

The International HapMap Project

The International HapMap Consortium*

*Lists of participants and affiliations appear at the end of the paper

The goal of the International HapMap Project is to determine the common patterns of DNA sequence variation in the human genome and to make this information freely available in the public domain. An international consortium is developing a map of these patterns across the genome by determining the genotypes of one million or more sequence variants, their frequencies and the degree of association between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe. The HapMap will allow the discovery of sequence variants that affect common disease, will facilitate development of diagnostic tools, and will enhance our ability to choose targets for therapeutic intervention.

Common diseases such as cardiovascular disease, cancer, obesity, diabetes, psychiatric illnesses and inflammatory diseases are caused by combinations of multiple genetic and environmental factors¹. Discovering these genetic factors will provide fundamental new insights into the pathogenesis, diagnosis and treatment of human disease. Searches for causative variants in chromosome regions identified by linkage analysis have been highly successful for many rare single-gene disorders. By contrast, linkage studies have been much less successful in locating genetic variants that affect common complex diseases, as each variant individually contributes only modestly to disease risk^{2,3}. A complementary approach to identifying these specific genetic risk factors is to search for an association between a specific variant and a disease, by comparing a group of affected individuals with a group of unaffected controls⁴. In the absence of strong natural selection, there is likely to be a broad spectrum of frequency of such variants, many of which are likely to be common in the population. A number of association studies, focused on candidate genes, regions of linkage to a disease or more large-scale surveys, have already led to the discovery of genetic risk factors for common diseases. Examples include type 1 diabetes (human leukocyte antigen (HLA)⁵), insulin⁶ and *CTLA4* (ref. 7)), Alzheimer's disease (*APOE*)⁸, deep vein thrombosis (factor V)⁹, inflammatory bowel disease (*NOD2* (refs 10, 11) and also 5q31 (ref. 12)), hypertriglyceridaemia (*APOA5*)¹³, type 2 diabetes (*PPARG*)^{14,15}, schizophrenia (neuregulin 1)¹⁶, asthma (*ADAM33*)¹⁷, stroke (*PDE4D*)¹⁸ and myocardial infarction (*LTA*)¹⁹.

One approach to doing association studies involves testing each putative causal variant for correlation with the disease (the 'direct' approach)². To search the entire genome for disease associations would entail the substantial expense of whole-genome sequencing of numerous patient samples to identify the candidate variants³. At present, this approach is limited to sequencing the functional parts of candidate genes (selected on the basis of a previous functional or genetic hypothesis) for potential disease-associated candidate variants. An alternative approach (the 'indirect' approach) has been proposed²⁰, whereby a set of sequence variants in the genome could serve as genetic markers to detect association between a particular genomic region and the disease, whether or not the markers themselves had functional effects. The search for the causative variants could then be limited to the regions showing association with the disease.

Two insights from human population genetics suggest that the indirect approach is able to capture most human sequence variation, with greater efficiency than the direct approach. First, ~90% of sequence variation among individuals is due to common variants²¹. Second, most of these originally arose from single historical mutation events, and are therefore associated with nearby variants that were present on the ancestral chromosome on which the mutation occurred. These associations make the indirect approach

feasible to study variants in candidate genes, chromosome regions or across the whole genome. Prior knowledge of putative functional variants is not required. Instead, the approach uses information from a relatively small set of variants that capture most of the common patterns of variation in the genome, so that any region or gene can be tested for association with a particular disease, with a high likelihood that such an association will be detectable if it exists.

The aim of the International HapMap Project is to determine the common patterns of DNA sequence variation in the human genome, by characterizing sequence variants, their frequencies, and correlations between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe. The project will thus provide tools that will allow the indirect association approach to be applied readily to any functional candidate gene in the genome, to any region suggested by family-based linkage analysis, or ultimately to the whole genome for scans for disease risk factors.

Common variants responsible for disease risk will be most readily approached by this strategy, but not all predisposing variants are common. However, it should be noted that even a relatively uncommon disease-associated variant can potentially be discovered using this approach. Reflecting its historical origins, the uncommon variant will be travelling on a chromosome that carries a characteristic pattern of nearby sequence variants. In a group of people affected by a disease, the rare variant will be enriched in frequency compared with its frequency in a group of unaffected controls. This observation, for example, was of considerable assistance in the identification of the genes responsible for cystic fibrosis²² and diastrophic dysplasia²³, after linkage had pointed to the general chromosomal region.

Below we provide a brief description of human sequence variation, and then describe the strategy and key components of the project. These include the choice of samples and populations for study, the process of community engagement or public consultation, selection of single-nucleotide polymorphisms (SNPs), genotyping, data release and analysis.

Human DNA sequence variation

Any two copies of the human genome differ from one another by approximately 0.1% of nucleotide sites (that is, one variant per 1,000 bases on average)^{24–27}. The most common type of variant, a SNP, is a difference between chromosomes in the base present at a particular site in the DNA sequence (Fig. 1a). For example, some chromosomes in a population may have a C at that site (the 'C allele'), whereas others have a T (the 'T allele'). It has been estimated that, in the world's human population, about 10 million sites (that is, one variant per 300 bases on average) vary such that both alleles are observed at a frequency of $\geq 1\%$, and that these 10 million common SNPs constitute 90% of the variation in the popu-

lation^{21,28}. The remaining 10% is due to a vast array of variants that are each rare in the population. The presence of particular SNP alleles in an individual is determined by testing ('genotyping') a genomic DNA sample.

Nearly every variable site results from a single historical mutational event as the mutation rate is very low (of the order of 10^{-8} per site per generation) relative to the number of generations since the most recent common ancestor of any two humans (of the order of 10^4 generations). For this reason, each new allele is initially associated with the other alleles that happened to be present on the particular chromosomal background on which it arose. The specific set of alleles observed on a single chromosome, or part of a chromosome, is called a haplotype (Fig. 1b). New haplotypes are formed by additional mutations, or by recombination when the maternal and paternal chromosomes exchange corresponding segments of DNA, resulting in a chromosome that is a mosaic of the two parental haplotypes²⁹.

The coinheritance of SNP alleles on these haplotypes leads to associations between these alleles in the population (known as linkage disequilibrium, LD). Because the likelihood of recombination between two SNPs increases with the distance between them, on average such associations between SNPs decline with distance. Many empirical studies have shown highly significant levels of LD, and often strong associations between nearby SNPs, in the human genome³⁰⁻³⁴. These strong associations mean that in many chromosome regions there are only a few haplotypes, and these account for most of the variation among people in those regions^{31,35,36}.

The strong associations between SNPs in a region have a practical value: genotyping only a few, carefully chosen SNPs in the region will provide enough information to predict much of the information about the remainder of the common SNPs in that region. As a result, only a few of these 'tag' SNPs are required to identify each of

the common haplotypes in a region^{35,37-39} (Fig. 1c).

