

## 140.668 SPECIAL TOPICS IN GENOMICS

(Biostatistics – 2<sup>nd</sup> term, 3 units)

### COURSE SYLLABUS

Instructor: Hongkai Ji

#### Contact

Office: 615 N. Wolfe St., Rm. E3638

Office hours: Thursday 3:00-4:00pm

Office Phone: 410-955-3517

Email: [hji@jhsph.edu](mailto:hji@jhsph.edu)

Course page: <http://www.biostat.jhsph.edu/~hji/courses/genomics/>

#### Class times

Monday 10:30 - 11:50am

Wednesday 10:30 - 11:50am

#### Location

Wolfe W4013

#### 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, hierarchical models, pooling information; (2) sequence motif discovery: EM, Markov Chain Monte Carlo, Gibbs motif sampler, *cis*-regulatory module sampler, phylogenetic footprinting; (3) tiling arrays: ChIP-chip, model-based background correction, data segmentation, hidden Markov models, hierarchical mixture models; (4) exon arrays: probe selection, gene-level expression, alternative splicing; (5) SNP arrays: genotyping, copy number variations, genome-wide association studies; (6) next-generation sequencing: ChIP-seq, RNA-seq, models and analysis; (7) gene regulation, epigenetics and epigenomics; (8) transposons, miRNA and others.

#### Key words

Biological: microarray, tiling array, exon array, SNP array, next-generation sequencing, ChIP-chip, ChIP-seq, transcription factor binding motif, 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 array and massively parallel sequencing technologies, sequence motifs, polymorphisms and miRNAs, (2) know the current challenges and open issues in computational genomics, (3) obtain skills to develop novel statistical approaches to study gene expression, gene regulation and other problems in computational biology.

## **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**

Attendance: 10%  
Participation: 30%  
Homework: 30%  
Final Project: 30%

## **Schedule**

- Oct 27 (Mon): Course overview, molecular biology, bioinformatics, roles of statistics (Ji)
- Oct 29 (Wed): Microarray analysis, normalization, preprocessing, probe effect, background correction (Irizarry)
- Nov 3 (Mon): Differential gene expression, gene set analysis, gene ontology (Irizarry)
- Nov 5 (Wed): Multiple testing, false discovery rate, pooling information (Ji)
- Nov 10 (Mon): Motif discovery, Gibbs motif sampler (Ji)
- Nov 12 (Wed): Cis-regulatory modules, module sampler, phylogenetic footprinting (Ji)
- Nov 17 (Mon): Tiling arrays, ChIP-chip, background correction (Ji)
- Nov 19 (Wed): Tiling array cont., data segmentation, hidden Markov models (Ji)
- Nov 24 (Mon): Exon arrays, probe selection, gene-level expression, alternative splicing (Ji)
- Nov 26 (Wed): SNP arrays, genotyping, copy number variations (Sharpf)
- Dec 1 (Mon): Association studies (Ruczinski)
- Dec 3 (Wed): Next-generation sequencing, data and models (Ji)
- Dec 8 (Mon): Combining technologies to study transcriptional regulation (Ji)
- Dec 10 (Wed): Epigenetics, DNA methylation (Irizarry)
- Dec 15 (Mon): Transposons and miRNAs (Wheelan)
- Dec 17 (Wed): Student presentations