

GGSB COMPUTATIONAL TRACK - COURSES

For detailed information on course time schedules visit the [Academic Information System](#)

For additional information please view the [Doctoral Training in Computational Genomics](#) website

GGSB COMPUTATIONAL TRACK: THREE [3] REQUIRED COURSES AND THREE [3] CORE ELECTIVES PLUS TWO [2] ADDITIONAL ELECTIVES PLUS TWO [2] ROTATIONS

THREE [3] REQUIRED COURSES:

STAT 24400 Statistical Theory and Methodology I. This sequence is a systematic introduction to the principles and techniques of statistics, as well as to practical considerations in the analysis of data, with emphasis on the analysis of experimental data. This course is the first quarter of a two-quarter systematic introduction to the principles and techniques of statistics, as well as to practical considerations in the analysis of data, with emphasis on the analysis of experimental data. This course covers tools from probability and the elements of statistical theory. Topics include the definitions of probability and random variables, binomial and other discrete probability distributions, normal and other continuous probability distributions, joint probability distributions and the transformation of random variables, principles of inference (including Bayesian inference), maximum likelihood estimation, hypothesis testing and confidence intervals, likelihood ratio tests, multinomial distributions, and chi-square tests. Examples are drawn from the social, physical, and biological sciences. The coverage of topics in probability is limited and brief, so students who have taken a course in probability find reinforcement rather than redundancy. Students who have already taken STAT 25100 may choose to take STAT 24410 (if offered) instead of STAT 24400. Students taking either STAT 24400 or STAT 24410 will have appropriate preparation for STAT 24500. *Autumn, Winter.*

AND

HGEN 48600 Fundamentals of Computational Biology: Models and Inference. Covers key principles in probability and statistics that are used to model and understand biological data. There will be a strong emphasis on stochastic processes and inference in complex hierarchical statistical models. Topics will vary but the typical content would include: Likelihood-based and Bayesian inference, Poisson processes, Markov models, Hidden Markov models, Gaussian Processes, Brownian motion, Birth-death processes, the Coalescent, Graphical models, Markov processes on trees and graphs, Markov Chain Monte Carlo. PQ: STAT 244 or equivalent. *Winter*

AND

HGEN 48800 Fundamentals of Computational Biology: Algorithms and Applications. This course will cover principles of data structure and algorithms, with emphasis on algorithms that have broad applications in computational biology. The specific topics may include dynamic programming, algorithms for graphs, numerical optimization, finite-difference, schemes, matrix operations/factor analysis, and data management (e.g. SQL, HDF5). We will also discuss some applications of these algorithms (as well as commonly used statistical techniques) in genomics and systems biology, including genome assembly, variant calling, transcriptome inference, and so on. *Spring*

PLUS THREE [3] CORE ELECTIVES CHOSEN FROM THE FOLLOWING LIST

HGEN 47000 Human Genetics I. This course covers classical and modern approaches to studying cytogenetic, Mendelian, and complex human diseases. Topics include chromosome biology, single gene and complex diseases, non-Mendelian inheritance, cancer genetics, human population genetics, and genomics. The format includes lectures and student presentations. *Autumn*

OR

MGCB 31400 Genetic Analysis of Model Organisms. Fundamental principles of genetics discussed in the context of current approaches to mapping and functional characterization of genes. The relative strengths and weaknesses of leading model organisms are emphasized via problem-solving and critical reading of original literature. *Autumn.*

OR

HGEN 47100 Introductory Statistical Genetics. This course focuses on genetic models for complex human disorders and quantitative traits. Topics covered also include linkage and linkage disequilibrium mapping genetic models for complex traits, and the explicit and implicit assumptions of such models. *Winter.*

OR

ECEV 35600 Population Genetics I. Kreitman, Steinrücken. Examines the basic theoretical principles of population genetics, and their application to the study of variation and evolution in natural populations. Topics include selection, mutation, random genetic drift, quantitative genetics, molecular evolution and variation, the evolution of selfish genetic systems, and human evolution. *Spring.*