As the extent of association between nearby markers varies dramatically across the genome^{30-32,34,35,40}, it is not efficient to use SNPs selected at random or evenly spaced in the genome sequence. Instead, the patterns of association must be empirically determined for efficient selection of tag SNPs. On the basis of empirical studies, it has been estimated that most of the information about genetic variation represented by the 10 million common SNPs in the population could be provided by genotyping 200,000 to 1,000,000 tag SNPs across the genome^{31,36,38,39}. Thus, a substantial reduction in the amount of genotyping can be obtained with little loss of information, by using knowledge of the LD present in the genome.

For common SNPs, which tend to be older than rare SNPs, the patterns of LD largely reflect historical recombination and demographic events⁴¹. Some recombination events occur repeatedly at 'hotspots'^{30,42}. The result of these processes is that current chromosomes are mosaics of ancestral chromosome regions²⁹. This explains the observations that haplotypes and patterns of LD are shared by apparently unrelated chromosomes within a population and generally among populations⁴³.

These observations are the conceptual and empirical foundation for developing a haplotype map of the human genome, the 'HapMap'. This map will describe the common patterns of variation, including associations between SNPs, and will include the tag SNPs selected to most efficiently and comprehensively capture this information.

The International HapMap Consortium

An initial meeting to discuss the scientific and ethical issues associated with developing a human haplotype map was held in Washington DC on 18-19 July 2001 (<http://www.genome.gov/10001665>). Groups were organized to consider the ethical issues, to develop the scientific plan and to choose the populations to

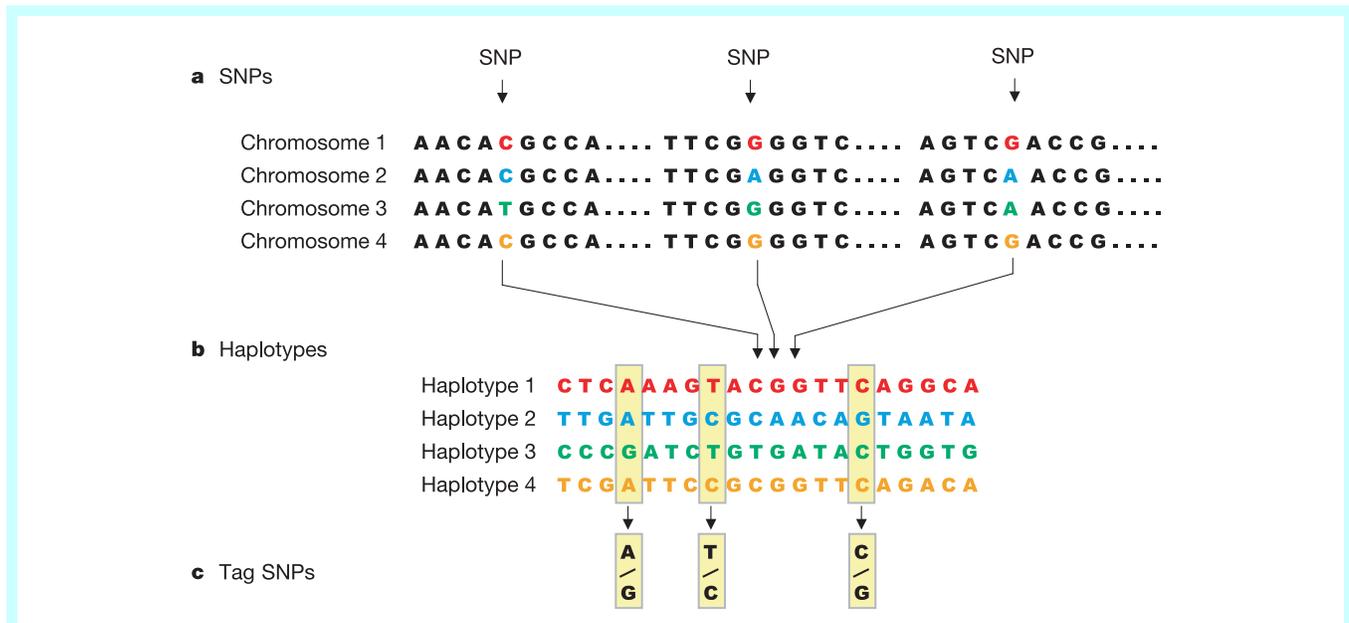


Figure 1 SNPs, haplotypes and tag SNPs. **a**, SNPs. Shown is a short stretch of DNA from four versions of the same chromosome region in different people. Most of the DNA sequence is identical in these chromosomes, but three bases are shown where variation occurs. Each SNP has two possible alleles; the first SNP in panel **a** has the alleles C and T. **b**, Haplotypes. A haplotype is made up of a particular combination of alleles at nearby SNPs. Shown here are the observed genotypes for 20 SNPs that extend across 6,000 bases of DNA. Only the variable bases are shown, including the

three SNPs that are shown in panel **a**. For this region, most of the chromosomes in a population survey turn out to have haplotypes 1-4. **c**, Tag SNPs. Genotyping just the three tag SNPs out of the 20 SNPs is sufficient to identify these four haplotypes uniquely. For instance, if a particular chromosome has the pattern A-T-C at these three tag SNPs, this pattern matches the pattern determined for haplotype 1. Note that many chromosomes carry the common haplotypes in the population.

include. The International HapMap Project (<http://www.hapmap.org/>) was then formally initiated with a meeting in Washington DC on 27–29 October 2002 (<http://www.genome.gov/10005336>). The participating groups and funding sources are listed in Table 1.

DNA samples and populations

Human populations are the products of numerous social, historical and demographic processes. As a result, no populations are typical, special or sharply bounded^{44,45}. As most common patterns of variation can be found in any population⁴⁶, no one population is essential for inclusion in the HapMap. Nonetheless, we decided to include several populations from different ancestral geographic locations to ensure that the HapMap would include most of the common variation and some of the less common variation in different populations, and to allow examination of various hypotheses about patterns of LD.

Studies of allele frequency distributions suggest that ancestral geography is a reasonable basis for sampling human populations^{44,47,48}. Pilot studies using samples from the Yoruba, Japanese, Chinese and individuals with ancestry from Northern and Western Europe have shown substantial similarity in their haplotype patterns, although the frequencies of haplotypes often differ^{31,44}. Given these scientific findings, coupled with consideration of ethical, social and cultural issues, these populations were approached for inclusion in the HapMap through a process of community engagement or consultation (see Box 1).

The HapMap developed with samples from these four large

populations will include a substantial amount of the genetic variation found in all populations throughout the world. The goal of the HapMap is medical, and the common patterns of variation identified by the project will be useful to identify genes that contribute to disease and drug response in many other populations. Samples from several other populations are being collected for studies that will examine how similar their haplotype patterns are to those in the HapMap. If the patterns found are very different, samples from some of these populations may be genotyped on a large scale to make the HapMap more applicable to them. Further follow-up studies in other populations, small and large, are likely to be undertaken by scientists in many nations for common disease gene discovery.

