



CHALLENGE

Increase biological discoveries in schizophrenia research through next-generation genetic sequencing.

SOLUTION

JMP Genomics from SAS dynamically links powerful statistical analysis with sophisticated graphics to provide a comprehensive picture of collected data.

RESULTS

Next-gen sequencing lets researchers gather information about sequence variations and expression in a single step – and yields more precise analysis results. It may someday lead scientists to the “promised land” of personalized medicine.

MORE INFORMATION

Learn more about the work of NCGR:

www.ncgr.org

Discover more about JMP Genomics:

www.jmp.com/genomics

Learn more about SAS: www.sas.com

Next-gen technology revolutionizes brain studies

Genetic sequencing leads to critical discoveries in schizophrenia research at the National Center for Genome Resources.

Faye Schilkey understands that the sum of complementary, revolutionary technologies is impossible to calculate. A new generation of exploration lays the foundation for yet another.

Schilkey is Associate Director of the New Mexico Genome Sequencing Center at the National Center for Genome Resources (NCGR) in Santa Fe. NCGR is a nonprofit research institute dedicated to improving human health and nutrition.

The New Mexico Genome Sequencing Center supports that effort by providing groundbreaking sequencing services using Illumina® Genome Analyzer, the internally designed Alpheus® software system and JMP® Genomics.

A tradition of innovation continues.

Gen two

NCGR was founded in 1994 to provide bioinformatics databases and analysis for Los Alamos National Laboratory, which had begun sequencing the human genome as part of the Human Genome Project. Tremendous volumes of data were being generated as scientists worked to define the genetic instructions for human life. NCGR's task was to manage, analyze and share that data.

In 2007, NCGR took things to the next level with next-generation sequencing.

“What this means,” says Schilkey, “is that now we can re-sequence an organism’s genome and compare it to a ‘reference sequence’ to discover differences that might be involved in diseases – and we can do it faster, and for less money, than ever before.”

NCGR researchers have used their bioinformatics expertise to create an integrated solution for discovery based on next-gen technology. Data generated using the Illumina Genome Analyzer II or another next-gen platform is analyzed and mined using Alpheus, then exported into JMP Genomics for further statistical analysis.

“We’ve created a pipeline for sequencing data,” Schilkey continues. “We now generate huge amounts of data, and Alpheus allows us to filter it. We go from tens of thousands of genes down to perhaps a few hundred that appear to warrant further study.”

Researchers then use JMP Genomics to analyze that data and present it.



**STATISTICAL
DISCOVERY.™
FROM SAS.**

Sequencing holds promise for understanding why some drugs fail in clinical trials, or help only a subset of those tested.

‘Functional genomics’

First-generation sequencing gave scientists the tools to define the DNA of an organism. The introduction of next-gen sequencing is “like a revolution as well,” Schilkey says. “It is giving birth to translational applications, and personalized medicine may not be too far off.”

In the future, a doctor might use a patient’s genome sequence to identify a genetic variation that could cause an allergic reaction to a specific vaccine. Sequencing holds promise for understanding why some drugs fail in clinical trials, or help only a subset of those tested. Knowing the genetic variations between patients could help physicians decide which individuals could be treated safely with the medicine and which ones could not.

“Next-generation sequencing allows for this and other functional genomics applications,” Schilkey explains.

Sequencing messenger RNA (mRNA), known as digital transcript expression (DTE), allows scientists to gather information about sequence and expression in a single step. DTE generates counts of sequences, which provide a more precise measurement of transcript abundance than microarray intensities. NCGR’s research into schizophrenia, led by NCGR President and CEO Dr. Stephen Kingsmore, is using this approach.

Cerebellar cortex tissue is isolated, post-mortem, from individuals with diagnosed cases of schizophrenia and a control group of others without the diagnosis. From that tissue, researchers are able to gather signals that indicate differences between the brains of the schizophrenia patients and the others. “What goes in is tissue from the cerebellar cortex, and what comes out are clues and answers really critical to the disease,” Schilkey explains.

“There is a tremendous discovery aspect to generation-two sequencing technology for mRNA studies,” Schilkey says, in that “whatever genes the tissue is expressing we’re finding, including new and novel genes not yet discovered. We are not limited by the predefined probes of currently known genes on a microarray chip.”

A lot of discovery naturally entails a lot of data.

Schilkey explains that after generating gigabases of data from sequencing the mRNA from the 14 schizophrenia cases and six in the control group, they needed a bioinformatics tool to filter, analyze and visualize this gen-two data.

Enter NCGR’s Alpheus – a Web-based analysis system designed for gigabase-scale resequencing efforts born in 2006 from an NCGR collaboration with mesothelioma researchers at Brigham and Women’s Hospital and Harvard Medical School. Alpheus can easily determine the gene expression and DNA differences in cases and controls and reduce genes of interest from tens of thousands to tens with significant differences.

JMP Genomics then plays a critical role, dynamically linking powerful statistical analysis with sophisticated graphics to provide a comprehensive picture of the data the NCGR research team has generated.

“A great feature of the JMP software is its ability to analyze and visualize so much statistical information at such an accelerated rate,” Schilkey says.

Right out of the box

In fact, she adds, “What was exciting about the schizophrenia project was we found we could use out-of-the-box expression tools in JMP Genomics with the Illumina data and, given the Illumina data’s sensitivity, could see significant differences right away. We could see outliers and case and control separation immediately.”

The NCGR team regularly uses the transformation, quality-control normalization, analysis of variance and annotation tools in JMP Genomics. These include kernel density plots, scatter plots, principal components analysis, volcano plots, heat maps and Venn diagrams. All these tools became vital to NCGR's schizophrenia research for discerning and discovering critical information about cases and controls.

As for the graphical attributes of JMP Genomics, Schilkey says, "We take the graphical features for granted. But to be able to visualize that separation is so wonderful. Important differences just pop right out.

"We also use some of the Venn diagrams in JMP to overlay gene ontology information, which sometimes helps us reduce sets from a couple of hundred to 25 or so that we think might be both statistically significant and biologically relevant."

Schilkey and her colleagues also use JMP to present their results, finding it to be a tool that takes those results and makes them easily understandable to a wide audience.

A 'turnkey operation'

NCGR deploys JMP Genomics in other projects, including soybean research to find expression differences that relate to important agronomic traits. Future projects that will use JMP Genomics include an examination of differential expressions of tissue-specific genes in pig pathogenesis.

"With Illumina, Alpheus and JMP Genomics, we've created a turnkey operation to arrive at significant results that then can be published. We've sequenced samples, we've used Alpheus to find the most important genes, regions or polymorphisms to look at, and we've then brought in JMP Genomics to explore the statistics."

Generation-two sequencing is bringing new discoveries to a great many fields of study. The potential is open-ended.

"The reason we often call it 'generation two' rather than just 'next generation,'" says Schilkey, "is because we think that the sequencing future holds generation-three, four, etc., and we'll be ready to turn that data into discoveries."

The pipeline has been laid to that future.

"We take the graphical features for granted. But to be able to visualize that separation is so wonderful. Important differences just pop right out."

Faye Schilkey



SAS Institute Inc. World Headquarters +1 919 677 8000 To contact your local JMP office, please visit www.jmp.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2008, SAS Institute Inc. All rights reserved. 103651_516149.1008