

OPINION

## Integrating ethics and science in the International HapMap Project

*The International HapMap Consortium\**

Genomics resources that use samples from identified populations raise scientific, social and ethical issues that are, in many ways, inextricably linked. Scientific decisions about which populations to sample to produce the HapMap, an international genetic variation resource, have raised questions about the relationships between the social identities used to recruit participants and the biological findings of studies that will use the HapMap. The sometimes problematic implications of those complex relationships have led to questions about how to conduct genetic variation research that uses identified populations in an ethical way, including how to involve members of a population in evaluating the risks and benefits posed for everyone who shares that identity. The ways in which these issues are linked is increasingly drawing the scientific and ethical spheres of genomics research closer together.

For many biomedical research projects, coordinating the work on the study design and the associated ethical issues presents serious challenges. Frequently, ethical concerns are identified and addressed only after the research goals have been stated and the study has been designed. Alternatively, in large projects, a separate team of ethics advisers might work solely on that aspect of the study.

Here, we describe how in a multi-centre, multi-nation collaborative project — the **International HapMap Project** — ethical considerations were identified and addressed from its inception<sup>1</sup>. The HapMap Project is a major international research effort to construct a

resource to facilitate future studies that relate human genetic variation to health and disease. The Project raises many ethical issues because it will allow researchers to compare patterns of variation among both individuals and populations. Throughout the Project, ethicists and social scientists have worked collaboratively with geneticists to address these issues (members of the International HapMap Consortium are listed in BOX 1).

This article provides an overview of the ethical, social and cultural issues raised by the International HapMap Project and describes how the Project is addressing them. The specific processes that were used to engage communities or consult with the public in the localities where people were approached to donate blood samples, as well as the principal findings that emerged from those processes, will be described in separate publications.

### Scientific rationale

Any two humans are approximately 99.9% identical in their DNA sequences<sup>2</sup>, but the 0.1% by which they vary contributes to differences in their risk of getting certain diseases and their responses to drugs, infectious agents, toxins and other environmental factors<sup>3</sup>. Finding the genetic variants that influence disease risk and drug response is necessary to understand how genetic and environmental factors interact to influence health. Although the road from scientific discovery to improved health outcomes can be long, understanding these factors should eventually lead to better methods of prevention, diagnosis and treatment.

The scientific design of the International HapMap Project has been described elsewhere<sup>1</sup> (see online links box for information **about the project**). The Project's rationale is based on the finding that many sets of adjacent SNPs have been passed down through the generations largely intact, resulting in strong associations among SNP alleles (variants) in chromosomal regions<sup>4</sup>. A set of associated SNP alleles in a chromosomal region is known as a haplotype. Most regions have only a few common haplotypes, which account for most of the genetic variation in these regions. So, much of the information about patterns of variation in a region can be obtained by looking at just a few 'tag' SNPs in that region instead of all of them<sup>5-7</sup>.

The HapMap will describe the common patterns of human genetic variation by identifying the chromosomal regions with sets of strongly associated SNPs (the haplotypes) and the SNPs that identify (tag) them. Researchers will use the tag SNPs from the HapMap to compare a group of people with a disease (for example, diabetes) with a group of people without it. In most regions, both groups will have similar haplotype frequencies. However, regions with alleles that affect the disease will have a higher frequency of haplotypes associated with the disease in the affected group. As a result, researchers will be able to scan the entire genome by genotyping only approximately 500,000 tag SNPs, and not all 10 million common SNPs. Similarly, researchers will be able to examine regions of interest that are identified from other studies by looking at only a fraction of the SNPs in those regions. The HapMap will therefore greatly reduce the cost of whole-genome scans for ASSOCIATION STUDIES and will facilitate more targeted association, candidate-gene, FAMILY-LINKAGE and ADMIXTURE-MAPPING STUDIES<sup>8</sup>.

