Tech News

A decade after the genome, bioinformatics comes of age

Before the emergence of genome-scale sequencing, the combination of biology and computer science seemed more of an intellectual curiosity than a research necessity. A decade ago, most life scientists didn't generate large enough volumes of quantitative data to warrant the use of information technology to facilitate their analyses.

But the proliferation of data from next-generation sequencing, proteomics, cellular systems biology, and advanced imaging studies makes it difficult to question the need for computational methods to analyze biological data these days. "When I first began in this field, I had enormous difficulty explaining to people what I did because it seemed so arcane," says Mark Musen, now the codirector of Stanford University's Biomedical Informatics Training Program. Now, he says, "people are taking for granted that information technology plays an important role in the work that takes place in academic medical centers, in industry, and in government."

To facilitate this growing need in biology, many graduate training programs in bioinformatics sprang up during the Human Genome Project to deal with sequencing data. But according to Brian Athey, Chair Designate of the University of Michigan Center for Computational Medicine and Bioinformatics, in the decade since the completion of that project, the field has evolved from one focused squarely on the genome to one encompassing relationships among a broader set of biological and medical data. And, he says, growing interest and funding resources for translational research (such as the NIH's Clinical Translational Science Awards) continue to blur the line between bioinformatics and fields that grew up to handle medical data, such as clinical informatics and biomedical informatics.

The field of bioinformatics has come of age, with scientists who seized opportunities for bioinformatics training in the post-genome era now making their mark on a variety of research questions. Their ability to see biological data through the lens of mathematics and computer science enables them to approach fundamental biological questions from a different vantage point.

Bioinformatics Training

Because of the areas that their field bridges, scientists trained in bioinformatics generally have some combination of undergraduate training in biology and mathematics or computer science. Stanford University biomedical informatics professor Atul Butte, who has an M.D. and a Ph.D. along with an undergraduate degree in computer science, tends to look for students with 'hybrid' undergraduate experience: that is, a major in one relevant area with courses in another, along with research experience in bioinformatics.

But even with such varied preparation, bioinformatics graduate programs recognize that entering students are usually stronger in one field or the other. Brian Pierce, who works on computational protein design at the University of Massachusetts Medical School in Worcester, MA, was a computer scientist with software development experience, but no formal biology or biochemistry training, when he decided to enter Boston University's graduate program in 2002.

Timothy Reddy, Postdoctoral Fellow, HudsonAlpha Institute for Biotechnology. Photo courtesy by Jason Gertz.

Computational expertise—in mathematics, probability, statistics or computer science—is generally harder for a graduate student to pick up on the fly, says Tom Tullius, who is interim director of the Bioinformatics Program at Boston University. Tullius and his colleagues generally look for students with strong computational skills and interests, and then help them develop biological expertise. The University of Michigan has developed a two-week intensive course for incoming students without laboratory training—a molecular biology boot camp of sorts—which offers them hands-on experience at the bench before graduate school begins. Stanford University's program in Biomedical Informatics has its roots in clinical informatics, Musen says, and incoming students are a diverse mix of computational biologists, experimental biologists, physicians, pharmacists, and nurses.

Among the bioinformatics graduate programs that emerged over the last 15 years, many started as master's-degree programs with a particular focus on applying informatics skills to solve problems in industry. That trend holds today: many master's-level trainees are either looking for jobs in industry or are industrial scientists looking to augment their bioinformatics skills, Tullius says. Ph.D. students generally take similar coursework, but combine that work with a series of laboratory rotations and eventually a thesis project. These scientists end up taking a number of career paths: industry, academia and government research labs.

After Lee Katz finished a master's degree in bioinformatics at Georgia Institute of Technology in 2005, he found that the field was new enough that many potential employers with whom he interviewed didn't completely understand his skill set. With an inkling that the field hadn't quite come of age, he decided that he might be wise to bide his time and gain research experience in Georgia Tech's Ph.D. program. The strategy paid off: in 2010 Katz was hired by a contractor for the Centers for Disease Control and Prevention in Atlanta to work on pathogen genomics and meningitis, a natural extension of his doctoral research.

Thriving at the Interface

Even though the value of bioinformatics is now recognized, researchers working in the field still face all the challenges extant at the interface of two different disciplines. They have to decide which tasks to focus on, and must also find the right mix of environment,

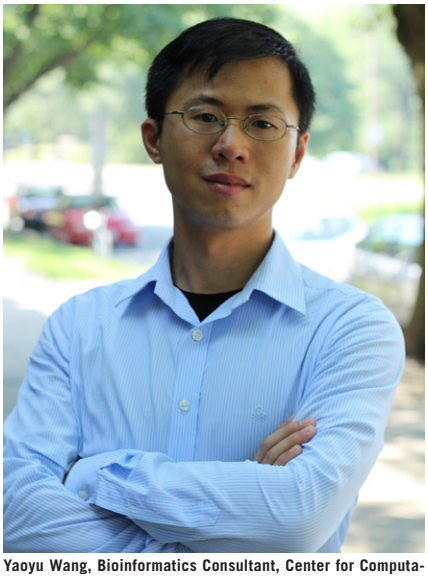
resources, and collaborators that are best-suited to their research goals.

