How do genetic mutations affect a person’s likelihood for developing certain kinds of cancer — or responding to certain kinds of cancer treatment? How can we estimate the effectiveness of a new drug before we test it on human beings? Where are viruses spreading — and how can we stop them? And how do we make sense of the mountains of electronic healthcare data being collected every second by sensors, scans and tests?

Our researchers develop the algorithms that look for patterns in genetic data and try to understand the chemical and physical interactions responsible for all human, animal and plant life. They also build computer models that predict how biological processes act and react under certain conditions.

Our investigations are not only leading to a better understanding of life — they’re also helping medical doctors create tools for better diagnosis and treatment of diseases.

In addition to performing groundbreaking research, we offer a highly regarded Ph.D. program in Computational Biology (jointly with the University of Pittsburgh). We also offer the Bachelor of Science in Computational Biology, as well as master’s programs in both computational biology and automated science.

Research Themes

Our work builds on the strong history of interdisciplinary research at Carnegie Mellon and cuts across the fields of computer science, mathematics, statistics, biology and chemistry. We aim to develop a deeper global understanding of the building blocks of life, and to develop tools for individualized diagnosis and treatment of cancer and other diseases. Our research themes include:

- **Automated Discovery:** We are developing artificial intelligence software that acts independently to build models from large datasets and choose the best experiments to do next.
- **Integrating Diverse Big Data:** We create new methods that can combine big datasets from different experiments to improve understanding of complex biological systems.
- **Genetic Sequencing:** We are developing a new understanding of how genetic processes work and what happens when they don’t.
Dept. of Health and Human Services  53%
State of Pennsylvania  14%
National Science Foundation  21%
Foundations  12%

* 3-yr. averages, FY16–FY18

Sponsored Research Funding Sources*  

$3.9M

Student Spotlights

Students Enrolled as of fall 2018

Ph.D.  34

Male  71%  Female  29%

Amir Alavi  
Ph.D. student in Computational Biology

Single-cell RNA-sequencing (scRNA-seq) allows researchers to determine the levels of each gene within an individual cell. Due to its great potential, scRNA-seq has taken the transcriptomics field by storm. Alavi aims to develop statistical methods to make comparative analysis of scRNA-seq data fast, accurate and scalable. He and collaborators have developed a system that relies on deep learning to obtain a compact representation of each cell’s expression, and then uses this new representation to enable other analyses such as nearest neighbor similarity analysis, visualization and clustering.

Natalie Sauerwald  
Ph.D. student in Computational Biology

The 3D folding of chromosomes within a cell’s nucleus has a major influence on key cellular processes and has been implicated in serious diseases, but we still don’t understand exactly how to describe this structure or how it exerts its influence. Sauerwald develops algorithms to analyze the 3D structure of chromosomes and understand the role it plays in regulating genomic functions. She has developed methods to determine how the genome’s shape changes across many conditions of healthy and diseased human cells, and continues to design innovative algorithms for uncovering this new aspect of genomics.

“... We can now use computer vision methods to study biological processes better than human biologists can with their eyes. We also can understand how a specific drug targets specific processes in a cell.”

Robert F. Murphy  
Head, Computational Biology Department  
Ray and Stephanie Lane Professor of Computational Biology  
Professor of Biological Sciences, Biomedical Engineering and Machine Learning