

Opinion

Bioinformatics: alive and kicking

Lincoln D Stein^{*†}

Addresses: ^{*}Ontario Institute for Cancer Research, Toronto, ON, M5G 0A3, Canada. [†]Cold Spring Harbor Laboratory, NY 11724, USA.
Email: lincoln.stein@gmail.com

Published: 17 December 2008

Genome Biology 2008, **9**:114 (doi:10.1186/gb-2008-9-12-114)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/12/114>

© 2008 BioMed Central Ltd

Abstract

Bioinformatics has become too central to biology to be left to specialist bioinformaticians. Biologists are all bioinformaticians now.

Six years ago I felt like the boy who hit a telephone pole with a wooden stick at the exact instant a power failure darkened all the lights across the US Northeast. In February 2003 I gave a keynote address for the second annual O'Reilly Bioinformatics Technology Conference called 'Bioinformatics: Gone in 2012' in which I predicted that bioinformatics as a discipline separate from mainstream biology would be gone in ten years. My talk was met with resentment, disappointment and stunned disbelief by an audience of computer geeks who had come to the conference for the express purpose of getting in on the hot new thing. Worse, this was the year in which biotech and pharma realized they had significantly overinvested in bioinformatics and started large-scale layoffs. In the light of a downsized bioinformatics market, the O'Reilly publishing house cancelled a series of planned bioinformatics textbooks, and never sponsored another Bioinformatics Technology Conference. It seemed as though my predictions had come true ten years early, and although I knew it was all coincidental, I couldn't suppress the sinking feeling that I was the villain who triggered the collapse.

As it happens, my predictions were quite wrong. Not only did bioinformatics recover nicely from its early-millennium swoon, but it looks like it is here to stay through 2012 and beyond. Is this a good thing? At the halfway point between my keynote address and the date of my dire predictions, let's have a look at my arguments and update them against what has happened since.

The core of my 2003 beef with bioinformatics was that it is a family of techniques and not a research discipline unto itself. Today, if you search for the definition of bioinformatics on Google, you get a family of explanations that boil down to

'using computers to manage, organize and analyze large amounts of biological information'. I don't find this a satisfying definition. Physicists, geologists and chemists use computers to manage, organize and analyze large amounts of data from their disciplines, and at that time they did not have disciplines named 'physicoinformatics', 'geoinformatics' and 'cheminformatics'. My argument was that information management was so fundamental to the biological sciences that bioinformatics would be absorbed into the mainstream biological curriculum just like the techniques of molecular biology, sequencing and macromolecular separations. In ten years, I felt, every biology graduate student and postdoc would have just as much facility with computer-based information management tools as they had in 2003 with multi-channel pipetors, electrophoresis units, ultracentrifuges and other stock-in-trade. My prediction was that bioinformatics would become one of a series of core courses taught in undergraduate and graduate biology programs, and that there would be a vanishing market for researchers who focus solely on biological data management.

I was half right. Today, bioinformatics lectures are offered by almost every undergraduate and postgraduate biology program in North America, Europe and Asia. Many colleges and universities go further and make bioinformatics courses part of the core biology curriculum. At the same time, educational institutions offering certificates and advanced degrees in bioinformatics have increased dramatically over the past decade. In 1998, a compilation of institutions offering bioinformatics training listed only ten degree-granting programs in the USA [1]. Ten years later, there are at least 74 such programs in the United States and Canada, and more than 150 worldwide [2,3]. At the same time, the

average biologist has become far more computer-savvy than he or she was in 2003. It is now routine for wet labs to maintain Wikis to organize their papers and protocols, and unexceptional to see an enterprising graduate student or postdoc create a relational database to manage the results from a complex set of experiments. Accessible web-based bioinformatics tools are commonplace, and many, in particular the University of California, Santa Cruz (UCSC) genome browser [4], encourage researchers to upload and analyze their own datasets.

