

Interview

Taking a Faster Approach to Genetic Discovery

Speed. The underlying advantage a researcher can gain when racing to make new discoveries. With access to one of the world's largest computing clusters, Dr. John Blangero of the Southwest Foundation for Biomedical Research has the means to quickly analyze enormous quantities of genetic data. Now he concentrates on ways to obtain data faster.

i: What is the focus of your research?

JB: Our research focuses on the localization and identification of genes influencing common complex diseases. Things like heart disease, diabetes, osteoporosis, and obesity. Even some psychiatric and infectious diseases. These phenotypes are particularly complex because they involve a lot of different genes, a lot of environmental factors, and interactions between the two. We've been using sequence and transcriptional variations to jumpstart the identification of genes likely involved.

i: What methods are you currently using?

JB: We're largely a discovery-based shop. Traditionally, we've been using a genetic type of approach to discovering genes. By that I mean we've been using the process of transmission genetics—large-family studies and classical linkage-based scans for regions of the genome involved in the disease. With the advent of large-scale expression studies, we've been able to use expression-based discovery tools. We think this has allowed us to greatly increase the speed at which we are finding the genes involved in these complicated diseases, which is what originally appealed to us about the Illumina® platforms. We want to use expression on a high-dimensional level in thousands of samples to quantitate observed unique phenotypes so that we

can gain a much bigger picture into total human biological variation. These days, the ability to do much more global analyses is what really gets us excited.

i: Why did you decide to switch to Illumina technology?

JB: As a major center for statistical genetic data analysis, we have had experience with virtually all of the existing platforms for transcriptome analysis. We specifically chose Illumina for our own studies because we found it to be less labor-intensive, more cost-effective, and accurate. The Illumina platform has allowed us to perform transcriptional profiling on an epidemiological scale that was not previously attempted.

We also made an active decision to switch to Illumina for almost all of our genotyping needs. For high-density analysis, we believe that Illumina provides the best overall solution. With the recent addition of the BeadXpress™ Reader, we are also using Illumina technology for those studies that require a much reduced number of SNPs to be genotyped, with the same labor, cost savings, and accuracy we have observed for the high-density assays.

i: Was there anything that surprised you?

JB: We have found the Illumina system very easy to use and troubleshoot. Importantly, we have been able to obtain high-quality



Dr. John Blangero, Director, AT&T Genomics Computing Center, Department of Genetics, Southwest Foundation for Biomedical Research.

"We specifically chose Illumina for our studies because we found it to be less labor-intensive, more cost-effective, and accurate."



results, even when starting material was of lesser quality. We like the fact that there is a similar level of robustness across all of the different applications. Also, we have always received excellent technical support and service.

i: Now that you are using Illumina platforms, what have you accomplished that was previously out of reach?

JB: Illumina technology has allowed our molecular genetics work to scale up rapidly and cost-efficiently. We now have no qualms about performing much larger scale analyses with far greater sample sizes than would have been possible a few years ago. It has enabled us to put together a research program that allows for joint genome-wide genetic and genome-wide transcriptome analysis. The results have been spectacular and have greatly increased our speed in discovering novel genes involved in complex diseases.

i: What research are you currently conducting using Illumina technology?

JB: We have used Illumina technology to perform large-scale transcriptional profiling in human family-based material to rapidly identify *cis*-regulated genes that correlate with various disease-related phenotypes. We're also using the genome-wide SNP assays for both conventional high-density association analysis of existing quantitative trait locus regions and for genome-wide association analysis. Specifically, we've used GoldenGate® Genotyping on the BeadArray™ and the BeadXpress Readers, the Human-6 Expression BeadChip, and the HumanHap550-Duo BeadChip. We have also used the Illumina-supplied Tecan Freedom Evo robotics for pre- and post-PCR and done whole-chromosome sequencing using the Illumina Genome Analyzer.

Our first intensive project was to perform transcriptional profiles on lymphocyte samples from a large set of Mexican American families. We analyzed 1,280 samples using the Human-6 Expression BeadChips; 97% of samples gave good

results. Given that these samples had been stored for over 15 years, we were extremely pleased with our success rate.

Currently, we are analyzing data from the HumanHap550 BeadChip in conjunction with transcriptional profiles obtained from another BeadChip. That's 550,000 SNPs times 22,000 good transcripts, so we are getting into very large numbers. Ultimately, this means that in a specific gene, we can identify the *cis* effects very easily and, even more exciting, the *trans* effects that influence transcription levels of our genes of interest. Now we have the potential to identify upstream regulators and then relate these to the more classical end phenotypes of the disease. But this is just a stepping stone towards the true end game, which is complete sequence data.

i: What advantages do you gain by combining genomic and transcriptome data sets?

JB: We're able to immediately go and query any gene that happens to be expressed in our particular tissue. We can look at it and find *trans* regulators that were previously unknown. All of a sudden you know something that is upstream from it, and could be manipulated, which is of great interest from a medical point of view. If I find a functional variation in a gene that interests me, in terms of expression levels, and take that same variant and look at it compared to the rest of the transcriptome, I can find things that are downstream. So you start to get a real causal pathway of the relationship between genes. And that's the name of the game in identifying pathways that may be susceptible to pharmacological intervention. It's a way to speed target discovery.

i: Where does sequencing come into play with all this?

JB: The idea is to ultimately capture all the variation you have in a sample. The only way you can do this is through whole-genome resequencing. So much rare variation seems to be important, we believe this will be the true answer.

“Illumina technology has allowed... [us] to scale up rapidly and cost-efficiently...results have been spectacular and have greatly increased our speed in discovering novel genes involved in complex diseases.”

i: What future research studies do you have planned?

JB: We'll be performing miRNA and methylation analyses to identify potential new biomarkers for common human diseases, whole human chromosome sequencing to identify cis-regulated expression quantitative trait loci, and pathogen genome sequencing to identify strain differences in host response.

We're especially excited about the potential to perform deep resequencing on all of our subject material. We're convinced that we need to find all of the rare variations when looking at common diseases and the only way we will be able to do this is with cost-effective high-throughput sequencing technology like that of our Illumina Genome Analyzer.

i: Once whole-genome resequencing is a standard task, what's next for you?

JB: We'll be going after causal variation in relation to disease. I expect that in the future all our epidemiological studies will involve the complete resequencing of the human genome and comparing any variation to the myriad of end phenotypes related to disease. Obviously, it will be a massive computational task. We have to come up with a clever way to prioritize things because obviously not all variants are equally likely to be functional. Still, I'd rather have all the information.

ADDITIONAL INFORMATION

Please visit www.illumina.com to learn more about Illumina's gene expression analysis solutions and the Genome Analyzer.

This newsletter is dedicated to publishing news, perspectives, and insights about and for the Illumina Community. We are committed to providing you with the content you want to see as a member of the Illumina Community. Please email us with suggestions for topics you'd like to read about at icommunity@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, and CSPro are registered trademarks or trademarks of Illumina. All other brands and names contained herein are the property of their respective owners.

Pub. No. 070-2008-003 Current as of 13 February 2008.

