![THOMSON REUTERS]

| Home | About Scientific | Press Room | Contact Us |

scienceWATCH.com
TRACKING TRENDS & PERFORMANCE IN BASIC RESEARCH

Interviews | Analyses | Data & Rankings

2008 : September 2008 - Author Commentaries : Olga G. Troyanskaya

AUTHOR COMMENTARIES - 2008

September 2008

🅿 PDF

Olga G. Troyanskaya
Featured Scientist from *Essential Science Indicators*SM

Dr. Olga Troyanskaya has been named a Rising Star in the field of Computer Science, according to an analysis published by ScienceWatch.com *in May. Her citation record in this field in* Essential Science Indicators *from Thomson Reuters includes 31 papers cited a total of 1,533 times between January 1, 1998 and April 30, 2008. She also has Highly Cited Papers in the field of Clinical Medicine. Dr. Troyanskaya is an Assistant Professor in the Department of Computer Science and the Lewis-Sigler Institute for Integrative Genomics at Princeton University.*

+enlarge

In the interview below, she talks with us about her highly cited work.

SW: Please tell us a little about your research and educational background.

My background is interdisciplinary—I have a Ph.D. in Biomedical Informatics and undergraduate degrees in both Computer Science and Biology. My research has always reflected this—I have been involved in bioinformatics research since undergraduate days, first working with Steven Salzberg, then at Johns Hopkins University and The Institute for Genomic Research, and then with Gad Landau and Alex Bolshoy at Haifa University in Israel.

The focus of my Laboratory for Bioinformatics and Computational Genomics at Princeton University is also at the intersection of computer science and molecular biology—in developing novel computational algorithms and systems to address biological problems.

SW: What do you consider the main focus of your research, and what drew your interest to this particular area?

My group's research is in the area of computational functional genomics, specifically in developing novel computational methods for the prediction of protein function, interactions, and regulation from diverse high-throughput biological data. We aim to develop algorithms and systems that can make accurate predictions based on modeling and analysis of noisy high-throughput data, as well as based on very large collections of data. Our research is closely integrated with biology—in fact, an important aspect of our work is the development of integrative technologies that combine computation and experiments.

This research area is especially interesting to me because the technologies my group

develops can make a substantial impact on the fields of biology, and in the long term, biomedicine. We develop methods that can analyze the vast amount of available data in biological context and generate accurate, novel predictions regarding the functions of unknown proteins or structures of key disease pathways. Then, by using such technologies to direct biological experiments, we hope to substantially accelerate the pace of biological discovery, including the elucidation of functions of previously unstudied proteins or the identification of potential disease-related proteins and pathways.

"One of the exciting aspects of my work is in how dynamic the field of computational functional genomics is."

SW: Many of your highly cited papers deal with the analysis of genes from microarray data. Would you talk a little about this aspect of your research—how you got started in it, and what some of your findings have been?

Gene expression microarrays were arguably the first technique that enabled researchers to produce fast and relatively cheap snapshots of the systems-level dynamics of gene regulation. I was excited about the potential of analyzing such data to discover function for unknown proteins and to start to examine the molecular basis of complex diseases on a systems level.

My first foray into microarray data analysis concerned a highly technical aspect of the field: with colleagues at Stanford University, including David Botstein and Russ Altman, we developed a method for accurately estimating missing values in microarray datasets. Such values occur, for example, when a specific expression level cannot be reliably determined in the microarray experiment. The missing value estimation method we developed, KNNimpute, is still widely used in the research community. Since then, our work has included the analysis of multiple clinical datasets, focusing on identifying clinically relevant biomarkers and finding chromosomal amplification and deletions (both from array CGH and gene expression microarray studies), pathway modeling, and most recently, analysis of very large microarray compendia.

I am excited about the potential of performing sophisticated computational analyses of the large existing collections of microarray and other functional genomics data to answer questions that are often very hard to address by a single study. This is one of the directions of our recent work, including developing a "Google"-type search engine for microarrays (SPELL) and probabilistic Bayesian systems for the analysis of diverse functional genomics data (bioPIXIE and MEFIT).

SW: Your most-cited paper in our database in the field of Computer Science is the 2003 *PNAS* paper, "A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*)." Would you give our readers some background on this paper—its goals and findings—and why it is so popular?

This manuscript describes a probabilistic Bayesian method for the integration of diverse genome-scale data into confidence-weighted functional relationship networks among proteins, which can then be used to predict protein function. We demonstrated the principle of probabilistic data integration for functional genomics data, and this is now a very large and active research area, in which my group is still very involved.

"...an important aspect of our work is the development of integrative technologies that combine computation and experiments."

Since then, we have developed a full-scale learning-based system for integration of heterogeneous biological data and prediction of protein function and functional relationship networks (bioPIXIE); this system is widely used by the yeast community. We have also introduced the idea of exploring such questions in the context of specific pathways or tissues, as proteins can have multiple functions in different biological contexts.

SW: How has this field changed since you first started working in it?

One of the exciting aspects of my work is in how dynamic the field of computational functional genomics is. New experimental techniques, biological questions, and computational approaches are constantly coming out, and a diverse group of highly interdisciplinary researchers aims to address these challenges through a variety of approaches. Compared with a decade ago, two key differences are perhaps in the increasing sophistication of the computational methods and in the much closer tie-in of most studies with the experimental biology.

SW: Where do you see this work going in five to ten years?

We are still far from understanding the full complexity of gene function and regulation on a systems level, and my aim is to continue developing integrated computational and experimental technologies for addressing this problem. Our goal is to map cellular regulatory structures and develop predictive models

for effects of genetic and environmental perturbations.

Looking at the long term, my hope is that computational methods will guide genome-scale explorations of complex molecular, cellular, and organismic systems at complementary levels of resolution, some day leading us to integrate our understanding of microscopic biology with macroscopic physiology and medicine.

Olga Troyanskaya, Ph.D.
Department of Computer Science
and
Lewis-Sigler Institute for Integrative Genomics
Princeton University
Princeton, NJ, USA

Olga G. Troyanskaya's most-cited paper with 390 cites to date:

Garber ME, *et al.*, "Diversity of gene expression in adenocarcinoma of the lung," *Proc. Nat. Acad. Sci. USA* 98(24): 13784-9, 20 November 2001. Source: *Essential Science Indicators* from Thomson Reuters.

Keywords: bioinformatics, interdisciplinary research, computer science, molecular biology, computational algorithms, computational functional genomics, protein function, gene expression microarray analysis.

PDF

back to top

2008 : September 2008 - Author Commentaries : Olga G. Troyanskaya

Scientific Home | About Scientific | Site Search | Site Map

Copyright Notices | Terms of Use | Privacy Statement