PE/TIGR provokes genome sequencing skeptics

The announcement in May by Perkin-Elmer (PE, Norwalk, CT) and The Institute for Genomic Research (TIGR, Rockville, MD) of the formation of a new commercial venture to "substantially" sequence the human genome in three years (*Nature Biotechnology* 16:491, 1998) has been met with some skepticism both from competing commercial operations and those involved with the publicly funded genome efforts.

The sequencing effort, to be led by Craig Venter at TIGR, will receive around \$200 million in funding from PE. Venter, like others at companies such as Incyte (Palo Alto, CA), intends to undertake whole genome "shotgun" sequencing—breaking the genome into short stretches, and randomly sequencing the fragments. Venter plans to repeat the process 10 times (10× shotgun sequencing), an approach that would produce overlapping fragments covering close to 100% of the entire genome. The process will be facilitated by new PE DNA sequencers that are still in development and due for release early next year. According to Venter, the sequence data will be made publicly available but with a delay to allow patenting of an estimated 100–300 genes of interest. PE/TIGR plan to sequence the 120 megabase *Drosophila melanogaster* genome by next year as a test of its strategy and sequencing technology.

Incyte and Human Genome Sciences (HGS; Rockville, MD) have been identifying and sequencing genes in the human genome for the last several years. Roy Whitfield, Incyte president and CEO, is not threatened by Venter's program. He says that Incyte's own plans—3× shotgun coverage yielding about 95% of the entire sequence in both coding and noncoding regions—will be completed in October 1999. Furthermore, as part of its collaboration with Molecular Dynamics (MD; Sunnyvale, CA) and Amersham Pharmacia Biotech (Uppsala, Sweden), Incyte has been using a new 96-capillary electrophoresis system developed by MD for more than six months. It has also been using 96-lane slab gel electrophoresis adapted from standard PE sequencing equipment. Whitfield contends that the automated instruments that PE is promising do nothing that Incyte's existing equipment does not already do. Incyte recently completed 3× coverage of the 17 megabase, eight chromosome-pair genome of *Candida albicans*, the dimorphic yeast that causes thrush. "To us it represents a minihuman genome," says Whitfield.

William Haseltine, chairman and CEO of HGS, also insists that the PE/TIGR venture—and perhaps much of the publicly funded effort—is somewhat redundant. HGS, Haseltine says, has identified more than 100,000 unique genes of an estimated 120,000 total. Expressed Sequence Tags (ESTs) and at least partial sequences of more than "95% of all the genes are already in hand, and the majority are available on the World Wide Web," asserts Haseltine. "The race to find human genes is over," he says, adding that "the Human Genome [Project (HGP)] juggernaut" has "systematically ignored" such claims.

The public science community worries that privately generated data—whether from

Incyte, HGS, or PE/TIGR—will not find its way into public databases in a timely manner. Delayed release for gene patenting purposes (as Venter suggests) and the mixing of public and private data before its release will taint any true reciprocity of data exchange. "It's very much a one-way window," says Michael Morgan, program director of genetics at the Wellcome Trust (London). His philosophy is that "If it's not in the public domain, it hasn't been done."

The HGP has deposited approximately 4.5% of completed, annotated sequence in the public databases, and as much as 10.5% of the raw sequence data. It is currently estimated that more than one third of the total genome sequence will be available by mid-

Genomica set to score twice

Beneath the headline genome sequencing squabbles, a number of companies are finding a niche developing software to mine the copious data produced by the gene sequencing and functional genomics enterprise. In the past year or so, companies such as NetGenics (Cleveland, OH), Molecular Informatics (Santa Fe, NM), and Structural Bioinformatics (San Diego, CA) have secured venture financing, signed deals with pharmaceutical companies, or released so-called bioinformatics products to simplify the analysis of genomics data.

Now, newcomer Genomica (Boulder, CO) is set to do all three at once. On June 4, the company signed its first deal, a multi-year agreement with the genomics firm Oxagen (Abingdon, UK), and expects to sign a bigger deal with a "large multinational" within days. "As you might expect, it's an exciting time around here right now," says Thomas G. Marr, the company's founder and CEO. "Not only have we just released the final version of our Discovery Manager product, but we are hoping to put it in the hands of our first two customers within the next couple weeks." According to Susan Strong, Genomica's sales and marketing director, those two placements should raise over \$1.0 million for the company. At the same time, Marr is on the road with underwriter Punk, Ziegel & Co. (New York), attempting to secure a second round of venture financing.

For Genomica, founded in 1995, this represents the end of an effort that was con-

ceived at Los Alamos National Laboratory (Los Alamos, NM), brought to term at Cold Spring Harbor Laboratory (Cold Spring Harbor, NY), and pushed into life at the urgings of the US NIH (Bethesda, MD). "My colleagues and I had developed these tools that were capable of managing the diverse set of data being generated by the Human Genome Project, integrating findings from animal and human studies, and then allowing users to mine and analyze that data using a friendly interface," explained Marr. "NIH, which had supported this work, wanted us to commercialize these tools and actually wrote to us asking us to form a company."

The result is Discovery Manager, a suite of tools that uses over 400 different object classes to represent various genomic entities and allows heterogeneous data from a variety of sources to be unified in one data set for more useful analysis. "It was important, according to our beta testers and potential users, that Discovery Manager could incorporate all of the proprietary legacy databases that the companies themselves have created over the years," said Strong.

One thing that sets Discovery Manager apart from other bioinformatics tools is its ability to incorporate results from epidemiological studies and clinical trials into the same database structure as the genomics data. According to scientists at firms that have reviewed the software package, other products—such as Molecular Informatics's BioMerge and BioLIMS, and NetGenics's SYNERGY package—are designed to perform sequence analysis and search for homologies among known genes and proteins.

Joseph Alper

Vicki Glaser is a freelance writer working in Allentown, PA.

Joseph Alper is a freelance writer working in Louisville, CO.