The project will study a total of 270 DNA samples: 90 samples (see Supplementary Information, part 1) from a US Utah population with Northern and Western European ancestry (samples collected in 1980 by the Centre d'Etude du Polymorphisme Humain (CEPH)⁴⁹ and used for other human genetic maps, 30 trios of two parents and an adult child), and new samples collected from 90 Yoruba people in Ibadan, Nigeria (30 trios), 45 unrelated Japanese in Tokyo, Japan, and 45 unrelated Han Chinese in Beijing, China. All donors gave specific consent for their inclusion in the project. Population membership was determined in ways appropriate for each culture: for the Yoruba by asking the donor whether all four grandparents were Yoruba, for the Han Chinese by asking the donor whether at least three of four grandparents were Han Chinese, and for the Japanese by self-identification. The CEPH samples are available

Table 1 Groups participating in the International HapMap Project

Country	Research group	Institution	Role	Per cent genome	Chromosomes	Funding agency		
Japan	Yusuke Nakamura	RIKEN, Univ. of Tokyo	Genotyping: Third Wave	25.1%	5, 11, 14, 15, 16, 17, 19	Japanese MEXT		
	Ichiro Matsuda	Health Sciences Univ. of Hokkaido, Eubios Ethics Inst., Shinshu Univ.	Public consultation, Samples					
United Kingdom	David Bentley	Sanger Inst.	Genotyping: Illumina	24.0%	1, 6, 10, 13, 20	Wellcome Trust		
	Peter Donnelly	Univ. of Oxford	Analysis			TSC, US NIH		
	Lon Cardon	Univ. of Oxford	Analysis			Wellcome Trust, TSC, US NIH		
Canada	Thomas Hudson	McGill Univ. and G�enome Qu�ebec Innovation Centre	Genotyping: Illumina	10.0%	2, 4p	Genome Canada, G�enome Qu�ebec		
China	Huanming Yang The Chinese HapMap Consortium	Changqing Zeng	Beijing Genomics Inst., Chinese National Human Genome Center at Beijing	Genotyping: PerkinElmer and Sequenom	4.8%	10%	3, 8p, 21	Chinese MOST, Chinese Academy of Sciences, Natural Science Foundation of China, Hong Kong Innovation and Technology Commission, University Grants Committee of Hong Kong
		Wei Huang	Chinese National Human Genome Center at Shanghai, Inst. of Biomedical Sciences (Taiwan)	Genotyping: Illumina	3.2%			
		Lap-Chee Tsui	Univ. of Hong Kong, Hong Kong Univ. of Sci. & Tech., Chinese Univ. of Hong Kong	Genotyping: Sequenom	2.0%			
	Houcun Zhang	Beijing Normal Univ.	Community engagement					
	Changqing Zeng	Beijing Genomics Inst.	Samples			Chinese MOST		
United States	Mark Chee	Illumina	Genotyping: Illumina	15.5%	30.9%	8q, 9, 18q, 22, X	US NIH	
	David Altshuler	Whitehead Inst.	Genotyping: Sequenom and Illumina	9.1%		4q, 7q, 18p, Y		
			Analysis					
	Richard Gibbs	Baylor College of Medicine, ParAllele	Genotyping: ParAllele	4.4%	12			
	Pui-Yan Kwok	UCSF, Washington Univ.	Genotyping: PerkinElmer	1.9%	7p			
	Aravinda Chakravarti	Johns Hopkins Univ.	Analysis					
	Mark Leppert	Univ. of Utah	Community engagement Samples			W.M. Keck Foundation, Delores Dore Eccles Foundation, US NIH		
Nigeria	Charles Rotimi	Howard Univ., Univ. of Ibadan	Community engagement Samples			US NIH		
	Lincoln Stein	Cold Spring Harbor Lab., New York	Data coordination centre			TSC		

from the non-profit Coriell Institute of Medical Research (<http://locus.umdnj.edu/nigms/>); cell lines and DNA from the new samples will be available from Coriell in early 2004 for future studies with research protocols approved by appropriate ethics committees. It is anticipated that other researchers will genotype additional SNPs in these samples in the future, and that these data will continuously improve the HapMap.

These samples will have population and sex identifiers without information that could link them to individual donors. As the goal of the project is solely to identify patterns of genetic variation, no medical or other phenotypic information will be included. About 50% more samples were collected than will be used, so that inclusion of a sample from any particular donor cannot be known.

Samples of 45 unrelated individuals should be sufficient to find 99% of haplotypes with a frequency of 5% or greater in a population. Studies of LD can use random individual samples, trios or larger pedigrees; each design has advantages (ease of sampling) and disadvantages (decreasing efficiency with increasing numbers of related individuals). Analysis of existing data and computer simulations suggested that unrelated individuals and trios have considerable power for estimating local LD patterns. The trios will provide useful information on the accuracy of the genotyping platforms being used for the project.

Choice of SNPs

A high density of SNPs is needed to describe adequately the genetic variation across the entire genome. When the project started, the average density of markers in the public database dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>)⁵⁰ was approximately one every kilobase (2.8 million SNPs) but, given their variable distribution, many regions had a lower density of SNPs.

Further SNPs were obtained by random shotgun sequencing from whole-genome and whole-chromosome (flow-sorted) libraries⁵¹, using methods developed for the initial human SNP map⁵², and also by collaboration with Perlegen Sciences³⁶ and through the purchase of sequence traces from Applied Biosystems⁵³ for SNP detection (see Supplementary Information, part 2). One useful result of this search for more SNPs is the confirmation of SNPs found previously. SNPs for which each allele has been seen independently in two or more samples ('double-hit' SNPs) have a higher average minor allele frequency than do 'single-hit' SNPs²⁸. This leads to substantial savings in assay development. On 4 November 2003, the number of SNPs (with a unique genomic position) in dbSNP (build 118) was 5.7 million, and the number of double-hit SNPs was over 2 million. By February of 2004, 6.8 million SNPs (with a unique genomic position) are expected to be in dbSNP and available for the project, including 2.7 million double-hit SNPs.

As the extent of LD and haplotypes varies by 100-fold across the genome^{30–32,34,35}, a hierarchical genotyping strategy has been adopted. In an initial round of genotyping, the project aims to genotype successfully 600,000 SNPs spaced at approximately 5-kilobase intervals and each with a minor allele frequency of at least 5%, in the 270 DNA samples. Priority is being given to previously validated SNPs, double-hit SNPs and SNPs causing amino-acid changes (as these may alter protein function). When these genotyping data are produced (by mid-2004; see below for details of data release), they will be analysed for associations between neighbouring SNPs. Additional SNPs will then be genotyped in the same DNA samples at a higher density only in regions where the associations are weak. Further rounds of analysis and genotyping will be carried out as required. It is expected that more than one million SNPs will be genotyped overall. This hierarchical strategy will permit regions of the genome with the least LD to be characterized at densities of up to one SNP per kilobase, maximizing the characterization of regions with associations only over short distances.

