The Columbia University Department of Systems Biology

Research • Education • Technology

Moving toward the future of biomedical science
**In the history of science, quantitative models have triggered golden eras of discovery in many scientific disciplines.**

Consider, for example, the ways in which modeling and computational simulation have improved our understanding of high-energy particles in physics, weather patterns in climatology, and the flow of capital in economics. In these fields and others, methodologies derived from the quantitative sciences have complemented empirical observation by providing powerful tools for generating and testing hypotheses without having to perform actual experiments. In each case, the subsequent availability of high-performance computing platforms led to further leaps forward, allowing simulations of models that were too complex to be solved analytically.

In biology and medical research, recent developments suggest that we have reached a similar turning point. Over the past decade, new technologies have made it possible to measure, on a genome-wide scale, many variables that determine cellular activity. This has led to an explosion in the amount of data that are now available — far too much to analyze using traditional methods. In response, scientists trained in systems and computational biology have begun to play an increasingly important role. Using algorithms and complex models of cellular regulation running on powerful supercomputers, these scientists can now mine the data for insights into both basic biology and the biology of human health. The computational models that they develop have generated new hypotheses and have led to many surprising discoveries. Perhaps most importantly, these approaches have improved the efficiency of biological research, pointing the way toward experiments that are most likely to succeed.

When the human genome sequence was published in 2001, many hoped that we had uncovered the Rosetta Stone that would enable us to conquer disease. A number of revolutionary advances since then have made it clear, however, that the genome is just a list of parts. How these parts can be combined or assembled to form distinct living systems is still largely unknown. Scientists rarely talk anymore about finding “the gene that causes x,” but more often about how physiologic traits and diseases emerge from the abnormal interaction of multiple gene factors and their cellular envi-
At Columbia, systems biology integrates various types of information in order to understand how cells behave at the level of genes, proteins, and cells. Research can include developing models of interactions between proteins and genes (far left), computationally predicting protein structure (top right), and performing experimental tests in the laboratory to validate these and other types of computational models (bottom right).

Andrea Califano
Chair, Department of Systems Biology

Barry Honig
Director, Center for Computational Biology & Bioinformatics

At Columbia, systems biology integrates various types of information in order to understand how cells behave at the level of genes, proteins, and cells. Research can include developing models of interactions between proteins and genes (far left), computationally predicting protein structure (top right), and performing experimental tests in the laboratory to validate these and other types of computational models (bottom right).

With its decision to create a Department of Systems Biology in 2013, Columbia University took a bold step to embrace this new approach. This commitment reflects the strength of the growing systems biology research community at Columbia as well as the ways in which the perspective and tools that systems biology offers have begun to inform research done across the university. In its strategic plan, 2020 Vision, the College of Physicians and Surgeons identified systems biology as one of the “scientific priorities that will define the future of health and biomedical science.” We are very excited at the prospect of growing the discipline at Columbia, and about the many productive collaborations on the horizon.

As we begin this new chapter, we would like to thank Lee Goldman, Dean of the Faculties of Health Sciences and Medicine, for his guidance and advocacy on our behalf, as well as the estate of Judith Sulzberger, whose generous financial support was essential for getting the department off the ground.

Only time will tell what role systems biology will play in the next generation of biological research. But wherever it goes, we are excited that Columbia will be there to help lead the way.

Environment. For these reasons, the coming decades of biology will require more than reading the code written in the genome; this information will need to be integrated with data about protein structure, gene expression levels, and chemical modifications that occur after a protein is coded that can affect its activity. Furthermore, these rich sources of data will need to be interpreted through the lens of what we call the “regulatory logic” of the cell; that is, the multilayered networks of molecular interactions that are specific to particular cell types and contexts. This is an ambitious agenda, but already biologists have begun to map these networks in an effort to understand how all the pieces fit together.

