

Perspective

Next-Generation Sequencing Opens Up New Opportunities

Bioinformatics experts at NCGR embrace the next-generation Illumina® sequencing technology.

Most groups start with data generation and move into data analysis. But that isn't the story of the National Center for Genome Resources (NCGR). Taking the opposite path, NCGR began with bioinformatics expertise in place, and grew into a data-generating powerhouse. NCGR chose the Genome Analyzer from Illumina to meet their needs for round-the-clock, easy-to-operate sequencing.

A BIOINFORMATICS BEGINNING

NCGR was founded nearly 15 years ago with the goal of sharing and analyzing DNA sequence information obtained from the Human Genome Project. As a provider of bioinformatics and software systems, they developed the very first sequence analysis system. Over time, however, data analysis wasn't enough. Four years ago, NCGR decided it was time to head in a new direction. To accomplish this, they brought in a new leadership team.

Dr. Stephen Kingsmore was part of that team. He began his career as a physician in Northern Ireland. Twenty years ago Dr. Kingsmore moved to the U.S. and began studies at Duke University in human genetics, which, at the time, was considered a new field. From there his career path took him to the University of Florida, Curagen Corporation, and a Yale spin-out company called Molecular Staging. Four years ago, Dr. Kingsmore decided it was time to return to the non-profit sector and found a home at the NCGR.

MOVING TO DATA GENERATION

The new leadership team immediately recognized the value of maintaining NCGR's original bioinformatics mission, but saw merit in expanding capabilities for data generation and NCGR-driven research. The next-generation sequencing revolution was concurrently taking shape. Instead of traditional long Sanger-sequencing reads, researchers were starting to generate lots of short-read sequences that required very different analysis methods. Using human messenger RNA sequence data from patients diagnosed with lung cancer (provided by Brigham and Women's Hospital), NCGR developed the first software system specifically designed to analyze short sequencing reads. Today, this system is known as the Alpheus Sequence Variant Detection Pipeline. This web-based application provides an intuitive environment for bench scientists to mine the billions of base pairs of sequence now available to them.

The next step for NCGR was to increase their data generation efforts. They began to evaluate different platforms available for next-generation sequencing. NCGR considered the Roche 454 sequencer, which had been on the market the longest, "but [we] had some concerns about the data quality and cost," says Dr. Kingsmore. Furthermore, NCGR decided the AB SOLiD technology was not sufficiently mature and chose to purchase the Illumina® Genome Analyzer system. "We were interested in sequencing RNA at very, very high throughput and depth. The work-



Dr. Stephen Kingsmore is the president and CEO of the National Center for Genome Resources (NCGR) in Santa Fe, New Mexico. NCGR is a non-profit research institute dedicated to improving human health and nutrition through genome sequencing and analysis.

flow for the Genome Analyzer was significantly simpler on the sample prep side than the rival technologies. If you need to generate 50+ sample libraries a week, or even if you have to generate five libraries a week, emulsion PCR-based methods are really tricky to gear up into production phase. For our organization, we just couldn't see using a front-end that involved emulsion PCR. Solid-phase bridge amplification has really panned out for us. It was a key concern and a key advantage."

FORGING AHEAD WITH THE GENOME ANALYZER

Seven months ago, the NCGR received its first Genome Analyzer. Demand drove them to order three more, with plans to run all four instruments 24 hours a day, seven days a week. They plan to remain true to their bioinformatics roots by pipelining and setting up web-based interfaces that will enable people to analyze the newly generated data. "We are doing two principal things: resequencing tissues or DNA samples looking for nucleotide variants to associate with traits or diseases, and digital transcript expression, where instead of mining for nucleotide variants, people want to look at differential expression of transcripts by counting molecules. Merging these data seems to be a powerful new way of understanding complex genetic disorders and traits."

Dr. Kingsmore and others at the NCGR are also pushing their own projects forward. One of the largest efforts, led by Dr. Greg May, involves improving crop production. To do so, the NCGR is studying the relative environmental tolerance of common variants of inbred plant strains. This is much more difficult in humans. Large areas of common variants are found

in people with similar attributes; however a series of newer variations, called rare variants, may lead to a particular phenotype.

The NCGR is using this "rare variant" approach in its schizophrenia studies. These rare variants can be identified and then incorporated onto a new array for large-scale experiments. "We want to bridge sequencing for discovery and array-based approaches for validation of those discoveries. We see the future being much more about personalized medicine fueled by multiplexed diagnostics and also about point-of-care testing...moving our content into clinical studies and validating it for use in the next generation of medical practice," emphasizes Dr. Stephen Kingsmore.

NCGR plans to continue using the Illumina technology and their bioinformatics expertise to pursue their mission: To benefit human health and nutrition through genome sequencing and analysis.

"The workflow for the Genome Analyzer was significantly simpler on the sample prep side than the rival technologies."

ADDITIONAL INFORMATION

To learn more about the Genome Analyzer system and customer-demonstrated applications, visit www.illumina.com/sequencing.

We are committed to providing you with the content you want as a member of the Illumina community. Please email us with comments and suggestions for topics at icomunity@illumina.com.

FOR RESEARCH USE ONLY

© 2008 Illumina, Inc. All rights reserved.

Illumina, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, iScan, and GenomeStudio are registered trademarks or trademarks of Illumina. All other brands and names contained herein are the property of their respective owners.

Pub. No. 770-2008-009 Current as of 1 May 2008

