

140.668 STATISTICS FOR GENOMICS

(Biostatistics – 4th term, 3 units)

COURSE SYLLABUS

Instructor: Jeffrey Leek

Contact

Office: 615 N. Wolfe St., Rm. E3624

Office hours: By Appointment

Office Phone: 410-955-1166

Email: jleek "at" jhsph.edu

Course page: <http://www.biostat.jhsph.edu/~jleek/teaching/2011/genomics>

TA: Simina Boca

TA's Email: sboca "at" jhsph.edu

Class times

Monday: 10:30 - 11:50am

Wednesday: 10:30 - 11:50am

R Lab: Friday 10:30-11:50am

Location

Class: Wolfe W2033

Lab: Wolfe E3607

Description

The course exposes students to active research in computational genomics and introduces advanced statistical methods for solving bioinformatics problems. Topics include (1) microarray analysis: normalization, preprocessing, differential gene expression, multiple hypothesis testing, false discovery rate; (2) SNP arrays: genotyping, copy number variations, genome-wide association studies; (3) tiling arrays: ChIP-chip, model-based background correction, data segmentation, hidden Markov models, hierarchical mixture models; (4) gene regulation, epigenetics and epigenomics; (5) next-generation sequencing: ChIP-seq, RNA-seq, models and analysis; (6) Flow cytometry, FACS, normalization and clustering; (7) LC-MS proteomics, peak detection, quantification, and downstream analysis. (8) Combining technologies to understand gene regulation; (9) genomic structural aberrations including transposons, miRNA and others.

Key words

Biological: microarray, tiling array, SNP array, next-generation sequencing, ChIP-chip, ChIP-seq, flow cytometry, mass-spectrometry, gene regulation, miRNA

Statistical: multiple testing, false discovery rate, EM, Markov Chain Monte Carlo, hierarchical model, mixture model, hidden Markov model

Course Learning Objectives

Upon successful completion of this course, students will be able to (1) understand and critique existing methodology for the analysis of various microarray, massively parallel sequencing, mass spectrometry, and flow cytometry technologies (2) know the current challenges and open issues in computational genomics, (3) obtain skills to develop novel statistical approaches to study molecular biology and function.

Prerequisites

Knowledge of statistical inference (biostatistics 140.673-674 or equivalent), programming skills in R, MATLAB, C or other languages. Knowledge of advanced statistical computing (biostatistics 140.778) is a plus.

Texts

Course slides and journal papers

Grading policy

Reading Assignments: 20% - Each student must turn in their own answers to the reading problems

Attendance + Participation: 30%

Final Project: 50%

All reading assignments will be graded on a check-plus system, check for complete, plus for above and beyond.

Final projects will be graded on the basis of completeness, accuracy of applied methods, interpretation of conclusions, and format.

Schedule

Mar 28 (Mon): Course overview, molecular biology, bioinformatics, roles of statistics

Gene Expression Microarrays:

Mar 30 (Wed): Microarray analysis, normalization, preprocessing, probe effect, background correction

April 4 (Mon): Differential gene expression, gene set analysis, gene ontology

April 6 (Wed): Multiple testing, false discovery rate, and artifacts

SNP Chips:

April 11 (Mon): SNP arrays, genotyping, copy number variations

April 13 (Wed): Association studies

Tiling Arrays:

April 18 (Mon): Tiling arrays, ChIP-chip, background correction

April 20 (Wed): Tiling array cont., data segmentation, hidden Markov models

Epigenetics:

April 25 (Mon): Epigenetics, DNA methylation

Next Generation Sequencing:

April 27 (Wed): Next generation sequencing technologies, alignment, base calling

May 2 (Mon): Next generation applications: RNA-seq, Chip-Seq, Epigenetics

Flow Cytometry:

May 4 (Wed): Flow cytometry, normalization, and clustering

Proteomics:

May 9 (Mon): LC-MS proteomics, peak detection, normalization

Combining Technologies:

May 11 (Wed): Genetics of gene expression

May 16 (Mon): Combining technologies to study transcriptional regulation

Genomic Structural Aberations:

May 18 (Wed): Transposons and miRNAs