

The rough guide to the genome



A new effort to map human genetic variation should provide a shortcut for researchers trying to uncover the roots of disease. Carina Dennis profiles the 'HapMap' project.

Searching for the genetic factors associated with common ailments such as heart disease and asthma is like trying to find a needle in a haystack. But an international consortium is now working on a way to speed the search, by packaging the haystack of human genetic variation into more manageable bundles.

The three-year, US\$100-million endeavour, called the International HapMap Project, is the biggest thing in genomics since the effort to sequence the entire human genome, which was completed in April this year. If that sequence is the 'Book of Life', then the HapMap will be a handy index, helping gene-hunters to zoom in quickly on informative chapters or pages. "The HapMap will provide the missing link between the DNA sequence of the genome and the way in which the genome influences the risk of disease," says Francis Collins, director of the National Human Genome Research Institute in Bethesda, Maryland, one of the HapMap's main financial backers.

Geneticists want to know why some of us are more likely than others to succumb to a particular disease. But they often have no idea where to start looking among the three billion letters, or bases, of the human genetic code. So researchers often sample populations that are affected by the disease, and study single nucleotide polymorphisms (SNPs) — points in the genome at which one base can differ between individuals. If a particular SNP is inherited with the disease, it is a strong indication that a gene that confers susceptibility lies somewhere nearby.

There are thought to be some ten million common SNPs scattered throughout the human genome. Testing thousands of people for millions of SNPs is beyond current technology — so in practice, gene-hunters use a few thousand SNPs at most. And with no prior knowledge of which SNPs will be most useful, it is a hit-and-miss business.

The HapMap project aims to use our understanding of the genetic shuffling that occurs during the production of sex cells to make gene searches more efficient. In the cell divisions that give rise to eggs or sperm, pairs of chromosomes line up and exchange portions of genetic material. This is not entirely random: the breaking, exchange and reas-



On the road: Francis Collins launches the HapMap project.

ing tend to occur at particular points. As a result, blocks of sequence have been inherited down the generations from our common ancestors without being broken up. The blocks vary widely in size, but most are thought to be 10,000 bases or more in length¹.

The DNA sequence of a block is known as a haplotype. Common haplotypes can be identified by sampling only a handful of key SNPs, and studies have indicated that, for each block, there are only a few common haplotypes in the population¹⁻³. The HapMap should therefore provide geneticists with the means to scan the entire genome rapidly for disease genes, perhaps by analysing as few as 300,000 SNPs¹. The map could also be used to study how people respond differently to prescription drugs, according to their particular genetic make-up (see page 760).

Like the publicly funded effort to sequence the human genome, the HapMap consortium has a commercial rival. The HapMap team has also made sure to consult with the communities from which DNA will be sampled, to avoid the charges of scientific 'colonialism' that have dogged earlier attempts to study human genetic diversity.

Three-pronged attack

The consortium will focus on DNA samples collected from three distinct populations: people of northern- and western-European descent; a Nigerian population called the Yoruba; and an Asian collection composed of Japanese and Han Chinese. Analysis of the samples from people of European ancestry, which were collected for earlier population-genetic studies, is already under way.

The HapMap consortium will produce a haplotype map based on the similarities in block structure and common haplotypes for the European, African and Asian populations. "The block architecture will be similar, but the common haplotypes are likely to vary between different populations in terms of composition and frequency," says Pui-Yan Kwok, a geneticist at the University of California, San Francisco, and a member of the HapMap consortium.

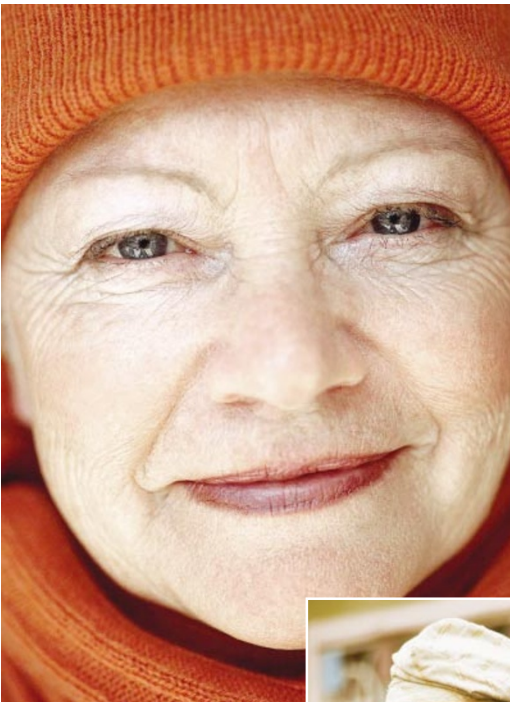
The idea is not to document the full extent of human genetic diversity, but to provide useful tools for gene discovery. If



it becomes obvious that other populations possess haplotypes that are not represented in the maps, this information can be incorporated at a later date. The Yoruba population, however, is expected to contain the greatest genetic diversity. "What you find in European and Asian populations, you almost always find in Africa," says Collins.

