Consortium publishes Phase II map of human genetic variation

October 17 2007

The International HapMap Consortium today published analyses of its second-generation map of human genetic variation, which contains three times more markers than the initial version unveiled in 2005. In two papers in the journal *Nature*, the consortium describes how the higher resolution map offers greater power to detect genetic variants involved in common diseases, explore the structure of human genetic variation and learn how environmental factors, such as infectious agents, have shaped the human genome.

Any two humans are more than 99 percent the same at the genetic level. However, it is important to understand the small fraction of genetic material that varies among people because it 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 information. Consequently, researchers refer to the map of human genetic variation as a haplotype map, or HapMap.

The International HapMap Consortium is a public-private partnership of researchers and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States. The U.S. component of the project is led by the National Human Genome Research Institute (NHGRI) on behalf of the 20 institutes, centers and offices of the National Institutes of Health (NIH) that contributed funding.

“Thanks to this consortium’s pioneering efforts to map human genetic

variation, we are already seeing a windfall of results that are shedding new light on the complex genetics of common diseases," said NHGRI Director Francis S. Collins, M.D., Ph.D. "This new approach to research, called genome-wide association studies, has recently uncovered new clues to the genetic factors involved in type 2 diabetes, cardiovascular disease, prostate cancer, multiple sclerosis and many other disorders. These results have opened up new avenues of research, taking us to places we had not imagined in our search for better ways to diagnose, treat and prevent disease."

The second-generation haplotype map, or Phase II HapMap, contains more than 3.1 million genetic variants, called single nucleotide polymorphisms (SNPs) — three times more than the approximately 1 million SNPs contained in the initial version. The more SNPs that are on the map, the more precisely researchers can focus their hunts for genetic variants involved in disease. The rapid growth of genome-wide association studies over the past year and half has been fueled by the HapMap consortium's decision to make its SNP datasets immediately available in public databases, even before the first and the second versions of the map were fully completed.

Researchers around the globe have now associated more than 60 common DNA variants with risk of disease or related traits, with most of the findings coming in the past nine months. As just one example, the Wellcome Trust consortium in England looked at 14,000 cases and 3,000 shared controls, finding variants associated with increased risk of bipolar disorder, coronary artery disease, Crohn's disease, rheumatoid arthritis, type 1 diabetes and type 2 diabetes.

"We are thrilled that the worldwide scientific community is taking advantage of this powerful new tool and we anticipate even more exciting findings in the future. The improved SNP coverage offered by the Phase II HapMap, along with better statistical methods, promises to

further increase the accuracy and reliability of genome-wide association studies," said Gil McVean, Ph.D., of the University of Oxford in England, who co-led the group that analyzed the HapMap data.

Another analysis leader, Mark Daly, Ph.D., of Massachusetts General Hospital and the Broad Institute of MIT and Harvard in Cambridge, Mass., said, "In addition to providing a critical backbone for standard genome-wide association studies, the Phase II HapMap identifies additional features of human genetic variation that will bolster efforts to pinpoint rarer disease mutations."

The Phase II HapMap was produced using the same DNA samples used in the Phase I HapMap. That DNA came from blood 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. No medical or personal identifying information was obtained from the donors, but the samples were labeled by population group.

To provide information on less common variations and to enable researchers to conduct genome-wide association studies in additional populations, NHGRI plans to extend the HapMap even further. Among the populations donating additional DNA samples are: Luhya in Webuye, Kenya; Maasai in Kinyawa, Kenya; Tuscans in Italy; Gujarati Indian in Houston; Chinese in metropolitan Denver; people of Mexican ancestry in Los Angeles; and people of African ancestry in the southwestern United States.

In its overview paper in Nature, the consortium estimates that the Phase II HapMap captures 25 percent to 35 percent of common genetic variation in the populations surveyed. The consortium also confirmed that use of Phase II HapMap data has helped to improve the coverage of various commercial technologies currently being used to identify disease-

related variants in genome-wide association studies. Researchers did note, however, that current technologies tend to provide better coverage in non-African populations than in African populations because of the greater degree of genetic variability in African populations.

The overview paper also reports that the Phase II HapMap has provided new insights into the structure of human genetic variation. One new finding was the surprising extent of recent common ancestry found in all of the population groups. Taking advantage of the map's increased resolution, the researchers identified stretches of identical DNA between pairs of donor chromosomes and then compared these stretches both within and across individuals. Their analysis showed that 10 to 30 percent of the DNA segments analyzed in each population showed shared regions indicating descent from a common ancestor within 10 to 100 generations.

In addition, the new map enabled researchers to quantify more precisely the rates of shuffling, or recombination, seen among different gene classes in the human genome. In their overview paper, researchers report that recombination rates vary more than six-fold among different gene classes. The highest rates of recombination were found among genes involved in the body's immune defense, while the lowest rates appear among genes for chaperones, which are proteins that play a crucial role in making sure other proteins are folded properly. In general, genes that code for proteins associated with the surface of cells and external functions, such as signaling, were found to be more prone to recombination than those that code for proteins internal to cells.

While the reasons for the varying recombination rates remain to be determined, the findings pose interesting evolutionary questions. In their paper, researchers suggest that one explanation may be that some recombinations in areas of the genome that affect responses to infectious agents or other environmental pressures may be selected for because

they provide a survival advantage.

A related study appearing in the same issue of Nature describes how the enhanced map can help pinpoint pivotal changes in the human genome that arose in recent history. These changes, now common among various populations worldwide, became prevalent through natural selection — meaning they were somehow beneficial to human health. Although these DNA variants may still be important, their biological significance remains largely unknown.

Using the Phase II HapMap data, a team led by researchers at the Broad Institute of MIT and Harvard identified hundreds of genomic regions that carry the hallmarks of recent positive natural selection. These regions are large, often extending for millions of nucleotides and including multiple genes. Thus, the researchers developed a set of computational guidelines to help locate the single letter changes that formed the focal points for evolutionary change.

The work uncovered several intriguing genetic variations that could provide novel insights into the biological forces underlying natural selection in humans. Two differences, which are common primarily in Asian populations, lie within the EDAR and EDA2R genes. In humans, these genes function together to form hair follicles and sweat glands, as well as other structures.

The researchers also identified DNA variations in African populations that may be linked to resistance to Lassa fever, a viral infection common in Western Africa. These changes lie in two genes, LARGE and DMD, which are involved in viral entry into cells. The findings help underscore one of the study's key themes — that multiple genes, acting together in the same biological process, often show signs of positive selection, both in humans and other organisms. Integrating these data may bolster efforts to understand the biological consequences of human genetic

variation.

“Human history and the genome have been dramatically shaped by environmental factors, diet and infectious disease,” said co-first author Pardis Sabeti, Ph.D., who is a postdoctoral fellow at the Broad Institute of MIT and Harvard. “The gene variants identified in our study open new windows on these evolutionary forces and provide a launching point for future biological studies of human adaptation.”

The effort to build the improved HapMap relied heavily on the high-throughput genotyping capacity of Perlegen Sciences, Inc., of Mountain View, Calif. The firm tested virtually the entire known catalog of human SNP variation on the HapMap samples, as well as contributed some of its own resources to make the map possible.

“The Phase II HapMap is truly an example of a public-private collaboration at its best. It’s wonderful that everyone pulled together to create this improved map, which is a priceless tool for all researchers seeking to use genomic information to improve human health, be they in government, academia or industry," said Kelly A. Frazer, Ph.D., formerly vice president of genomics at Perlegen and now director of genomic biology at Scripps Genomic Medicine Program, in La Jolla, Calif.

Source: National Human Genome Research Institute