

Analysis of genomic data: R Bioconductor

Omics data are quickly becoming ubiquitous in research and clinical studies, particularly in cancer, with several hospitals now routinely measuring genomic and transcriptomic profiles of cancer patients, in their efforts to move towards personalized medicine.

While many bioinformatic and statistical methods exist to analyze such data, the size and complexity of omics data can be daunting for researchers approaching the field. Genomic data are characterized by high-dimensionality and complexity, technology-specific biases, and require domain-specific knowledge and bespoke informatic tools to be successfully analyzed.

This half-day course introduces participants to Bioconductor, an R-based open science and open development project for the analysis and comprehension of high-throughput biological data, such as RNA and DNA sequencing. We will introduce the Bioconductor project and how it relates to other R packages, explain how to work with DNA sequences in R, and how to analyze RNA sequencing datasets, using The Cancer Genome Atlas as an example. At the end of the course, participants should be able to explore DNA mutations and copy number alterations, and to perform and interpret a differential expression analysis.

Course outline

The course consists of two parts. The first session is a mix of theory and practice. Participants will be introduced to the basics of sequencing data and Bioconductor. We will also discuss how to perform a differential expression analysis and how to integrate genomic and clinical data. The second session will be hands-on, where participants apply what they have learned to a real dataset. There will be a brief rejoinder at the end of the session to discuss the analyses.

Specifically, the course will cover the following topics:

1. Data import and management in R/Bioconductor
2. Exploratory Data Analysis and Quality Control (EDA/QC)
3. Data normalization
4. Differential expression analysis
5. Integration of genomic and clinical data

By the end of this course, you should be able to:

- Have a working knowledge of DNA and RNA sequencing
- Perform an exploratory analysis of genomic and transcriptomic data
- Perform a differential expression analysis
- Interpret and visualize the results

Target audience

The course is aimed at biostatisticians or medical researchers working with biological or clinical data that want to learn how to include genomics and transcriptomics data in their analyses. Participants are expected to have basic knowledge of the R statistical software; no prior knowledge of Bioconductor is required.

Instructors:

Davide Risso, University of Padua