

Overview & Syllabus

Statistics, the science of data analysis, is *the* applied mathematics of the 21st century...and genomics is the new physics. This era is characterized by massive data sets, automated measurement, instant communication, and raw computational power. We are in the midst of an intellectual ferment comparable to the dawn of quantum mechanics. Yesterday it was line spectra and blackbody distributions which upset the tidy Newtonian applecart; today it is sequence data and array technology which is toppling theory after cherished theory.

If we look more closely, we see that the birth of this new science is intimately tied to the ability to measure that which was previously unmeasurable. New sequencing and assay technologies have revolutionized biology by giving us billions of bits of data in hours rather than tens of bits in years. In one sense, it is no surprise that measurement birthed genomics: new tools for quantitative measurement are always the point of the spear in any discipline, as it is data analysis which is at the center of the scientific enterprise. Whether we have in mind Chargaff's rule or Hund's, simply counting that which no one had thought to count before is the surest way to a new scientific insight.

It is here that statistics enters. Counting leads to combinatorics, fractions of counts lead to probabilities, and counts of counts – histograms – lead inexorably to distributions. While counting sounds trivial, it is not. If you wish to count multiple things at a time, to massage data into countable form, or to visualize that which you've counted, you must know how to program.

My main objectives in this class are to introduce you to tools for the analysis of huge biological data sets and to give you experience with a number of different problems at the forefront of research. The course will be divided into four segments: tools, comparative genomics, functional genomics, and population genomics. For each unit, we will have some fantastic guest lecturers this quarter: first authors of papers in *Science*, *Nature*, *PNAS*, and other prestigious journals. Note that this syllabus is not set in stone; depending on the level of feedback I may speed things up or slow things down.

There will be four two week homework assignments (10% each), one for each unit. Each assignment will involve programming. We will not gloss over technical details when relevant, but the focus will be the application of statistics and computation to biology. There will also be a final project for which you will give a 10-15 minute presentation (depending on class size) in the last week of class, along with a short 2-3 page paper with one figure. This paper will be due on the last day of finals week. The final presentation and paper will be worth about 30% each. The latest version of the syllabus, homeworks, and lecture notes will be available at the class webpage, which is jinome.stanford.edu/stat366.

Date	Room	Material	Unit	Assignment
Jan. 9	S360	Introduction	Tools & Basic Bioinfo.	HW1A Due Jan. 18
Jan. 11	S361	Unix, Perl		
Jan. 16	S360	R, Databases, Pairwise Alignment, BLAST		HW1B Due Jan. 25
Jan. 18	S361	Phylogeny, Multiple Alignment, PSSMs		
Jan. 23	S361 9:30 am	HMMs, Gene Finding, Motifs	Comparative Genomics	HW2 Due Feb.8
Jan. 25	S361	Modern Directions in Alignment (Tom Do)		
Jan. 30	S360	ENCODE Project (George Asimenos)		
Feb. 1	S361	Applications of Genome Alignment: Motif Finding, Evolutionary Constraint, Ancestral Reconstruction		
Feb. 6	S361	Functional Genomic Assays, Interaction Networks, Basic Graph Theory, Random Networks	Functional Genomics	HW3 Due Mar. 1
Feb. 8	-	NO CLASS (BMI Retreat)		
Feb. 13	S361	Network Integration & Functional Enrichment		
Feb. 15	S361	Protein Recommender & Network Alignment (Jason Flannick)		
Feb. 20	S361	Stochastic Genetic Circuits; Data-Driven Circuit Inference		
Feb. 22	S361	Deterministic Genetic Circuits (Onn Brandman)		
Feb. 27	S361	Principles of Molecular Evolution	Population Genomics	HW4 Due Mar. 15
Mar. 1	S361	Tests for Selection & Modern Pop. Gen. (Eric Wang)		
Mar. 6	S361	Haplotype Map, Structural & Expression Variation		
Mar. 8	S362	Pharmacogenomics, Whole Genome Association		
Mar. 13	S361	Final Presentations	-	Final Paper Due Mar. 23
Mar. 15	S361	Final Presentations	-	