competences:

1. Learning technical aspects related to the generation of data from high-throughput molecular assays.

2. Applying bioinformatics tools for the quantitative analysis of DNA and RNA measurements obtained from high-throughput platforms.

3. Basic skills in analysing next-generation sequencing data with the R/Bioconductor platform.

4. Communicating scientific research by means of a presentation and a short manuscript, including the use of computational tools that facilitate reproducing the obtained results.

Specific competences:

1. Understanding the most relevant aspects of the experimental methodology.

2. Acquiring basic concepts of experimental design (replication, randomization, blocking).

3. Developing an intuition for critical aspects in the data obtained by high-throughput molecular assays.

4. Learning raw data quality assessment and control procedures.

5. Performing basic analyses of differential expression.

6. Assessing functional enrichment of differentially expressed genes.

7. Performing basic steps to call, annotate and filter variants from DNA sequencing data.

8. Understanding the basic concepts and analysis techniques of protein expression profiling with mass-spectrometry proteomics assays.

9. Understanding the basic concepts and analysis techniques of metabolite profiling with high-throughput metabolomics assays.

10. Learning basic computational and statistical techniques to integrate different kinds of "omics" data.

11. Using R and markdown to generate reproducible documents describing the statistical analysis of high-throughput "omics" data.

Learning outcomes

1. Basic intuition for the exploration of high-throughput molecular assays data.

2. Ability to perform a basic analysis of differential expression.

3. Ability to perform a basic functional enrichment analysis.

4. Understanding of the fundamental concepts and tools behind reproducible research in bioinformatics.

Prerequisites

1. Basic level of computer programming.

2. Fundamental concepts of statistics.

Contents

Contents section 1:

1.1 Introduction to high-throughput genomics technologies.

1.2 Introduction to R and Bioconductor.

1.3 Quality assessment and normalization of RNA-seq data.

1.4 Experimental design and batch effect identification.

Contents section 2:

2.1 Differential expression analysis of RNA-seq data.

2.3 Functional annotations.

2.2 Functional enrichment analysis.

2.3 GSEA and GSVA approaches to functional analysis.

2.3 Reproducible research.

Contents section 3:

3.1 Quality assessment, read mapping and summarization of RNA-seq data.

3.2 Variant calling, annotation and prioritization from DNA-seq data.

3.3 Protein expression profiling and analysis with mass-spectrometry proteomics assays.

3.4 Metabolite expression profiling and analysis with high-throughput metabolomics assays.

3.5 Integrative analysis of multiple 'omics' data sets

Teaching methods

Approach and general organization of the subject

All sessions of this course, except for the one about the introduction to high-throughput genomics technologies and the one about reproducibility in research, are "hands-on" using a computer. The subject is roughly organized into the three blocks described above, with usually two lecture-free weeks between each block 1 and 2, to facilitate starting working in a data analysis project that students need to complete to pass the subject.

Training activities

The students are expected to follow the data analysis steps described in each "hands-on" lecture, answering questions by themselves that are intertwined throughout the slides. At the end of each "hands-on" session there will be one or more exercises proposed to the student to help him/her consolidating the concepts illustrated during the session. The students will have to develop a data analysis project during the entire term which will be delivered in two parts. One at an earlier deadline in the middle of the term and the other at the end of the term. At the end of each week where students had lectures, they have to answer a short quiz about the most important concepts taught during the lectures of that week.

Evaluation

Assessment system

Weekly quizzes

Data Analysis project

Presentation of the data analysis project

Final exam

Grading system

Weekly quizzes (10%)

Data Analysis project (60%)

Presentation of the data analysis project (10%)

Final exam (20%)

A minimum performance of 50% on each item is required to pass the subject.

Bibliography and Information Resources

F. Hahne et al. Bioconductor case studies. Springer, New York, 2008.

Y. Xie. Dynamic documents with R and knitr. CRC Press, Boca Raton, 2014.

H. Wickham. R packages. O'Reailly, Sebastopol, 2015.

C. Gandrud. Reproducible research with R and RStudio. CRC Press, Boca Raton, 2015.

J.R. González and A. Cáceres. Omic association studies with R and Bioconductor. Chapman and Hall/CRC, 2019.