Systems biology is the discipline that is best poised to understand biology from the perspective of complexity. It appreciates and embraces the challenge of understanding how large numbers of gene products work together to produce observable traits and diseases, rather than studying one gene at a time, and has produced a growing repertoire of tools designed to dissect complexity at different levels of biological activity. Moreover, in just a little over a decade, systems biology has begun generating important discoveries that are relevant for both basic biology and biomedical research.
The mission of the Columbia University Department of Systems Biology is to develop new methods for understanding the biological world from a systems perspective, particularly at the genomic and molecular levels. Our faculty shares a common interest in combining high-throughput experimentation, quantitative analysis, and innovative technology development. Researchers use computational approaches and laboratory experimentation in an iterative way, developing predictive models of biological systems, and then validating them in the laboratory.
Some facts about the Columbia University Department of Systems Biology

> With more than two-dozen faculty members, a community of more than 250 individuals working in department labs, and an outstanding record of publications, the Department of Systems Biology is among the largest programs in this field.

> The Department is highly multidisciplinary. Our faculty’s research interests include bioinformatics, biophysics, cancer biology, chemical biology, genetics and genomics, machine learning, microbiology, molecular evolution, stem cell and developmental biology, structural biology, synthetic biology, and virology and immunology.

> The Department is home to six centers of excellence in systems and computational biology, including the Center for Multiscale Analysis of Genomic and Cellular Networks (MAGNet). MAGNet is one of eight National Centers for Biomedical Computing and one of 12 Integrated Cancer Biology Programs funded by the National Cancer Institute.

> Department members conduct collaborative research with investigators in the Herbert Irving Comprehensive Cancer Center, other institutes at Columbia University, and at other universities. In addition to producing new insights into basic biology, many of these collaborations utilize approaches from systems and computational biology to study human diseases such as cancer, infectious diseases, metabolic disorders, and psychiatric illnesses.

> The Department manages the JP Sulzberger Columbia Genome Center, which provides a state-of-the art infrastructure for next-generation genome sequencing and high-throughput molecular screening. The Genome Center supports researchers inside the department, across the Columbia University community, and at other institutions. The Genome Center is also the official high-throughput screening and chemistry core for NYSTEM, the New York State foundation for stem cell research.

> We maintain several high-performance computing systems. In 2013 a new and improved cluster with a maximum performance of 212 TFlops was installed, making Columbia home to one of the world’s largest computing environments dedicated to research in systems and molecular biology.

> We oversee graduate education and postdoctoral training in systems and computational biology to promote their use in biological research. Our educational efforts include a training grant from the National Institutes of Health. Upon graduation, our alumni have a strong track record of launching successful careers in both academia and industry.

> Researchers in the Department of Systems Biology develop innovative software tools and databases for the study of various facets of biology. Through an open source, web-based platform called geWorkbench, we make these resources available to researchers anywhere.

> Columbia University investigators have played a leadership role in promoting the exchange of ideas within the global systems biology community by organizing key conferences and meetings that attract hundreds of investigators each year, including the Dialogue for Reverse Engineering Assessments and Methods (DREAM) and the RECOMB/ISCB Conference on Regulatory and Systems Genomics.
Clinical Trials for Cancer, One Patient at a Time

Columbia University Medical Center (CUMC) researchers are developing a new approach to cancer clinical trials in which the goal is to design and test therapies in a single patient, rather than in a group of patients. Instead of focusing on mutated genes that have already been associated with cancer, these “N-of-1 trials” use systems biology methods developed in the laboratory of Andrea Califano to analyze the molecular regulatory networks that drive an individual patient’s tumor cells. By reverse engineering these regulatory networks, Califano identifies genes and gene pairs — called “master regulators” — that are critical bottlenecks within these networks and are essential for the survival of the tumor. If FDA-approved drugs that inhibit these genes exist, they are then tested in a mouse model of the patient’s tumor and, if successful, they are considered as potential therapeutic agents for the patient. Identifying the molecular networks that drive cancer and pinpointing drugs able to target these networks could then be used to develop more precise diagnoses and treatments for other patients with similar molecular signatures. “Eventually,” Califano says, “we hope to be able to treat patients based on common vulnerabilities of the cancer cellular machinery, of which genetic mutations are only indirect evidence. Genetic alterations are clearly responsible for tumorigenesis, but control points in molecular networks may be better therapeutic targets.”

