

Perspective

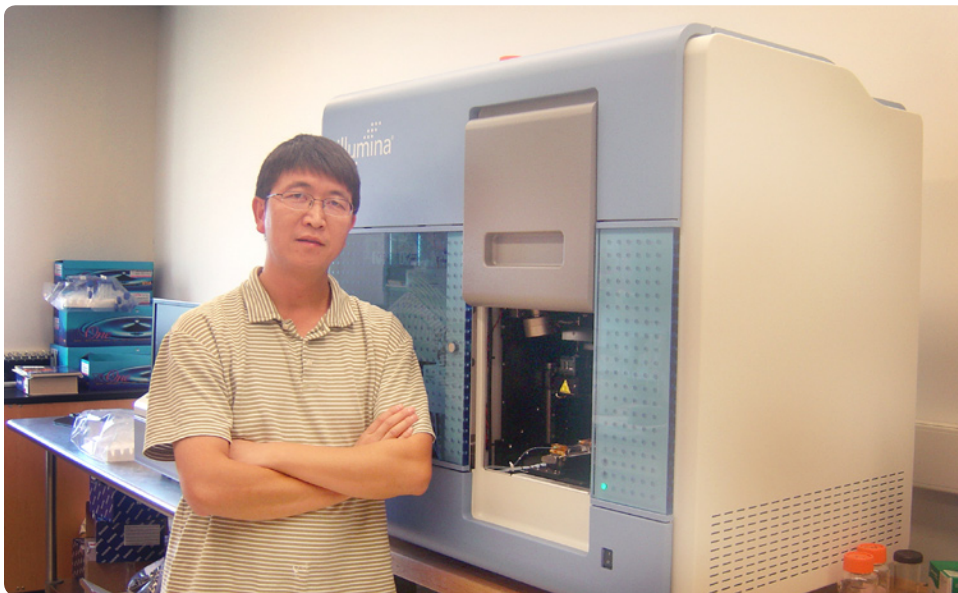
How to Sequence DNA at 200 mph

One researcher's search for breakthrough sequencing speed and accuracy ends with the Genome Analyzer.

In the past few years, researchers have experienced an explosion of new and improved technologies that provide them with significantly more information in drastically less time than ever before. This is especially true of DNA sequencing methods, which have moved from slow-moving gels and hands-on data analysis to automated sequencing and high-powered bioinformatics tools. Even after initial developments, however, many projects remained too expensive and time-consuming to perform. All this changed for Dr. Yuan Gao with Solexa® Sequencing Technology and the Illumina® Genome Analyzer.

A PERSONAL BACKGROUND IN SEQUENCING

As a doctoral student at the University of Tennessee at Memphis and St. Jude Children's Research Hospital in the early 90s, one of Dr. Gao's projects was to clone and sequence a cdc2-related kinase. Automated DNA sequencers were still a few years away, so Dr. Gao suffered through the time-consuming traditional sequencing process of running slab gels with P³² and S³⁵ radioisotopes, followed by manual sequence assembly and comparison with various databases to identify coding sequences. But he knew there had to be a better way. He began to write his own bioinformatics



DR. YUAN GAO WITH HIS GENOME ANALYZER

Dr. Yuan Gao is an Assistant Professor in the Center for the Study of Biological Complexity and Department of Computer Science at Virginia Commonwealth University. He earned a B.S. degree in Biology from Beijing University, an M.S. degree in Biochemistry from the University of Tennessee at Memphis, and a Ph.D. in Computer Science at the University of Memphis. Subsequently, he did his post-doctoral study in George Church's Lab at Harvard. His research interests include systems biology, genome sequencing, and algorithm design and analysis. He is currently on the fast track to analyzing billions of bases.

programs to help with the interrogation of DNA sequence data.

Pursuing his interest in computational biology, Dr. Gao went on to obtain his Ph.D. in Computer Science from the University of Memphis. After spending over four years in the Computational Biology Center at the IBM T.J. Watson Research Center to work on his dissertation on pattern discovery from biological sequences, Dr. Gao realized that there was a lack of high-quality, high-throughput data and that the future lay in integrated research combining “wet” and “dry” lab procedures.

Dr. Gao did his post-doctoral research with George Church at Harvard Medical School. Among other things, Dr. Church’s lab was well known for pioneering next-generation DNA sequencing methods. During the time that Dr. Gao worked with Dr. Church, the lab was granted the patent for “polony sequencing” and began building the first prototype polony sequencer. This system would eventually become part of the foundation for ABI’s SOLiD technology. One of the projects Dr. Gao contributed to while in Dr. Church’s lab was the writing of algorithms and software that would facilitate automated image capture, image analysis, and short-read mapping.

