Medscape Medical News

ASHG 2008: 1000 Genomes Project Vastly Expands Map of Human Genome Variation

Jacquelyn K. Beals, PhD

November 17, 2008 (Philadelphia, Pennsylvania) — The 1000 Genomes Project is using a variety of next-generation sequencing technologies to create an extensive public database of single-nucleotide polymorphisms (SNPs) and copy-number variations (CNVs) in designated human population samples.

The project, which focuses on lower-frequency polymorphisms, has already generated a quantity of data equivalent to GenBank *each week* for the past 2 months. The international public–private consortium is using and evaluating a variety of sequencing technologies, including Illumina, Roche 454, and ABI SOLiD sequencing platforms. To date, 3.8 terabases have been deposited in the public database.

Dr. David Altshuler, MD, PhD, from the Department of Genetics and Medicine, Massachusetts General Hospital, and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard University, in Boston, provided an overview of the 1000 Genomes Project in a presentation here at the American Society of Human Genetics 58th Annual Meeting.

"The first generation of genomewide association studies was well powered only for SNPs with minor allele frequencies greater than 5%," said Dr. Altshuler. However, next-generation sequencing techniques now enable detection of lower-frequency polymorphisms. A goal of the 1000 Genomes Project, in each human sample, is to identify nearly all SNPs and detectable CNVs with minor allele frequencies greater than or equal to 1%.

Three pilot studies were undertaken in the first year of the 2-year project: pilot 1 entails 4 times coverage (the number of bases surveyed or the amount of nucleotide information obtained) of the genomes of 180 people; pilot 2 will obtain 20 times coverage of the genomes of 2 parent–parent–child trios; and pilot 3 will accomplish the targeted sequencing of 1000 genes. Data produced by the pilot studies already exceeds 1000 times coverage of the human genome.

Dr. Altshuler also participated in a press briefing on genomewide association studies and the 1000 Genomes Project, moderated by Francis S. Collins, MD, PhD, former director of the National Human Genome Research Institute, National Institutes of Health (NIH), in Bethesda, Maryland.

"Obviously [genomewide association study] is a powerful way to find common variants, but not a powerful way to find uncommon variants," said Dr. Collins in his introductory comments, "because very few people in your cases and controls will have those. So one of the deficits here, in what [this study] can tell us, is in that range of genetic risk factors that are perhaps only present in 1% or 2% of people, as opposed to 20% to 30% of people. Those have largely escaped detection so far, and yet they could turn out to be really important." The 1000 Genomes Project is designed to obtain data on these lower-frequency polymorphisms.

Medscape Pathology & Lab Medicine asked Dr. Altshuler the reasoning behind pilot 2. "Some of our samples will actually be parents and 1 of their offspring," he said. "Obviously, children get their DNA from their parents, and the ability to have 2 looks at that DNA — 1 from the parent and 1 from the child — to see the properties of it being transmitted from parent to child, allows you to test all sorts of types of errors you can get into, and also allows you to generate a higher quality of information about the chromosomes that are being transmitted."

He added that "we don't actually know that much about the process of mutation in human DNA. And the reason is because the rate is so extraordinarily low — about 1 per 100 million DNA letters per generation — that the only way you could actually directly measure it...would be to sequence the DNA of parents and of their children," Dr. Altshuler noted.

"Even in the whole genome, there'd normally be only a couple of dozen events. So to get lots and lots of those events, you need to do that in lots of people. That wouldn't motivate us to do the design solely for that purpose," said Dr. Altshuler, "but that is something we're going to get out of that design. We're going to learn about human mutation, which, amazingly enough, is too rare to have ever been directly observed in that way in large numbers."

Dr. Collins added: "Interestingly, when the Department of Energy decided that they wanted to support the Human Genome Project, even before NIH did back in 1985, it was just this issue of trying to get a handle on the human mutation rate, and how it was affected by things like low-level radiation. So here we are, coming around now, finally able to answer the question!"

Having gotten off to a successful start, the 1000 Genomes Project now has a target of sequencing 1200 people to 4 times coverage, with data collection completed by the winter of 2009. Quarterly releases of data will begin in January 2009. Raw data are already available on the 1000 Genomes Project Web site.

Dr. Altshuler and Dr. Collins have disclosed no relevant financial relationships.

American Society of Human Genetics 58th (ASHG) Annual Meeting: Abstract 84. November 13, 2008.