A fundamental question for many young bioinformatics researchers is finding the right balance between computational work and experimental work. Even though Tim Reddy started out as an undergraduate computer science major, his postdoctoral work at the HudsonAlpha Institute for Biotechnology in Huntsville, Alabama is focused on how stress hormones change gene regulation—a question with implications for single-cell organisms and chronic diseases in humans such as diabetes. Reddy evenly splits his research efforts, spending approximately half of his time doing experimental biology while the other half he devotes to computational biology. Still, he says, such an even split is relatively unusual these days. But as more experimental biologists pick up computational skills and more computational researchers learn more biology, he suspects that the distinctions between the two fields will become less defined.

After completing his Ph.D. at Boston University, Brian Pierce took a job at Pfizer in La Jolla, CA, in 2008, working in a bioinformatics role to understand genomics and microarray data related to vaccine development. Although he enjoyed the work, he ultimately decided to return to computational protein design. In 2010, he took a position as a research assistant professor at the University of Massachusetts Medical School working alongside Ph.D. advisor Zhipeng Weng on protein design for new immunotherapeutics.

Bioinformatics researchers also rely on having sufficient information technology resources, says Yaoyu Wang, a bioinformatics analyst at the Dana-Farber Cancer Institute in Boston, MA. As a postdoc, Wang worked in a lab that studied the evolution of HIV in patients. One of the difficulties was the amount of data produced but the limited infrastructure and lack of bioinformatics colleagues with whom to discuss computational issues.

The more nontraditional the research question, the more important it becomes to possess the right network of collaborators. Kaustubh Supekar, currently a postdoc at Stanford University, is working at the interface of neuroscience and informatics, combining classical machine learning with imaging and neuroscience to look at ways of correlating various brain imaging scans with medical diagnoses. Someday this form of research could help doctors use imaging data to diagnose complex brain disorders such as autism. The work is challenging: doctors have a limited number of brain scans of children with autism, but each brain scan contains immense amounts of data.



Yaoyu Wang, Bioinformatics Consultant, Center for Computational Cancer Biology, Dana-Farber Cancer Institute. Image courtesy of Yaoyu Wang.

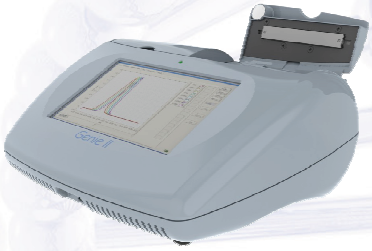
With so many different disciplines required to conduct his research, "I cannot do everything," Supekar says, whose collaborators include radiological imaging experts, electrical engineers, neurologists, and psychiatrists. He can, however, speak the language of each of these disciplines. It is this ability to translate ideas from one field to another that enables him to see patterns and relationships, thereby leading to new hypotheses.

Even with the opportunities available in this research field, bioinformatics also presents career challenges. Supekar wants to continue working in academia, but due to the diversity of expertise his work requires, he'll need to carefully weigh the benefits of infrastructure and personnel at a particular institution against the resources he'll need to carry out the work. "There are no departments where I can just go and apply for faculty jobs," he says. Supekar will have to find an institution with strong neuroscience and computer science programs and the right combination of neuroscientists, engineers, computer scientists, and clinicians to continue the work he's started.

Moving from the Periphery to the Center

The role of bioinformatics researchers in understanding biology continues to evolve since the Human Genome Project. At first many researchers built online databases designed as research tools such as University of California–Santa Cruz's genome browser (www.genome.ucsc.edu), the European database Ensembl (www.ensembl.org) and the series of databases available at the National Center for Biotechnology Innovation (www.ncbi.nlm.nih.gov), Butte says. But moving forward, it's

Genie™ II



Molecular detection by Isothermal amplification of DNA or RNA in under 10 minutes

Highly sensitive optics for detection of fluorescence and luminescence at temperatures of up to 99°C