Biomarker Identified for Predicting Early Prostate Cancer Aggressiveness

A team led by Cory Abate-Shen, Michael Shen, and Andrea Califano found that measuring the expression levels of three genes associated with aging can be used to predict the aggressiveness of seemingly low-risk prostate cancer. Use of this three-gene biomarker, in conjunction with existing cancer-staging tests, could help physicians better determine which men with early prostate cancer can be safely followed with “active surveillance” and spared the risks of prostate removal or other invasive treatment. Using a technique called gene set enrichment analysis, the team identified 19 genes that are enriched in a mouse model of prostate cancer in which the cancers are invariably indolent. They then used a decision-tree learning model, a type of computer algorithm, to identify three genes—FGFR1, PMP22, and CDKN1A—that together can accurately predict the outcome of seemingly low-risk tumors. Tumors that test negative for the biomarker are deemed aggressive. A blinded retrospective study of 43 patients identified all 14 who ultimately developed advanced prostate cancer. More than 200,000 new cases of prostate cancer are diagnosed each year in the U.S. According to Dr. Abate-Shen, “The three-gene biomarker could take much of the guesswork out of the diagnostic process and ensure that patients are neither overtreated nor undertreated.” The researchers plan to evaluate the test in a larger, prospective clinical trial.
Connections Found between Genetic Networks for Schizophrenia and Autism

Dennis Vitkup and colleagues performed network-based analyses of rare genetic mutations to map the gene networks that lead to schizophrenia. When they compared one schizophrenia network to a previously discovered autism network, they discovered that different copy number variants in the same genes can lead to either schizophrenia or autism. The overlapping genes are important for axon guidance, synapse function, and cell migration — processes in the brain that have been shown to play a role in the development of these two diseases. These gene networks are particularly active during prenatal development, suggesting that the foundations for schizophrenia and autism are laid very early in life. Dr. Vitkup believes that there may be many more genes to be found that are tied to schizophrenia, but predicts that they will function within the networks that his team has described. “Until a few years ago,” he explained, “people were looking for just a handful of genes responsible for autism and schizophrenia, so the idea that many hundreds of genes are involved is a big change in thinking... Our study and the studies of our collaborators suggest that in the search for the causes of complex genetic disorders, it will be more productive to look for common pathways and gene circuits than for a handful of causal genes. This type of network analysis gives us a way to begin to make sense of what’s happening.”

Uniting Structural and Systems Biology

When the Center for Multiscale Analysis of Genomic and Cellular Networks (MAGNet) was founded in 2005, one of its goals was to integrate the methods of structural biology with those of systems biology. Determining whether two proteins were structurally capable of interacting, researchers hoped, would both improve the predictive value of models of molecular interactions by giving another layer of complementary evidence, and lead to new types of predictions that could not be made using other methods. In a paper published in Nature, Barry Honig, Andrea Califano, and colleagues reported on a novel approach for using information about protein structures to make predictions about protein-protein interactions on a genome-wide scale. Their approach capitalizes on methods the Honig lab has developed over the last 15 years for inferring a protein’s activity by predicting structural similarities between it and other proteins with known local structures and functions. This work culminated in the development of an algorithm called Predicting Protein-Protein Interactions (PrePPI), which has helped to predict 300,000 interactions in humans that are highly likely to occur in nature, a 10-fold increase over what had previously been identified. “This approach is new — both for structural and systems biology,” Honig says. “We can now expand the utility of structural information to generate predictive models of protein-protein interactions.”
Studies Reveal Genes that Drive Glioblastoma

Columbia University Medical Center researchers identified 18 new genes responsible for driving glioblastoma multiforme, the most common—and most aggressive—form of brain cancer in adults. Using a combination of high-throughput DNA sequencing and a new method of statistical analysis developed by co-author Raul Rabadan, an assistant professor in the Department of Systems Biology, the team generated a short list of candidate gene mutations that were highly likely to drive cancer, as opposed to mutations that have no effect. In addition to the 18 new genes, the analysis identified 15 driver genes that had been previously identified in other studies, confirming the accuracy of the technique. Once the Rabadan group’s analysis identified these driver genes, laboratory studies validated their predictions. Considering these results along with those from a previous study in which the team identified a gene fusion that causes a subset of glioblastomas, Rabadan and collaborators Antonio Lavarone and Anna Lasorella point out that approximately 15% of these deadly brain tumors could now be targeted with drugs that have already been approved by the FDA. As Lasorella remarks, “There is no reason why these patients couldn’t receive these drugs now in clinical trials.”

