**Contact:**

**University of Maryland School of Medicine Public Affairs**

**(410) 706-5260**

**UNIVERSITY OF MARYLAND SCHOOL OF MEDICINE RESEARCHERS DEVELOP NEW WAY TO DECODE LARGE AMOUNTS OF BIOLOGICAL DATA**

*Finding Could Radically Speed Up Understanding of Genomic Information*

**Baltimore, Md., May 25, 2016** – In recent years, the amount of genomic data available to scientists has exploded. With faster and cheaper techniques increasingly available, hundreds of plants, animals and microbes have been sequenced in recent years. However, this ever-expanding trove of genetic information has created a problem: how can scientists quickly analyze all of this data, which could hold the key to better understanding many diseases, and solving other health and environmental issues.

Now, a University of Maryland School of Medicine researcher has helped develop an innovative computing technique that, on very large amounts of data, is both faster and more accurate than current methods. To spur research, a program using this technique is being offered for free to the biomedical research community.

“This is a whole new approach, with multiple opportunities for further development,” said **Andrew F. Neuwald**, PhD, Professor of Biochemistry & Molecular Biology and a Senior Scientist at the Institute for Genome Sciences (IGS) at the University of Maryland School of Medicine.

A description of the new method was published today in *PLOS Computational Biology*. Dr. Neuwald collaborated on the work with Stephen F. Altschul, PhD, a senior investigator at the National Center for Biotechnology Information at the National Institutes of Health.

Genomic sequence data encodes information regarding the structure and function of proteins, which comprise the basic cellular machinery and thus determine the structure and function of all microbes, plants and animals.

The new program is called GISMO, an acronym for “Gibbs Sampler for Multi-Alignment Optimization”. Gibbs sampling, a statistical technique for solving highly complex problems, is a central feature of the approach. In this case, sampling is used to find biological signals – relevant patterns that can help scientists better understand how organisms work. Neuwald says the approach improves upon conventional sequence alignment programs, which, unlike GISMO, can easily mistake random patterns in the data for biologically valid signals.

Current widely-used methods typically compare each sequence to every other sequence; this takes a prohibitively long time to compute for sets of a hundred thousand or more related protein sequences, which are now available for analysis. Neuwald describes these methods as “bottom up.” He and Dr. Altschul developed a technique that is “top down”; instead of comparing sequences to each other, it compares each sequence to an evolving statistical model. This approach is not only faster, but is also better at finding biologically relevant signals, which can, for example, help researchers unravel the mechanisms underlying cancer and inherited diseases. This technique becomes progressively faster than other methods as the size of the data set becomes larger.

Dr. Neuwald has a varied background, in molecular biology, computer science and Bayesian statistics and has been working on this technique for years. Dr. Altschul, whose formal training is in mathematics, was the first author on two landmark publications describing the popular sequence database search programs BLAST and PSIBLAST. They confirmed GISMO’s superior performance on large, diverse sequence sets by testing it against five widely used conventional methods. Dr. Neuwald is excited about GISMO’s potential: “Because researchers have been finding ways to speed up and improve conventional methods for decades and because GISMO takes such a new and different approach, I am confident that we can make GISMO even faster and more accurate going forward.”

GISMO is available for free to the biomedical research community through the IGS (<http://gismo.igs.umaryland.edu/>).

“This work combines two of the most fertile and dynamic areas in current science: big data and genomic analysis,” said UM SOM Dean **E. Albert Reece**, MD, PhD, MBA, who is also the vice president for Medical Affairs, University of Maryland, and the John Z. and Akiko K. Bowers Distinguished Professor. “In its creative linkage of analytics, statistics and genetics, this work points the way forward for a number of fields.”

**About the Institute for Genome Sciences**

The Institute for Genome Sciences, founded in 2007, is an international research center within the University of Maryland School of Medicine. Comprised of an interdisciplinary, multidepartment team of investigators, the Institute uses the powerful tools of genomics and bioinformatics to understand genome function in health and disease, to study molecular and cellular networks in a variety of model systems, and to generate data and bioinformatics resources of value to the international scientific community.  
igs.umaryland.edu

**About the University of Maryland School of Medicine**

The University of Maryland School of Medicine was chartered in 1807 and is the first public medical school in the United States and continues today as an innovative leader in accelerating innovation and discovery in medicine. The School of Medicine is the founding school of the University of Maryland and is an integral part of the 11-campus University System of Maryland. Located on the University of Maryland’s Baltimore campus, the School of Medicine works closely with the University of Maryland Medical Center and Medical System to provide a research-intensive, academic and clinically based education. With 43 academic departments, centers and institutes and a faculty of more than 3,000 physicians and research scientists plus more than $400 million in extramural funding, the School is regarded as one of the leading biomedical research institutions in the U.S. with top-tier faculty and programs in cancer, brain science, surgery and transplantation, trauma and emergency medicine, vaccine development and human genomics, among other centers of excellence. The School is not only concerned with the health of the citizens of Maryland and the nation, but also has a global presence, with research and treatment facilities in more than 35 countries around the world. <http://medschool.umaryland.edu/>