### Project history and organization

Public discussion of a human haplotype map began in 2001 at a meeting attended by genetics and genomics researchers, experts in the ethical, legal and social implications

Box 1 | **The International HapMap Consortium**

The International HapMap Consortium is a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to develop a public resource that will help researchers find genes that are associated with human health and disease. The following authors are members of this group (institutions and authors are listed alphabetically (except for chairs/co-chairs, who are listed first); author affiliations are detailed in the online [supplementary information S1](#) (box)):

**Ethical, Legal, and Social Issues Group**

Vanderbilt University: Ellen Wright Clayton (co-chair) | University of Montréal: Bartha M. Knoppers (co-chair) | Chinese Academy of Social Sciences: Renzong Qiu | Genetic Interest Group: Alastair Kent | Howard University: Georgia M. Dunston | Kyoto University: Kazuto Kato | Nagasaki University: Norio Niikawa | University of Oklahoma: Morris W. Foster | University of Ibadan: Isaac F. Adewole | Wellcome Trust: Jessica Watkin

**Community Engagement/Public Consultation and Sample Collection Groups**

Beijing Normal University and Beijing Genomics Institute: Houcan Zhang and Changqing Zeng | Health Sciences University of Hokkaido, Shinshu University and Eubios Ethics Institute: Ichiro Matsuda, Yoshimitsu Fukushima, Darryl R. Macer and Eiko Suda | Howard University and University of Ibadan: Charles N. Rotimi, Clement A. Adebamowo, Toyin Aniagwu, Patricia A. Marshall, Olayemi Matthew, Chibuzor Nkwodimmah and Charmaine D. M. Royal | University of Utah: Mark F. Leppert and Missy Dixon

**Populations and Ethical, Legal, and Social Issues Initial Planning Group**

David L. Valle (chair), Ellen Wright Clayton (co-chair), Lynn B. Jorde (co-chair), John W. Belmont, Aravinda Chakravarti, Mildred K. Cho, Troy Duster, Morris W. Foster, Marla Jasperse, Bartha M. Knoppers, Pui-Yan Kwok, Julio Licinio, Jeffrey C. Long, Patricia A. Marshall, Pilar N. Ossorio, Vivian Ota Wang, Charles N. Rotimi, Charmaine D. M. Royal, Patricia Spallone and Sharon F. Terry

**Genotyping centres**

Baylor College of Medicine and ParAllele BioScience: Richard A. Gibbs, John W. Belmont, Paul Hardenbol, Thomas D. Willis and Fuli Yu | Broad Institute: David Altshuler and Stacey B. Gabriel | Chinese HapMap Consortium: Huanming Yang, Lan-Yang Ch'ang, Wei Huang, Bin Liu, Yan Shen, Paul Kwong-Hang Tam, Lap-Chee Tsui, Mary Miu Yee Waye, Jeffrey Tze-Fei Wong, Changqing Zeng and Qingrun Zhang | Illumina: Mark S. Chee, Luana M. Galver, Semyon Kruglyak, Sarah S. Murray and Arnold R. Oliphant | McGill University and Génome Québec Innovation Centre: Alexandre Montpetit, Thomas J. Hudson, Fanny Chagnon, Vincent

Ferretti, Martin Leboeuf, Michael S. Phillips and Andrei Verner | University of California at San Francisco and Washington University: Pui-Yan Kwok, Shenghui Duan, Denise L. Lind, Raymond D. Miller, John P. Rice, Nancy L. Saccone, Patricia Taillon-Miller and Ming Xiao | University of Tokyo and RIKEN: Yusuke Nakamura, Akihiro Sekine, Koki Sorimachi, Toshihiro Tanaka, Yoichi Tanaka, Tatsuhiro Tsunoda and Eiji Yoshino | Wellcome Trust Sanger Institute: David R. Bentley, Panos Deloukas and Don Powell

**Analysis groups**

Broad Institute: David Altshuler, Mark J. Daly and Stephen F. Schaffner | Cold Spring Harbor Laboratory: Lincoln D. Stein, Fiona Cunningham, Ardavan Kanani and Gudmundur A. Thorisson | Johns Hopkins University School of Medicine: Aravinda Chakravarti, Peter E. Chen, David J. Cutler, Carl S. Kashuk and Shin Lin | University of Oxford: Peter Donnelly, Jonathan Marchini, Gilean A. T. McVean and Simon R. Myers | University of Oxford/Wellcome Trust Centre for Human Genetics: Lon R. Cardon, Gonçalo R. Abecasis, Andrew Morris and Bruce S. Weir | US National Institutes of Health: James C. Mullikin, Stephen T. Sherry and Michael Feolo

