

# ADVANCES IN DRUG DEVELOPMENT

Current Developments in Oncology Drug Research

Section Editor: Mark J. Ratain, MD

---

## Using the HapMap in Drug Development

Lisa D. Brooks, PhD  
Program Director  
Genetic Variation Program  
National Human Genome Research Institute  
National Institutes of Health

### **H&O** What is the HapMap?

**LB** The HapMap is a description of the patterns of human genetic variation, in particular highlighting associations among different variants.

There are 3 billion sites in the human genome and approximately 20 million of these sites vary among individuals. A basic question asked about any disease or condition or feature affected by genes is: what variants contribute to that phenotypic effect? There are two ways to answer this question. First, all 20 million variants in the human genome could be tested. This approach would be extremely difficult because all of the variants are not known and extremely expensive because this testing would be done by genotyping, which is costly. However, the number of variants that need to be tested can be decreased through knowing what variants are associated with each other. Knowing the genotype at one variant makes it possible to predict the genotype at another, nearby variant.

The human genome is not a compilation of independent variants. Rather, the relatively minor amount of recombination that has taken place through the generations of human existence means that many variants are inherited together in what are known as “haplotype blocks.” There are only a handful of haplotypes in any given region of the genome, and any one block could contain several single nucleotide polymorphisms (SNPs) or other variants, such as small insertions or deletions, larger insertions or deletions, or copy number variants.

The HapMap was created as a tool to efficiently locate genomic regions that affect any particular trait. Through the collection of data from four different populations, we were able to elucidate the associations of SNPs located

nearby one another, revealing strongly associated SNPs in haplotype blocks.

Haplotype blocks vary in size. Assuming that there are 20 SNPs strongly associated with each other in a haplotype block, the strong associations mean that the SNPs contain redundant information. Perhaps just three contain most of the information on the variation in that region. This is the major benefit of the HapMap: the three tag SNPs that are selected are just about as useful to study as all 20 SNPs, making such research far more cost-effective.

### **H&O** How are the HapMap data being used?

**LB** The data from the HapMap are being used in case-control studies in many different arenas. A basic experiment would involve two groups of individuals, one group that has the disease of interest and a control group, which does not. Each group would contain 1,000 or so individuals; this number is not exact, but it needs to be large enough to control for various confounding factors and to ensure statistical power.

Using the HapMap tag SNPs, the individuals in both groups are genotyped, and the investigators scan for regions in the genome that differ in frequency between the cases and controls. These regions may contain genes that are somehow associated with the disease in question. In actual practice, most researchers obtain panels of SNPs from genotyping providers. These companies have created panels of SNPs based on the HapMap data, and in general, researchers rely on these preexisting panels in order to find out information across the genome.

Once a researcher finds a region of interest, in which there is a difference in SNP frequency between cases and controls, the region needs to be studied in more detail to figure out which genes and variants contribute to the phenotype causally, rather than simply being associated with it because of close linkage.

In general, there is a high degree of variability among individuals for their drug-metabolizing enzymes. For this reason, a drug may be effective for one individual but not another. A case-control study can be done to compare a group of people that respond well to a drug with a group that does not; this response would be the phenotype of interest.

**H&O** Do patterns of variation appear to be similar across the populations sampled in the HapMap?

**LB** Yes. Interestingly, the basic patterns are fairly similar among all populations. This similarity makes sense when one considers that all current populations come from one ancestral population. For samples with non-African ancestry, approximately 400,000 SNPs are needed, and samples with African ancestry require approximately 600,000–700,000 SNPs. The difference in requirement for these populations is due to the fact that there is more variation within African populations and there has historically been more recombination in African populations; for non-African populations a bottleneck occurred when they moved out of Africa, which increased the lengths of the blocks. The blocks are therefore shorter and the associations are weaker in populations with African ancestry, so more tag SNPs are required in order to be reflective of the entire genome.

**H&O** How is the HapMap being used in drug development?

**LB** The HapMap plays two key roles in drug development. First, as described above, if one finds a gene region that is associated with, for example, breast cancer, it is possible to then find out what genes and variants are involved and in what way, whether it is due to a truncated protein or a problem with regulation, or some other factor. In other words, the data indicating which haplotype blocks contribute to risk support subsequent research into the biology of the disease. Once the biology of the disease is understood, drug targets can be identified. If an abnormal protein product is identified, one can then ask whether something can be done to change the abnormality.

In addition, as mentioned above, the data can be used to further pharmacogenomic research. It will be possible to identify the optimal drug dose based on knowing an individual's variants and how those variants affect the metabolism of a particular drug.

Currently, anyone who is interested in mapping any phenotype would use the HapMap. Recently, a gene associated with macular degeneration, known as complement factor H, was identified. Researchers at the National Cancer Institute are creating the Cancer Genome Atlas, which is becoming an indispensable resource for the development of drugs to treat cancer, as the connections between genetic variations and the development of cancer are increasingly revealed.

**H&O** Are there any cautions regarding the application of HapMap data into research?

**LB** One important point is that in any given study, one needs to be certain that the differences in allele frequencies observed between cases and controls are due to differences in disease risk factors, rather than simply to differences in the proportions of ethnicities. Most variants are shared among populations. For example, the A, B, AB, and O blood types appear in virtually all populations. However, the variants differ in frequency among populations. If, in a case-control study, one arm contained a high percentage of individuals of a particular ethnicity, and the other arm contained a high percentage of individuals with a different ethnicity, the differences in allele frequency could be due to the population differences in allele frequencies, rather than to genetic variation for variants contributing to disease risk. It is essential to control for population comparisons in order to be certain that the differences observed are due to susceptibility differences and not to population allele frequency differences.

**H&O** What is the place of the HapMap in terms of the shift toward viewing different types of cancer in terms of genetic variation?

**LB** The role of the HapMap in these new understandings pertains to the molecular diagnosis of any particular disease. What is generally known as breast cancer may be a collection of many different diseases. There may be many different genes that, when not working properly, result in breast cancer; however, since different genes are involved, different treatments might be indicated. Using the HapMap, one may see that there are 10 genes involved in a particular disease, and that some individuals have variants in one gene that lead to risk and others have variants in a different gene. Even though these individuals show the same phenotype, the underlying mechanism—and therefore treatments—would be different. By enabling us to determine what genes are involved, the HapMap may lead to advances in molecular diagnosis.

### Suggested Reading

The International HapMap Project. <http://www.hapmap.org>.

The International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437:1299-1320.

The International HapMap Consortium. The International HapMap Project. *Nature*. 2003;426:789-796.

The International HapMap Consortium. Integrating ethics and science in the International HapMap Project. *Nature Rev Genet*. 2004;5:467-475.

Marsh S, McLeod HL. Pharmacogenomics: from bedside to clinical practice. *Hum Mol Genet*. 2006;15(suppl 1):R89-R93.

Ribas G, Gonzales-Neira A, Salas A, et al. Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Hum Genet*. 2006;118:669-679.