

U.S. Department of Health and Human Services  
NATIONAL INSTITUTES OF HEALTH NIH News  
National Human Genome Research Institute (NHGRI) <<http://www.nhgri.nih.gov/>>  
For Immediate Release: Wednesday, September 1, 2010

CONTACT: Geoff Spencer, NHGRI, 301-402-0911, <e-mail:[spencerg@mail.nih.gov](mailto:spencerg@mail.nih.gov)>

### THIRD GENERATION MAP OF HUMAN GENETIC VARIATION PUBLISHED Map adds data from seven global populations

An international consortium today published a third-generation map of human genetic variation, called the HapMap, which includes data from an additional seven global populations, increasing the total number to 11 populations. The improved resolution will help researchers interpret current genome studies aimed at finding common and rarer genetic variants associated with complex diseases.

Any two humans are more than 99 percent the same at the genetic level. But, the small fraction of genetic material that varies among people can help explain individual differences in susceptibility to disease, response to drugs or reaction to environmental factors. Variation in the human genome is organized into local neighborhoods called haplotypes, which usually are inherited as intact blocks of DNA sequence information. Consequently, researchers refer to the map of human genetic variation as a haplotype map, or HapMap.

"The generated HapMap provides an important foundation for studies aiming to find genetic variation related to human diseases. It is now routinely used by researchers as a valuable reference tool in our quest to use genomics for improving human health," said Eric D. Green, M.D., Ph.D., director of the National Human Genome Research Institute (NHGRI), a part of the National Institutes of Health, which provided major funding for the HapMap Project.

The most common genetic differences among people are SNPs, single-nucleotide polymorphisms. Each SNP reflects a specific position in the genome where the DNA spelling differs by one letter in different people. SNPs serve as landmarks across the genome. The initial version of the HapMap contained approximately 1 million SNPs, and the second-generation map brought that total to more than 3.1 million SNPs. Over the last few years, researchers conducting genome-wide association studies have relied on publicly available data from the HapMap to discover hundreds of common genetic variants associated with complex human diseases, such as cardiovascular disease, diabetes, cancer and many other health conditions.

The first- and second-generation versions of HapMap resulted from the analysis of DNA collected from 270 volunteers from four geographically diverse populations: Yoruba in Ibadan, Nigeria; Japanese in Tokyo; Han Chinese in Beijing; and Utah residents with ancestry from northern and western Europe.

The third-generation HapMap, reported in the Sept. 2 issue of the journal *Nature*, is the largest survey of human genetic variation performed thus far. It has data on 1,184 people, including the initial HapMap samples. Additional human samples were collected from the original populations and from seven new populations: individuals of African ancestry from the Southwestern United States; Chinese individuals from metropolitan Denver; Gujarati Indians from Houston; Luhya people from Webuye, Kenya; Maasai people from Kinyawa, Kenya; individuals of Mexican ancestry from Los Angeles; and individuals from Tuscany, Italy. No medical or personal identifying information was obtained from the HapMap donors. However, the samples are identified by the population from which they were collected.

Researchers analyzed approximately 1.6 million SNPs in about 500 samples from the four original populations and more than 650 samples from the seven new populations. In addition, the consortium sequenced 10 regions totaling about 1 million base pairs in 692 samples from this set in 10 of the 11 populations. More than 800 copy-number variants, where people have different numbers of copies of genomic regions, were also added to the resource.

As expected, the increased number of samples allows detection of variants that are much rarer than could be found by the earlier HapMaps. Because of human population history, lower-frequency variation is shared less among populations. For instance, 77 percent of the detected SNPs were new, revealing that many more variants remain to be found, especially rare variants.

In addition, the larger scale of the new dataset reinforces the result found by previous smaller studies that non-African diversity is largely a subset of African diversity. The researchers also found that the third-generation HapMap increases the power to identify signals of natural selection-variants that increased rapidly in frequency very recently in some populations because they were somehow beneficial to human health.

The researchers assessed the latest generation HapMap for its ability to predict SNPs in other populations. They found that using one population to predict another population's variants works for common variants and for some less-common variants in related populations. However, it does not work well for rare variants in related populations, meaning that rare variants are likely to make much more population-specific contributions to disease. This finding underscores the value of efforts already underway that use efficient 'next-generation' DNA sequencing technologies to sequence large numbers of whole genomes within various populations to find rare variants that contribute to disease.

Many of the HapMap researchers are part of the 1000 Genomes Project, an international public-private consortium launched in 2008 that is building an even more detailed map of human genetic variation. Project researchers are currently using next-generation DNA sequencing technologies to build a public database containing information from the complete genomes of 2,500 people from 27 populations around the world, many of which were studied in the HapMap project. Disease researchers will be able to use the catalogue, which is being developed over the next two years, in their studies of the contribution of common and rarer genetic variation to illness.

The Consortium that produced this latest HapMap included researchers Baylor College of Medicine in Houston; the Broad Institute in Cambridge, Mass.; and the Wellcome Trust Sanger Institute in Hinxton, Cambridge, England. Collaborating researchers worked at Arizona State University, Tempe; Baylor College of Medicine, Houston; Case Western Reserve University, Cleveland; Howard University, Washington, D.C.; the Institute for Oncological Study and Prevention, Florence, Italy; Moi University, Eldoret, Kenya; the University of California, Los Angeles; the University of California, San Francisco; the University of Houston-Clear Lake; and the University of Oklahoma, Norman. Funding was provided by the NHGRI, the National Institute on Deafness and Other Communication Disorders, and the Wellcome Trust.

"The HapMap project has been a stellar example of how improved technologies and dedicated work on obtaining DNA samples from individuals in major populations allows us to provide more detailed resources for studies of human disease," said Richard Gibbs, Ph.D., director of the Human Genome Sequencing Center at the Baylor College of Medicine.

The International HapMap Consortium devoted considerable time and resources to try to ensure that the map was designed, developed and used in a manner that is sensitive to a wide range of ethical and social issues. In addition to getting informed consent from individual sample donors, a careful process of community engagement was conducted with each group approached to participate in the project.

Researchers can access HapMap data through the NIH National Center for Biotechnology Information at <http://hapmap.ncbi.nlm.nih.gov/downloads/index.html.en#release>.

NHGRI is one of 27 institutes and centers at NIH, an agency of the Department of Health and Human Services. NHGRI's Division of Extramural Research supports grants for research and for training and career development. For more, visit [www.genome.gov](http://www.genome.gov).

NIDCD supports and conducts research and research training on the normal and disordered processes of hearing, balance, smell, taste, voice, speech and language and provides health information, based upon scientific discovery, to the public. For more information about NIDCD programs, see the Web site at [www.nidcd.nih.gov](http://www.nidcd.nih.gov).

The National Institutes of Health (NIH) -- The Nation's Medical Research Agency -- includes 27 Institutes and Centers and is a component of the U.S. Department of Health and Human Services. It is the primary federal agency for conducting and supporting basic, clinical and translational medical research, and it investigates the causes, treatments, and cures for both common and rare diseases. For more information about NIH and its programs, visit [www.nih.gov](http://www.nih.gov).

##

This NIH News Release is available online at:  
<<http://www.nih.gov/news/health/sep2010/nhgri-01.htm>>.