**SNP discovery**

Baylor College of Medicine: Richard A. Gibbs, John W. Belmont, Erica Sodergren and George M. Weinstock | Broad Institute: Bruce W. Birren | Washington University: Richard K. Wilson and Lucinda L. Fulton | Wellcome Trust Sanger Institute: Jane Rogers

**Scientific management**

Chinese Academy of Sciences: Hua Han | Chinese Ministry of Science and Technology: Hongguang Wang | Genome Canada: Martin Godbout and John C. Wallenburg | Génome Québec: Paul L'Archevêque and Guy Bellemare | Japanese Ministry of Education, Culture, Sports, Science and Technology: Kazuo Todani, Takashi Fujita and Satoshi Tanaka | The SNP Consortium: Arthur L. Holden and Eric H. Lai | US National Institutes of Health: Francis S. Collins, Jean E. McEwen, Vivian Ota Wang, Lisa D. Brooks, Mark S. Guyer, Elke Jordan, Jane L. Peterson, Jack Spiegel, Lawrence M. Sung and Lynn F. Zacharia | Wellcome Trust: Karen Kennedy, Michael G. Dunn, Richard Seabrook, Mark Shillito, Barbara Skene and John G. Stewart

**Methods Initial Planning Group**

Eric S. Lander (chair), Eric H. Lai (co-chair), Deborah A. Nickerson (co-chair), Gonçalo R. Abecasis, David Altshuler, David R. Bentley, Michael Boehnke, Lon R. Cardon, Mark J. Daly, Panos Deloukas, Julie A. Douglas, Stacey B. Gabriel, Richard R. Hudson, Thomas J. Hudson, Leonid Kruglyak, Pui-Yan Kwok, Yusuke Nakamura, Robert L. Nussbaum, Charmaine D. M. Royal, Stephen F. Schaffner, Stephen T. Sherry, Lincoln D. Stein and Toshihiro Tanaka

(ELSI) of genetics and genomics research, consumers, community members and funding agency representatives (see online links box for [2001 HapMap agenda and participants](#)). Although most participants agreed that haplotype mapping would help speed up the discovery of genes that affect at least some diseases, others expressed concerns about the ethical and social implications of such an initiative.

A recommendation was made to proceed but with a commitment to integrate ethical, social and cultural considerations into all phases of the Project. To this end, two initial

planning groups were established: a Methods Group to consider the technical requirements and a Populations/ELSI Group to consider ethical and sampling issues (see online links box for [2001 HapMap working groups roster](#)). The latter group included genomics and population genetics researchers and ELSI experts. The inclusion of ethicists in the initial meeting and the immediate formation of the combined working groups to decide fundamental questions of study design are examples of the deliberate integration of scientific and ethical considerations from the beginning of the Project.

The Populations/ELSI Group was given responsibility for two interrelated scientific and ethical questions: how to sample human genetic variation to identify common haplotypes, and whether to name the populations from which the donors came. Initially, Group members were from the United Kingdom, Canada and the United States — the initial countries that gave financial support to the Project. After Japan and China had secured financial support, both the Methods Group and the Populations/ELSI Group were succeeded by new groups with members from each participating country.

The current ethics group, the International HapMap ELSI Group, makes recommendations to the Project's Steering Committee (which includes some members of the ELSI Group). Some ELSI Group members also serve on the Project's Communications Group to ensure that attention is paid to ethical issues in Project publications and that the descriptions of the scientific design and findings reflect sensitivity to social and cultural issues. In some of the countries that are involved, committees that make recommendations to the funding agencies also oversee the project. Local ethics committees had final authority over the way the samples were collected.

### Sampling strategy

Most evidence indicates that modern humans appeared in Africa more than 150,000 years ago<sup>9</sup>. Some of the descendants of this group remained in Africa, whereas others migrated, eventually reaching all parts of the world. Much of the genetic variation that is present in the original ancestral population in Africa was therefore carried into the descendant populations and is now in all world populations<sup>10</sup>. Any one population includes approximately 90% of the genetic variation that is present throughout the world<sup>11,12</sup>.

Population expansions, BOTTLENECKS, FOUNDER EFFECTS and natural selection have influenced the frequency of variants and haplotypes in populations in different parts of the world<sup>12,13</sup>. So, most common haplotypes are expected to be found in all human populations, but the frequencies of particular haplotypes will vary among populations<sup>4</sup>. Mutation and recombination have created new haplotypes since the migration out of Africa and, therefore, some variants and haplotypes are found only in individuals with ancestry from particular geographical regions.