A Computational Tool for Understanding Cancer Cell Heterogeneity

Dana Pe’er and colleagues have developed a computational method that enables scientists to visualize and interpret high-dimensional data produced by single-cell measurement technologies such as mass cytometry. The method, called viSNE (visual interactive Stochastic Neighbor Embedding) has particular relevance to cancer research and therapeutics. Researchers now understand that cancer within an individual can harbor subpopulations of cells with different molecular characteristics. Groups of cells may behave differently from one another, including in how they respond to treatment. The ability to study single cells, as well as to identify and characterize subpopulations of cancerous cells within an individual, could lead to more precise methods of diagnosis and treatment. “Our method not only will allow scientists to explore the heterogeneity of cancer cells and to characterize drug-resistant cancer cells, but also will allow physicians to track tumor progression, identify drug-resistant cancer cells, and detect minute quantities of cancer cells that increase the risk of relapse,” Pe’er says.

References


Centers of Excellence

The Department of Systems Biology has been recognized as a leader in advancing research in systems biology and bioinformatics. We currently receive support through six major center of excellence grants. These grants have been instrumental in supporting our research and expanding our computational infrastructure.

Center for Multiscale Analysis of Genetic and Cellular Networks (MAGNet)
MAGNet is one of eight National Centers for Biomedical Computing, an accredited center within the National Cancer Institute’s Integrated Cancer Biology Program, and one of 12 interdisciplinary Centers for Cancer Systems Biology. Research at MAGNet is focused on developing new computational tools for identifying molecular interactions in the cell and discovering how these interactions produce cellular phenotypes.

Protein Structure Initiative (PSI)
The Protein Structure Initiative is a National Institutes of Health program aimed at improving researchers’ ability to determine three-dimensional protein structures and to interpret them in terms of biological function. The Department of Systems Biology functions as the bioinformatics core for the Northeast Structural Genomics Consortium, one of the major centers funded by the PSI.

Library of Integrated Network-Based Cellular Signatures (LINCS)
LINCS is an NIH initiative that is developing a standardized library of observations of molecular activities that occur when specific types of cells are exposed to specific chemicals. Researchers in the Department of Systems Biology have two LINCS grants — one supports the profiling of molecular signatures that occur when cells are exposed to multiple chemical compounds in synergy (such as in combination therapies for cancer), and the second funds efforts to use systems approaches to explain mechanisms of drug activity.

Cancer Target Discovery and Development Center (CTD²)
CTD² is a program of the National Cancer Institute that supports the development of new ways to use the rapidly growing body of cancer genomics data to systematically and efficiently develop new precision treatments for cancer. Researchers in the Department of Systems Biology are using systems biology approaches to study tumor progression and the development of resistance to chemotherapy and targeted therapies.

International Serious Adverse Event Consortium (iSAEC)
iSAEC is a nonprofit organization whose mission is to identify DNA variants that confer a risk for serious, drug-induced adverse events. The consortium’s goal is to improve the biomedical community’s ability to predict which patients are at higher risk when taking specific medications, and to tailor treatment options accordingly. This information could also be helpful in designing safer drugs. The Department of Systems Biology serves as the consortium’s data analysis and coordination center.
Our students arrive with diverse interests in the biological sciences, computer science, chemistry, mathematics, physics, engineering, and other related fields. During graduate studies, students engage in research that often results in high-profile publications. Because of the increasing importance of integrated, data-intensive research to the future of the biological sciences, graduates with expertise in systems and computational biology are currently in high demand. Upon graduation, our alumni have a strong track record of launching successful careers in both academia and industry.
Infrastructure

