COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

300 George Street, Suite 501, 203.737.6020
http://cbb.yale.edu
M.S., Ph.D.

Directors of Graduate Studies
Mark Gerstein (Bass 432A, 203.432.6105, mark.gerstein@yale.edu)
Hongyu Zhao (300 George St., Suite 503, 203.785.3613, hongyu.zhao@yale.edu)

Professors James Aspnes (Computer Science), Joseph Chang (Statistics & Data Science), Kei-Hoi Cheung (Emergency Medicine), Ronald Coifman (Mathematics; Computer Science), Donald Engelman (Molecular Biophysics & Biochemistry), Richard Flavell (Immunobiology), Alison Galvani (Public Health), Mark Gerstein (Biomedical Informatics; Molecular Biophysics & Biochemistry; Computer Science), Antonio Giraldez (Genetics), Murat Gunel (Neurosurgery; Genetics), William Jorgensen (Chemistry), Douglas Kankel (Molecular, Cellular & Developmental Biology), Haifan Lin (Cell Biology; Genetics), Elias Lolis (Pharmacology), Shuangge Ma (Public Health), Andrew Miranker (Molecular Biophysics & Biochemistry), Anna Pyle (Molecular Biophysics & Biochemistry), Lynne Regan (Molecular Biophysics & Biochemistry; Chemistry), Valerie Reinke (Genetics), Gordon Shepherd (Neuroscience), Abraham Silberschatz (Computer Science), Dieter Söll (Molecular Biophysics & Biochemistry; Chemistry), Günter Wagner (Ecology & Evolutionary Biology), Heping Zhang (Public Health; Statistics & Data Science), Hongyu Zhao (Public Health; Genetics), Steven Zucker (Computer Science; Electrical Engineering; Biomedical Engineering)

Associate Professors Chris Cotsapas (Neurology), Forrest Crawford (Public Health), Thierry Emonet (Molecular, Cellular & Developmental Biology), Steven Kleinstein (Pathology), Yuval Kluger (Pathology), Michael Krauthammer (Pathology), Jun Lu (Genetics), James Noonan (Genetics), Corey O’Hern (Mechanical Engineering & Materials Science; Physics), Jeffrey Townsend (Public Health), Zuoheng (Anita) Wang (Public Health)

Assistant Professors Murat Acar (Molecular, Cellular & Developmental Biology), Julien Berro (Molecular Biophysics & Biochemistry), Damon Clark (Molecular, Cellular & Developmental Biology), Smita Krishnaswamy (Genetics)

FIELDS OF STUDY
Computational biology and bioinformatics (CB&B) is a rapidly developing multidisciplinary field. The systematic acquisition of data made possible by genomics and proteomics technologies has created a tremendous gap between available data and their biological interpretation. Given the rate of data generation, it is well recognized that this gap will not be closed with direct individual experimentation. Computational and theoretical approaches to understanding biological systems provide an essential vehicle to help close this gap. These activities include computational modeling of biological processes, computational management of large-scale projects, database development and data mining, algorithm development, and high-performance computing, as well as statistical and mathematical analyses.

SPECIAL ADMISSIONS REQUIREMENTS
Applicants are expected (1) to have a strong foundation in the basic sciences, such as biology, chemistry, and mathematics, and (2) to have training in computing/informatics, including significant computer programming experience. The Graduate Record Examination (GRE) General Test is required, and the GRE Subject Test in Biochemistry, Cell and Molecular Biology; Biology; Chemistry; Computer Science; or other relevant discipline is recommended. Alternatively, the Medical College Admission Test (MCAT) may be substituted for the GRE tests. Applicants for whom English is not their native language are required to submit results from the Test of English as a Foreign Language (TOEFL).

To enter the Ph.D. program, students apply to an interest-based track within the interdepartmental graduate program in Biological and Biomedical Sciences (BBS), http://bbs.yale.edu.

INTEGRATED GRADUATE PROGRAM IN PHYSICAL AND ENGINEERING BIOLOGY (PEB)
Students applying to one of the interest-based tracks of the Biological and Biomedical Sciences program may simultaneously apply to be part of the PEB program. See the description under Non-Degree-Granting Programs, Councils, and Research Institutes for course requirements, and http://peb.yale.edu for more information about the benefits of this program and application instructions.