Population differences in haplotype frequencies are important for discovering genes that are related to health and disease. For example, an association study can have false-positive results if haplotypes in chromosomal regions other than those with the causal alleles differ in frequency between individuals with the disease and healthy controls<sup>14</sup>; this problem often occurs when different proportions of cases and controls have been recruited from different populations. In addition, populations that have experienced recent bottlenecks might have longer haplotype lengths, potentially making initial identification of associated chromosomal regions more efficient<sup>15</sup>. On the other hand, shorter haplotypes (as is often seen in African populations, reflecting their longer history) might allow finer identification of the causative regions.

### Box 2 | Individual inclusion criteria

Sample donors had to be adults (as legally defined in each country) and competent to provide informed consent. The Yoruba participants were required to have four Yoruba grandparents (as identified by the donor). The Han Chinese participants were required to have at least three Han grandparents. In Japan, where it is considered culturally insensitive to ask donors specifically about the origins of their grandparents, individuals wishing to donate were simply told that the aim was to collect samples from people whose grandparents were from Japan.

Asking individuals to identify themselves, or their grandparents, as having a particular ancestry will inevitably result in some imprecise or inaccurate assignments. In fact, as in many population genetics studies, the ways in which the populations themselves were defined for the Project were inexact. For example, the Centre d'Etude du Polymorphisme Humain (CEPH) samples were collected in 1980, when most researchers did not consider it necessary to ask potential donors how they self-identified or where their ancestors originated; it was merely generally known that most individuals in that community were of primarily northern and western European descent.

The lack of precise population definitions and ancestral geographical self-assignments is not problematic for developing the HapMap because exact demarcations are not necessary for the way that the HapMap will be used, and the Project does not aim to define populations. No claims are made about the genetic 'purity' of the sample sets or the populations of which donors are members; such claims would be scientifically spurious, as human populations are the products of countless social, historical and demographic processes, and therefore can never be sharply defined.

Because the goal of the HapMap Project is to develop a tool that will help researchers to discover the genetic contributors to health and disease, a population-sampling strategy was chosen to maximize the downstream benefits of the Project for all populations — both sampled and un-sampled populations. For the purposes of the Project, a 'population' is defined as a group of people with a shared ancestry and therefore a shared history and pattern of geographical migration. (It is important to note, however, that on the basis of this definition, many individuals can claim membership of more than one population, while some of those who claim a particular population identity do not share the same biological ancestry; this situation makes doing genetic variation research with identified populations both scientifically and ethically complex.) Although the globally common haplotypes could have been identified with samples from any one population, studying samples from several populations with different ancestral geographies, reflecting different population histories, should make the HapMap most useful for studies in many populations.

For both scientific and ethical reasons, Project planners recommended choosing one or more populations with ancestry from Africa, Asia and Europe, and at the same time keeping open the possibility of adding populations with ancestry from other parts of the world at a later date. The results of an initial study of samples that had already been collected from the Yoruba people (from Nigeria), Japanese and Chinese individuals, and residents of the United States (Utah residents with ancestry from northern and western Europe,

collected in 1980 by the Centre d'Etude du Polymorphisme Humain, and known as the CEPH samples) showed substantial similarities in the haplotype patterns of these populations, but differences in the frequencies and lengths of many haplotypes<sup>4</sup>. This indicated that a HapMap developed with samples from these populations or from others with similar geographical ancestry would probably include much of the common genetic variation that is present in the world, along with some more regionally specific variation.

As part of a separate project, the US National Institutes of Health (NIH) are currently arranging sample collection from several other populations, including two in Africa, one in Europe and several in the United States that have complex histories of admixture. These samples will be genotyped across a few genomic regions to see how useful the HapMap will be for a wider range of populations. If any of these other populations have haplotype patterns that are substantially different from those in the populations initially studied to develop the HapMap, the samples from those populations could be analysed across the genome, which will become feasible as genotyping costs continue to decline. Over time, the HapMap might also be augmented with data from other investigators.

