UNIVERSITEIT
GENT

Studiefiche

Vanaf academiejaar 2024-2025

Statistische genomica (C003401)

Cursusomvang *(nominale waarden; effectieve waarden kunnen verschillen per opleiding)*

Studiepunten 5.0 **Studietijd 150 u**

Aanbodsessies en werkvormen in academiejaar 2024-2025

A (semester 1)	Engels	Gent	groepswerk hoorcollege werkcollege

Lesgevers in academiejaar 2024-2025

Clement, Lieven	WE02	Verantwoordelijk lesgever
Martens, Lennart	GE31	Medelesgever

Aangeboden in onderstaande opleidingen in 2024-2025	stptn	aanbodsessie
Educatieve Master of Science in de wetenschappen en technologie(afstudeerrichting wiskunde)	5	A
Master of Science in de wiskunde	5	A
Master of Science in Statistical Data Analysis	5	A
Uitwisselingsprogramma wiskunde (niveau master)	5	A

Onderwijstalen

Engels

Trefwoorden

Statistical genomics; generalized linear and additive models and their application within bio-informatics; count regression; multiple testing; next-generation sequencing; mass spectrometry; single-cell.

Situering

To enable the master to a) understand, read and communicate on ‘omics experiments, b) select appropriate statistical methods and software tools for analyzing different types of ‘omics data, c) implement, preprocess, analyze and interpret ‘omics experiments with statistical data-analysis pipelines and software, d) gain insight in how different analysis steps affect statistical inference in bioinformatics problems, e) access, retrieve and analyze data from genomic resources on the web, f) extract information related to high-throughput data analysis pipelines from research papers.

Inhoud

This course covers essential statistical concepts and methods for extracting biological insights from high-throughput ‘omics data. The focus lies on the various aspects of typical statistical data-analysis pipelines for high-throughput technologies: i.e. preprocessing, normalization, primary annotation and downstream analysis of ‘omics profiles. Theory lectures will be followed by hands-on exercises using the R programming language and software packages from the Bioconductor ecosystem.
In the first module, the basic concepts will be introduced on proteomics data. The module covers the important aspects of

- Peptide identification
- Differential proteomics
- Data exploration and quality control using plots and models
- Preprocessing: Background correction, Normalization, Summarization, Filtering
- Dealing with batch effects and other confounders
- Statistical Concepts

- Linear models
- Penalized regression and regularization
- Multiple testing
- Trade-off between biological relevance/effect size vs statistical significance
- Empirical Bayes Methods

The second module will extend these concepts towards digital high-throughput technologies, i.e. next generation sequencing (NGS) based omics. The main focus is on bulk and single-cell transcriptomics, and the topics incldue

- NGS Data exploration
- Preprocessing/normalization: accounting for batch effects, library size, etc.
- Dimensionality reduction
- Generalized linear models
- Overdispersion

The third module is topic specific. Depending on the interest of the students, the literature and the project, specific platforms and/or applications will be covered: e. g. single cell trranscriptomics, metagenomics, etc.

Begincompetenties

Having successfully completed introductory courses in basic probability, statistics and linear models. Experience with the statistical programming language R.

Eindcompetenties

1 Have knowledge of methods and concepts related to the analysis of genomics data.
2 The student can identify different sources of variability in high-throughput ‘omics experiments.
3 The student is able to choose statistical methods that are appropriate for a) assessing biological/biomedical research questions using ‘omics data and b) the data characteristics of specific hightroughput technologies.
4 The student can correctly interpret the analysis results.
5 The student can preprocess, analyze, visualize and interpret ‘omics experiments using existing statistical data-analysis pipelines and software.
6 The student can correctly report the analysis results. Formulate the conclusions of the data analysis with respect to the subject matter research question.
7 Assess statistical significance in the context of multipe testing.
8 The student can port statistical concepts introduced in the lecture to other high-throughput platforms and/or applications.
9 The student can critically evaluate and interpret statistical methods used in primary research articles.
10 The student can extract information on the statistical algorithms in high-throughput data analysis pipelines from research papers.

Creditcontractvoorwaarde

Toelating tot dit opleidingsonderdeel via creditcontract is mogelijk mits gunstige beoordeling van de competenties

Examencontractvoorwaarde

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

Didactische werkvormen

Groepswerk, Werkcollege, Hoorcollege

Toelichtingen bij de didactische werkvormen

Exercises: PC-labs using R/Bioconductor.

Studiemateriaal

Type: Handouts

Naam: open-source e-cursus
Richtprijs: Gratis of betaald door opleiding
Optioneel: nee
Taal : Engels
Beschikbaar op Ufora : Nee
Online beschikbaar : Ja

Referenties

Statistical Genomics is a fast evolving field. Relevant literature and review papers will be supplied on Ufora.

Vakinhoudelijke studiebegeleiding

The students exercise concepts and methods introduced during lectures by analyzing real ‘omics data sets in PC-labs. The students will be closely supervised during the PC-labs, and while preparing their projects.

Evaluatiemomenten

periodegebonden en niet-periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Schriftelijke evaluatie met open vragen

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Schriftelijke evaluatie met open vragen

Evaluatievormen bij niet-periodegebonden evaluatie

Werkstuk

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is mogelijk

Toelichtingen bij de evaluatievormen

The project work relates to research questions involving real ‘omics data. The students are allowed to work in groups of 3-4 persons. The result of the project work is a written report that should satisfy scientific and professional standards. The project will be evaluated by the lecturers and teaching assistants, as well as through peer evaluation.
A second term examination for the project is possible. Feedback on the project will be provided and the students that have to take the second term exam can update their report accordingly.

Eindscoreberekening

The total mark is a weighted average of:

- Homework (2/20)
- Project (5/20)
- Exam (13/20)