Scientists

A Scientist and an Advocate

Steven Salzberg

Professor, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

Steven Salzberg's creation of software tools for sequence analysis and genome assembly caught our attention. Curious to know more, *BioTechniques* contacted him to find out about the ambition, character, and motivation that led to his success.

What first drew you to the field of bioinformatics?

While working on a Ph.D. in computer science at Harvard University in the late 1980s, I heard about the Human Genome Project (HGP). Immediately, it struck me that this was going to be one of the most exciting scientific endeavors in the world for years to come and I wanted to be part of it. I started to educate myself about genomics, genetics, and molecular biology by sitting in on Stephen Jay Gould's evolution courses and reading text books, all while following the HGP progress.

Once enough sequence was obtained for researchers to locate and identify genes, I recognized an opportunity to contribute. I started applying methods I knew from work in natural language processing to the problem of gene finding. Around the same time, I first met Hamilton Smith, a Johns Hopkins University colleague and Nobel Prize winning molecular biologist who discovered restriction enzymes. He was interested in computer science as a hobby and was curious about how I was approaching the question. He later introduced me to researchers at The

Institute for Genomic Research (TIGR); I started collaborating with them and finally joined TIGR in 1997.

What has been your most significant scientific contribution so far?

My group at TIGR was responsible for the computational analysis of the bacterial, archaeal, viral, parasitic, and *Arabidopsis* genomes first sequenced and published during the late 1990s and early 2000s. Each of those papers was a landmark in its own area. Our efforts focused on genome assembly, as well as finding and annotating genes using statistical techniques developed in my group. We created a number of successful and widely adopted computational biology techniques. Even today, 14 years after its publication, our GlimmerM program is still the most accurate gene finder for bacteria and archaea.

Currently, we are focused on aligning short reads from next-generation sequencing datasets, including RNA-Seq, which involves alignment of intron spanning transcripts to multiple locations on the genome, assembly of the reads, and quantification of the transcripts. I think it is fair to say that the software we developed, Bowtie, TopHat, and Cufflinks, have become the standard in the field.

During your career, what has been your biggest surprise?

At TIGR, we were in the midst of sequencing the anthrax genome at the time of the 2001 attacks, so the FBI approached us to aid in their investigation. Unlike our previous and subsequent genome projects, our goal with the anthrax genome was identifying mutations that could be used as forensic markers. The strains we sequenced appeared 100% identical across the genome and it was very frustrating looking for differences. But finally, we discovered a small duplication of about 1000bp in one of the strains. It was an exact tandem duplication and the assembler had collapsed the two copies on top of each other so the sequence looked identical to the reference. We only identified the duplication after combing through the raw data looking for individual reads that didn't fit with the assembly. We later found two other duplications. In the end, the FBI used

those three duplications to trace the attack strain to a particular isolate in a specific vial at Fort Detrick.

You have also written a significant number of editorials addressing misleading health claims. What motivated you to devote your efforts here?

In 2004, my group at TIGR began to sequence thousands of strains of the influenza virus in an effort to track sequence changes so that we could improve the design of new vaccines. Through an interaction with a journalist, I became aware of the anti-vaccine movement, which is based on several beliefs that are wildly inaccurate. While we were working day and night to improve influenza vaccines, a growing movement was simultaneously advocating avoiding vaccination.

The overarching goal of doing biomedical research is to develop better treatments so people live longer, healthier lives. I decided that in addition to my research, another way I might achieve this goal was to educate the public on the benefits of vaccines and counter the pseudo-science promoted through the anti-vaccine movement, starting with the thoroughly discredited theory that vaccines cause autism. Once I started writing about this, I encountered other kinds of quack treatments offered by true believers and scam artists alike. My hope is that I can reach people while they are still open minded so they will look at the evidence and stop doing things to harm themselves.

I write about topics for the scientific community as well, voicing my opposition to gene patents as well as my support for open access publishing, open source software, and sharing scientific data as freely and quickly as possible. As scientists, our goal is to solve scientific problems and move our fields ahead; the best way to do that is to be open and share.

Interviewed by Kristie Nybo, Ph.D. Image courtesy of Steven Salzberg.

BioTechniques 53:71 (August 2012)
doi 10.2144/000113898

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