### Choosing the communities

On the basis of the recommendation of the working groups that the HapMap be developed, at least initially, with samples from populations with African, Asian and European ancestry, the Project is studying 270 DNA samples from 4 populations: 30 trios (two parents and an adult child) from the Yoruba people of

Ibadan, Nigeria; 45 unrelated Japanese in the Tokyo area; 45 unrelated Han Chinese in Beijing; and 30 trios from the Utah population represented by the CEPH samples (BOX 2). (Although Japanese and Chinese samples in the pilot study had similar haplotype frequencies, Japanese and Han Chinese samples are both being included because of the interest of funding agencies in Japan and China to use samples from their own majority populations.) Although the samples come from large populations with imprecise boundaries, they are appropriate for the Project because the purpose is to create a resource that can be used in populations throughout the world, and not to 'define' particular populations or to study population relatedness.

The HapMap could have been developed with some of the same sets of stored samples that were analysed in the already-mentioned preliminary study by Gabriel *et al.*<sup>4</sup>, or with stored samples that had already been collected from other populations; however, the consent processes that were used for most previously collected samples (except for the CEPH samples) were judged to be inadequate for the HapMap Project because they had not included discussions about sharing samples with other investigators, about samples being used for genetic variation research that was not disease-specific or about the possibility that such research might raise group-based concerns. So, new samples from Nigeria, Japan and China were collected specifically for the Project, with community-consultation and consent processes that addressed those issues. For the CEPH, all living donors of the previously collected samples were contacted to obtain consent to have their samples included in the Project, and the local Institutional Review Board (IRB) gave permission to include samples from deceased CEPH donors.

Including the CEPH samples reflected researchers' desire to build on the foundation of many earlier studies that used those samples, including the human genetic linkage map<sup>16</sup>. In addition, the Utah researchers who collected the CEPH samples had maintained long-term trust relationships with most donors, making it feasible to contact them about the Project.

Because the HapMap could be developed scientifically with samples from any populations in Africa and Asia, the decisions about which specific communities to approach for the new samples that were required by the Project were based on ethical and practical considerations. Researchers at Howard University in Washington DC and at the University of Ibadan in Nigeria had already established research collaborations and had built a

relationship of trust with the Yoruba people from the Aba Alamu community in Ibadan. Chinese investigators collected samples in the Beijing Normal University residential community because it represents a diverse, yet socially cohesive, population that is composed of mostly Han people from nearly every province in China who have a range of educational, occupational and socioeconomic backgrounds. In addition, many members of that academic community were familiar with research and research ethics that formed the foundation for discussions about the HapMap Project. Japanese researchers collected samples from five different communities in Tokyo (drawing people from many parts of Japan) in which participants were accustomed to being recruited for biomedical studies.

---

**“Although there are differences among populations ... it is important that the findings of the HapMap Project not be over-simplified to perpetuate social and historical stereotypes.”**

---

#### Identifying the populations

Although the HapMap will include no personal identifiers or medical information about sample donors, each sample will be identified by the population from which it came. The scientific rationale for identifying the populations is that differences in haplotype frequencies and lengths among populations will be important for how data from the Project are used. In chromosomal regions in which the populations have similar haplotype frequencies, the HapMap and tag SNPs will be the same for all the populations; in regions in which the populations differ in haplotype frequencies, the HapMap and tag SNPs might differ among the populations. Having the population information will make it possible to choose the most efficient sets of tag SNPs to use in future association studies in particular populations. For example, knowing that one particular set of HapMap samples came from the Japanese would inform which tag SNPs might be most useful in a drug-response study in people of Japanese descent or in people from other populations found to have similar haplotype frequencies.

From an ethical point of view, removing population identifiers could create a false sense of protection from collective risks. As the researchers and institutions involved in the HapMap Project are named in grants and publications, it would be easy to guess the populations from which donors were recruited. It would also not be difficult to discern from previously collected data sets the identity of these populations. Rather than allow donors to assume that their population identities were protected or allow other researchers to infer those identities (and construct their own, perhaps inaccurate, interpretations of how those identities relate to genetic variation), naming the populations was thought to be more ethically appropriate. This approach gave donors the ability to evaluate the implications of inclusion and to have some input into how they wanted their population named. Naming the populations will also allow HapMap researchers and ethicists to provide better context for others for interpreting the biological significance of genetic findings that are associated with particular population identities.

