# Analysis of -omics data

Coordinator: David Bickel

Instructors: David Bickel, Alexandre Blais, Daniel Figeys, Ilya Ioshikhes, Julian Little, others TBA

## Course description

BCH5101, Analysis of –omics data (3 credits)

Theoretical and practical aspects of various methods currently used to analyze the mountain of –omics data. Methods: sequence alignment and database searches, sequence analysis and bioinformatics of gene regulation, DNA microarray and sequencing technologies to identify transcription factor binding sites, analysis of proteomics data, statistical analysis of preprocessed gene expression and protein/metabolite abundance data, epidemiology applications. Improve students' ability to critically read the literature and to make informed choices of methods for the analysis of their own data.

## Prerequisites

BCH2333 and BCH3170 or approval of coordinator

## Background and rationale

In the post-genomic era, sophisticated computational and statistical methods of analyzing transcriptomics and proteomics data are increasingly used to generate hypotheses and to draw scientific conclusions. Consequently, students need familiarity with such methods in order to critically read much of the literature and often in order to interpret their own data in graduate studies and in future research careers.

## Evaluation

The oral student seminar given at the end of the class is worth 30% of the grade, and the accompanying written report is worth 30% of the grade. Unannounced quizzes (30% of the grade) motivate students to stay updated on the assignments, which include recent review/research articles that mention the data analysis methods studied. Students may elect to work in small teams for the seminar and report but not for the quizzes. The remaining 10% of the grade is for participation since working out problems in class has proven to be an effective method of learning data analysis.

## Syllabus (tentative draft)

Each class period lasts 3 hours, and there will be a total of 13 class periods, including lecturing, problem solving, and student seminars. Tentative topics follow in the sequence of presentation, with each bullet point corresponding to approximately 1.5 hours of class time. Drs. Ioshikhes and Bickel will contribute less pending the availability of other instructors. Review articles and research articles will be assigned to give example applications of the lecture topics.

**Ilya Ioshikhes: genomics data analysis**

- Introduction to bioinformatics: main areas. Sequence and structure databases.
- Sequence comparisons I: pair-wise sequence alignment and database searches.
- Sequence comparisons II: multiple sequence alignment.
- Advanced sequence analysis: gene and promoter searches.
- Bioinformatics of gene regulation I: approaches to TFBS mapping.
- Bioinformatics of gene regulation II: miRNA; nucleosome mapping and other applications.

**Alexandre Blais: transcriptomics data analysis**

- DNA microarray and sequencing technologies to identify transcription factor binding sites.
- DNA microarray and sequencing technologies to identify transcription factor binding sites.
- Transcription factor binding site identification from large scale genomic datasets: going from genes to regulatory elements, and going from regulatory elements to genes.

**Daniel Figeys: proteomics data analysis**

- Introduction to analyzing proteomics data.
- Protein identification and quantitation.
- Methods for interpreting protein-protein interaction data measured on current platforms.
- Methods for reconstructing protein networks.

**David Bickel: metabolomics, transcriptomics, and proteomics data analysis**

- Introduction to probability and biostatistics; common mistakes.
- Parameter estimation I: single feature (only one gene, metabolite, or protein).
- Parameter estimation II: multiple features using a posterior mean.
- Hypothesis testing I: single feature (only one gene, metabolite, or protein).
- Hypothesis testing II: illocal false discovery rate for multiple features.
- Hypothesis testing III: local false discovery rate for multiple features.

**Julian Little: epidemiological and clinical issues with –omics data**

- Epidemiological issues in genome-wide association studies (GWAS) I: Basic concepts
- Epidemiological issues in GWAS II: Disease prediction and gene-environment effects
- Evaluation of “-omic” profiling I: Clinical validity
- Evaluation of “-omic” profiling II: Clinical utility