

DOI:
10.1038/nrg2235

 GENOMICS

HapMap Phase II unveiled



Phase II of the [International HapMap Project](#) has just been published, adding substantial amounts of human genetic variation data to the Phase I release and providing important insights for the design of gene association studies.

Phase II involved genotyping 3.1 million SNPs from the same populations as Phase I. The increased SNP density — to about one SNP per kilobase — provides higher-resolution fine-scale genetic maps, and facilitates detection and localization of recombination hotspots. Because the newly identified SNPs have lower minor allele frequencies than those from Phase I, rare variation is better represented. Importantly for genome-wide association studies, ~25–35% of all common SNPs are thought to have been genotyped, correlating with up to 95% of all common SNPs. Although currently available arrays capture 68–88% of this variation, analytical approaches that use the HapMap data can further increase the value of these arrays.

The new data provide two important insights into the structure

of linkage disequilibrium. First, the apparent long-range similarity among haplotypes reveals recent common ancestry and inbreeding within populations. The high-quality haplotype data is a further asset for genome-wide association studies, as haplotype-sharing approaches can complement standard SNP association tests and should be especially useful to identify potential rare variants that are associated with complex diseases. Second, despite high SNP density, no tag could be identified for up to 1% of all high-frequency SNPs (a tag is a SNP that provides information on other variants nearby). Most of these untaggable SNPs lie in recombination hotspots, suggesting that regions of interest that contain hotspots might require additional sequencing to identify variants.

Another new insight is that hotspots account for ~60% of recombination in the human genome and for 6% of its sequence. Recombination is lower in transcribed genic regions and peaks 5' of transcription start sites, although most hotspots do not

lie in promoters. Average recombination rates vary among gene classes, with defense and immunity genes having the highest and chaperone-encoding genes having the lowest rates. The authors propose that recombination might be favoured in regions that are exposed to recurrent selection (for example, from the environment or by pathogens).

Now that the fleshed out HapMap is available, what is the future of the HapMap project? Additional samples and populations will need to be sequenced and genotyped to provide information on rarer variants. Other important goals will involve generating molecular phenotypes for the HapMap samples and integrating SNP information with structural variation.

Magdalena Skipper

ORIGINAL RESEARCH PAPER

The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007)

WEB SITES

International HapMap Project :

<http://www.hapmap.org/>

dbSNP: <http://www.ncbi.nlm.nih.gov/SNP/>