SPECIAL REQUIREMENTS FOR THE PH.D. DEGREE
With the help of a faculty advisory committee, each student plans a program that includes courses, seminars, laboratory rotations, and independent reading. Students are expected to gain competence in three core areas: (1) computational biology and bioinformatics, (2) biological sciences, and (3) informatics (including computer science, statistics, and applied mathematics). While the courses taken to satisfy the core areas of competency may vary considerably, all students are required to take the following courses: CB&B 562, CB&B 740, and CB&B 752. A typical program will include ten courses. Completion of the core curriculum will typically take three to four terms, depending in part on the prior training of the student. With approval of the CB&B director of graduate studies (DGS), students may take one or two undergraduate courses to satisfy areas of minimum expected competency. Students will typically take two to three courses each term and three research rotations (CB&B 711, CB&B 712, CB&B 713) during the first year. After the first year, students will start
working in the laboratory of their Ph.D. thesis supervisor. Students must pass a qualifying examination normally given at the end of the second year or the beginning of the third year. There is no language requirement. Students will serve as teaching assistants in two term courses. In addition to all other requirements, students must successfully complete CB&B 601, Fundamentals of Research: Responsible Conduct of Research (or another course that covers the material) prior to the end of their first year of study. In their fourth year of study, all students must successfully complete B&BS 503, RCR Refresher for Senior BBS Students.

M.D./PH.D. STUDENTS
Students pursuing the joint M.D./Ph.D. degrees must satisfy the course requirements listed above for Ph.D. students. With approval of the DGS, some courses taken toward the M.D. degree can be counted toward the ten required courses. Such courses must have a graduate course number, and the student must register for them as graduate courses (in which grades are received). Laboratory rotations are available but not required. One teaching assistantship is required.

MASTER’S DEGREE
M.S. (en route to the Ph.D.) To qualify for the awarding of the M.S. degree a student must (1) complete two years (four terms) of study in the Ph.D. program, with ten required courses taken at Yale, (2) complete the required course work for the Ph.D. program with an average grade of High Pass or higher, (3) successfully complete three research rotations, and (4) meet the Graduate School’s Honors requirement.

Terminal Master’s Degree Program The CB&B terminal master’s program has limited availability and is intended primarily for postdoctoral fellows supported by training grants and for students with sponsored funding, e.g., from industry. The curriculum requirements are the same as in the CB&B Ph.D. program, except that there are no requirements for fulfilling laboratory research rotations, serving as a teaching assistant, or completing a Ph.D. dissertation. Terminal M.S. students will be expected to complete an M.S. project, including a project report. Completion of the terminal M.S. degree will typically take four terms of full-time study. Applicants should contact the CB&B registrar before submitting an M.S. application.

COURSES
Additional courses focused on the biological sciences and on areas of informatics are selected by the student in consultation with CB&B faculty.

CB&B 523b / ENAS 541b / MB&B 523b / PHYS 523b, Biological Physics  Simon Mochrie
The course has two aims: (1) to introduce students to the physics of biological systems and (2) to introduce students to the basics of scientific computing. The course focuses on studies of a broad range of biophysical phenomena including diffusion, polymer statistics, protein folding, macromolecular crowding, cell motion, and tissue development using computational tools and methods. Intensive tutorials are provided for MATLAB including basic syntax, arrays, for-loops, conditional statements, functions, plotting, and importing and exporting data.

CB&B 555a / CPSC 553a / GENE 555a, Machine Learning for Biology  Smita Krishnaswamy
This course introduces biology as a systems and data science through open computational problems in biology, the types of high-throughput data that are being produced by modern biological technologies, and computational approaches that may be used to tackle such problems. We cover applications of machine-learning methods in the analysis of high-throughput biological data, especially focusing on genomics and proteomics data, including denoising data; nonlinear dimensionality reduction for visualization and progression analysis; unsupervised clustering; and information theoretic analysis of gene regulatory and signaling networks. Students’ grades are based on programming assignments, a midterm, a paper presentation, and a final project.

CB&B 561a / MB&B 561a / MCDB 561a / PHYS 561a, Introduction to Dynamical Systems in Biology  Thierry Emonet and Kathryn Miller-Jensen
Study of the analytic and computational skills needed to model genetic networks and protein signaling pathways. Review of basic biochemical concepts including chemical reactions, ligand binding to receptors, cooperativity, and Michaelis-Menten enzyme kinetics. Deep exploration of biological systems including: kinetics of RNA and protein synthesis and degradation; transcription activators and repressors; lysoxygen/lysis switch of lambda phage and the roles of cooperativity and feedback; network motifs such as feed-forward networks and how they shape response dynamics; cell signaling, MAP kinase networks and cell fate decisions; bacterial chemotaxis; and noise in gene expression and phenotypic variability. Students learn to model using MATLAB in a series of in-class hackathons that illustrate biological examples discussed in lectures.