OR

HGEN 31100 Evolution of Biological Molecules. The course connects evolutionary changes imprinted in genes and genomes with the structure, function and behavior of the encoded protein and RNA molecules. Central themes are the mechanisms and dynamics by which molecular structure and function evolve, how protein/ RNA architecture shapes evolutionary trajectories, and how patterns in present-day sequence can be interpreted to reveal the interplay data of evolutionary history and molecular properties. Core concepts in macromolecule biochemistry (folding and stability of proteins and RNA, structure-function relationships, kinetics, catalysis) and molecular evolution (selection, mutation, drift, epistasis, effective population size, phylogenetics) will be taught, and the interplay between them explored. *Winter*

OR

BCMB 32200 Biophysics of Biomolecules. This course covers the properties of proteins, RNA, and DNA, as well as their interactions. We emphasize the interplay between structure, thermodynamics, folding, and function at the molecular level. Topics include cooperativity, linked equilibrium, hydrogen exchange, electrostatics, diffusion, and binding. *Spring*

OR

HGEN 46900 Human Variation and Disease. This course focuses on principles of population and evolutionary genetics and complex trait mapping as they apply to humans. It will include the discussion of genetic variation and disease mapping data. *Spring.*

OR

HGEN 47300 Genomics and Systems Biology. This lecture course explores technologies for high-throughput collection of genomic-scale data, including sequencing, genotyping, gene expression profiling, and assays of copy number variation, protein expression and protein-protein interaction. In addition, the course will cover study design and statistical analysis of large data sets, as well as how data from different sources can be used to understand regulatory networks, i.e., systems. Statistical tools that will be introduced include linear models, likelihood-based inference, supervised and unsupervised learning techniques, methods for assessing quality of data, hidden Markov models, and controlling for false discovery rates in large data sets. Readings will be drawn from the primary literature. Evaluation will be based primarily on problem sets. *Spring.*

OR

MGCB 32000 Quantitative Analysis of Biological Dynamics. The basic focus of the course will be quantitative approaches to understanding organization and dynamics at the molecular, subcellular and cellular levels, and will rest on three pillars - modern imaging and image analysis, quantitative analysis and presentation of data, mathematical modeling and computer simulations *Spring*

ADDITIONAL REQUIRED COURSES

GRADED LAB ROTATIONS TO BE TAKEN IN WINTER, SPRING & SUMMER QUARTERS:

BSDG 40100 Non-Thesis Research: Biological Sciences. Laboratory rotation – (10 weeks) *Winter, Spring*

BSDG 40100 Non-Thesis Research: Biological Sciences. Laboratory rotation – (5 weeks) *Summer.*

BSDG 40102 Non-Thesis Research: Biological Sciences. Third Laboratory rotation (Optional) (Second 5 weeks of quarter) *Summer.*

ADDITIONAL GGSB REQUIREMENT:

GENE 31800 Current Seminar Topics in Genetics. This course will expose student to current research topics in genetics for the bi-monthly GGSB Invited Seminar Series. **This is a required ½ credit course for all GGSB students and will be graded Pass/Fail.** *Autumn, Winter, Spring*

ADDITIONAL DIVISIONAL REQUIREMENTS:

GENE 31900 Introduction to Research (Allstars). Lectures on current research by departmental faculty and other invited speakers. A required course for all first-year graduate students. *Autumn*

BSDG 55100 Responsible, rigorous, and reproducible conduct of research: R3CL. Required of all BSD first-year doctoral students. The course is designed to stimulate thinking and facilitate discussion about the purpose and necessity of ethical conduct with respect to scientific and academic practices; to create personal awareness of the ethical dilemmas and choices that may be encountered in the course of a career in the sciences; to increase awareness and understanding of the importance of reproducible, rigorous, and transparent research; and to provide practical information regarding policies and procedures related to conduct in the Division of Biological Sciences at the University of Chicago. Winter quarter only. *Winter*

MGCB 32100 Senior Graduate Ethics. A second training in the ethical conduct of research is required for students still registered four years after their initial training. Senior ethics training content is more closely aligned with research areas and so this training is coordinated by the individual graduate programs.