With its strong biomedical focus, the HapMap project is quite different from the Human Genome Diversity Project, an ill-fated exercise first proposed more than a decade ago. That project's founders were interested in anthropological questions. They wanted to sample DNA from many populations to understand their genetic relationships to one another, and to link the findings to studies of language and other cultural phenomena. Many groups representing indigenous peoples objected, arguing that it would exploit vulnerable populations and intrude into people's beliefs about their own origins. Fatally wounded by these accusations, the project never got off the ground.

Given the potential for controversy, the HapMap consortium's ethical experts are going to great lengths to obtain informed



World view: the HapMap initiative will gather genetic data from African, Asian and ancestrally European populations.

across the genome at intervals of around 5,000 bases. Further SNPs will then be identified where needed to define haplotype blocks. The team is also running a pilot study in parallel to scrutinize every known SNP in ten selected regions of the genome, each some 500,000 bases long. This should reveal whether the map can be produced using fewer SNPs, or whether more will be required.

In keeping with the precedent set by the Human Genome Project, the HapMappers will release their data as soon as they are available. "That way, other people can work on the data at the same time and come up with new ideas for analysing them," says Peter Donnelly, a statistician at the University of Oxford, UK, and a HapMap collaborator.

Patent parasites?

But this policy of openness has led to concerns about what Collins calls "parasitic intellectual property claims". The project's leaders fear that some people could combine the consortium's data with their own, and then patent the findings in a way that restricts others' ability to work freely with the HapMap data.

Earlier rushes to patent SNPs fell by the wayside when it became clear that patent offices did not consider them to pass the test of having a clear 'utility'. But as knowledge of genetic variation grows, some may claim that haplotypes are patentable tools for gene discovery. "Haplotypes, more than SNPs, come closer to patentability," says Lawrence Sung of the University of Maryland School of Law

in Baltimore, who is advising the HapMap consortium. So users of the HapMap database, run by the Cold Spring Harbor Laboratory in New York state, will be required to click on an agreement to certify that they will not take any action that would exclude others from using the data.

Whether or not haplotypes are patentable, one company is trying to turn them to profit. Perlegen, based in Mountain View, California, claims already to have made its own haplotype map. Two years ago, the company unveiled its data for human chromosome 21 (ref. 3). Pharmaceutical firms including GlaxoSmithKline and Bristol-Myers Squibb have now enlisted Perlegen to sift through their clinical DNA samples to track down genetic markers that might explain why individual patients respond differently to drugs.

The relationship between Perlegen and the HapMap consortium has so far been more cordial than competitive. And both sides are adamant that the company's work does not render the public effort redundant. "Often the only way we know a scientific result is right is when multiple people can arrive at the same answer," says David Cox, Perlegen's chief scientist.

Flustered gene-hunters whose quarries remain hidden in the vastness of the human genome can only hope that the HapMap will help to answer their prayers. ■

Carina Dennis is Nature's Australasian correspondent.

1. Gabriel, S. B. *et al. Science* **296**, 2225–2229 (2002).
2. Dawson, E. *et al. Nature* **418**, 544–548 (2002).
3. Patil, N. *et al. Science* **294**, 1719–1723 (2001).
4. Reich, D. E., Gabriel, S. B. & Altshuler, D. *Nature Genet.* **33**, 457–458 (2003).
5. Carlson, S. C. *et al. Nature Genet.* **33**, 518–521 (2003).

▶ www.hapmap.org

consent from the study populations, and to involve communities in decision-making. They have given extensive thought to the question of how to explain the project's rationale in language that ordinary people can understand. All DNA donors will be asked to give consent for their samples to be used not just for the HapMap project, but also for future studies of genetic variation. For each sample population, a community advisory group will be set up to ensure that these future studies are consistent with the consent form.

Other challenges are technical — the first being to work out which of the millions of SNPs identified so far will be most useful in constructing the maps. Some SNP variants, for instance, occur at frequencies that are too low to be readily studied in the sample populations^{4,5}. The HapMappers also need to ensure that they have a good, even coverage of SNPs across the entire genome. So initial work on the project, which began in October last year, has concentrated on addressing these issues. "We have beefed up SNP density almost threefold across the genome," says Collins.

The consortium plans to construct the initial map by studying 600,000 SNPs scattered