Genotyping

Each genotyping centre is responsible for genotyping all the samples for all the selected SNPs on the chromosome regions allocated (Table 1). Among the centres, a total of five high-throughput genotyping technologies are being used, which will provide an opportunity to compare their accuracy, success rate, throughput and cost. Access to several platforms is an advantage for the project, as a SNP assay that fails on one platform may be developed successfully using another method in order to fill a gap in the HapMap. All platforms will be evaluated using a common set of performance criteria to ensure that the quality of data produced for the project meets a uniformly high standard.

Genotype quality is being assessed in three ways. First, at the beginning of the project, all centres were assigned the same randomly selected set of 1,500 SNPs for assay development and genotyping in the 90 CEPH DNA samples being used for the project. Genotyping centres produced data that were on average more than 99.2% complete and more than 99.5% accurate (as compared to the consensus of at least two other platforms). Second, every genotyping experiment includes samples for internal quality checks, with each 96-well plate containing duplicates of five different samples, and one blank. In addition, the data from trios provide a check for consistent mendelian inheritance of SNP alleles. For all the populations, the data from the unrelated samples provide a check that the SNPs are in Hardy–Weinberg equilibrium (a test of genetic mating

Box 1

Community engagement, public consultation and individual consent

As no personally identifiable information will be linked to the samples, the risk that an individual will be harmed by a breach of privacy, or by discrimination based on studies that use the HapMap, is minimal. However, because tag SNPs for future disease studies will be chosen on the basis of haplotype frequencies in the populations included in the HapMap, the data will be identified as coming from one of the four populations involved, and it will be possible to make comparisons between the populations. As a result, the use of population identifiers may create risks of discrimination or stigmatization, as might occur if a higher frequency of a disease-associated variant were to be found in a group and this information were then overgeneralized to all or most of its members⁶⁴. It is possible that there are other culturally specific risks that may not be evident to outsiders⁶⁵. To identify and address these group risks, a process of community engagement, or public consultation, was undertaken to confer with members of the populations being approached for sample donation about the implications of their participation in the project^{66,67}. The goal was to give people in the localities where donors were recruited the opportunity to have input into the informed consent and sample collection processes, and into such issues as how the populations from which the samples were collected would be named. Community engagement is not a perfect process, but it is an effort to involve potential donors in a more extended consideration of the implications of a research project before being asked to take part in it⁶⁸. Community engagement and individual informed consent were conducted under the auspices of local governments and ethics committees, taking into account local ethical standards and international ethical guidelines. As in any cross-cultural endeavour, the form and outcome of the processes varied from one population to another. A Community Advisory Group is being set up for each community to serve as a continuing liaison with the sample repository, to ensure that future uses of the samples are consistent with the uses described in the informed consent documents. A more detailed article discussing ethical, social and cultural issues relevant to the project, and describing the processes used to engage donor populations in identifying and evaluating these issues, is in preparation.

patterns). Although a small proportion of SNPs may fail these checks for biological reasons, they more typically fail if a genotyping platform makes consistent errors, such as undercalling heterozygotes. Third, a sample of SNP genotypes deposited by each centre will be selected at random and re-genotyped by other centres. These stringent third-party evaluations of quality will ensure the completeness and reliability of the data produced by the project.

Data release

The project is committed to rapid and complete data release, and to ensuring that project data remain freely available in the public domain at no cost to users. The project follows the data-release principles of a 'community resource project' (<http://www.wellcome.ac.uk/en/1/awtpubrepdat.html>).

All data on new SNPs, assay conditions, and allele and genotype frequencies will be released rapidly into the public domain on the internet at the HapMap Data Coordination Center (DCC) (<http://www.hapmap.org/>) and deposited in dbSNP. Individual genotype and haplotype data initially will be made available at the DCC under a short-term 'click-wrap' licence agreement. This strategy has been adopted to ensure that data from the project cannot be incorporated into any restrictive patents, and will thus remain freely available in the long term. The only condition for data access is that users must agree not to restrict use of the data by others and to share the data only with others who have agreed to the same condition. When haplotypes are defined in a region, then the individual genotypes, haplotypes and tag SNPs in that region will be publicly released to dbSNP, where there are no licensing conditions. Project participants have agreed that their own laboratories will access the data through the DCC and under the click-wrap licence, ensuring that all scientists have equal access to the data for research.

The consortium believes that SNP, genotype and haplotype data in the absence of specific utility do not constitute appropriately patentable inventions. Specific utility would involve, for example, finding an association of a SNP or haplotype with a medically important phenotype such as a disease risk or drug response. The project does not include any phenotype association studies. However, the data-release policy does not block users from filing for appropriate intellectual property on such associations, as long as any ensuing patent is not used to prevent others' access to the HapMap data.

Data analysis

The project will apply existing and new methods for analysis and display of the data. LD between pairs of markers will be calculated using standard measures such as D' (ref. 54), r^2 (refs 55, 56) and others. Various methods are being evaluated to define regions of high LD and haplotypes along chromosomes. Existing methods include 'sliding window' LD profiles^{57,58}, LD unit maps⁵⁹, haplotype blocks^{31,35} and estimates of meiotic recombination rates along chromosomes^{35,60–62}. After analysis of the LD in the first phase of the project, regions in which there is little or no LD will be identified and ranked for further SNP selection and genotyping. Methods to select optimal collections of tag SNPs will be developed and evaluated (see above). The project will thus provide views of the data and tag SNPs that will be useful to the research community. As all data and analysis methods will be made available, other researchers will also be able to analyse the data and improve the analysis methods.

To assist optimization of SNP selection and analysis of LD and haplotypes, a pilot study is underway to produce a dense set of genotypes across large genomic regions. Ten 500-kilobase regions of the genome (see Supplementary Information, part 3) will be sequenced in 48 unrelated HapMap DNA samples (16 CEPH (currently being sequenced), 16 Yoruba, 8 Japanese and 8 Han Chinese). All SNPs identified, as well as any additional SNPs in the public databases, will be genotyped in all of the 270 HapMap DNA

samples, and the genotype data will be released following the guidelines described above. This study will provide dense genotype data for developing methods for SNP selection and for assessing the completeness of the information extracted, to guide the later stages of genotyping.

When the HapMap is used to examine large genomic regions, the problem of multiple comparisons will arise from testing tens to hundreds of thousands of SNPs and haplotypes for disease associations. This will lead to difficulty in separating true from false-positive results. Thus, new statistical methods, replication studies and functional analyses of variants will be important to confirm the findings and identify the functionally important SNPs.