PLUS TWO [2] ADDITIONAL ELECTIVES CHOSEN FROM THE FOLLOWING LIST

BIOS 20186 Fundamentals of Cell and Molecular Biology. This course is an introduction to molecular and cellular biology that emphasizes the unity of cellular processes amongst all living organisms. Topics are the structure, function, and synthesis of nucleic acids and protein; structure and function of cell organelles and extracellular matrices; energetics; cell cycle; cells in tissues and cell-signaling; temporal organization and regulation of metabolism; regulation of gene expression; and altered cell functions in disease states. *Spring*

OR

STAT 34300 Applied Linear Statistical Methods. This course introduces the theory, methods, and applications of fitting and interpreting multiple regression models. Topics include the examination of residuals, the transformation of data, strategies and criteria for the selection of a regression equation, nonlinear models, biases due to excluded variables and measurement error, and the use and interpretation of computer package regression programs. The theoretical basis of the methods, the relation to linear algebra, and the effects of violations of assumptions are studied. Techniques discussed are illustrated by examples involving both physical and social sciences data. *Autumn*

OR

STAT 37790 Topics in Statistical Machine Learning "Topics in Statistical Machine Learning" is a second graduate level course in machine learning, assuming students have had previous exposure to machine learning and statistical theory. The emphasis of the course is on statistical methodology, learning theory, and algorithms for large-scale, high dimensional data. The selection of topics is influenced by recent research results, and students can take the course in more than one quarter. *Autumn*

OR

CMSC 35490 Special Topics in Machine Learning. Learned emulators leverage neural networks to increase the speed of physics simulations in climate models, astrophysics, high-energy physics, and more. Recent empirical results have illustrated that these emulators can speed up traditional simulations by up to eight orders of magnitude. However, little is understood about these emulators. While it is possible that recent results are representative of what is possible in most settings, a more likely scenario is that these approaches are more effective for some simulators than others, and that learned emulators achieve strong average-case performance but fail to capture rare but important phenomena. In this graduate seminar course we will provide an overview and investigate recent literature on this topic, focusing on the following questions: 1. Introduction to learned

emulators: how do they work, where have they been successful so far and what are the goals in this field? 2. Two different paradigms of learned emulation: physics vs. data driven. What are the advantages and pitfalls of each? 3. Robustness of emulation to noise: what is known so far? 4. Parameter estimation: how to handle parameter uncertainty? We will provide a list of papers covering the above topics and students will be evaluated on in-class presentations. *Autumn*

OR

CMSC 37720 Computational Systems Biology. Introductory concepts of systems biology, computational methods for analysis, reconstruction, visualization, modeling and simulation of complex cellular networks including biochemical pathways for metabolism, regulation and signaling. Students will have the opportunity to explore systems of their own choosing and will participate in developing algorithms and tools for comparative genomic analysis, metabolic pathway construction, stoichiometric analysis, flux analysis, metabolic modeling and cell simulation. A particular focus of the course will be on furthering our understanding of the computer science challenges in the engineering of prokaryotic organisms. The course requires written assignments, programming assignments and a final course project. *Autumn* .

OR

ECEV 32000 Introduction to Scientific Computing for Biologists. The course will cover basic concepts in computing for an audience of biology graduate students. The students will receive basic training in the use of version control systems, databases and regular expressions. They will learn how to program in python and R and how to use R to produce publication-grade figures for their manuscripts, and how to typeset scientific manuscripts and theses using LaTeX. All the examples and exercises will be biologically motivated and will make use of real data. The approach will be hands-on, with lecturing followed by exercises in class. *Winter*

OR

STAT 30900. Mathematical Computation I: Matrix Computation. This is an introductory course on numerical linear algebra, which is quite different from linear algebra. We will be much less interested in algebraic results that follow from axiomatic definitions of fields and vector spaces but much more interested in analytic results that hold only over the real and complex fields. The main objects of interest are real- or complex-valued matrices, which may come from differential operators, integral transforms, bilinear and quadratic forms, boundary and coboundary maps, Markov chains, correlations, DNA microarray measurements, movie ratings by viewers, friendship relations in social networks, etc. Numerical linear algebra provides the mathematical and algorithmic tools for analyzing these matrices. Topics covered: basic matrix decompositions LU, QR, SVD; Gaussian elimination and LU/LDU decompositions; backward error analysis, Gram-Schmidt orthogonalization and QR/complete orthogonal decompositions; solving linear systems, least squares, and total least squares problem; low-rank matrix approximations and matrix completion. We shall also include a brief overview of stationary and Krylov subspace iterative methods; eigenvalue and singular value problems; and sparse linear algebra. *Autumn*

