Navigating the genome:
Interpreting the functional impact of genetic variation
BIO 425/525

Spring 2018, 2 credits, Tuesdays and Thursdays 11am - 12.10pm, @Comp Lab
gokcumenlab.org/Bio-425525

Instructor

Omer Gokcumen, Assistant Professor (omergokc@buffalo.edu), **gokcumenlab.org**
Office Hours: Thursday 3.30pm - 5pm or by appointment @Cooke 641

Course Description

The primary goal of this course to teach the students how to navigate in the several recent -omics datasets, including genome browsers, large multi-tissue transcriptional datasets, human genetic variation maps, as well as human disease association studies. Specifically, throughout the course, the students will learn to use several available computational tools to visualize overall genetic variation in humans, identify a subset of variants relevant to a particular disease/phenotype/adaptation, and excavate the potential functional impact of such variation. The course directly complements existing courses offered in Biological Sciences, including *Human Evolutionary Genomics*, *Bioinformatics*, *Human Genetic Disease*, among others. The course is aimed at students in biological sciences and related fields with a keen interest in human genetic variation and its biomedical and evolutionary impact of human genetic variation.

Course Objectives

After taking this course, the student will be able to:
- Learn to navigate multiple -omics browsers and datasets, including human reference genome, 1000 Genomes Project database, ENCODE Project database, GTeX browser, Human Selection Browser, GALAXY, among others.
- The students will learn and apply fundamental statistical concepts of genome-wide studies, especially multiple hypothesis correction and enrichment analyses.
- Get accustomed to the extent of human genetic variation and its functional impact.
- Understand the modern genetic methodologies through which genetic variation is linked with biological function.

Prerequisites: Bio 319 - Genetics or permission of the instructor.

Texts and Reading Materials

Links to relevant scientific papers, databases and computational tools will be provided on the course website. The material is mainly methodological papers, describing the databases used in the class. Human Evolutionary Genetics" by Jobling et al. is recommended as a companion book covering more theoretical aspects of the course. This book is **not** required. However, it is a good book and explains some of the rather dense concepts in a straightforward and clear manner. It is available through the bookstore, and a copy of it will be reserved in the library for the class for reference. For those who are interested in further reading: "Gene: An intimate history" by Siddharta Mukherjee is a nice review of the history of medical genetics.

Outcome Measures and Grading

At the heart of this course will be hands-on in-class exercises on computers. The main out multiple homework assignments will be given, which are mostly based on replication of in-class exercises with different subset of genetic variants.

- **75%** of the grade will be based on the results of 4 homework assignments, where the students are required to prepare a report of an analysis they will make on a given problem (e.g., the functional impact of a variant, enrichment analysis of a subset of conserved genetic elements, etc.).

- **25%** of the grade will be based on unannounced quizzes before the discussion to measure the preparedness of the students. The questions in these quizzes will involve main points highlighted in the assigned papers for that day or main themes highlighted during previous in-class discussions. The worst quiz score will be discarded for final assessment.

- *Reference to the university undergraduate Incomplete Policy and any additional instructor requirements and comments regarding incomplete grades.*

- *If you require classroom or testing accommodations due to a disability, please contact Accessibility Resources, located at 25 Capen Hall. AR can be reached by phone at (716) 645-2608 or by email at stu-accessibility@buffalo.edu. Please inform me as soon as possible about your needs so that we can coordinate your accommodations.*

- The grades will not be given on a curve. If everybody does great work, than everybody will get an A. **Remember**, this is a high level course. **WORK HARD!**

The letter grades will be assigned in the following manner:

94-100%=A; 85-93%=A-; 75-84%=B+, 65-74%=B, 60-64%=B-, 55-59%=C+, 50-54%=C, 45-49%=C-, 40-44=D+, below 40%=D-. Non Attendance or extreme poor performance will lead to an F.