JP Sulzberger Columbia Genome Center

In cooperation with the Herbert Irving Comprehensive Cancer Center, the Department of Systems Biology manages the Judith P. Sulzberger Columbia Genome Center, a state-of-the-art environment for conducting high-throughput biomedical research. The Genome Center supports researchers in the Department of Systems Biology, the Columbia University research community, and other academic and industrial institutions. The Genome Center consists of two facilities:

> The **Genome Sequencing and Analysis Center** performs next-generation genome, exome, and RNA sequencing, and offers assistance in performing advanced data analysis. Faculty and staff at the Genome Center specialize in developing advanced bioinformatics methods for interpreting large collections of genomic data generated at the Center.

> The **High-Throughput Screening Center** supports basic and translational research by designing, optimizing, and running assays that can include thousands or even millions of experiments running in parallel. Our staff is highly experienced in designing innovative customized assays for exploring specific questions of scientific interest. In collaboration with the Columbia University Chemical Probe Synthesis Facility and Organic Chemistry Collaborative Center, the Columbia Genome Center is the high-throughput screening and chemistry core facility for NYSTEM, the New York State foundation for stem cell research.

Advanced Research Computing Services

Advanced Research Computing Services (ARCS) is a core facility operated by the Columbia University Center for Computational Biology and Bioinformatics. We maintain one of the world’s largest academic computing environments devoted to research in molecular and systems biology. ARCS provides professional computational services including high-performance computing, data storage, virtual server hosting, colocarion and server hosting, and IT services for the Department. We also support high-performance computing for other institutes at Columbia, including the Mailman School of Public Health and Mind, Brain, and Behavior Initiative.

Our computing resources are housed in two data centers totaling more than 3,000 sq. ft. of floor space. In 2013, a $2 million high-performance computing grant from the National Institutes of Health supported a major system upgrade. The new computing cluster now boasts 212 teraflops of performance, a figure that is nearly nine times the total computing capacity of our previous computing platform. It now comprises 6,336 CPU-cores, over 70,000 CUDA-cores (GPU), and 22 TB of total system memory.
Remembering Judith Sulzberger

Judith P. Sulzberger (1923-2011) was a visionary physician-philanthropist who applied her medical knowledge to advance research at the Columbia University College of Physicians & Surgeons (P&S) and elsewhere. A former member of the Board of Directors of the New York Times, she served on the CUMC Board of Visitors, the P&S Alumni Steering Committee, and the Health Sciences Advisory Council.

When genome studies at Columbia began to grow in importance to biomedical research, Dr. Sulzberger provided the guidance, the seed money, and ongoing support for what would become the Judith P. Sulzberger Columbia Genome Center. Her philanthropy also supported the launch of the Columbia Initiative in Systems Biology (now the Department of Systems Biology), as well as a large number of other initiatives at Columbia.

Founding Faculty

Andrea Califano
Chair, Department of Systems Biology
Director, Columbia Genome Center

Barry Honig
Director, Center for Computational Biology and Bioinformatics

Cory Abate-Shen
Dimitris Anastassiou
Harmen Bussemaker
Virginia Cornish
Adolfo Ferrando
Aris Floratos
Dana Pe'er
Itskik Pe'er
Raul Rabadan
Sagi Shapira
Lawrence Shapiro
David E. Shaw
Michael Shem
Yufeng Shen
José Silva
Peter Sims

Brent Stockwell
Milan Stojanovic
Nicholas Tatonetti
Saeed Tavazoie
Dennis Vitkup
Harris Wang
Chris Wiggins
Chaolin Zhang

Contact Us

Columbia University Department of Systems Biology
Irving Cancer Research Center
1130 St. Nicholas Avenue, 8th Floor
New York, NY 10032
Phone: 212-851-5208

To learn more about our research and programs, visit us online at systemsbiology.columbia.edu.

© 2013, Columbia University.

Front cover image: Global mapping of cancer gene expression changes to the human metabolic network. Image by Jie Hu and Dennis Vitkup using the iPath tool.