

UNLOCK INNOVATION



HOW DO WE ADVANCE BIOMEDICAL SCIENCE AND LAUNCH THE NEXT BIG INNOVATION?

CREATIVE / COLLABORATIVE / TRANSFORMATIONAL

WE ARE IN AN ERA OF UNPRECEDENTED CHALLENGES AND OPPORTUNITIES. EVEN AS WE MAINTAIN AN ENVIABLE POSITION IN RESEARCH SUPPORT, WE STRIVE TO PUSH THE BOUNDARIES OF WHAT SCIENCE AND MEDICINE CAN ACCOMPLISH.

PITT TO LEAD NATIONAL CENTER OF EXCELLENCE TRANSLATING BIG DATA TO KNOWLEDGE

The National Institutes of Health (NIH) has awarded the University of Pittsburgh an \$11 million, four-year grant to lead a Center of Excellence for Big Data Computing, which will help scientists capitalize more fully on large amounts of available data and make data science a more prominent component of biomedical research.

The highly competitive process for grants under the new NIH Big Data to Knowledge initiative—known as BD2K—attracted proposals from 136 institutions around the nation. Pitt’s Center for Causal Modeling and Discovery, one of only 11 such centers to be funded, aims to develop new tools and approaches to turn the tremendous amount of information available to physicians and scientists—including data from electronic health records, digital images, and molecular analyses of genes, proteins, and metabolites—into discoveries that will benefit human health.

“Individual biomedical researchers now have the technology to generate an enormous quantity and diversity of data. Adequately analyzing these data to discover new biomedical knowledge remains a major challenge, however,”

said Gregory Cooper, MD, PhD, professor and vice chair of the Department of Biomedical Informatics, School of Medicine, and principal investigator on the project. “Our goal is to make it much easier for researchers to analyze big data to discover causal relationships in biomedicine.”

The new center is a multidisciplinary collaboration of researchers from Pitt, Carnegie Mellon University, the Pittsburgh Supercomputing Center, and Yale University. Within its successful application, Pitt proposed innovative collaborations with multiple universities.

“As part of a national consortium, this Center of Excellence will put Pitt on the map as a home of big data science,” said Arthur S. Levine, MD, senior



COOPER

vice chancellor for the health sciences and John and Gertrude Petersen Dean of Medicine. “Our strengths in this field have stimulated collaborations with leading institutions, including Harvard and Stanford, and now we will be able to further develop such partnerships in many more meaningful ways.”

The center includes a team that will develop and implement causal modeling and discovery algorithms, or processes, to support the data analyses of three separate investigative groups, each focusing on a distinct biomedical problem whose answer lies in a sea of data: cell signals that drive the development of cancer, the molecular basis of lung disease susceptibility and severity, and the functional connections within the human brain (the “connectome”).

Each project will act as a test bed for the development, rigorous testing, and refinement of analytic tools. When successful, the algorithms and software likely can be applied to other biomedical research questions. The center will provide free, open-source software that scientists all over the world can use with their own datasets to uncover causal biomedical relationships. Their feedback will further enhance the algorithms and software.

“The good news is that we have so much data. But the bad news is that we have so much data,” said Jeremy M. Berg, PhD, codirector of the center, Pitt’s associate senior vice chancellor for science strategy and planning in the health sciences, and Pittsburgh Foundation Professor of Personalized Medicine. “Our challenge is to find strategies that enable us to sort through all this collected information efficiently and effectively to find meaningful relationships that lead us to new insights in health and disease.”

“The center also will be a training ground for the next generation of data scientists who will advance and accelerate the development and broader use of big data science models and methods,” said center codirector Ivet Bahar, PhD, who is Distinguished Professor, John K. Vries Professor, and chair of the Department of Computational and Systems Biology, School of Medicine. “We will create new educational materials as well as workshops and online tutorials to facilitate the use of causal modeling and discovery algorithms by the broader scientific community and to enable efficient translation of knowledge between basic biological and applied biomedical sciences.”

“THE FUTURE OF BIOMEDICAL RESEARCH IS ABOUT ASSIMILATING DATA ACROSS BIOLOGICAL SCALES FROM MOLECULES TO POPULATIONS. AS SUCH, THE HEALTH OF EACH ONE OF US IS A BIG DATA PROBLEM. ENSURING THAT WE ARE GETTING THE MOST OUT OF THE RESEARCH DATA THAT WE FUND IS A HIGH PRIORITY FOR NIH.”

PHILIP E. BOURNE, PHD, NIH ASSOCIATE DIRECTOR FOR DATA SCIENCE

BD2K