Conclusion

The goal of the International HapMap Project is to develop a research tool that will help investigators across the globe to discover the genetic factors that contribute to susceptibility to disease, to protection against illness and to drug response. The HapMap will provide an important shortcut to carry out candidate-gene, linkage-based and genome-wide association studies, transforming an unfeasible strategy into a practical one. In its scope and potential consequences, the International HapMap Project has much in common with the Human Genome Project, which sequenced the human genome⁶³. Both projects have been scientifically ambitious and technologically demanding, have involved intense international collaboration, have been dedicated to the rapid release of data into the public domain, and promise to have profound implications for our understanding of human biology and human health. Whereas the sequencing project covered the entire genome, including the 99.9% of the genome where we are all the same, the HapMap will characterize the common patterns within the 0.1% where we differ from each other.

For the full potential of the HapMap to be realized, several things must occur. The technology for genotyping must become more cost efficient, and the analysis methods must be improved. Pilot studies with other populations must be completed to confirm that the HapMap is generally applicable, with consideration given to expanding the HapMap if needed so that all major world populations can derive the greatest benefit. To use the tools created by the HapMap, later projects must establish carefully phenotyped sets of affected and unaffected individuals for many common diseases in a way that preserves confidentiality but retains detailed clinical and environmental exposure data. Longitudinal cohort studies of hundreds of thousands of individuals will also be invaluable for assessing the genetic and environmental contributions to disease.

Careful and sustained attention must also be paid to the ethical issues that will be raised by the HapMap and the studies that will use it. By consulting members of donor populations about the consent process and the implications of population-specific findings before sample collection, the project has helped to advance the ethical standard for international population genetics research. Future population genetics projects will continue to refine this approach. It will be an ongoing challenge to avoid misinterpretations or misuses of results from studies that use the HapMap. Researchers using the HapMap should present their findings in ways that avoid stigmatizing groups, conveying an impression of genetic determinism, or attaching incorrect levels of biological significance to largely social constructs such as race.

The HapMap holds much promise as a powerful new tool for discovery—to enhance our understanding of the hereditary factors involved in health and disease. Realizing its full benefits will involve the close partnership of basic science researchers, population geneticists, epidemiologists, clinicians, social scientists, ethicists and the public. □

doi:10.1038/nature02168.

1. King, R. A., Rotter, J. I. & Motulsky, A. G. *The Genetic Basis of Common Diseases* Vol. 20 (eds Motulsky, A. G., Harper, P. S., Scriver, C. & Bobrow, M.) (Oxford Univ. Press, Oxford, 1992).
2. Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).
3. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genet.* **33** (Suppl.), 228–237 (2003).
4. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
5. Dorman, J. S., LaPorte, R. E., Stone, R. A. & Trucco, M. Worldwide differences in the incidence of type 1 diabetes are associated with amino acid variation at position 57 of the HLA-DQ β chain. *Proc. Natl Acad. Sci. USA* **87**, 7370–7374 (1990).
6. Bell, G. I., Horita, S. & Karam, J. H. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* **33**, 176–183 (1984).
7. Nisticò, L. *et al.* The *CTLA-4* gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Hum. Mol. Genet.* **5**, 1075–1080 (1996).
8. Strittmatter, W. J. & Roses, A. D. Apolipoprotein E and Alzheimer's disease. *Annu. Rev. Neurosci.* **19**, 53–77 (1996).
9. Dahlbäck, B. Resistance to activated protein C caused by the factor V R³⁰⁶Q mutation is a common risk factor for venous thrombosis. *Thromb. Haemost.* **78**, 483–488 (1997).
10. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
11. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
12. Rioux, J. D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nature Genet.* **29**, 223–228 (2001).
13. Pennacchio, L. A. *et al.* An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* **294**, 169–173 (2001).
14. Deeb, S. S. *et al.* A Pro12Ala substitution in PPAR γ 2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nature Genet.* **20**, 284–287 (1998).
15. Altshuler, D. *et al.* The common PPAR γ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genet.* **26**, 76–80 (2000).
16. Stefansson, H. *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am. J. Hum. Genet.* **71**, 877–892 (2002).
17. Van Eerdewegh, P. *et al.* Association of the *ADAM33* gene with asthma and bronchial hyperresponsiveness. *Nature* **418**, 426–430 (2002).
18. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nature Genet.* **35**, 131–138 (2003).
19. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin- α gene that are associated with susceptibility to myocardial infarction. *Nature Genet.* **32**, 650–654 (2002).
20. Collins, F. S., Guyer, M. S. & Chakravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580–1581 (1997).
21. Kruglyak, L. & Nickerson, D. A. Variation is the spice of life. *Nature Genet.* **27**, 234–236 (2001).
22. Kerem, B. *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073–1080 (1989).
23. Hästbacka, J. *et al.* Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.* **2**, 204–211 (1992).
24. Li, W. H. & Sadler, L. A. Low nucleotide diversity in man. *Genetics* **129**, 513–523 (1991).
25. Wang, D. G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
26. Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
27. Halushka, M. K. *et al.* Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genet.* **22**, 239–247 (1999).
28. Reich, D. E., Gabriel, S. B. & Altshuler, D. Quality and completeness of SNP databases. *Nature Genet.* **33**, 457–458 (2003).
29. Pääbo, S. The mosaic that is our genome. *Nature* **421**, 409–412 (2003).
30. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
31. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
32. Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
33. Abecasis, G. R. *et al.* Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.* **68**, 191–197 (2001).
34. Dawson, E. *et al.* A first-generation linkage disequilibrium map of human chromosome 22. *Nature* **418**, 544–548 (2002).
35. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. High-resolution haplotype structure in the human genome. *Nature Genet.* **29**, 229–232 (2001).
36. Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
37. Johnson, G. C. L. *et al.* Haplotype tagging for the identification of common disease genes. *Nature Genet.* **29**, 233–237 (2001).
38. Carlson, C. S. *et al.* Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nature Genet.* **33**, 518–521 (2003).
39. Goldstein, D. B., Ahmadi, K. R., Weale, M. E. & Wood, N. W. Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet.* **19**, 615–622 (2003).
40. Taillon-Miller, P. *et al.* Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nature Genet.* **25**, 324–328 (2000).
41. Chakravarti, A. Population genetics—making sense out of sequence. *Nature Genet.* **21**, 56–60 (1999).
42. Chakravarti, A. *et al.* Nonuniform recombination within the human β -globin gene cluster. *Am. J. Hum. Genet.* **36**, 1239–1258 (1984).
43. Tishkoff, S. A. *et al.* Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**, 1380–1387 (1996).

44. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, 1994).
45. Foster, M. W. & Sharp, R. R. Race, ethnicity, and genomics: social classifications as proxies of biological heterogeneity. *Genome Res.* **12**, 844–850 (2002).
46. Barbujani, G., Magagnoli, A., Minch, E. & Cavalli-Sforza, L. L. An apportionment of human DNA diversity. *Proc. Natl Acad. Sci. USA* **94**, 4516–4519 (1997).
47. Rosenberg, N. A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
48. Jorde, L. B. *et al.* Microsatellite diversity and the demographic history of modern humans. *Proc. Natl Acad. Sci. USA* **94**, 3100–3103 (1997).
49. Dausset, J. *et al.* Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**, 575–577 (1990).
50. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
51. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
52. The International SNP Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
53. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
54. Lewontin, R. C. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* **49**, 49–67 (1964).
55. Hill, W. G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231 (1968).
56. Ohta, T. & Kimura, M. Linkage disequilibrium due to random genetic drift. *Genet. Res.* **13**, 47–55 (1969).
57. Dawson, K. J. The decay of linkage disequilibrium under random union of gametes: how to calculate Bennett's principal components. *Theor. Popul. Biol.* **58**, 1–20 (2000).
58. Langley, C. H. & Crow, J. F. The direction of linkage disequilibrium. *Genetics* **78**, 937–941 (1974).
59. Maniatis, N. *et al.* The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by haplotype analysis. *Proc. Natl Acad. Sci. USA* **99**, 2228–2233 (2002).
60. Hudson, R. R. Estimating the recombination parameter of a finite population model without selection. *Genet. Res.* **50**, 245–250 (1987).
61. Fearnhead, P. & Donnelly, P. Estimating recombination rates from population genetic data. *Genetics* **159**, 1299–1318 (2001).
62. McVean, G., Awadalla, P. & Fearnhead, P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241 (2002).
63. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
64. Clayton, E. W. The complex relationship of genetics, groups, and health: what it means for public health. *J. Law Med. Ethics* **30**, 290–297 (2002).
65. Foster, M. W. & Sharp, R. R. Genetic research and culturally specific risks: one size does not fit all. *Trends Genet.* **16**, 93–95 (2000).
66. Sharp, R. R. & Foster, M. W. Involving study populations in the review of genetic research. *J. Law Med. Ethics* **28**, 41–51 (2000).
67. Marshall, P. A. & Rotimi, C. Ethical challenges in community-based research. *Am. J. Med. Sci.* **322**, 241–245 (2001).
68. Juengst, E. T. Commentary: what "community review" can and cannot do. *J. Law Med. Ethics* **28**, 52–54 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank many people who contributed to this project: J. Beck, C. Beiswanger, D. Coppock, J. Mintzer and L. Toji at the Coriell Institute for Medical Research for transforming the samples, distributing the DNA and cell lines, and storing the samples for use in future research; J. Greenberg and R. Anderson of the NIH National Institute of General Medical Sciences (NIGMS) for providing funding and support for cell-line transformation and storage in the NIGMS Human Genetic Cell Repository at the Coriell Institute; K. Wakui at Shinshu University for assistance in transforming the Japanese cell lines; N. Carter and D. Willey at the Wellcome Trust Sanger Institute for flow sorting the chromosomes and for library construction, respectively; M. Deschesnes and B. Godard for assistance at the University of Montréal; C. Darmond-Zwaig, J. Olivier and S. Roumy at McGill University and Génome Québec Innovation Centre; C. Allred, B. Gillman, E. Kloss and M. Rieder for help in implementing data flow protocols; S. Olson for work on the website explanations; S. Adeniyi-Jones, D. Burgess, W. Burke, T. Citrin, A. Clark, D. Cowhig, P. Epps, K. Hofman, A. Holt, E. Juengst, B. Keats, J. Levin, R. Myers, A. Obuoforibo, F. Romero, C. Tamura and A. Williamson for providing advice on the project to NIH; A. Peck and J. Witonsky of the National Human Genome Research Institute (NHGRI) for help with project management; E. DeHaut-Combs and S. Saylor of NHGRI for staff support; M. Gray for organizing phone calls and meetings; the people of Tokyo, Japan, the Yoruba people of Ibadan, Nigeria, and the community at Beijing Normal University, who participated in public consultations and community engagements; and the people in these communities who were generous in donating their blood samples. This work was supported in part by Genome Canada, Génome Québec, the Chinese Ministry of Science and Technology, the Chinese Academy of Sciences, the Natural Science Foundation of China, the Hong Kong Innovation and Technology Commission, the University Grants Committee of Hong Kong, the Japanese Ministry of Education, Culture, Sports, Science and Technology, the Wellcome Trust, the SNP Consortium, the US National Institutes of Health (FIC, NCI, NCR, NEL, NHGRI, NIA, NIAAA, NIAID, NIAMS, NIBIB, NIDA, NIDCD, NIDCR, NIDDK, NIEHS, NIGMS, NIMH, NINDS, OD), the W.M. Keck Foundation and the Delores Dore Eccles Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.B. (drb@sanger.ac.uk) or M.F. (fost1848@mshmailhub.oulan.ou.edu).

The International HapMap Consortium (Participants are arranged alphabetically by institution, and then by principal investigator and alphabetically by last name.)

Genotyping centres: Baylor College of Medicine and ParAllele BioScience Richard A. Gibbs¹, John W. Belmont¹, Paul Hardenbol², Thomas D. Willis² & Fuli Yu¹; **Chinese HapMap Consortium** Huanming Yang³, Lan-Yang Ch'ang⁴, Wei Huang⁵, Bin Liu³, Yan Shen⁶, Paul Kwong-Hang Tam⁷, Lap-Chee Tsui⁸, Mary Miu Yee Waye⁹, Jeffrey Tze-Fei Wong¹⁰, Changqing Zeng³ & Qingrun Zhang³; **Illumina** Mark S. Chee¹¹, Luana M. Galver¹¹, Semyon Kruglyak¹¹, Sarah S. Murray¹¹ & Arnold R. Oliphant¹¹; **McGill University and Génomique Québec Innovation Centre** Alexandre Montpetit¹², Thomas J. Hudson¹², Fanny Chagnon¹², Vincent Ferretti¹², Martin Leboeuf¹², Michael S. Phillips¹² & Andrei Verner¹²; **University of California at San Francisco and Washington University** Pui-Yan Kwok¹³, Shenghui Duan¹⁴, Denise L. Lind¹³, Raymond D. Miller¹⁴, John P. Rice¹⁴, Nancy L. Saccone¹⁴, Patricia Taillon-Miller¹⁴ & Ming Xiao¹³; **University of Tokyo and RIKEN** Yusuke Nakamura^{15,16}, Akihiro Sekine¹⁶, Koki Sorimachi¹⁶, Toshihiro Tanaka¹⁶, Yoichi Tanaka¹⁶, Tatsuhiko Tsunoda¹⁶ & Eiji Yoshino¹⁷; **Wellcome Trust Sanger Institute** David R. Bentley¹⁸, Panos Deloukas¹⁸, Sarah Hunt¹⁸ & Don Powell¹⁸; **Whitehead Institute/MIT Center for Genome Research** David Altshuler^{19,20} & Stacey B. Gabriel¹⁹

