

MATH-474

**Statistics for genomic data analysis**

Goldstein Darlene

Cursus	Sem.	Type
Computational and Quantitative Biology		Opt.
Ing.-math	MA2, MA4	Opt.
Mathématicien	MA2	Opt.
Statistics	MA2, MA4	Opt.

Language of teaching	English
Credits	5
Session	Summer
Semester	Spring
Exam	During the semester
Workload	150h
Weeks	14
<b>Hours</b>	<b>4 weekly</b>
Lecture	2 weekly
Exercises	2 weekly
<b>Number of positions</b>	

**Remark**

Cours donné en alternance tous les deux ans

**Summary**

After a short introduction to basic molecular biology and genomic technologies, this course covers the most useful statistical concepts and methods for the analysis of genomic data.

**Content**

- Molecular biology and technology background
- R software and BioConductor packages
- Robust regression/High-density oligo array signal quantification/Quality assessment for Affymetrix GeneChips
- Empirical Bayes method for identifying differentially expressed genes
- Linear models for designed experiments
- Hypothesis testing, ROC curves, multiple hypothesis testing
- Gene set testing
- Cluster analysis
- Classical and machine learning methods for classification
- Sequence data (NGS) analysis
- Generalized linear modeling for differential expression (NGS)
- Additional topics as time permits: e.g. Meta-analysis, genome-wide association studies (GWAS)

**Keywords**

statistics; statistical methods; data analysis; DNA; RNA; mRNA; genomics; genomic data; microarray; sequencing data; NGS; NGS technologies; machine learning; R statistical software; BioConductor

**Learning Prerequisites****Important concepts to start the course**

Elementary statistics

Previous experience with R is helpful (but not necessary)

**Learning Outcomes**

By the end of the course, the student must be able to:

- Apply appropriate methods to analyze genomic data
- Carry out targeted analyses of genomic data
- Design genomic experiments

### **Transversal skills**

- Access and evaluate appropriate sources of information.
- Write a scientific or technical report.

### **Teaching methods**

Lectures and computer practical exercises

### **Expected student activities**

Regular attendance in class, practical exercises, prepare a short report (max. 10 pages) on an analysis of genomic data using tools and methods from the course

### **Assessment methods**

Evaluation is based on a written report of a genomic data analysis project.

Dans le cas de l'art. 3 al. 5 du Règlement de section, l'enseignant décide de la forme de l'examen qu'il communique aux étudiants concernés.

### **Resources**

#### **Moodle Link**

- <https://go.epfl.ch/MATH-474>