Naming the populations does, however, have important ethical and social ramifications. For example, the HapMap will make it more efficient to study population history and make inferences about population relatedness. It would be impractical, and even undesirable, to limit the type of research that can be done with the HapMap to 'purely' biomedical studies, because population-history research can be useful biomedically and because the line between biomedical and population-history research is often imprecise. For the particular populations that are included in the HapMap, as for most large, loosely defined populations, the fact that the HapMap will facilitate population-history studies raises few concerns. For others, however, such as many American-Indian tribes and small isolated indigenous groups, population-history findings from genetic studies of members' samples could conflict with religious or cultural understandings about their origins, or legal or political claims that relate to land or items of cultural patrimony<sup>17</sup>. For these groups, the decision to construct the HapMap with samples from named populations could signal renewed emphasis on a disfavoured aspect of genetic variation research.

Indeed, it is mainly for these reasons that no samples from members of American-Indian tribes are included among those being initially analysed for the Project. In 2003, the National Human Genome Research Institute (NHGRI) convened a meeting with several leaders in the

American–Indian health-research community to explore the extent of interest among American–Indians in participating in this type of research. Most of the attendees were not interested in tribal participation in such a study at this time, citing concerns that the HapMap will facilitate population–history studies and comparisons among populations. They did, however, suggest that it might be appropriate to reconsider participation later, on the basis of what is learned from studies of haplotype patterns in other populations and the likelihood that discoveries of genes associated with diseases and drug responses in those populations would or would not benefit people of American–Indian ancestry.

Despite the many non-biological factors that contribute to population identities, the way that a population is labelled in the HapMap and described in publications will have implications for all members of the population, as all of them (and all members of closely related populations) might be affected by the interpretation and use of findings of future studies that use the HapMap<sup>18</sup>. If, for example, future studies lead to the discovery of genetic variants associated with obesity, the frequencies of those variants could be determined for each population in the HapMap sample sets. If a higher frequency of obesity-associated variants were found in the samples from one population and this information was then erroneously applied to all or most of its members and to members of closely related populations, entire populations could be stigmatized or suffer discrimination<sup>19</sup>, especially in places where individuals with ancestry from those populations are a minority.

This risk of group stigmatization is inherent in any study of samples from identified populations. Nevertheless, the limitations and ambiguities of population identifiers must continually be emphasized. For example, the individuals sampled from the residential community at Beijing Normal University do not represent all people in China, where there are 56 officially recognized ethnicities. Nor do the people sampled in Ibadan, Nigeria, represent all Africans or even all Yoruba people. Such limitations will be noted explicitly in Project publications that report the study's findings, and researchers who do future studies with these samples or with Project data will also need to be aware of these complexities when designing and reporting their studies. Although there are differences among populations in the frequencies of some genetic variants, it is important that the findings of the HapMap Project not be over-simplified to perpetuate social and historical stereotypes.

### Informed consent and privacy

ELSI experts and geneticists developed a general template for documents to obtain informed consent from the donors of the new samples and for re-consent from the living donors of the CEPH samples. The consent forms were then modified by the researchers who interacted with the communities to make them culturally appropriate before being submitted for approval by local ethics committees. The new samples were collected with population and sex identifiers, but without links to individual donors. Although the collectors of the CEPH samples do retain links to individuals, which made it possible to seek the re-consent to have the samples used for the HapMap, these links are held in strict confidence and will not be shared with HapMap Project researchers or users, or with the repository where the samples are stored.

---

**“Community consultation for the HapMap was done under the auspices of local governments and ethics committees, taking into account international and local ethical guidelines.”**

---

The HapMap will include no medical or other phenotypic information about the sample donors. Such information is unnecessary because the HapMap will simply be a description of patterns of genetic variation; it cannot be used by itself to find disease-associated genes. To further diminish individual privacy risks, more samples were collected from members of each population than were made into cell lines or used so that no one — not even the sample donors — can know whether any particular person's sample was actually used to develop the HapMap.

Nonetheless, extensive genomic information about each individual donor will eventually be contained in the HapMap database, in the form of millions of SNP genotypes for each person. Moreover, the data will be accessible to anyone with an Internet connection through the HapMap web site (see online links box) and through the National Center for Biotechnology Information public database, dbSNP (see online links box). However, it will be extremely difficult for anyone to link any genomic data in the

HapMap database to a specific person. This could happen in only two ways. One is if someone thought a certain person's data were included in the HapMap database, and if they obtained blood or another tissue sample from that person, genotyped it and compared the information with the data in the HapMap database. Alternatively, a match could be discovered if somebody compared the information in the HapMap database with genetic information known to be from a donor whose data were already in another database.