Attendance and late submissions

- The attendance is mandatory and consistent absence will result in failure. This is an applied class. As such, without attendance, it has no value.
- For every day after the deadline of a homework 10% penalty will be applied to the grading of that paper. Submissions more than 3 days will not be accepted without a legitimate, solid excuse.

Integrity

If you plagiarize or cheat, you will directly receive an F for the course. There is no tolerance for this particular issue. Please see this webpage, if you have questions - http://undergrad-catalog.buffalo.edu/policies/course/integrity.shtml

Learning outcomes

Learn to navigate multiple -omics browsers and datasets, and analysis tools.	- Participation in class exercises. - Homeworks.
Learn and apply fundamental statistical concepts of genome-wide studies, especially multiple hypothesis correction and enrichment analyses.	- Class readings - Participation in class exercises. - Homeworks. - In-class Quizzes
Get accustomed to the extent of human genetic variation and its functional impact.	- Class readings - Participation in class exercises. - Homeworks. - In-class Quizzes
Understand the modern genetic methodologies through which genetic variation is linked with biological function.	- Class readings - Participation in class exercises. - Homeworks. - In-class Quizzes

Course Schedule

Week	Tuesday	Thursday	Reading
Jan 30, Feb 1	What makes a genome?	*Repeats and intergenic regions: Introduction to UCSC Genome Browser?*	(*1*)
Feb 6, 8	Locating genes in the genome? Gene annotation processes/databases.	In-class exercise: Finding chromosomal location, size, exonic structure and isoforms of a disease-related gene.	(*2*)
Feb 13, 15	In class exercise: Visualizing next generation sequencing using Integrated Genome Browser.	In-class exercise: Visualizing a neanderthal haplotype block in a modern genome.	(*4*)
Feb 20, 22	Human genetic variation? In class exercise: Visualizing 1,000 Genomes Project data.	Finding variants that are linked to human disease. Mendelian disorders. dbVar. In class exercise: locating extremely rare variants using EXaC database.	(*5*)
Feb 27, Mar 1	Intro to GWAS - In class exercise: Multiple hypothesis testing	Finding variants that are linked to human disease. Understanding GWAS - GWAS database.	(*6*)
Mar 6, 8	Genomic structural variation (Database of genomic variation).	Overlap analyses - Intro to GALAXY.	(*7*)
Mar 13, 15	In-class exercise - Enrichment/depletion of exonic sequences with a subset of genetic variants.	Finding links to function: Enrichment analyses. Intro to Gene ontology and Pathways.	(*8*)
Mar 20, 22	In-class exercise - Finding expression patterns of a gene in multiple tissues and developmental stages.	Intro to ENCODE: In class exercise: Visualizing different methylation signals across the genome.	(*9*)
Mar 27, 29	Deviation from neutrality: Intro to selection browsers.	In class exercise: Subsetting based on a combination of properties (e.g., common variants that affect expression in liver).	(*10*)
Apr 3, Apr 5	Intro to microbiome databases.	Wrap-up and discussion of systems level analyses.	(*11*)

REFERENCES

1. J. Wang, L. Kong, G. Gao, J. Luo, A brief introduction to web-based genome browsers. *Brief. Bioinform.* **14**, 131–143 (2013).

2. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

3. ENCODE Project Consortium *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**, 57–74 (2012).

4. K. Prüfer *et al.*, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. **505**, 43–49 (2014).

5. 1000 Genomes Project Consortium *et al.*, A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).

6. D. Welter *et al.*, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).

7. B. Giardine *et al.*, Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451–1455 (2005).

8. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

9. M. B. Gerstein *et al.*, Architecture of the human regulatory network derived from ENCODE data. *Nature*. **489**, 91–100 (2012).

10. K. Lindblad-Toh *et al.*, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. **478**, 476–482 (2011).

11. Human Microbiome Project Consortium, Structure, function and diversity of the healthy human microbiome. *Nature*. **486**, 207–214 (2012).