Community engagement/public consultation and sample-collection groups: Beijing Normal University and Beijing Genomics Institute Houcan Zhang²¹ & Changqing Zeng³; **Health Sciences University of Hokkaido, Eubios Ethics Institute and Shinshu University** Ichiro Matsuda²², Yoshimitsu Fukushima²³, Darryl R. Macer²⁴ & Eiko Suda²⁴; **Howard University and University of Ibadan** Charles N. Rotimi²⁵, Clement A. Adebamowo²⁶, Toyin Anigwue²⁶, Patricia A. Marshall²⁷, Olayemi Matthew²⁶, Chibuzor Nkwodimma²⁶ & Charmaine D. M. Royal²⁵; **University of Utah** Mark F. Leppert²⁸ & Missy Dixon²⁸

Analysis Groups: Cold Spring Harbor Laboratory Lincoln D. Stein²⁹, Fiona Cunningham²⁹, Ardavan Kanani²⁹ & Gudmundur A. Thorisson²⁹; **Johns Hopkins University School of Medicine** Aravinda Chakravarti³⁰, Peter E. Chen³⁰, David J. Cutler³⁰ & Carl S. Kashuk³⁰; **University of Oxford** Peter Donnelly³¹, Jonathan Marchini³¹, Gilean A. T. McVean³¹ & Simon R. Myers³¹; **University of Oxford, Wellcome Trust Centre for Human Genetics** Lon R. Cardon³², Gonçalo R. Abecasis³³, Andrew Morris³² & Bruce S. Weir³⁴; **US National Institutes of Health** James C. Mullikin³⁵, Stephen T. Sherry³⁶ & Michael Feolo³⁶; **Whitehead Institute/MIT Center for Genome Research** David Altshuler^{19,20}, Mark J. Daly¹⁹ & Stephen F. Schaffner¹⁹

Affiliations for participants: 1, Baylor College of Medicine Human Genome Sequencing Center, Department of Molecular and Human Genetics, 1 Baylor Plaza, Houston, Texas 77030, USA; 2, ParAllele BioScience, 384 Oyster Point Boulevard, Suite 8, South San Francisco, California 94080, USA; 3, Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 100300, China; 4, Institute of Biomedical Sciences, 128 Yen-Jiou Yuan Road, Sec. 2, Taipei, Taiwan, 115; 5, Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China; 6, Chinese National Human Genome Center at Beijing, Yongchang North Road 3-707, Beijing Economy and Technology Development Zone, 100176, China; 7, University of Hong Kong, Genome Research Centre, 6/F, Laboratory Block, 21 Sassoon Road, Pokfulam, Hong Kong; 8, University of Hong Kong, 10/F, Knowles Building, Pokfulam Road, Hong Kong; 9, The Chinese University of Hong Kong, Department of Biochemistry, Room 608, 6/F Mong Man Wai Building, Shatin, Hong Kong; 10, Hong Kong University of Science and Technology, Department of Biochemistry, Clear Water Bay, Knowlton, Hong Kong; 11, Illumina, 9885 Towne Centre Drive, San Diego, California 92121, USA; 12, McGill University and Génomique Québec Innovation Centre, 740 Dr Penfield Avenue, Montréal, Québec H3A 1A4, Canada; 13, University of California, Cardiovascular Research Institute, 505 Parnassus Avenue Long 1332A, Box 0130, San Francisco, California 94143, USA; 14, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA; 15, University of Tokyo, Institute of Medical Science, 4-6-1 Sirokanedai, Minato-ku, Tokyo 108-8639, Japan; 16, RIKEN SNP Research Center, 1-7-22 Suehiro-cho, Tsurumi-ku Yokohama, Kanagawa 230-0045, Japan; 17, RIKEN Technology Transfer and Research Coordination Division, 2-1 Hirosawa, Wako, Saitama, Japan; 18, Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 19, Whitehead Institute/MIT Center for Genome Research, 1 Kendall Square, Cambridge, Massachusetts 02139, USA; 20, Massachusetts General Hospital, 50 Blossom Street, Wellman 831, Boston, Massachusetts 02114, USA; 21, Beijing Normal University, 19 Xijiekouwai Street, Beijing 100875, China; 22, Health Sciences University of Hokkaido, Ezuko Institution for Developmental Disabilities, Ezumachi 575, Kumamoto, Japan; 23, Shinshu University School of Medicine, Department of Medical Genetics, Matsumoto 390-8621, Japan;

Ethical, Legal and Social Issues: Chinese Academy of Social Sciences Renzong Qiu³⁷; **Genetic Interest Group** Alastair Kent³⁸; **Howard University** Georgia M. Dunston²⁵; **Kyoto University** Kazuto Kato³⁹; **Nagasaki University** Norio Niikawa⁴⁰; **University of Montréal** Bartha M. Knoppers⁴¹; **University of Oklahoma** Morris W. Foster⁴²; **Vanderbilt University** Ellen Wright Clayton⁴³ & Vivian Ota Wang⁴³; **Wellcome Trust** Jessica Watkin⁴⁴

SNP Discovery: Baylor College of Medicine Richard A. Gibbs¹, John W. Belmont¹, Erica Sodergren¹ & George M. Weinstock¹; **Washington University** Richard K. Wilson⁴⁵ & Lucinda L. Fulton⁴⁵; **Wellcome Trust Sanger Institute** Jane Rogers¹⁸; **Whitehead Institute/MIT Center for Genome Research** Bruce W. Birren⁴⁶

Scientific Management: Chinese Academy of Sciences Hua Han⁴⁷; **Chinese Ministry of Science and Technology** Hongguang Wang⁴⁸; **Genome Canada** Martin Godbout⁴⁹ & John C. Wallenburg⁵⁰; **Génomique Québec** Paul L'Archevêque⁵¹ & Guy Bellemare⁵¹; **Japanese Ministry of Education, Culture, Sports, Science and Technology** Kazuo Todani⁵², Takashi Fujita⁵³ & Satoshi Tanaka⁵²; **The SNP Consortium** Arthur L. Holden⁵⁴ & Eric H. Lai⁵⁵; **US National Institutes of Health** Francis S. Collins⁵⁶, Lisa D. Brooks⁵⁶, Jean E. McEwen⁵⁶, Mark S. Guyer⁵⁶, Elke Jordan⁵⁷, Jane L. Peterson⁵⁶, Jack Spiegel⁵⁸, Lawrence M. Sung⁵⁹ & Lynn F. Zacharia⁵⁶; **Wellcome Trust** Karen Kennedy⁴⁴, Michael G. Dunn⁴⁴, Richard Seabrook⁴⁴, Mark Shillito⁶⁰, Barbara Skene⁴⁴ & John G. Stewart⁴⁴

