GHENT
UNIVERSITY

# Course
# Specifications

From the academic year 2024-2025 up to and including the academic year

## Statistical Genomics (C003401)

**Course size** *(nominal values; actual values may depend on programme)*

**Credits 5.0** **Study time 150 h**

**Course offerings and teaching methods in academic year 2025-2026**

| | | | |
|---|---|---|---|
| A (semester 1) | English | Gent | group work<br>lecture<br>seminar |

**Lecturers in academic year 2025-2026**

| | | |
|---|---|---|
| Clement, Lieven | WE02 | lecturer-in-charge |
| Bouwmeester, Robbin | GE31 | co-lecturer |
| Vanderaa, Christophe | WE02 | co-lecturer |

| Offered in the following programmes in 2025-2026 | crdts | offering |
|---|---|---|
| Master of Science in Teaching in Science and Technology(main subject Mathematics) | 5 | A |
| Master of Science in Mathematics | 5 | A |
| Master of Science in Statistical Data Analysis | 5 | A |
| Exchange Programme in Mathematics (master's level) | 5 | A |

**Teaching languages**

English

**Keywords**

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

**Position of the course**

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.

**Contents**

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.

**Initial competences**

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

**Final competences**

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.

**Conditions for credit contract**

Access to this course unit via a credit contract is determined after successful competences assessment

**Conditions for exam contract**

This course unit cannot be taken via an exam contract

**Teaching methods**

Group work, Seminar, Lecture

**Extra information on the teaching methods**

Exercises: PC-labs using R/Bioconductor.

**Study material**

Type: Handouts

Name: open-source e-course
Indicative price: Free or paid by faculty
Optional: no
Language : English
Available on Ufora : No
Online Available : Yes

**References**

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

**Course content-related study coaching**

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.

**Assessment moments**

end-of-term and continuous assessment

**Examination methods in case of periodic assessment during the first examination period**

Written assessment with open-ended questions

**Examination methods in case of periodic assessment during the second examination period**

Written assessment with open-ended questions

**Examination methods in case of permanent assessment**

Assignment

**Possibilities of retake in case of permanent assessment**

examination during the second examination period is possible

**Extra information on the examination methods**

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.

**Calculation of the examination mark**

The total mark is a weighted average of:

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