Medical Xpress

1000 Genomes Project releases pilot data

June 21 2010

The completion of three pilot projects designed to determine how best to build an extremely detailed map of human genetic variation begins a new chapter in the international project called 1,000 Genomes, said the director of the Baylor College of Medicine Human Genome Sequencing Center, which is a major contributor to the effort.

"Mapping all the shared normal variation in human populations is a critical step to interpreting medically actionable genetic changes," said Dr. Richard Gibbs, also a professor in the department of molecular and human genetics at BCM.

The 1,000 Genomes project began in 2008 with the kickoff of three pilot programs. Completion of the pilots launches the full-scale effort to build a public database of human genetic variation from the genomes of 2,500 people from 27 populations around the world. With the announcement, groups involved in the project placed their final data in freely available databases that can be used and accessed by the worldwide research community.

"The 1000 Genomes Project has a simple goal: peer more deeply into the genetic variations of the human genome to understand the genetic contribution to common human diseases," said Dr. Eric D. Green, director of the National Human Genome Research Institute, which provides major funding to the effort. "I am excited about the progress being made on this resource for use by scientists around the world and look forward to seeing what we learn from the next stage of the project."

Recent studies looking for variations that contribute to common human ailments, such as heart disease and diabetes, indicate that a host of rare variations account for much of the burden of disease in the human population. Complex and detailed maps such as those to be assembled from the project provide a potent tool for identifying those rare variations.

The pilot program tested the viability of three strategies. BCM designed and coordinated the strategy that involved targeting the sequencing to gene coding regions. This project provided the most complete data for the exons (or coding regions) of 1,000 genes, as it was designed to deeply sample the DNA in each of nearly 700 people. An estimated 2 percent of the human genome is composed of protein-coding genes.

"We also developed new methods to target variation in genes, and showed that this approach gave maximum information about this important class of human variation", said Dr. Fuli Yu, an assistant professor in the BCM Human Genome Sequencing Center and coordinator of the study.

The project's fast pace was made possible only by next-generation sequencing technology, which can produce thousands or million of sequences rapidly. The techniques involved allow researchers to evaluate all the rare variants found in areas of the genome known to be associated with human disease.

Another of the pilot projects involved using a variety of sequencing technologies to sequence the genomes of six people (two nuclear families including parents and one daughter) at high coverage (meaning in exacting detail). Each sample was sequenced from 20 to 60 times, uncovering a more complete picture of DNA variation in these families. Using different technologies scientists also obtained a better understanding of the strengths of each sequencing platform.

The other pilot project sequenced the genomes of 179 people in less detail - subjecting each sample to an average of approximately four sequencing passes. Researchers then combined the data from different people to discover which genetic variants they share. This technique will provide valuable information in uncovering those genomic variations shared among people or populations.

More information: Researchers can obtain the data freely through the 1000 Genomes website (www.1000genomes.org) or from the NCBI at ftp://ftp-trace.ncbi.nih.gov/1000genomes/ or the EBI at: ftp://ftp.1000genomes.ebi.ac.uk/ . Researchers with limited computing power will be able to access the data through Amazon Web services through the company's Elastic Compute Cloud (AmazonEC2). The database contains all forms of variation found in the genome from single changes called single nucleotide polymorphisms (SNPs), to small insertions and deletions (of genetic material) to the large changes in the structure and number of copies of chromosomes called copy number variations.

Provided by Baylor College of Medicine

Citation: 1000 Genomes Project releases pilot data (2010, June 21) retrieved 16 April 2024 from https://medicalxpress.com/news/2010-06-genomes.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.