Computational Personal Genomics  
Spring 2012 Course Information  
MIT: 6.881  
Lectures: Tue, 4-6 PM, Room 32-144  
Units: 12 (2-0-10). Prerequisites: 6.047 or permission

Course staff

Lecturer:
• Manolis Kellis: 32-D524 (Stata Center), manoli@mit.edu, 617-253-2419

Teaching Staff:
• Luke Ward: 32-D526 (Stata Center), lukeward@mit.edu, 617-715-4881
• Matthew Eaton: 32-D528 (Stata Center), mleaton@mit.edu, 617-253-6079

Course secretary:
• Teresa Cataldo: 32-G475 (Stata Center), cataldo@csail.mit.edu, 617-452-5005

Course overview

This seminar course will investigate the current state of computational challenges in personal genomics. We will work together to study seminal papers in the current literature, apply existing methods to complete genomes or genotypes, and explore open problems and existing research directions. The ultimate goal is understanding what can be learned computationally from a personal genome and more generally how sequence differences between individuals lead to phenotypic differences in gene expression, disease predisposition, or response to treatment.

Each class will be divided into a discussion hour of relevant literature led by assigned students, and a practical mini-lab hour getting students started on an application of the learned techniques to complete genomes and genotypes.

Electronic resources

• Course webpage: http://compbio.mit.edu/6.881
• Papers: assigned papers will be available on the website.
• Emails: You can reach the course staff through the stellar website or directly by email.

Grading

Your grade in this course will be based on the following:

• Participation (20%)
• Mini labs (20%)
• Presentation of assigned paper (30%)
• Term project (30%)
Paper reading and participation

Each week we will read one or two articles: a member of the class will prepare a presentation and lead a discussion, and everyone is expected to participate. Upon reading the articles, each member of the class is expected to submit their reactions to the paper(s), including questions, insights and criticisms (due Monday at midnight). The presenter will receive these to fine-tune their paper presentation, or to include open questions as discussion points.

Weekly mini labs

After each paper discussion, we will hold a weekly short demo on how to apply the learned methods to existing datasets. We will help students get set up during class, and programming assignments will help students continue these after class. The mini labs will include programming or scripting for applying existing methods, and the code and results will be due by midnight the following Monday.

Term project

You will complete a final project that is planned out across the whole term. It should focus in depth on one or more of the topics discussed in the class. You may either work alone or with one partner. Teams will be expected to undertake more ambitious projects.

The most rewarding project topics are usually the most challenging (and possibly riskiest!). This might involve defining a biological problem, identifying relevant datasets, designing and implementing new algorithms, applying the methods, and interpreting the results. Alternatively, you can select a less risky project such as comparing several computational biology algorithms for solving the same problem and implementing, applying, and rigorously evaluating the results. You might also analyze a relevant conference or journal article, including criticism, corrections, and/or improvements. An element of the final project grading will depend on the challenge and originality of your project, so spend enough time to choose carefully.

Project proposals will be due on Monday 3/12 at midnight, and the final projects are due Wednesday 5/16 at 11:59 PM, with in-class presentations on Thursday 5/17 at 4 PM. The goal of the proposal is to get you thinking about your project and to make sure it is appropriate both in its subject matter and its complexity. The goal of the final project is to get you to dive more in depth into one of the topics of the course towards independent research in personal genomics.

Textbooks

There are no required textbooks for the course. There are some resources available through the MIT library that may help as an independent outside source for any points that need clarification:


Collaboration policy

You are encouraged to collaborate on both the mini labs and the final project. However:

- You must work independently on each problem before you discuss it with others.
- You must write the solutions on your own.
• You must acknowledge your collaborators and outside sources.

Schedule

What follows is an approximate schedule for the class. The topics and reading assignments are subject to change based on the needs of the class.

<table>
<thead>
<tr>
<th>Week &amp; topic</th>
<th>Papers</th>
<th>Programming project</th>
</tr>
</thead>
</table>
• 1000 Genomes Consortium (2010) A map of human genome variation from population-scale sequencing. | Calculate LD over a small stretch of the genome. |
| 2) 2/14 – Overview of medical genetics: from linkage studies to GWAS to WGS. | • Lander and Schork (1994) Genetic dissection of complex traits.  
• Altshuler et al. (2008) Genetic mapping in human disease.  
• Choi et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. | Find segments of IBD between relatives. |
| 3) 2/28 – Ancient selection: mammalian comparative genomics. | • Ng and Henikoff (2006) Predicting the effects of amino acid substitutions on protein function.  
• Lindblad-Toh et al. (2011) A high-resolution map of human evolutionary constraint using 29 mammals. | Use chimp to label your SNPs as ancestral/derived; are there genes where the derived alleles cluster? |
| 4) 3/6 – Human origins, introgression during speciation: Chimps, Denisovans, Neanderthal. | • Patterson et al. (2006) Genetic evidence for complex speciation of humans and chimpanzees.  
• Reich et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia.  
• Green et al. (2010) A draft sequence of the Neandertal genome. | Count your Neandertal/Denisovan alleles; in which genes do they lie? Are they randomly distributed? |
<table>
<thead>
<tr>
<th>Week &amp; topic</th>
<th>Papers</th>
<th>Programming project</th>
</tr>
</thead>
</table>
- Price et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies.  
- Gravel et al. (2011) Demographic history and rare allele sharing among human populations. | Perform PCA on Hapmap individuals; paint ancestry on a genome of an individual from an admixed population. |
- Yi et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. | Discover ancestry-informative or highly-differentiated alleles; does your genome carry mutations there not typical of your population? |
- Gibson (2012) Rare and common variants: twenty arguments.  
- Zuk et al. (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. | Use the GWAS catalog or other tools to determine your relative risk for disease. |
- Sakabe et al. (2012) Transcriptional enhancers in development and disease. | Use predicted active regulator motifs and HaploReg to see which genes are downstream of your homozygous conserved enhancer motif disruptions. |
<table>
<thead>
<tr>
<th>Week &amp; topic</th>
<th>Papers</th>
<th>Programming project</th>
</tr>
</thead>
</table>
| 9) 4/24 – Genetics of regulation: eQTLs, cQTLs, mQTLs. | • Pickrell et al. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing.  
• Kasowski et al. (2010) Variation in transcription factor binding among humans. | Predict the expression level of genes in one of your cell types, relative to those of other individuals. |
| 10) 5/1 – Pathway analysis of GWAS. | • Lango et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height  
• Xu et al. (2011) Exome sequencing supports a de novo mutational paradigm for schizophrenia. | Calculate the genetic contribution to your height; calculate your exome or enhancer mutational burden |
| 12) 5/17 – Final presentations | N/A | N/A |