

Personal Genomics for Bioinformaticians

Instructor: Melissa Gymrek

Audience: UCSD grad students (Ph.D. and M.Sc. students)

Objective: Given someone's genome, what could you learn?
(Students are encouraged to analyze their own 23&Me data).

Student backgrounds:

- Computer Science and Engineering
- Electrical Engineering
- Bioinformatics
- Biomedical Sciences

Modules:

- Introduction to personal genomics
- Ancestry
- Complex traits
- NGS
- Mutation hunting

Course Resources

Course website:

https://gymreklab.github.io/teaching/personal_genomics/personal_genomics_2017.html

Syllabus:

https://s3-us-west-2.amazonaws.com/cse291personalgenomics/genome_course_syllabus_winter17.pdf

Materials for generating problem sets:

<https://github.com/gymreklab/personal-genomics-course-2017>

Contact: mgymrek@ucsd.edu

Bioinformatics/Genomics Courses @UCLA

- <http://www.bioinformatics.ucla.edu/graduate-courses/>
- <http://www.bioinformatics.ucla.edu/undergraduate-courses/>
- All courses require 1 year of programming, a probability course
- 3 courses are combined Graduate and Undergraduate Courses
 - Undergraduate Courses enrollment are primarily students from Computer Science and Undergraduate Bioinformatics Minor (Mostly Life Science Students)
 - Graduate Enrollment is mostly from Bioinformatics Ph.D. program, Genetics and Genomics Ph.D. program, CS Ph.D. Program.
- 3 courses are only Graduate Courses
 - Enrollment is mostly Bioinformatics Ph.D. program and CS Ph.D. Program
- Courses are taught by faculty in Computer Science or Statistics (or have joint appointments there)

Bioinformatics/Genomics Courses @UCLA


- CM221. Introduction to Bioinformatics
 - Taught by Christopher Lee
 - Focuses on Classical Bioinformatics Problems and Statistical Inference
- CM222. Algorithms in Bioinformatics
 - Taught by Eleazar Eskin
 - Focuses on Algorithms in Sequencing
- CM224. Computational Genetics
 - Taught by Eran Halperin
 - Focuses on Statistical Genetics and Machine Learning Methods
- CM225. Computational Methods in Genomics
 - Taught by Jason Ernst and Bogdan Pasaniuc
 - Has component of reading recent papers
- CM226. Machine Learning for Bioinformatics
 - Taught by Sriram Sankararaman
 - Focuses on both Machine Learning Methods and Statistical/Population Genetics
- Statistics 254. Statistical Methods in Computational Biology
 - Taught by Jessica Li
 - Focuses on Statistical Methods

Algorithms in Bioinformatics Course Project

Algorithms in Bioinformatics

Order results by...



ID	Name	Genome	Chromosome	Difficulty	SNP	Indel	CopyNumber	Inversion	STR	ALU	TotalScore
91	douglasyao_UG	1	1	m	87.19	46.99	33.86	0.0	63.46	0.0	46.3
185	number of the beast	1	1	m	96.66	66.62	0.0	0.0	0.0	0.0	32.66
120	tnqls1029	1	1	m	92.06	70.2	0.0	0.0	0.0	0.0	32.45
161	K.Song	1	1	m	87.06	39.27	0.0	0.0	0.0	0.0	25.27
195	cb39	1	1	m	86.58	37.41	0.0	0.0	0.0	0.0	24.8
182	pat	1	1	m	85.03	35.87	0.0	0.0	0.0	0.0	24.18
199	Ashe Center	1	1	m	93.65	26.52	0.0	0.0	0.0	0.0	24.03
200	DSSG	1	1	m	82.42	36.51	0.0	0.0	0.0	0.0	23.78
121		1	1	m	85.08	23.38	0.0	0.0	0.0	0.0	21.69
135	nlapier2	1	1	m	82.51	25.35	0.0	0.0	0.0	0.0	21.57
189	yetti_fire_bomb	1	1	m	85.08	21.8	0.0	0.0	0.0	0.0	21.38
137	blhill	1	1	m	71.04	20.73	0.0	0.0	0.0	0.0	18.35
168	vinetty	1	1	m	72.03	0.0	0.0	0.0	0.0	0.0	14.41
176	dreilleyj	1	1	m	0.0	0.66	0.0	0.0	0.0	0.0	0.13
190	1Pie...	1	1	m	0.03	0.0	0.0	0.0	0.0	0.0	0.01

