BIMM 104: **Genome Informatics**

Course Instructor: Dr. Barry J. Grant (bjgrant@umich.edu)
Class Website: https://bioboot.github.io/bimm104_w18/

Overview: This upper division 4-unit course covers modern developments in genomics with an emphasis on the informatics foundations of current genomic science.

Major topic areas include: Advances in high-throughput sequencing technologies; Bioinformatic analysis methods and workflows; Genome re-sequencing and variation analysis; Genome wide association studies; Transcriptomics; Comparative genomics; Pharmacogemomics; and Personal genomics. Two guest lectures from genomic scientists at Illumina Inc., Synthetic Genomics Inc., Human Longevity Inc., or the La Jolla Institute for Allergy and Immunology will feature subjects chosen by student voting preferences.

Students will also have the opportunity to discuss the social impacts and ethical implications of how genomic sequence information is used in our society. Example student led presentation topics include: The importance of genomic data de-identification. How to best balance privacy concerns with the desire to share and re-use data generated by taxpayer funded research? Should doctors know and preemptively act based on our genetic disease predispositions? Do YOU want to know your genetic disease predispositions? Should patients be given access to their own data and be able to share it? Should my child be 'tracked' in school according to her DNA sequence? How are racial DNA differences impacting racial prejudices?

To fully participate in this course students will be expected to use their own laptop computers during weekly discussion groups to access bioinformatics software and data resources that are freely available online.

Recommended preparation: BICD 100 (Genetics) is highly recommended and may be taken concurrently. This course is designed for biology majors with no programing or advanced computing skills. **Prerequisites**: BILD 1 (The Cell).

Schedule: Lectures are on Tuesday and Thursday at 10:30 - 11:50 am in Center Hall, University Center. (Map Bldg #984). Discussion groups meet once a week for 50 mins in Galbraith Hall, Revelle. (Map Bldg #141). A detailed lecture schedule with topic outlines is provided below.

Class announcements: All announcements regarding the course will be by email to your UCSD address.

Office hours & location: TBD – For now email me for a time and we will make it happen.

Textbook: There is no textbook for the course. **Lecture notes, homework assignments, grading criteria, pre-class screen casts** and required **reading material** will be available from our public facing course website: https://bioboot.github.io/bimm104_w18/

Lecture Schedule (Draft, with number of topics to be reduced):

Winter 2018 BIMM 104: **Genome Informatics** Lectures (TuTh) 10:30 - 11:50 am		
Tu, 01/09	**Welcome to *Genome Informatics*** (Course Introduction, instructional approach, leaning goals & expectations)	1
Th, 01/11	**What is a genome?** (Primer on key concepts and vocabulary including genome replication, genes, exons/introns/splicing, transcription, nucleosomes and repetitive sequences)	2
Tu, 01/16	**Sequencing technologies past, present and future.** (Sanger, Shotgun, PacBio, Illumina, toward the $500 human genome)	3
Th, 01/18	**Major bioinformatics resources for genomics.** (Databases, tools and visualization resources from NCBI, EBI & UCSC).	4
Tu, 01/23	**Alignment method foundations of genomic analysis 1** (Classic Needleman-Wunsch, Smith-Waterman and BLAST heuristic approaches)	5
Th, 01/25	**Alignment method foundations of genomic analysis 2** (Short read aligners, indexing and working with high-throughput sequencing data)	6
Tu, 01/30	**Genomic analysis workflows** (The Galaxy platform for quality control and analysis; FASTQ, SAM and BAM file formats; Sample workflow with FASTQC and bowtie2)	7
Th, 02/01	**Genome assembly** (The Genome Reference Consortium; *De novo* genome sequencing and genome assembly)	8
Tu, 02/06	**Genome annotation** (Genome annotation, gene finding and functional annotation)	9
Th, 02/08	**Genome re-sequencing and variation** (Aligning to reference genomes; SNP and indel calling; Structural and Copy Number Variations; Germline vs Somatic variants; Population vs Personal variants)	10
Tu, 02/13	**Mid Term: Find a gene project assignment** (Principles of database searching and sequence analysis)	11
Th, 02/15	**Genome wide association studies (GWAS)** (Odds ratios, Manhattan plots, SNP arrays and imputation)	12
Tu, 02/20	**Transcriptomics** (RNA-Seq aligners, counts and FPKMs, differential expression tests)	13
Th, 02/22	**Transcriptional regulation** (ChIP-Seq, peak calling, eQTLs, ENCODE project highlights and criticisms, transcription factor binding site identification)	14
Tu, 02/27	**Interpreting gene lists** (Functional databases KEGG, InterPro, GO ontologies and functional enrichment)	15

