

Perspective

# Improving the Human Condition 50,000 Sequences at a Time

Dr. Jay Shendure's effort to bring highly multiplexed targeted resequencing to the mainstream.

How would your research change if you could study the genetic basis of a disease by comparing the functional genomes of thousands of affected individuals? What if sequencing all the protein-coding regions of our genomes was a standard step in diagnostic genomic medicine? Two years ago, with access only to conventional sequencing, highly multiplexed targeted resequencing applications were just ideas. Today, the Illumina® Genome Analyzer is helping scientists turn these speculations into realities. For the price of a latte, the Genome Analyzer can sequence a megabase of DNA, making large-scale targeted resequencing the next paradigm-shifting discovery method. Now that the sequencing power is here, an equally affordable and reliable front-end method for efficiently amplifying thousands of genomic regions is needed. Overcoming this rate-limiting step is one of Dr. Jay Shendure's many scientific interests.

## A LEADER IN SEQUENCING TECHNOLOGY

Bringing targeted resequencing to a broad user base is a natural fit for Jay Shendure, M.D., Ph.D., whose mission is to develop and use DNA sequencing technology to study the genetic basis of human disease. For the last five years, this Technology Review 2006 Young Innovator has either been building next-generation sequencers or trying to figure out how best to use them.

As a graduate student in George Church's lab, Jay Shendure was a leading researcher

on the development of "polony sequencing" which uses emulsion PCR and sequencing by ligation—technology that was ultimately sold to Applied BioSystems and underpins their SOLiD sequencing system. Even with his familiarity with other sequencing platforms, Dr. Shendure decided to use the Genome Analyzer for his targeted resequencing research.

## CHOOSING THE GENOME ANALYZER

Last summer Dr. Shendure evaluated his options for which commercial platform to put in his lab. He says the low cost per base, ease of operation, and simplified workflow of the Illumina system appealed to him. He adds, "If all other things were equal, the ease of the automated cluster generation workflow relative to emulsion PCR would still tip me towards the Genome Analyzer." As proof of the system's simplicity, he tells us how Choli Lee, a bright young student fresh out of college, was able to have the Genome Analyzer up and running with minimal training. With this system, Dr. Shendure says that the rate-limiting step in his lab was not going to be running the sequencer, but being challenged with the science.

Dr. Shendure acquired his Genome Analyzer last September, and quickly set out to resequence 55,000 human protein-coding sequences, approximately one-fourth of the total number of exons in the genome. He explains, "It's clear that targeted resequencing would be a great way to knock two orders of magnitude off the price of



Jay Shendure, M.D., Ph.D., Assistant Professor, Department of Genome Sciences, University of Washington.

"With the Genome Analyzer we've been able to test a lot of optimizations very quickly which speaks to the ease of use, throughput, and fast workflow of the system."

genome analysis, [allowing you to] forget about the 99% of the genome that you are less interested in and focus on the 1% that is the most amenable to functional validation.”

#### FROM 10,000 TO 50,000 TARGETS

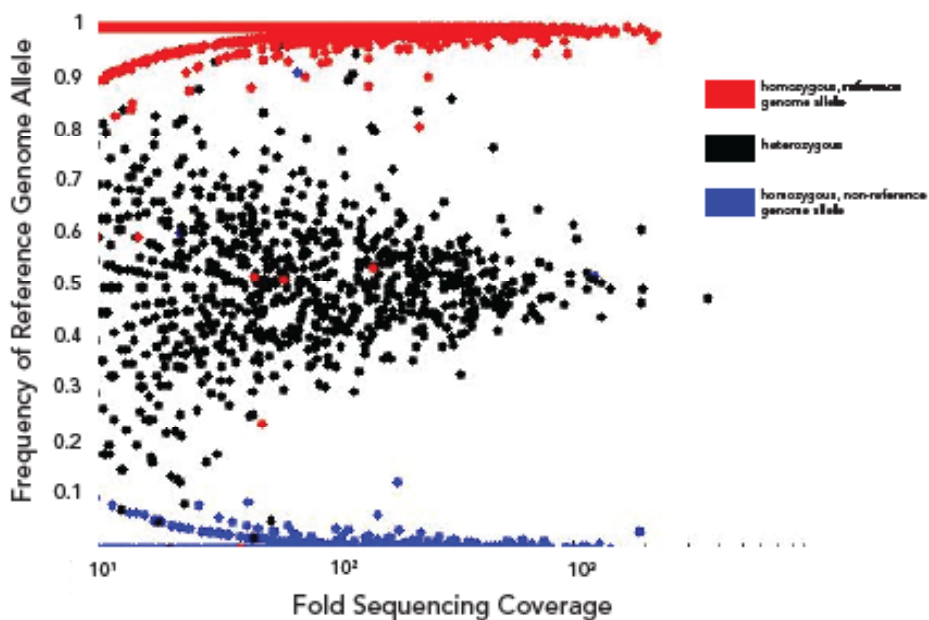
With sequencing in place, the next step was to figure out a novel way to specifically isolate and amplify thousands of exons. Conventional PCR-based targeting strategies are cost-prohibitive and inefficient when multiplexing tens of thousands of reactions. To get around this problem, Dr. Shendure’s team tried an innovative approach. They synthesized molecular inversion probes (MIPs) on a programmable microarray designed to target sequences