With all these training and online resources available, one would think there would be less need for card-carrying bioinformaticians, and my personal experience suggests that this is the case. Eight years ago, at the height of the bioinformatics bubble, pharmaceutical companies and other industry players were offering big premiums to qualified bioinformaticians. However, as of 2008, *The Scientist's* annual salary survey reported a median income of US\$85,000 for all of the life sciences in the United States [5], while the OpenWetWare bioinformatics career survey [6] found a median income of just US\$70,000 for self-identified bioinformaticians in North America. Granted, the two surveys are not comparable, but it does suggest that the salad days of six-figure salaries for entry-level bioinformaticians are unlikely to return.

On the other hand, bioinformatics as a named discipline is stronger than ever. A decade ago at the annual Cold Spring Harbor Biology of Genomes meeting, the bioinformatics session would be offered early on Sunday morning (the last day of the meeting) and was sparsely attended. Now bioinformatics pervades the entire meeting; every talk has a strong bioinformatics or computational biology component, and the talks that are heavy in computational biology are always among those that are most heavily attended. A major contributor to this trend is the breathtaking growth in the size and complexity of datasets. Six years ago the largest dataset imaginable was the human genome, with its 3 billion base pairs and 100 million raw sequencing reads. With advances in sequencing technology, it is now possible for a single machine to produce 1.7 billion base pairs over a two- to three-day period, and sequence a human genome at high coverage in just about a month. This revolution in sequencing technology has spawned such projects as the 1000 Genomes Project [7] and the International Cancer Genome Consortium [8], each of which will generate datasets thousands of times larger than the original Human Genome Project. Other aspects of biology have experienced similar technological leaps; for example, advances in fluorescently tagged markers and digital imaging now allow the temporal and spatial dynamics of gene expression to be followed in single cells in living organisms. In neurobiology, innovations in electrophysiology and optics allow the coordinated electrical activity of hundreds of neurons in a living animal's brain to be followed simultaneously. The Allen Institute for

Brain Science in Seattle, Washington, has produced a database of gene-expression information in the mouse brain [9] that is simply too large for the traditional practice of making local copies. Very serious computer science is needed to extract knowledge from such datasets.

My argument against bioinformatics based on analogy with chemistry and geology also didn't withstand the test of time. A few years after I gave the talk, the terms 'geoinformatics' and 'chemoinformatics' appeared on the scene, and show no sign of disappearing. Perhaps I should trademark 'physicoinformatics' before it is too late?

So bioinformatics isn't disappearing. But who is giving these bioinformatics talks, and making and analyzing these large databases? By and large these are not people who call themselves bioinformaticians. Instead, we are witnessing the rise of a new generation of computational biologists who spend part of their time at the bench and part of their time at the computer. Particularly eye-opening for me has been my recent experience at the Ontario Institute for Cancer Research, where I have been recruiting principal investigators for the new Informatics and Biocomputing Department. Almost all the young investigators that I have interviewed have asked about bench space, laboratory equipment and supplies. Clearly these researchers see themselves as biologists first and foremost; for them bioinformatics is a technique to be used, not a speciality to follow. A career limited to computational data management and analysis alone is too confining a niche for them; they want to take control of the datasets they generate, and temper theoretical models with empirical tests. Even I am seeing the writing on the wall, and have started to spec out the equipment for a modest wet lab of my own.

So here is my revised prognosis for the next five years:

Bioinformaticians: gone by 2012. Bioinformatics: stronger than ever.

References

1. **Bioinformatics: Academic/Degree Programs** [http://biotech.icmb.utexas.edu/pages/bioinform/biprograms_us.html]
2. **ISCB: Degree and Certificate Programs** [<http://www.iscb.org/iscb-degree-certificate-programs>]
3. **A list of bioinformatics courses and degrees worldwide** [<http://www.nslj-genetics.org/bioinfotraining>]
4. **UCSC Genome Browser** [<http://genome.ucsc.edu>]
5. **2008 Life Sciences Salary Survey** [<http://www.the-scientist.com/2008/9/11/45/1>]
6. **Biogang:Projects/Bioinformatics Career Survey 2008** [http://openwetware.org/wiki/Biogang:Projects/Bioinformatics_Career_Survey_2008]
7. **1000 Genomes Project** [<http://www.1000genomes.org/page.php>]
8. **International Cancer Genome Consortium** [<http://www.icgc.org>]
9. **Allen Institute for Brain Science** [<http://www.brain-map.org>]