MOVING FORWARD WITH SOLEXA TECHNOLOGY

In 2006, Dr. Gao found a great place to continue his interests in systems biology, genome sequencing, and data mining, as an Assistant Professor in the Center for the Study of Biological Complexity and the Department of Computer Science at Virginia Commonwealth University. It was here that he established his High Throughput Genomics and Systems Biology Lab and first purchased the Illumina Genome Analyzer. The market contenders at the time were 454, ABI, and Solexa. ABI’s system was unavailable, and not scheduled for release until 2008. Unable to put his research on hold, Dr. Gao needed to choose between the 454 and Solexa platforms.

“Possibly the two most important reasons for my endorsement and eventual purchase of the Genome Analyzer were its ease of preparing samples and the 40 million number of reads per chip, which is very important considering that my main interest is resequencing applications. Coming from both biology and computer science, I developed a love-hate relationship for manual vs. automation. The cluster station in the Genome Analyzer amplifies single molecules just as emulsion PCR (emPCR) does, but requires minimal human intervention,” states Dr. Gao. He continues on to say that, “we knew back then that data from 454 was going to cost 10x as much as data generated on the Genome Analyzer.” In addition, the low sample input requirement, only 100 ng DNA, which is simply not possible in the ePCR process due to inefficiencies in emPCR, appealed to Dr. Gao. After considering all of these factors: ease of use, lower cost per sample, low starting material requirement, higher throughput, and application flexibility, Dr. Gao chose the Illumina platform.

Today, Dr. Gao feels like he has “left the dark ages of sequencing” and compares current advances in sequencing technology to the strides made in the PC realm described by Moore’s law. “This principle states that the clock speed of CPU doubles every 18 months, and I suspect that there is also such a Moore’s law in next-generation sequencing technologies. If this rule holds true, by next summer the Genome Analyzer will be generating over 3 Gb/flow cell, followed by 6 Gb/flow cell, and then possibly 12 Gb/flow cell.” He has already achieved 2.6 Gb from a single-read run on the Genome Analyzer, more than twice the amount stated in the Genome Analyzer’s performance specifications. Dr. Gao can certainly use this level of power in his current research pursuits.

Collaborating with multiple researchers on projects that, even a year ago, would have been time- and cost-prohibitive, Dr. Gao is constantly looking for ways to

“Possibly the two most important reasons for my endorsement and eventual purchase of the Genome Analyzer were its ease of preparing samples and the 40 million number of reads per chip.”



generate and understand more information, more quickly. From a purely computational view, he is already able to cover an unprecedented amount of information—billions of bases—and convert them into usable knowledge for his biology-oriented counterparts. For example, he uses data from bacteria and parasite DNA run on the Genome Analyzer to develop assembly tools to *de novo* assembly of short reads for small genomes. Likewise, his experiments with novel bacterial genomes generate data that taxes conventional assembly programs. Dr. Gao also uses the Genome Analyzer to study insulin-stimulated pancreatic beta cell microRNA expression and alternatively spliced transcripts in cancer and other diseased tissues. Methylation studies are under way with the Genome Analyzer, and Dr. Gao will explore epigenetic studies of chromosome 21.

Perhaps one of Dr. Gao's most technologically advanced uses for the Genome Analyzer is his ongoing project with George Church. With Dr. Gao's help, Dr. Church is sequencing 1% of the genome for 1,000 people—about 30 million bases per person¹. This project, known as the Personal Genome Project, is an intermediate step on the way to the \$1,000 genome.

Dr. Yuan Gao is understandably excited about the current revolution happening in DNA sequencing. Summarizing his current experience, Dr. Gao says “generating data with the Genome Analyzer is just like driving a Ferrari going 200 mph. It's so fast that you lose all reference and just hang on.” It certainly is an amazing time to be in the driver's seat.

REFERENCE

- (1) Porreca GJ, Zhang K, Li, J, Xie B, Austin D, et al. (2007) Multiplex amplification of large sets of human exons. *Nat Meth* in press.

ADDITIONAL INFORMATION

To learn more about the Genome Analyzer, visit our website at www.illumina.com.

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.

FOR RESEARCH USE ONLY

© 2007 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. 770-2007-012 11Oct07