Th, 03/01	**Comparative genomics** (Clustering, ortholog identification, PhyloP and PhastCons measures of conservation, power of model organisms)	16
Tu, 03/06	**Pharmacogenomics** (Mapping genome variants to drug pharmacodynamics, pharmacokinetics and adverse reactions, PharmGKB) **N.B.** Find a gene assignment due today!	17
Th, 03/08	**Guest lecture 1** (Student selected topic from: **Metagenomics** / **Epigenomics** / **Personal genomics** / **Genome evolution** / **Genome editing and synthetic genomics** / **Social impacts and ethical implications of continuing genomic advances**)	18
Tu, 03/13	**Guest lecture 2** (Second student selected industry based genomic scientist presentation*)	19
Th, 03/15	**Course summary** (Summary of learning goals, student course evaluation time and exam preparation.	20
Th, 03/22	**Final exam!**	

Course Objectives:
At the end of this course students will:

- Understand the process by which genomes are currently sequenced and the bioinformatics processing and analysis required for their interpretation.
- Be familiar with the research objectives of the genomics related sub-disciplines of Transcriptomics, GWAS, Comparative genomics, and Pharmacogenomics.
- Be able to use and evaluate online bioinformatics resources (including major genomic databases, genome browsers and select quality control and analysis tools) to solve problems in genome science.

In short, students will develop a solid foundational knowledge of genome informatics and be able to evaluate new genomic information using online bioinformatic tools and resources.

Specific Learning Goals
Teaching toward the specific learning goals below is expected to occupy 60%-70% of class time. The remaining course content is at the discretion of the instructor with student body input. This includes student selected topics for peer presentation as well as two student selected guest lectures from industry based genomic scientists.

All students who receive a passing grade should be able to:

1. Appreciate and describe in general terms the rapid advances in sequencing technologies and the new areas of investigation that these advances have made accessible.
2. Productively use major bioinformatics resources for human and model organism genomic data at NCBI, EBI and UCSC.
3. Understand that sequence alignment is the most fundamental operation underlying genome informatics and indeed much of bioinformatics.
4. Be able to describe how dynamic programming works for pairwise sequence alignment and appreciate the differences between global and local alignment along with their major application areas.
5. Understand how genomes are annotated and genes predicted using bioinformatics approaches.
6. Employ bioinformatics methods via the Galaxy server framework and use online tools to interpret gene lists and annotate potential gene functions.
7. Understand how single-nucleotide polymorphisms (SNPs) and indels are identified. Appreciate the existence of structural variants and copy number variations along with their potential significance.
8. Be able to describe how transcripts are analyzed and their abundance quantified by RNA-Seq and related approaches.
9. Appreciate how transcription factor binding and histone modifications can be studied by ChIP-Seq.
10. Justify the value of studying many genomes and comparing the genomes of different organisms.
11. Be able to describe examples of how one person's reaction to a drug may be different from another's depending on their genetic makeup.
12. Have an understanding of the recent developments in targeted genome editing and synthetic genomics along with their current and potential future applications (*Guest lecture dependent*).
13. Have an appreciation for the social impacts and ethical implications of how genomic sequence information is used in our society (*Discussion group and presentation activities*).
14. Have an appreciation for the high level of demand for personnel with genome informatics skills and the high level of excitement surrounding genomics in both academia and industry.

Homework assignments, mid-term project, end of term presentation and final exam:
Weekly homework will consist of online knowledge assessment quizzes and application assignments together with pre-class reading and video screen-casts.

Specific grading criteria (assessment rubrics) for each week will be given at the time of assignment. Weekly grades will be posted on BlackBoard. Each student is responsible for checking BlackBoard to ensure that a grade has been entered for their submissions. Documents submitted by email do not always arrive at their intended destination and late submissions will

not be accepted after one week past the original due date. Collectively homework performance will account for 35% of the course grade.

A total of 20% of the course grade will be assigned based on both the mid-term "*find-a-gene project assignment"* (10%) and end of term "*social impacts and ethical implications presentation assignment*" (10%). The purpose of the mid-term assignment is for you to grasp the principles of database searching and sequence analysis that we cover in the course (see additional details online). For the later assignment each student will prepare a 5-10 PowerPoint slide presentation on a topic of personal interest related to the ethical implications of genomic data availability. This presentation will be uploaded to a discussion section on Blackboard and students are required to post at least two questions to the discussion. Further details will be given in class.

There will be one final exam that accounts for 45% of the final grade for the course.

Ethics Code.
You are encouraged to collaborate with your fellow students. However, all material submitted to the instructor must be your own work.

> *"Cheating is wrong. Cheating hurts our community by undermining academic integrity, creating mistrust, and fostering unfair competition. The university will punish cheaters with failure on an assignment, failure in a course, permanent transcript notation, suspension, and/or expulsion. Offenses may be reported to medical, law, or other professional or graduate schools when a cheater applies.*
>
> *Violations can include cheating on exams, plagiarism, reuse of assignments without permission, improper use of the Internet and electronic devices, unauthorized collaboration, alteration of graded assignments, forgery and falsification, lying, facilitating academic dishonesty, and unfair competition. Ignorance of these rules is not an excuse".*

You can read a full description of the ethics code at [Ethics URL here].