COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

300 George Street, Suite 501, 203.737.6029
http://cbb.yale.edu
M.S., Ph.D.

Directors of Graduate Studies
Mark Gerstein (Bass 432A, 203.432.6105, mark.gerstein@yale.edu)
Steven Kleinstein (300 George St., Suite 505,
203.785.6685, (steven.kleinstein@yale.edu)

Professors Marcus Bosenberg (*Dermatology*; *Pathology*), Cynthia Brandt (*Emergency Medicine*; *Anesthesiology*), Kei-Hoi Cheung (*Emergency Medicine*), Ronald Coifman (*Mathematics*; *Computer Science*), Chris Cotsapas (*Neurology*), Stephen Dellaporta (*Molecular, Cellular, and Developmental Biology*), Richard Flavell (*Immunobiology*), Joel Gelernter (*Genetics*; *Neuroscience*), Mark Gerstein (*Biomedical Informatics*; *Molecular Biophysics and Biochemistry*; *Computer Science*; *Statistics and Data Science*), Antonio Giraldez (*Genetics*), Jeffrey Gruen (*Genetics*; *Investigative Medicine*; *Pediatrics*), Murat Gunel (*Neurosurgery*; *Genetics*), Ira Hall (*Genetics*), Amy Justice (*Internal Medicine*; *Public Health*), Naftali Kaminski (*Internal Medicine*), Steven Kleinstein (*Pathology*; *Immunobiology*), Yuval Kluger (*Pathology*), Harlan Krumholz (*Internal Medicine*; *Investigative Medicine*; *Public Health*), Haifan Lin (*Cell Biology*; *Genetics*), Shuangge (Steven) Ma (*Public Health*), Andrew Miranker (*Molecular Biophysics and Biochemistry*; *Chemical and Environmental Engineering*), James Noonan (*Genetics*), Corey O'Hern (*Mechanical Engineering and Materials Science*; *Applied Physics*; *Physics*), Lajos Pusztai (*Internal Medicine*), Anna Pyle (*Molecular Biophysics and Biochemistry*), David Stern (*Pathology*), Hemant Tagare (*Radiology and Biomedical Imaging*; *Biomedical Engineering*), Jeffrey Townsend (*Public Health*; *Ecology and Evolutionary Biology*), John Tsang (*Immunobiology*), Heping Zhang (*Biostatistics*), Hongyu Zhao (*Public Health*; *Genetics*), Steven Zucker (*Computer Science*; *Electrical Engineering*; *Biomedical Engineering*)

Associate Professors Julien Berro (*Molecular Biophysics and Biochemistry*), Forrest Crawford (*Public Health*), Smita Krishnaswamy (*Genetics*), Jun Lu (*Genetics*), Ted Melnick (*Biostatistics*; *Emergency Medicine*), Kathryn Miller-Jensen (*Engineering and Applied Science*), John Murray (*Psychiatry*; *Neuroscience*; *Physics*), Renato Polimanti (*Psychiatry*), Edward Stites (*Laboratory Medicine*), Andrew Taylor (*Emergency Medicine*), Zuoheng (Anita) Wang (*Public Health*)

Assistant Professors Arnaud Augert (*Pathology*), David Braun (*Medical Oncology*), Leying Guan (*Biostatistics*), Jeffrey Ishizuka (*Medicine*; *Pathology; Immunobiology*), Samah Jarad (*Emergency Medicine*), Monkol Lek (*Genetics*), Bluma Lesch (*Genetics*), Benjamin Machta (*Physics*), Robert McDougal (*Biostatistics*), C. Brandon Ogbunu (*Ecology and Evolutionary Biology*) Steven Reilly (*Genetics*), Wade Schulz (*Laboratory Medicine*), Serena Tucci (*Anthropology*), David van Dijk (*Cardiology*), Rex Ying (*Computer Science*), Jack Zhang (*Molecular Biophysics and Biochemistry*)

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.

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

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 biomedical informatics, (2) biological sciences, and (3) informatics (including computer science, applied mathematics, statistics, and data science). 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 740 and CB&B 752 along with either CB&B 562 or CB&B 750. CB&B requires a minimum of ten course credits. 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 no later than the end of the third year. There is no foreign 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 course credits. 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 (2) complete the required course work for the Ph.D. program with an average grade of High Pass or higher, with ten required course credits taken at Yale including three successful research rotations and (3) meet the Graduate School's Honors requirement of at least two Honors grades.

