



DEPARTMENT OF ONCOLOGY-PATHOLOGY

K7F3102, Omics Data Analysis: From Quantitative Data to Biological Information, 3 credits (hec)

Analys av omikdata: Från kvantitativ data till biologisk information, 3 högskolepoäng
Third-cycle level / Forskarnivå

Approval

This syllabus was approved by the The Committee for Doctoral Education on 2023-11-24, and was last revised on 2024-03-15. The revised course syllabus is valid from autumn semester 2024.

Responsible department

Department of Oncology-Pathology, Faculty of Medicine

Prerequisite courses, or equivalent

Prior knowledge of statistics and the statistical programming language R is not a requirement, but would prove very useful during the course.

Purpose & Intended learning outcomes

Purpose

During recent years, omics data has become an integral part of many biomedical and clinical research projects. This broad introductory course aims at bridging the gap between classical biomedical research, omics technologies, and bioinformatics. The course will enable students to get an introduction to omics technologies and basic knowledge of omics data analysis workflows.

Intended learning outcomes

After completing the course, the student will be able to:

- Explain the principles of high-throughput technologies (e.g. genomics, transcriptomics, proteomics) and perform a basic omics data analysis workflow
- Judge and account for strengths and weaknesses in study design, experimental planning, and sample selection
- Perform basic quality control of data by the use of histograms, boxplots, principal

component analysis (PCA), etc

- Explain and perform data normalisation and other forms of data transformation
- Differentiate, choose, and be able to apply, statistical methods such as t-test, linear regression models, and multiple testing correction
- Explain the principles and applications of, and be able to apply, dimensionality reduction methods, such as PCA, t-distributed stochastic neighbour embedding (t-SNE), and uniform manifold approximation and projection (UMAP) to omics data
- Use tools for hierarchical clustering and enrichment analyses
- Use tools for gene ontology (GO) annotation or enrichment
- Create informative and clear visualisations of omics data

Course content

The course is organised in five segments:

1. Lectures

- Introduction to omics technologies and data structures
- The omics data analysis workflow: from quantitative data to biological information (with emphasis on analysis of quantitative omics data, such as proteomics and transcriptomics)
- Omics experimental design and sample selection
- Introduction to data transformation and normalisation
- Introduction to basic statistics in omics data analysis: significance tests, p values, and multiple testing correction (e.g. false discovery rate)
- Overview of dimensionality reduction methods: PCA, t-SNE, and UMAP
- Introduction to GO and enrichment analyses
- Introduction to correlation analysis and linear regression
- Basics of clustering methods
- Brief overview of online databases used in omics research (focused on cancer research)
- Introduction to machine learning

2. Data visualisation workshop

3. Literature study of major statistical concepts in omics research

4. Hands-on exercises on the use of online databases and tools for omics analysis

5. Workshops on coding in R and data analysis

- Introduction to the statistical programming language R
- Hands-on data analysis on student's dataset of choice or one of the provided datasets at Canvas

Forms of teaching and learning

The teaching activities for the course will be based on lectures, workshops, and data analysis cases. The students will participate in a literature study, with discussions in seminar groups, and perform an independent data analysis exam project. The students will also be able to download and use some of the software used in the workshops.

For the final data analysis workshop students can work on their own dataset, or one of the

datasets that we will provide for them. The statistical programming language R will be used extensively in the course.

Language of instruction

The course is given in English

Grading scale

Pass (G) /Fail (U)

Compulsory components & forms of assessment

Compulsory components

- Attendance during lectures and data analysis demonstrations (extra written literature study can be used to compensate absence)
- Attendance to the literature study discussion seminar
- Attendance to the examination seminar
- Written examination assignment

Forms of assessment

The course assessment is based on two types of assignments: a literature study with a critical view on an omics data analysis subject (performed in groups) and an individual written omics data analysis project illustrating the different topics covered during the course.

Course literature

Recommended literature will consist of recent original and review articles, which will be provided to students at the Canvas learning platform during the course.

Prior to the course, we recommend that the students install R and R Studio and familiarise themselves with the basics of coding in R through the package 'swirl'.

Recommended books prior to the course:

1. Modern Statistics for Modern Biology, by Susan Holmes and Wolfgang Huber, Cambridge University Press (2019). ISBN: 9781108705295. License: CC BY-NC-SA
2. Haddock, S. H. D., & Dunn, C. W. (2011). Practical computing for biologists. Sinauer Associates.