Two independent heating blocks

Compact, lightweight and fully portable with internal rechargeable battery

Operation without a computer via a large touch-screen panel

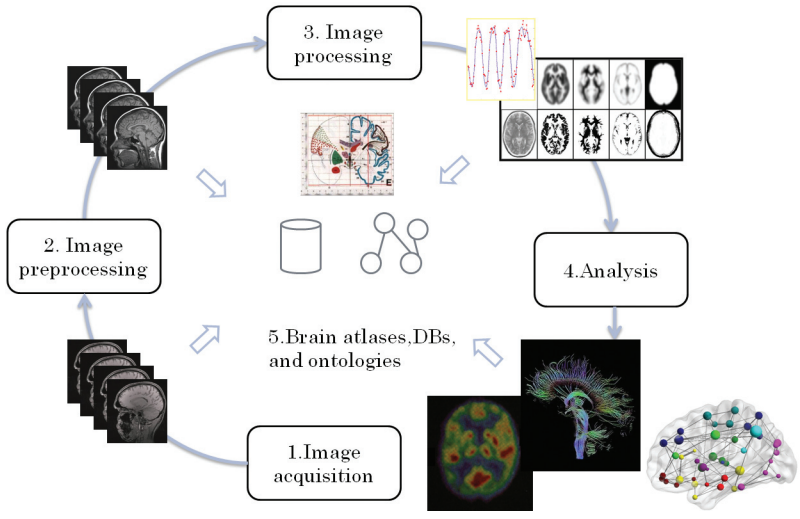
Operates with two specially-designed strips of 8 tubes, which can be used independently or together for a 16 sample run

Low cost and easy to use

OptiGene

www.optigene.co.uk
info@optigene.co.uk
Tel: +44(0)1403 274980
Fax: +44(0)1403 271017

Unit 5, Blatchford Road,
Horsham, West Sussex,
England RH13 5QR



In the field of neuroimaging informatics researchers design novel methods for: (1) Image acquisition: increasing spatial and temporal resolution, (2) Image preprocessing: improving quality of acquired data, (3) Image Processing: segmenting, registering, and visualizing preprocessed data, (4) Image analysis: extracting and quantifying information related to a hypothesis, and (5) Image storage: assembling raw and processed brain imaging data and knowledge from databases, brain atlases, and ontologies. Image courtesy of Kaustubh Supekar.

not enough for bioinformatics researchers to build databases that serve as resources for other researchers. The push now, he says, is for bioinformatics researchers to make fundamental discoveries about biology based on computational tools.

And with the ability for biologists to produce such large amounts of quantitative data, the field is reaching a type of inflection point where it's possible to generate more data than can be reasonably stored, Reddy says. "The question is no longer 'can you do this experiment?' but 'if you can do this experiment, what are the questions you should be asking?'"

With this increasing data deluge, it becomes tempting for experimentalists to think of their computational colleagues as service-providers, people who can help them analyze orphan data and deal with particular problems. Musen notes that this issue remains a growing pain for the field, adding that there is a tension between the furthering of biological questions and the development of new methods and algorithms. "It's only recently that we've been successful in making it clear that bioinformatics is doing things that are important in their own right," he says. "Ultimately, those objects of study will be important separate from whatever biological results they yield."

Nevertheless, biology's ever-present need for the support of informatics has led to the formation of consulting-based bioinformatics teams. Wang works for one such group at Dana-Farber Cancer Center, which is a hybrid between an academic lab and a consulting firm. The team includes specialists for information technology, sequencing, and bioinformatics, as well as administrators to facilitate the necessary collaborations. Academic laboratories that need computational support for translational research projects can approach the group at Dana-Farber, and they'll set up a team to work on the task at hand.

Within a month of completing his bioinformatics Ph.D. at the University of Michigan in 2006, Daniel Rhodes founded Compendia Bioscience in Ann Arbor, Michigan. The company grew out of his graduate work to build a database that assembled publicly available cancer genomics data, Oncomine. Compendia Bioscience has licensed that technology to develop software and consulting services for companies working on cancer drug development.

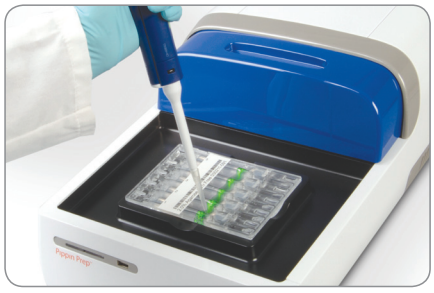
Rhodes sees an increasing need for bioinformaticians who also have deep knowledge of a particular biological problem, and who are full team members in designing solutions to that problem. "Are bioinformaticians at the center [of a research problem], or at the periphery serving the center?" he asks. He argues that collaborations proceed more smoothly when researchers with bioinformatics expertise are embedded within project teams, rather than providing a requested data analysis. "We're scientists engaged in the problem we're trying to solve and trying to craft our approach to the needs of that problem."

Written by Sarah Webb.

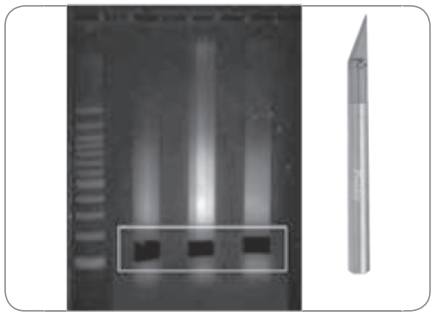
BioTechniques 51:157-161 (September 2011)
doi: 10.2144/000113731

To purchase reprints of this article, contact: biotechniques@fosterprinting.com

exacto.



not exacto.



NGS Library Construction.
Pippin Prep™

888.744.2244
www.sagescience.com
sage science