Terminal Master's Degree Program Students can be admitted for a terminal M.S. degree in Computational Biology and Bioinformatics with the goal of gaining competency in three core areas: (1) computational biology and biomedical informatics, (2) biomedical sciences, (3) informatics (including computer science, statistics, and applied mathematics). This is a two-year program. Students must complete nine courses, including at least three graduate courses in CBB, two graduate courses in the biological sciences, two graduate courses in areas of informatics, and two additional courses in any of the three core areas. In addition, M.S. students must take a one-term graduate seminar on research ethics and attend a CBB seminar series.

Terminal M.S. degree students are also expected to complete an M.S. project, write a research paper describing it, and defend the project in a seminar where they present the project and answer questions about the project as well as breadth knowledge of their coursework and track of study. The paper is evaluated by the student's research supervisor and a second reader from the CBB faculty. Students are expected to identify a faculty member to supervise the M.S. project by the end of the first year or early in the second year. Part-time study in this program is possible, but the degree must be completed within five years. Part-time students are expected to start the M.S. project after they have taken half of the required courses.

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 523a / ENAS 541a / MB&B 523a / PHYS 523a, Biological Physics Yimin Luo
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 / AMTH 553a / CPSC 553a / GENE 555a, Unsupervised Learning for Big Data Staff
This course focuses on machine-learning methods well-suited to tackling problems associated with analyzing high-dimensional, high-throughput noisy data including: manifold learning, graph signal processing, nonlinear dimensionality reduction, clustering, and information theory. Though the class goes over some biomedical applications, such methods can be applied in any field. Prerequisites: knowledge of linear algebra and Python programming.

CB&B 562b / AMTH 765b / ENAS 561b / INP 562b / MB&B 562b / MCDB 562b / PHYS 562b, Modeling Biological Systems II Joe 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: a 200-level biology course or permission of the instructor.

CB&B 601b, Fundamentals of Research: Responsible Conduct of Research Staff
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 634a, Computational Methods for Informatics Robert McDougal
This course introduces the key computational methods and concepts necessary for taking an informatics project from start to finish: using APIs to query online resources, reading and writing common biomedical data formats, choosing appropriate data structures for storing and manipulating data, implementing computationally efficient and parallelizable algorithms for analyzing data, and developing appropriate visualizations for communicating health information. The FAIR data-sharing guidelines are discussed. Current issues in big health data are discussed, including successful applications as well as privacy and bias concerns. This course has a significant programming component, and familiarity with programming is assumed. Prerequisite: CPSC 223 or equivalent, or permission of the instructor.

CB&B 638a, Clinical Database Management Systems and Ontologies Kei-Hoi Cheung and George Hauser
This course introduces database and ontology in the clinical/public health domain. It reviews how data and information are generated in clinical/public health settings. It introduces different approaches to representing, modeling, managing, querying, and integrating clinical/public health data. In terms of database technologies, the course describes two main approaches—SQL database and non-SQL (NoSQL) database—and shows how these technologies can be used to build electronic health records (EHR), data repositories, and data warehouses. In terms of ontologies, it discusses how ontologies are used in connecting and integrating data with machine-readable knowledge. The course reviews the major theories, methods, and tools for the design and development of databases and ontologies. It also includes clinical/public health use

cases demonstrating how databases and ontologies are used to support clinical/public health research.

CB&B 645b / S&DS 645b, Statistical Methods in Computational Biology Hongyu Zhao
Introduction to problems, algorithms, and data analysis approaches in computational biology and bioinformatics. We discuss statistical issues arising in analyzing population genetics data, gene expression microarray data, next-generation sequencing data, microbiome data, and network data. Statistical methods include maximum likelihood, EM, Bayesian inference, Markov chain Monte Carlo, and methods of classification and clustering; models include hidden Markov models, Bayesian networks, and graphical models. Offered every other year. Prerequisite: S&DS 538, S&DS 542, or S&DS 661. Prior knowledge of biology is not required, but some interest in the subject and a willingness to carry out calculations using R is assumed.

CB&B 711a and CB&B 712b and CB&B 713b, Lab Rotations Staff
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, Introduction to Health Informatics Andrew Taylor
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 750b, Core Topics in Biomedical Informatics Samah Jarad
The course focuses on providing an introduction to common unifying themes that serve as the foundation for different areas of biomedical informatics. It is designed for students with programming experience 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, predictive analytics, information extraction, deep learning, and other related topics.