The risk that individual sample donors will suffer a breach of privacy or discrimination on the basis of their genetic information, although not absent, is very small. As genotyping costs decline and individuals are more commonly genotyped for clinical purposes, it might one day become more feasible to identify individual donors whose genotypes are part of public databases such as the HapMap. It is, however, unlikely to be worth the cost and effort to try to obtain personal information in this way.

### Community engagement

Because the HapMap will allow comparisons of patterns of genetic variation among populations, the initial Populations/ELSI Group recognized that some meaningful process of community engagement, or public consultation (the term preferred in Japan), would be necessary before individual informed consent could be obtained, and that this process should continue after the samples were collected. Whereas a ‘population’ refers to a group of individuals who have a common geographical ancestry, a ‘community’ is a group with a multitude of local units of social organization within a population. Importantly, individuals might consider themselves to belong to many communities (both the collective of individuals at Beijing Normal University and the residents of the area where samples were collected in Japan include people from many places in each country) or to share a broader identity that subsumes many communities (such as the ‘Han Chinese’, ‘Yoruba’ or ‘Japanese’ identities).

The need for meaningful community involvement in any type of biomedical research with named populations has become increasingly recognized<sup>20–26</sup>. Acknowledgement of this need grew noticeably following discussion of the proposed Human Genome Diversity Project, an effort intended to sample human genetic variation globally for various scientific purposes — primarily anthropological — that generated considerable controversy<sup>27,28</sup>. Much of the controversy arose as a result of misunderstandings brought about by insufficient community involvement<sup>29</sup>.

Extensive discussions have taken place in the international bioethics community about the need for community involvement in biomedical research with named populations. The need for community engagement is especially strong for genetic variation research such as the HapMap, which is not focused on any particular disease<sup>30,31</sup>; for this reason, it was made a central part of the Consortium plan for the HapMap project<sup>1</sup>. International ethical guidelines now support the active involvement of communities in such studies<sup>32–36</sup>. The Human Genetic Cell Repository of the NIH National Institute of General Medical Sciences, at the non-profit Coriell Institute for Medical Research, where the HapMap samples are stored, strongly recommends (and requires for US populations) some form of community consultation before it will accept samples from identified populations for its genetic variation panels (see online links box for [Human Genetic Cell Repository submission information](#)).

Although the opinions of an entire population cannot be generalized from the views expressed by a small set of individuals in one specific locality<sup>22</sup>, the goal of community engagement or public consultation for the HapMap Project was more circumscribed: that was, to give people in the communities that were being approached for participation an opportunity to share with investigators their views on the ethical, social and cultural issues that the Project raises for them and their communities, and to provide some input into the way their samples would be collected and described.

Community engagement is distinct from community consent, in which community leaders, or even an entire community, can veto a research project<sup>37</sup>. For example, American–Indian tribes, as sovereign nations, require formal tribal approval (a form of community consent) before investigators can recruit research participants within tribal jurisdictions or specifically because they are tribal members<sup>38</sup>. Although this requirement is not present for any communities that were approached for the HapMap, had significant opposition in a particular community been encountered, no samples would have been collected there. In Ibadan, Nigeria, where there are organized community structures, and community leaders must be consulted before research is done, those processes were recognized and followed.

Community consultation for the HapMap was done under the auspices of local governments and ethics committees, taking into account international and local ethical guidelines<sup>32,33,35,36,39–43</sup>. Separate articles will describe in greater detail the community-consultation processes in each of the localities and what was learned from them. Community engagement and sample collection in Nigeria were supported by the NHGRI and carried out by US investigators with collaborators in Nigeria. In China and Japan, although the funding agencies had not initially budgeted for community-consultation activities, flexibility was found to allocate funds for this purpose. All of the community-engagement or public-consultation teams included individuals with expertise in genetics, bioethics and social science.

The principal investigators for each community-consultation team are members of the Project's Steering Committee so that information flows between the researchers and the participating localities.