CB&B 562b / AMTH 765b / ENAS 561b / INP 562b / MB&B 562b / MCDB 562b / PHYS 562b, Dynamical Systems in Biology  Damon Clark, Thierry Emonet, and Jonathon Howard
This course covers advanced topics in computational biology. How do cells compute, how do they count and tell time, how do they oscillate and generate spatial patterns? Topics include time-dependent dynamics in regulatory, signal-transduction, and neuronal networks; fluctuations, growth, and form; mechanics of cell shape and motion; spatially heterogeneous processes; diffusion. This year, the course spends roughly half its time on mechanical systems at the cellular and tissue level, and half on models of neurons and neural systems in computational neuroscience. Prerequisite: MCDB 561 or equivalent, or a 200-level biology course, or permission of the instructor.
CB&B 601b / IBIO 601b, Fundamentals of Research: Responsible Conduct of Research  Carla Rothlin and Walther Mothes
A weekly seminar presented by faculty trainers on topics relating to proper conduct of research. Required of first-year CB&B students, first-year Immunobiology students, and training grant-funded postdocs. Pass/Fail.

CB&B 647b / BIS 645b / GENE 645b, Statistical Methods in Human Genetics  Hongyu Zhao
Probability modeling and statistical methodology for the analysis of human genetics data are presented. Topics include population genetics, single locus and polygenic inheritance, linkage analysis, quantitative trait analysis, association analysis, haplotype analysis, population structure, whole genome genotyping platforms, copy number variation, pathway analysis, and genetic risk prediction models. Prerequisites: genetics; BIS 505; S&DS 541 or equivalent; or permission of the instructor.

CB&B 711a and CB&B 712b and CB&B 713b, Lab Rotations  Hongyu Zhou
Three 2.5–3-month research rotations in faculty laboratories are required during the first year of graduate study. These rotations are arranged by each student with individual faculty members.

CB&B 740a, Clinical and Translational Informatics  Richard Shiffman and Michael Krauthammer
The course provides an introduction to clinical and translational informatics. Topics include (1) overview of biomedical informatics, (2) design, function, and evaluation of clinical information systems, (3) clinical decision making and practice guidelines, (4) clinical decision support systems, (5) informatics support of clinical research, (6) privacy and confidentiality of clinical data, (7) standards, and (8) topics in translational bioinformatics. Permission of the instructor required.

CB&B 745b / AMTH 745b / CPSC 745b, Advanced Topics in Machine Learning and Data Mining  Smita Krishnaswamy and Guy Wolf
An overview of advances in the past decade in machine learning and automatic data-mining approaches for dealing with the broad scope of modern data-analysis challenges, including deep learning, kernel methods, dictionary learning, and bag of words/features. This year, the focus is on a broad scope of biomedical data-analysis tasks, such as single-cell RNA sequencing, single-cell signaling and proteomic analysis, health care assessment, and medical diagnosis and treatment recommendations. The seminar is based on student presentations and discussions of recent prominent publications from leading journals and conferences in the field. Prerequisite: basic concepts in data analysis (e.g., CPSC 545 or 563) or permission of the instructor.

CB&B 750b, Core Topics in Biomedical Informatics  Kei-Hoi Cheung and Cynthia Brandt
The course focuses on providing an introduction to common unifying themes that serve as the foundation for different areas of biomedical informatics, including clinical, neuro-, and genome informatics. The course is designed for students with significant computer experience and course work who plan to build databases and computational tools for use in biomedical research. Emphasis is on understanding basic principles underlying informatics approaches to interoperation among biomedical databases and software tools, standardized biomedical vocabularies and ontologies, biomedical natural language processing, modeling of biological systems, high-performance computation in biomedicine, and other related topics.

CB&B 752b / CPSC 752b / MB&B 752b / MCDB 752b, Biomedical Data Science: Mining and Modeling  Mark Gerstein
Biomedical data science encompasses the analysis of gene sequences, macromolecular structures, and functional genomics data on a large scale. It represents a major practical application for modern techniques in data mining and simulation. Specific topics to be covered include sequence alignment, large-scale processing, next-generation sequencing data, comparative genomics, phylogenetics, biological database design, geometric analysis of protein structure, molecular-dynamics simulation, biological networks, normalization of microarray data, mining of functional genomics data sets, and machine-learning approaches to data integration. Prerequisites: biochemistry and calculus, or permission of the instructor.