

BC professor makes sense of the DNA data deluge

Boston Business Journal - by [Jesse Noyes](#) Journal staff



W. Marc Bernsau

Gabor Marth, associate professor of biology at Boston College, is developing software that can better analyze data produced by genome decoding machines.

As technology gradually makes sequencing human genomes faster and more accessible, the need to make sense of the data has become increasingly important.

That's where Gabor Marth comes in.

Marth, an associate professor of biology at **Boston College**, is developing software that takes the heaps of data produced by genome-decoding machines and analyzes it with greater accuracy and speed than previous products. The software, called PyroBayes, could lead to better understandings of how DNA affects or causes diseases and potentially provide medical breakthroughs, Marth said.

To date only a few people have had their whole genomes sequenced. But projects under way such as the 1,000 Genomes program and attempts to commercialize the process are aimed at driving costs down while seeking new discoveries about the human genome.

As faster DNA-reading machines arrive in labs, the need to analyze and understand the mountains of data produced is growing. Marth's software carries out two processes, one called realignment and the other called variation discovery, he said.

The way the software works: First, a genome-sequencing machine produces the 3 billion bases of DNA found in humans, then Marth takes that data and compares it with a previously ordered whole human genome to create a linear sequence. Marth compared this to trying to piece together an entire book while only be supplied a few words at a time in a random order.

Whole genome-sequencing "machines can only read genomes in very small bits, basically like a word," Marth said. "At the end of the day you will see that you've read the book several times, but you haven't read the whole thing through."

Marth's software looks for small differences between the genomes, what he said would amount to "typos" in comparing separate copies of the same book. Those differences can result in significant changes in an individual's health and researchers are hoping identifying mutations will lead to diagnosing, and potentially treating, human diseases.

That is the purpose of the 1,000 Genomes project, an effort under way by several scientific research centers, including the **National Institutes of Health** (NIH). The purpose of the

project is "to provide catalogues across the whole genome to get at almost all variants in the genome," said Lisa Brooks, program director of the genetic variation program at NIH.

Marth's software is playing a key role in the project, and NIH has awarded his lab more than \$1.3 million to develop the software over four years.

But Marth said he also plans to commercialize his software for companies looking to develop therapies for various diseases. "Software that could do these things or on this scale did not exist or was so inefficient or inaccurate that it was almost not useful," Marth said.

Previous software packages that carried out similar functions ranged in cost from \$5,000 to \$25,000.

Kristen Stoops, director of business development for informatics at Cambridge-based **Helicos Biosciences Corp.** (Nasdaq: HLCS), a company that develops genome sequencing machines, said that software like Marth's "is critical to be able to understand the data that our technology is producing."

"Gabor and his lab have really taken the leading role in the academic community on the development of these tools," she added.

Jesse Noyes can be reached at jnoyes@bizjournals.com.

All contents of this site © American City Business Journals Inc. All rights reserved.