3 sequencing companies join 1000 Genomes Project

June 11 2008

Leaders of the 1000 Genomes Project announced today that three firms that have pioneered development of new sequencing technologies have joined the international effort to build the most detailed map to date of human genetic variation as a tool for medical research. The new participants are: 454 Life Sciences, a Roche company, Branford, Conn.; Applied Biosystems, an Applera Corp. business, Foster City, Calif.; and Illumina Inc., San Diego.

The 1000 Genomes Project, which was announced in January 2008, is an international research consortium that is creating a new map of the human genome that will provide a view of biomedically relevant DNA variations at a resolution unmatched by current resources. Organizations that have already committed major support to the project are: the Beijing Genomics Institute, Shenzhen, China; the Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K.; and the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health. The NHGRI-supported work is being done by the institute's Large-Scale Sequencing Network, which includes the Human Genome Sequencing Center at Baylor College of Medicine, Houston; the Broad Institute of MIT and Harvard, Cambridge, Mass.; and the Washington University Genome Sequencing Center at Washington University School of Medicine, St. Louis.

"The additional sequencing capacity and expertise provided by the three companies in the pilot phase will enable us to explore the human genome with even greater depth and speed than we had originally envisioned, and

will help us to optimize the design of the full study to follow," said Richard Durbin, Ph.D., of the Wellcome Trust Sanger Institute, who is co-chair of the consortium. "It is a win-win arrangement for all involved. The companies will gain an exciting opportunity to test their technologies on hundreds of samples of human DNA, and the project will obtain data and insight to achieve its goals in a more efficient and cost-effective manner than we could without their help."

The genetic blueprints, or genomes, of any two humans are more than 99 percent the same. Still, the small fraction of genetic material that varies among people holds valuable clues to individual differences in susceptibility to disease, response to drugs and sensitivity to environmental factors.

The 1000 Genomes Project builds upon the International HapMap Project, which produced a comprehensive catalog of human genetic variation – variation that is organized into neighborhoods called haplotypes. The HapMap catalog laid the foundation for the recent explosion of genome-wide association studies that have identified more than 130 genetic variants linked to a wide range of common diseases, including type 2 diabetes, coronary artery disease, prostate and breast cancers, rheumatoid arthritis, inflammatory bowel disease and a number of mental illnesses.

The HapMap catalog, however, only identifies genetic variants that are present at a frequency of 5 percent or greater. The catalog produced by the 1000 Genomes Project will map many more details of the human genome and how it varies among individuals, identifying genetic variants that are present at a frequency of 1 percent across most of the genome and down to 0.5 percent or lower within genes. The 1000 Genomes Project's high-resolution catalog will serve to accelerate many future studies of people with specific illnesses.

"In some ways, this application of the new sequencing technologies is like building bigger telescopes," said NHGRI Director Francis S. Collins, M.D., Ph.D. "Just as astronomers see farther and more clearly into the universe with bigger telescopes, the results of the 1000 Genomes Project will give us greater resolution as we view our own genetic blueprint. We'll be able to see more things more clearly than before and that will be important for understanding the genetic contributions to health and illness."

The HapMap was based mainly on genotyping technology, in which genetic markers were used to broadly scan the genome. In contrast, the 1000 Genomes Project catalog will be built on sequencing technology, in which the genome is examined at the level of individual DNA letters, or bases. The increased resolution will enable the 1000 Genomes' map to provide researchers with far more genomic context than the HapMap, including more precise information about the genetic variants that might directly contribute to disease.

"We find that there is a lot of value in participating in international consortia; they produce large datasets that are valuable to the scientific and medical communities while promoting the rapid release of the data" said Illumina Vice President and Chief Scientist David Bentley, Ph.D., who participated in the International HapMap Project.

To enhance the production of the 1000 Genomes map, each of the three biotech companies has agreed to sequence the equivalent of 75 billion DNA bases as part of the pilot phase. The human genome contains about 3 billion bases. Consequently, each company will contribute the equivalent of 25 human genomes over the next year, and additional sequence data over the project's expected three-year timeline. In addition, Applied Biosystems will contribute an additional 200 billion bases of human sequence through its collaboration with Baylor.

"This project is clearly the most ambitious and comprehensive study to date of the human genome. Our participation continues our commitment to partner with the scientific community to explore the genetic factors involved in human disease," said Francisco de la Vega, distinguished scientific fellow and vice president for SOLiD ™ System Applications and Bioinformatics at Applied Biosystems.

Michael Egholm, Ph.D., vice president of R&D at 454, said, "We are proud to contribute to the 1000 Genomes Project as we further our ongoing support of researchers worldwide and their goal of deepening our understanding of human genome complexity. By applying innovative technology to these complex challenges, this project will deliver the highest standard of data quality and analysis."

In its first phase, expected to last about a year, the 1000 Genomes Project is conducting three pilots that will be used to decide the best strategies for achieving the goals of the full-scale effort. The first pilot involves sequencing the genomes of six people (two nuclear families) at high resolution; the second involves sequencing the genomes of 180 people at lower resolution; and the third involves sequencing the coding regions of 1,000 genes in about 1,000 people.

The full-scale project will involve sequencing the genomes of at least 1,000 people, drawn from several populations around the world. The project will use samples from donors who have given informed consent for their DNA to be analyzed and placed in public databases. Most of these samples have already been collected, and any additional samples will come from specific populations. The data will contain no medical or personal identifying information about the donors.

Given the rapid pace of sequencing technology development, the cost of the entire effort is difficult to estimate, but is expected to be about $60 million. The sequence data provided by the three companies are

estimated to be worth approximately $700,000 for the pilot phase, and the firms are expected to contribute much more sequencing to the full project.

Already, the 1000 Genomes Project has generated such vast quantities of data that the information is taxing the current capacity of public research databases. Since the first phase was begun in late January, project participants have produced and deposited some 240 billion bases of genetic information with the European Bioinformatics Institute and the National Center for Biotechnology Information, a part of the U.S. National Library of Medicine. Data generated by the 1000 Genomes Project also will be distributed from a mirror site at BGI Shenzhen.

Along with their contributions of sequencing capacity, the companies, like all other project participants, have agreed to comply with the open access policies established by the 1000 Genomes Project Steering Committee. Those policies include rapid public release of the data, including project participants having no early access to the data; an intellectual property policy that precludes any participants from controlling the information produced by the project; regular progress reporting; and coordination of scientific publications with the rest of the consortium.

Source: National Human Genome Research Institute