Computational Methods in Genomics

Instructor: Jason Ernst and Bogdan Pasaniuc

Audience: UCLA grad students (Ph.D. and M.Sc. students)

Objective: Introduction to current topics in bioinformatics, genomics, and computational genetics and preparation for computational interdisciplinary research in genetics and genomics

Student backgrounds:

- Bioinformatics
- Computer Science
- Other (Math, EE, Biology, Medicine)

Modules:

- Clustering and Classification
- Hidden Markov Models
- Ancestry Inference
- Sequencing to map disease genes
- Regulatory Motifs (EM, Gibbs Sampling, Deep Learning)
- Heritability and Risk Prediction
- Graphical Models

Statistical Methods for Computational Biology

Instructor: Jessica Li

Audience: UCLA grad students (Ph.D. and M.Sc. students)

Objective: How would you use statistical methods to answer biological questions?

Student backgrounds:

- Bioinformatics
- Biomedical Sciences
- Statistics
- Computer Science

Topics:

- Hypothesis Testing
- Regression and Classification
- Clustering
- Dimension Reduction
- Simulation
- Statistical Inference
- Measures of Association

Course Resources

Course website:

<https://ccle.ucla.edu/course/view/17S-STATSM254-1?section=8>

(Restricted access; permission can be requested by email)

Contact: jli@stat.ucla.edu

Fundamentals of Comp Bio: Models and Inference

Instructor: John Novembre and Matthew Stephens

Audience: Grad students

Objective: Learn the fundamental ideas and algorithms in probability and statistics necessary to design new inference methods. Be able to understand three important notable methods papers in complete detail.

Student backgrounds:

- Statistics
- Human Genetics; Ecology and Evolution
- Biophysics

Modules:

- Introduction to Graphical Models
- Intro to Likelihood & Bayesian inference
- Mixture models and EM algorithms
- Poisson Processes, Markov models, HMMs
- Gibbs sampling, MCMC
- Multivariate Normal & Gaussian Processes
- Variational inference

Course Resources

Syllabus:

<https://workflowy.com/s/DDXV.Nj4kaDHBIE>

“Five Minute Stats”: Brief intro material for many topics taught in course

<http://stephens999.github.io/fiveMinuteStats/>

Contact: jnovembre@uchicago.edu

Machine Learning for Bioinformatics

Instructor: Sriram Sankararaman

Audience: UCLA graduate students (Ph.D. and Masters)

Objective: To formulate the biological question as problems in statistical inference, to understand the assumptions and tradeoffs underlying these formulations, to develop efficient inference algorithms and to assess the quality of their inferences.

A second goal is to get students started on independent research. An open-ended research project will form a major part of the course.

Student backgrounds:

- Computer Science
- Statistics
- Bioinformatics
- Human genetics
- Electrical Engineering

Modules:

- Introduction to genomics
- Introduction to statistics
- Supervised learning: association studies, heritability
- Unsupervised learning: clustering, mixture models, PCA, admixture models, EM
- Graphical models, conditional independence, HMMs
- Kernels and neural networks
- Genomic privacy

Course website

<http://web.cs.ucla.edu/~sriram/courses/cm226.fall-2016/html/index.html>

Syllabus

<http://web.cs.ucla.edu/~sriram/courses/cm226.fall-2016/admin/syllabus.pdf>

Computational Genomics @ Tel Aviv University

- Developed by Ron Shamir. Teachers: Shamir, Sharan, Wolfson, Gat-Viks
- Mandatory in the bioinformatics BSc (double major in CS & LS + 5 bioinf courses)
- Open to seniors and grads
- Audience: CS + courageous LS students
- Emphasis: alg & stat spectrs, bio context
- 3hr lecture + 1hr recitation / week
- Grade: 50% assignments, 50% class exam



CG @TAU: Syllabus

1. Introduction
2. Pairwise alignment - theory
3. Pairwise alignment - heuristics
4. Multiple alignment
5. Phylogeny
6. HMM
7. EM
8. Motif finding
9. Gene expression analysis
10. Suffix trees
11. Structural bioinfo
12. NGS algorithms
13. Bayesian networks

Lecture notes for 20 CG lectures (400+ pp), class presentations: <http://www.cs.tau.ac.il/~rshamir/algmb/algmb-archive.htm>



Computational tools for genetics

Teacher: Sagiv Shifman

Teacher assistant: Shahar Shohat

Audience: HUJI grad students (Ph.D. and M.Sc. students)