flanking exons, essentially creating probes analogous to PCR primers. They released the oligos into solution to form a pool of MIPs that hybridized to genomic DNA. Subsequent rounds of extension, ligation, and amplification isolated the targeted sequences which were then subjected to high-throughput sequencing using the Genome Analyzer. In October 2007, Dr. Shendure and his colleagues demonstrated that they were able to capture 10,000 out of the 55,000 targeted exons in a single multiplex reaction using this method<sup>1</sup>.

Since publishing their results, Dr. Shendure reports that his team has made significant improvements to their methodology. He acknowledges that in their November paper<sup>1</sup> they ran into two

“...we’ve been able to increase the number of targets we can amplify from 10,000 to 50,000 out of 55,000 targets (>90%) in the space of a month or two.”

FIGURE 1. VALIDATION OF EXON RESEQUENCING DATA BY COMPARISON TO HAPMAP DATA



MIP-based multiplex capture directed at 55,000 exons was performed on 750 ng of genomic DNA (NA12248; CEPH). Capture products were converted to a shotgun library and sequenced on a Genome Analyzer. Shown here are 3,126 positions for which there is 10× or greater coverage (with no quality cutoff) and for which HapMap genotyping data is available. The fold coverage for a given position in the resequencing data (x-axis) is plotted against the frequency in the resequencing data of the reference genome allele (y-axis). Colors correspond to HapMap genotypes for this individual. Data courtesy of Dr. Jay Shendure.

main problems: (1) they were able to efficiently capture only ~20% of the targets, and (2) heterozygote calling proved difficult. “The difficulties we had were a fault of the targeting technology, not the sequencing technology. With the Genome Analyzer we’ve been able to test a lot of optimizations very quickly, which speaks to the ease of use, throughput, and fast workflow of the system. Having the ability to rapidly turn around experiments, we’ve been able to increase the number of targets we can amplify from 10,000 to 50,000 out of 55,000 targets (>90%) in the space of a month or two. As a side benefit of the optimizations, we’ve also effectively solved the heterozygote problem (Figure 1). The data have been really exciting,” explains Dr. Shendure.

Dr. Shendure says that the Genome Analyzer’s accuracy has proved beneficial for his targeted resequencing experiments. He reports, “The consensus accuracies for resequencing applications need to be extremely high. One thing that wasn’t a guarantee was whether we could use the consensus of multiple reads, generated by a second-generation platform, over a given position to accurately make resequencing base calls as heterozygous or homozygous. What’s clear from the data that we’ve seen so far from the Genome Analyzer is that the answer is: yes, you can.”

#### LOOKING AT THE FUTURE

Now that his Genome Analyzer is running smoothly, Dr. Shendure is pursuing a number of other sequencing projects. “With collaborators, we’ve been doing ChIP-Seq experiments, bacterial genome resequencing, and developing protocols for shotgun sequencing of mRNA and barcoded multiplex promoter bashing. Every run is usually a mish-mash of a couple different projects,” he says. No matter what he’s doing, Dr. Shendure remains excited about developing discovery applications based on high-throughput sequencing. “The sequencing technology clearly works and has made a ton of progress and will continue to evolve. In the same way that this happened for Sanger sequencing, we need to put more effort in developing tools for the front end and for the back end. Getting these tools out there so they are available to a broad user base is going to take a little bit of time, but it’s going to be a lot of fun.”

#### REFERENCE

- (1) Porreca GJ, Zhang K, Li JB, Xie B, Austin D, et al. (2007) Multiplex amplification of large sets of human exons. *Nature Methods* 4(11): 931-936.

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 [icommunity@illumina.com](mailto: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. 070-2008-001 Current as of 13 February 2008

#### ADDITIONAL INFORMATION

To learn more about the Genome Analyzer system and customer-demonstrated applications, visit [www.illumina.com/sequencing](http://www.illumina.com/sequencing).

