United States Department of Commerce
NIST Computational Biology

The Computational Biology Project aims to analyze and characterize image-processing techniques used by cellular biology researchers; provide guidance to researchers in selecting the appropriate techniques for their research; and bring computational and measurement science expertise to the cellular biology community to help them to effectively deal with the large amounts of images generated by their research.

Lead agency:

National Institute of Standards and Technology (NIST)

Agency Mission:

To promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life.

Principal Investigator:

Alden Dima
Computer Scientist
Information Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8970
Gaithersburg, MD 20899-8970

Partner agency:

Karen Kafadar
University of Colorado

General Description:

Research on aging, like all biological research today, is being facilitated by automation that provides instrumentation control and data acquisition. At the same time, there have been advances in imaging and other sensor-based technologies. Researchers are now able to quickly collect large amounts of multimodal image-based data that serves as the primary output of their experiments and as the source of their measurements. These measurements ultimately provide the information required to decipher complex cellular processes including those related to aging.

Unfortunately, biological researchers are left with huge amounts of image data to process and analyze using techniques that are usually outside of their field of expertise. In addition, the large amounts of images require large amounts of secondary information (metadata) for their correct interpretation, handling and storage; gone are the days when a few sentence fragments in a lab notebook could jog memory and guarantee understandable and repeatable results.

Scientific literature has many references to and descriptions of image processing techniques, but experience shows that many techniques have limited applicability; a method that works well for optical character recognition may well fail miserably in cell biology. Indeed, even within a well-

defined field, certain techniques work well only on certain types of images; two images from different data channels of a microscope may require fundamentally different techniques. As image-based measurement becomes increasingly vital to biological research, the measurement uncertainty associated with image processing is increasingly becoming an issue.

Today, the biological researcher is expected to be an expert in his/her field of research as well as a savvy user of image processing software and techniques. There is a deluge of available options, and typically the research chooses tools and techniques that they have been exposed to and feel comfortable with. There is little guidance available, and much of the biological literature seems to give little information about the methods used for analyzing experimental data and their associated parameters. This situation essentially distracts biological researchers from fulfilling the central goals of their research, such as understanding the biology of aging and developing new treatments for aging-related diseases. Given that the U.S. population is aging and that scientific resources are not unlimited, research that aims to improve the ability of biological researchers to handle their critical image-based data will ultimately facilitate the development of new treatments for aging-related diseases.

A basic tenet of the Computational Biology project is that image processing and analysis techniques, despite their implementation in software, are fundamentally measurements and not simply calculations. As such, they can be characterized and understood in a manner similar to other measurement techniques. This suggests that the measurement uncertainty associated with the use of software-based image processing and analysis methods can and should be determined. It also suggests that clear guidance can be given to researchers to aid them in choosing the correct image processing techniques and will facilitate the interpretation of their research data.

The Computational Biology Project aims to analyze and characterize image-processing techniques used by cellular biology researchers; provide guidance to researchers in selecting the appropriate techniques for their research; and bring computational and measurement science expertise to the cellular biology community to help them to effectively deal with the large amounts of images generated by their research.

Excellence: What makes this project exceptional?

Though this project is very new, it is exceptional in its approach to addressing some of the key issues hindering cell biology research that depends on large quantities of multimodal image data, including research on the biology of aging. Typically, a cell biology research project will focus on an area of interest, perhaps one or more particular cell processes related to aging. The research staff will consist mostly of biological researchers who will run experiments and collect data including images from microscopes.

Given the complexity of cellular processes, at some point, the researchers will run into bottlenecks that slow down the pace of the research; perhaps they've collected huge amounts of complex images that need to be analyzed using techniques with which they are uncomfortable. They may bring in expertise from outside of the project's central area of interest to provide additional support. Over time, the project's success will become increasingly dependent on measurement and analysis techniques that are outside the scope of the biological researchers' specialized domain. The project will tend to start spending more time focusing on these ancillary issues instead of making progress on its primary research goal. Given the amount of biomedical research occurring, many projects will be in effect competing for the same outside support to solve a similar set of issues.

If outside expertise is not available, the biological researchers may find themselves in the difficult position of becoming competent in technical fields outside of their primary area of expertise. They may make technical decisions that ultimately hinder their ability to progress with their primary research goals such as understanding the biology of aging and developing new treatments for aging-related diseases.

The Computational Biology Project strategy is to effectively invert the problem to directly tackle the standards, measurements, and informatics issues that can slow down biological research. For example, one key issue with biological image data is the selection of the technique used to extract key features from the rest of the image (segmentation algorithms). Using cellular microscopy images generated by NIST biological researchers, our computational scientists, mathematicians, and statisticians intend to analyze and characterize common segmentation algorithms so that we can publish guidance for biological research community as a whole. This will effectively mitigate the need for each project to address this issue and brings NIST measurement expertise to bear on the problem.

***Significance*:** How is this research relevant to older persons, populations and/or an aging society?

As the population ages, an increasing emphasis will be placed on research related to the treatment of aging-related diseases – much of which will share infrastructure, resources, and expertise with other cell biology research.

Effectiveness: What is the impact and/or application of this research to older persons?

A key component in the treatment of aging-related diseases is the understanding of the complex cellular processes related to aging. This understanding, in turn, depends on sophisticated laboratory and data analysis techniques. By addressing some of the key issues related to the analysis of image data generated by cell biology research, the Computational Biology Project will improve the efficacy of cell biology research in general and as a result, research into the biology of aging will also benefit. This should, in turn, aid in the development of new treatments for aging-related diseases.

Innovativeness: Why is this research exciting or newsworthy?

From our perspective, this research is exciting because it brings NIST's traditional measurement expertise to bear in a new domain (cellular biology) and has the potential to make significant impact in terms of scientific and medical advancements. As our work progresses, we hope that it will become newsworthy in terms of the other advancements that were enabled by it.