Objective: Learn how to use R to analyze genetic data including statistical tests and graphs

Student backgrounds:

- Biology
- Some with additional background in:
 - Psychology
 - Computer science

Modules:

- Introduction to R programming (4 lessons)
- Statistical analysis with R (parametric, non-parametric, linear models, simulations, randomization, multiple testing, clustering, RNA-seq)

Course Resources

Course website:

Moodle – available only to enrolled students

Syllabus:

<http://shnaton.huji.ac.il/index.php/NewSyl/88895/2/2017/>

Contact: Sagiv.Shifman@mail.huji.ac.il

Evolution as a Learning Algorithm

Instructor: Sagi Snir, U Haifa, Israel

Audience: U Haifa, grad CS students

Background: Over the last decade, new approaches have been proposed to describe and model evolutionary processes based on theories from computational fields, especially from computational learning. Examples include Valiant's evolvability theory, the role of sex through mixability theory, and so on. This new perspective draws great attention both in the computer science theory community and in the evolutionary community and creates collaborations of their own.

Objective: The course, beyond the ability to read and understand advanced articles, will provide the student with basic tools for computational learning, algorithms, evolution, and evolutionary algorithms.

Modules: PAC Learning, Evolvability

Papers Suggested/Presented

1. *Evolvability* by Leslie G. Valiant
2. *Evolvability of Real Functions* by Paul Valiant.
3. *A mixability theory for the role of sex in evolution* by Adi Livnat, Christos Papadimitriou, Jonathan Dushoff, and Marcus Feldman
4. *Evolvability from Learning Algorithms* by Vitaly Feldman
5. *Distribution-Independent Evolvability of Linear Threshold Functions* by Vitaly Feldman
6. *Distribution Free Evolvability of Polynomial Functions over all Convex Loss Functions* by Paul Valiant
7. *Attribute-Efficient Evolvability of Linear Functions* by Elaine Angelino and Varun Kanade
8. *Evolution with Drifting Targets* by Kanade, Valiant, Vaughan
9. *Go With the Winners Algorithms* by Aldous and Vazirani
10. *Complex Adaptations and the Evolution of Evolvability* by Wagner and Altenberg
11. *Algorithms, games, and evolution* Erick Chastain, Adi Livnat, Christos Papadimitriou, and Umesh Vazirani

Contact: ssagi@research.haifa.ac.il

Advances in NGS Data Analysis & Applications

Instructor: Alex Zelikovsky

Audience: GSU grad students (Ph.D. and M.Sc. students)

Objective: Introduction to problems and current methods for NGS data analysis

Student backgrounds:

- Computer Science
- Bioinformatics/Computer Science
- Bioinformatics/Biology

Modules:

- Introduction to NGS technology
- Error correction
- Genome and Epigenome sequencing
- Differential expression of genes and pathways
- Sequencing of viral quasispecies
- Metagenomics

Course Resources

Course information:

<http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118169484.html>

Contact: alexz@cs.gsu.edu

Computational and Statistical Personal Genomics (52634)

Instructor: Or Zuk, Hebrew University Statistics department

Audience: 3rd year and graduate statistics and CS students.

Goals: In the course we will learn, implement and apply statistical methods for analyzing genetic data with emphasis on variation in human populations.

Topics:

- Introduction: basic concepts in genetics – genomes, genes, DNA -> RNA, mutations, variation
- High throughput sequencing technology: read alignment, variant calling (SNPs, indels, structural variation)
- Genome Wide Association Studies: Testing for association, risk prediction, detecting and correcting for population structure (PCA, mixed models), rare variants tests
- Population Genetics: Wright-Fisher model, the coalescent, selection
- Demography : Mixture models for populations, detecting Identity-by-descent and local ancestry (HMM), inferring demographic history from genomes

CGSI Teaching Bioinformatics Lunch - Discussion

1. What strategies do you use to teach bioinformatics to students with multidisciplinary backgrounds?
2. What compute resources/datasets have you found most useful?
3. How do you deal with increasingly large class sizes?
4. What is the most effective way to structure assignments?
5. How do you incorporate good coding practices (e.g. documentation, version control, reproducibility)?
6. How to improve representation of women and minorities in bioinformatics?

bit.ly/cgsi-teaching

Resources

- <https://gradescope.com/>
- <https://gitlab.com>
- [AWS Educate](#)