Initial Planning Groups: Populations and Ethical, Legal and Social Issues Group David L. Valle (chair)⁶¹, Ellen Wright Clayton (co-chair)⁴³, Lynn B. Jorde (co-chair)²⁸, John W. Belmont¹, Aravinda Chakravarti³⁰, Mildred K. Cho⁶², Troy Duster^{63,64}, Morris W. Foster⁴², Marla Jasperse⁶⁵, Bartha M. Knoppers⁴¹, Pui-Yan Kwok¹³, Julio Licinio⁶⁶, Jeffrey C. Long⁶⁷, Patricia A. Marshall²⁷, Pilar N. Ossorio⁶⁸, Vivian Ota Wang⁴³, Charles N. Rotimi²⁵, Charmaine D. M. Royal²⁵, Patricia Spallone^{44,69} & Sharon F. Terry⁷⁰; **Methods Group** Eric S. Lander (chair)⁴⁶, Eric H. Lai (co-chair)⁵⁵, Deborah A. Nickerson (co-chair)⁷¹, Gonçalo R. Abecasis³³, David Altshuler^{19,20}, David R. Bentley¹⁸, Michael Boehnke³³, Lon R. Cardon³², Mark J. Daly¹⁹, Panos Deloukas¹⁸, Julie A. Douglas⁶⁷, Stacey B. Gabriel¹⁹, Richard R. Hudson⁷², Thomas J. Hudson¹², Leonid Kruglyak⁷³, Pui-Yan Kwok¹³, Yusuke Nakamura^{15,16}, Robert L. Nussbaum⁷⁴, Charmaine D. M. Royal²⁵, Stephen F. Schaffner¹⁹, Stephen T. Sherry³⁶, Lincoln D. Stein²⁹ & Toshihiro Tanaka¹⁶

24, University of Tsukuba, Eubios Ethics Institute, P.O. Box 125, Tsukuba Science City 305-8691, Japan; 25, Howard University, National Human Genome Center, 2216 6th Street NW, Washington, District of Columbia 20059, USA; 26, University of Ibadan College of Medicine, Ibadan, Oyo State, Nigeria; 27, Case Western Reserve University School of Medicine, Department of Bioethics, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; 28, University of Utah, Eccles Institute of Human Genetics, Department of Human Genetics, 15 North 2030 East, Salt Lake City, Utah 84112, USA; 29, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA; 30, Johns Hopkins University School of Medicine, McKusick–Nathans Institute of Genetic Medicine, 600 North Wolfe Street, Baltimore, Maryland 21287, USA; 31, University of Oxford, Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, UK; 32, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; 33, University of Michigan, Department of Biostatistics, Center for Statistical Genetics, 1420 Washington Heights, Ann Arbor, Michigan 48109, USA; 34, North Carolina State University, Bioinformatics Research Center, Campus Box 7566, Raleigh, North Carolina 27695, USA; 35, US National Institutes of Health, National Human Genome Research Institute, 50 South Drive, Bethesda, Maryland 20892, USA; 36, US National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information, 8600 Rockville Pike, Bethesda, Maryland 20894, USA; 37, Chinese Academy of Social Sciences, Center for Applied Ethics, 2121, Building 9, Caoqiao Xinyuan 3 Qu, Beijing 100054, China; 38, Genetic Interest Group, 4D Leroy House, 436 Essex Road, London N1 3QP, UK; 39, Kyoto University, Institute for Research in Humanities, Ushinomiya-cho, Sakyo-ku, Kyoto 606-8501, Japan; 40, Nagasaki University Graduate School of Biomedical Sciences, Department of Human Genetics, Sakamoto 1-12-4, Nagasaki 852-8523, Japan; 41, University of Montréal, The Public Law Research Centre (CRDP), P.O. Box 6128, Downtown Station, Montréal, Québec H3C 3J7, Canada; 42, University of Oklahoma, Department of Anthropology, 455 West Lindsey Street, Norman, Oklahoma 73019, USA; 43, Vanderbilt University, Center for Genetics and Health Policy, 507 Light Hall, Nashville, Tennessee 37232, USA; 44, Wellcome Trust, 183 Euston Road, London NW1 2BE, UK; 45, Washington University School of Medicine, Genome Sequencing Center, Box 8501, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA; 46, Whitehead Institute/MIT Center for Genome Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA; 47, Chinese Academy of Sciences, 52 Sanlihe Road, Beijing 100864, China; 48, Chinese Ministry of Science and Technology, 15B Fuxing Road, Beijing 100862, China; 49, Genome Canada, 150 Metcalfe Street, Suite 2100, Ottawa, Ontario K2P 1P1, Canada; 50, McGill University, Office of Technology Transfer, 3550 University Street, Montréal, Québec H3A 2A7, Canada; 51, Génome Québec, 630 Boulevard René-Lévesque Ouest, Montréal, Québec H3B 1S6, Canada; 52, Ministry of Education, Culture, Sports, Science and Technology, 3-2-2 Kasumigaseki, Chiyodaku, Tokyo, Japan; 53, Hiraki and Associates, Toranomon No. 5 Mori Building, 17-1, Toranomon 1-Chome, Minato-Ku, Tokyo 105-0001, Japan; 54, The SNP Consortium, 3 Parkway North, Deerfield, Illinois 60015, USA; 55, GlaxoSmithKline, 5 Moore Drive, Research Triangle Park, North Carolina 27709, USA; 56, US National Institutes of Health, National Human Genome Research Institute, 31 Center Drive, Bethesda, Maryland 20892, USA; 57, Foundation for the National Institutes of Health, 1 Cloister Court, Bethesda, Maryland 20892, USA; 58, US National Institutes of Health, Office of Technology Transfer, 6011 Executive Boulevard, Rockville, Maryland 20852, USA; 59, University of Maryland School of Law, 500 West Baltimore Street, Baltimore, Maryland 21201, USA; 60, Herbert Smith, Exchange House, Primrose Street, London EC2A 2HS, UK; 61, Johns Hopkins University School of Medicine, Howard Hughes Medical Institute and the McKusick–Nathans Institute of Genetic Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA; 62, Stanford Center for Biomedical Ethics, 701A Welch Road, Palo Alto, California 94304, USA; 63, New York University, Department of Sociology, 269 Mercer Street, New York, New York 10003, USA; 64, University of California, Berkeley, Department of Sociology, 2420 Bowditch, Berkeley, California 94720, USA; 65, University of New Mexico Health Sciences Center, 214 East Nizhoni Boulevard, Gallup, New Mexico 87301, USA; 66, University of California, Los Angeles, School of Medicine, 695 Charles E. Young Drive South, Los Angeles, California 90095, USA; 67, University of Michigan, Department of Human Genetics, 1241 East Catherine Street, Ann Arbor, Michigan 48109, USA; 68, University of Wisconsin School of Law, 975 Bascom Mall, Madison, Wisconsin 53706, USA; 69, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK; 70, Genetic Alliance, 4301 Connecticut Avenue NW, Suite 404, Washington, District of Columbia 20008, USA; 71, University of Washington, Department of Genome Sciences, Box 357730, Seattle, Washington 98125, USA; 72, University of Chicago, Department of Ecology and Evolution, 1101 East 57th Street, Chicago, Illinois 60637, USA; 73, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA; 74, US National Institutes of Health, National Human Genome Research Institute, 49 Convent Drive, Bethesda, Maryland 20892, USA