



Pitt Gets \$11 Million from NIH to Lead Center of Excellence in National Big Data Research Consortium

The National Institutes of Health has awarded the University of Pittsburgh an \$11 million, 4-year grant to lead a Big Data to Knowledge Center of Excellence, an initiative that will help scientists capitalize more fully on large amounts of available data and to make data science a more prominent component of biomedical research. McGowan Institute for Regenerative Medicine affiliated faculty member [Ivet Bahar, Ph.D.](#), Distinguished Professor and JK Vries Chair, Department of Computational and Systems Biology, Pitt School of Medicine, is a co-director of the new Center for Causal Modeling and Discovery which will be part of an elite national team addressing the challenges of Big Data in biomedicine.



“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, M.D., senior vice chancellor for the health sciences and John and Gertrude Petersen Dean of the School 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.”

Much of science focuses on understanding the “why” or “how” in nature, and now the challenge is to find these answers within terabytes and petabytes of data, or what is now known as “Big Data,” said Gregory Cooper, M.D., Ph.D., professor and vice chair of the Department of Biomedical Informatics, Pitt School of Medicine and co-director of the new Center.

“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,” Dr. Cooper said. “Our goal is to make it much easier for researchers to analyze big data to discover causal relationships in biomedicine.”

According to center co-director Jeremy Berg, Ph.D., associate senior vice chancellor for science strategy and planning in the health sciences and director of Pitt’s Institute for Personalized Medicine, researchers now have access to a tremendous amount of information from electronic health records, digital images, and molecular analyses of genes, proteins, and metabolites.

“The good news is that we have so much data. But the bad news is that we have so much data,” Dr. Berg said. “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.”



A collaboration of researchers at Pitt, Carnegie Mellon University (CMU), the Pittsburgh Supercomputing Center, and Yale University, the new Center will develop and disseminate tools that can find causal links in very large and complex biomedical data. Faculty in CMU's Department of Philosophy, led by Clark Glymour, Ph.D., Alumni University Professor and founding chair, are key partners in this data science effort; and Nicholas Nystrom, Ph.D., director of strategic applications at the Pittsburgh Supercomputing Center, will work to optimize these tools for a high-performance computing environment.

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, these 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 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 Dr. Bahar. "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."

Other collaborators include the California Institute of Technology, Rutgers University, University of Crete, and the University of North Carolina.

"Data creation in today's research is exponentially more rapid than anything we anticipated even a decade ago," said NIH Director Francis S. Collins, M.D., Ph.D. "Mammoth data sets are emerging at an accelerated pace in today's biomedical research and these funds will help us overcome the obstacles to maximizing their utility. The potential of these data, when used effectively, is quite astounding."

Read more...

[UPMC/University of Pittsburgh Schools of the Health Sciences Media Relations News Release](#)



[Carnegie Mellon University News Release](#)

[National Institutes of Health: FY14 Awards](#)

[Pittsburgh Post-Gazette](#)

[EurekAlert!](#)

[Pittsburgh Business Times](#)

[Pittsburgh Tribune-Review](#)

[News-Medical.net](#)

[Back to Home Page](#)