

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

HapMap Project launched

ArticleInfo		
ArticleID	:	4899
ArticleDOI	:	10.1186/gb-spotlight-20031219-01
ArticleCitationID	:	spotlight-20031219-01
ArticleSequenceNumber	:	251
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2003-12-19 OnlineDate : 2003-12-19
ArticleCopyright	:	BioMed Central Ltd2003
ArticleGrants	:	
ArticleContext	:	130594411

Cathy Holding

Email: cholding@hgmp.mrc.ac.uk

A major international project to produce a complete map of common patterns of differences in the human genome - haplotypes - has been launched. The 'HapMap' is the first systematic approach to understanding diseases with a multigenic component.

In the December 18/25 [Nature](#), the International HapMap Consortium publishes details of the aims and methods of the \$100 million collaborative effort - a similar scale to the [Human Genome Project](#) - that combines the efforts of major genome sequencing centers including the [National Human Genome Research Institute](#) (NHGRI) and Baylor College of Medicine in the United States, the United Kingdom's [Sanger Institute](#) and Oxford University, the Chinese HapMap Consortium, and teams from universities in Japan, Africa, China, and Canada in association with the US company [Illumina](#). The Consortium intends to genotype more than a million sequence variants and to analyze their frequencies and degrees of association in a total of 270 DNA samples from Northern and Western European, Chinese, Nigerian, and Japanese populations.

Francis Collins, director of NHGRI - the primary funder for the HapMap - told us, "This is truly an international project with strong leadership from six countries. The HapMap will provide a critical resource that has previously been missing and will allow connecting variation in the genome with risks of human disease."

David Bentley, head of Human Genetics at the Wellcome Trust Sanger Institute, said, "The paper is a fully fleshed-out plan for research that summarizes conclusions from pilot data, so that people can get as clear a view as possible of what is to be done, in case it affects their research plans.

"The project has been in the planning stage for 2-3 years and has drawn on a lot of preliminary data. Now it is about to take off. It is aimed at as wide an audience as possible, and it should encourage feedback."

"This is the first comprehensive look at how we vary. It lays down a very important foundation that will be useful for tackling the very big problems of finding disease genes. It won't be the whole answer, but a very important first part," Bentley told us.

"It leads to discovery of candidates for SNPs that cause functional changes in proteins. A subset of SNPs that have been independently seen more than once are called double-hit SNPs. We have found that they are significantly more likely to convert into polymorphic assays, said Lincoln Stein, bioinformaticist at [Cold Spring Harbor Laboratory](#).

"We convert double hit SNPs into working assays and use them to genotype DNA from the members of the population panel of 270 individuals. We release these to the scientific community as three pieces of information: the assay design details, which are very valuable as they provide the recipe; the allele frequency information; and the genotype on an individual that is used to reconstruct linkage disequilibrium," he continued.

Richard Gibbs, professor and director of the Baylor College of Medicine [Human Genome Sequencing Center](#), explained, "Genotype here means a SNP study on an individual. We are beginning to define

genotypes, and loosely, we talk about the individual markers at an SNP site. If the project had done a million genotypes, this would represent 10,000 SNPs tested in a hundred people."

Collins said, "We are going to generate vast amounts of data. Already the HapMap Project has generated 30 million genotypes, and the figure is likely to go much higher - but the ideal way to present this information to users hasn't been figured out yet. We think users will be most interested in a gold standard set of tag SNPs for a region, or for the whole genome. These tag SNPs must be chosen so that they optimally represent variation across the entire genome. The ideal way of choosing this set is still a work in progress. This is not just 'turn-the-crank' science."

The ethics of obtaining and publishing the information were closely considered. "In contrast to the human genome sequence, which was all anonymous following informed consent, the HapMap data necessitate a level of identification," said Bentley. "Everybody has to buy into that. Everybody has to be aware. Information connected to ethnic origins is not acceptable to some people. There was a detailed consultation process, and it was important to get it right."

"I think it really is a valuable project, and it is wonderful to be participating in it. It is the next step beyond the human genome sequence that will build a bridge between sequence and disease gene discovery," said Gibbs.

References

1. *Nature*, [<http://www.nature.com/>]
2. Human Genome Project Information, [http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml]
3. National Human Genome Research Institute, [<http://www.nhgri.nih.gov/>]
4. Wellcome Trust Sanger Institute, [<http://www.sanger.ac.uk/>]
5. Illumina, [http://www.illumina.com/tech_life_hapmap_participants.htm]
6. Cold Spring Harbor Laboratory, [<http://www.cshl.org/>]
7. Baylor College of Medicine Human Genome Sequencing Center, [<http://www.hgsc.bcm.tmc.edu/>]