OR

BIOS 20187 Fundamentals of Genetics. The goal of this course is to integrate recent developments in molecular genetics into the structure of classical genetics with an emphasis on recent advances in genetics and genomics. Topics include Mendelian inheritance, genotype-phenotype relationships, linkage analysis, modern gene mapping techniques, gene expression, model systems genetics and analysis of genetic pathways. *Autumn*

OR

STAT 24500 Statistical Theory/Method-2. This course is the second quarter of a two-quarter systematic introduction to the principles and techniques of statistics, as well as to practical considerations in the analysis of data, with emphasis on the analysis of experimental data. This course continues from either STAT 24400 or STAT 24410 and covers statistical methodology, including the analysis of variance, regression, correlation, and some multivariate analysis. Some principles of data analysis are introduced, and an attempt is made to present the analysis of variance and regression in a unified framework. Statistical software is used. *Winter*.

OR

STAT 32950 Multivariate Statistical Analysis: Applications and Techniques. This course focuses on applications and techniques for analysis of multivariate and high dimensional data. Beginning subjects cover principal component analysis, factor model, canonical correlation, multi-dimensional scaling, discriminant analysis, clustering, and common techniques of dimension reduction. Further topics on statistical learning for high dimensional data and complex structures include penalized regression models (LASSO, ridge, elastic net),

sparse PCA, independent component analysis, Gaussian mixture model, and Expectation-Maximization methods. Theoretical derivations will be presented with emphasis on motivations, applications, and hands-on data analysis. *Spring*

OR

ECEV 42900 Theoretical Ecology. An introduction to mathematical modeling in ecology. The course will begin with linear growth and Lotka-Volterra models, and proceed to partial differential equations. The course's perspective will emphasize numerical computations and fitting models to data. *Winter*

OR

STAT 24610 Pattern Recognition This course treats statistical models and methods for pattern recognition and machine learning. Topics include a review of the multivariate normal distribution, graphical models, computational methods for inference in graphical models in particular the EM algorithm for mixture models and HMM's, and the sum-product algorithm. Linear discriminative analysis and other discriminative methods, such as decision trees and SVM's are covered as well. (Spring)

OR

STAT 30210 Bayesian Analysis and Principles of Statistics This course continues the development of Mathematical Statistics, with an emphasis on Bayesian analysis and underlying principles of inference. Topics include Bayesian Inference and Computation, Frequentist Inference and interpretation of p values and confidence intervals, Decision theory, admissibility and Stein's paradox, the Likelihood principle, Exchangeability and De Finetti's theorem, hierarchical modelling, multiple comparisons and False Discovery Rates. The mathematical level will generally be at that of an easy advanced calculus course. We will assume familiarity with standard statistical distributions (e.g., Normal, Poisson, Binomial, Exponential), with the laws of probability, expectation, conditional expectation, etc., and exposure to common statistical concepts such as p values and confidence intervals. Familiarity with the R statistical language will also be expected, and homework assignments will include programming problems in R. (Spring)

OR

STAT 35500 Statistical Genetics. This is an advanced course in statistical genetics. Prerequisites are Human Genetics 47100 and Statistics 24400 and 24500. Students who do not meet the prerequisites may enroll on a P/NP basis with consent of the instructor. Prerequisites are either Human Genetics 47100 or statistics preparation at the level of Statistics 24400 and 24500. This is a discussion course and student presentations will be required. Topics vary and may include, but are not limited to, statistical problems in linkage mapping, association mapping, map construction, and genetic models for complex traits. *Spring*

OR

STAT 37710 Machine Learning. This course provides hands-on experience with a range of contemporary machine learning algorithms, as well as an introduction to the theoretical aspects of the subject. Topics covered include: the PAC framework, Bayesian learning, graphical models, clustering, dimensionality reduction, kernel methods including SVMs, matrix completion, neural networks, and an introduction to statistical learning theory. *Spring*

A complete list of courses is available on the [Academic Information System](#) and the [GGSB Program website](#)