### Limitations and problems

Community consultation has many limitations<sup>44</sup>. The form and outcome of community consultation for the HapMap Project varied by location, and practical problems arose that required flexibility. In Nigeria, it took more than six months to obtain the necessary ethics committee approvals from each collaborating institution, which reduced the time available for pre-sample-collection community-engagement activities; follow-up activities, including interviews, focus groups, town meetings and a community survey, continued after the samples were collected. In China, the occurrence of the SARS epidemic in the middle of the process similarly required compressing many of the community-engagement activities into a shorter time frame, although investigators still completed all of the planned activities, including numerous interviews and focus groups, and received input from a range of people from many sectors of society. In Japan, a funding agency's concerns that the study would proceed for more than a year without any Japanese samples being studied (genotyping had begun on the previously collected CEPH samples while community engagement and sample collection proceeded at the three other sites) led to some streamlining of the process during the pre-sample collection phase. On the other hand, some concerns were expressed in Japan that the need to obtain the samples on the same timetable as the Yoruba and Han Chinese samples did not allow optimal time to engage the public fully<sup>45</sup>. In fact, at all the sites where new samples were collected, tension occurred between the need for adequate time to engage communities and the need to obtain samples on the timetable driven by scientific and funding practicalities.

Nevertheless, the genotyping centres and funding agencies made considerable compromises to accommodate the needs of the community-consultation processes. The community work increased the Project's expense. The genotyping centres had to wait more than one and a half years after the Project's official launch to receive DNA from all the samples. This delay required them to alter aspects of their original scientific plan, it increased the genotyping costs and created new and unexpected challenges. For example, the delay in sample availability resulted in much genotyping being done on the CEPH samples (which

### Box 3 | Sample storage

Cell lines were made from the blood samples and will be maintained at the National Institute of General Medical Sciences [Human Genetic Cell Repository](#) at the non-profit Coriell Institute (see online links box), which currently houses the world's largest collection of human cell lines available for general research. The cell lines and DNA from the samples will be made available to academic, government and commercial researchers around the world who have protocols that are approved by relevant ethics committees and that are determined by the Coriell Institute's Institutional Review Board to be consistent with the terms of the informed consent documents. A duplicate set of the Han Chinese samples will be retained in China, as required by Chinese law.

The samples stored at the Coriell Institute will be of enormous interest to researchers for many years in many types of study. The consent forms authorize a range of such studies, including population-history and cell line gene-expression research. Some such studies might raise risks that cannot be foreseen — a point made in the consent documents and discussed with donors. In general, however, the risks associated with such future research are unlikely to be different in kind from those raised by the HapMap itself. For example, as the samples are not associated with any medical information, they cannot be used by themselves to find genes related to potentially stigmatizing diseases or traits.

As no personal identifiers are included with any of the new samples (and the links to the Centre d'Etude du Polymorphisme Humain donors' identities are not available to the repository), it will be essentially impossible to contact donors to seek their re-consent for future studies. The absence of personal identifiers will also make it, in effect, impossible for individual donors to withdraw their samples.

Box 4 | **Sample collection among the Yoruba**

It was important from a scientific point of view to include in the International HapMap Project at least one population that had recent African ancestry, as these populations contain more genetic variation than other world populations. It would have been much cheaper and easier for the Project to have studied a set of African samples that had already been collected, to have sought samples from one of the large US African immigrant communities or to have sought samples from African-Americans (who have haplotypes and haplotype frequencies similar to the Yoruba)<sup>4</sup>. These approaches, however, would have raised ethical issues of their own. Collecting samples in Nigeria was therefore not an instance of exploiting a vulnerable population to get samples that were too difficult to obtain elsewhere. The decision was also informed by international guidelines that specifically endorse the inclusion of developing countries in genetics research<sup>33,36,46</sup>.

Donors in Nigeria were each given an equivalent of ~US \$8.00 and multivitamins worth ~US \$4.00 to compensate them for their time and travel — a standard amount for participation in research involving blood draws in that part of Africa. Prospective donors were not told that they would be compensated until after they had arrived to donate, to guard against the possibility that they would be induced to participate by the prospect of material benefit. The Project also provided some immediate benefits to the participating community. Staff on the local research team received training in research ethics issues, including procedures to strengthen community processes for obtaining informed consent. Funds were provided for further primary and preventive health-care services in the community, and more substantial support for local health-care infrastructure and equipment to enhance local research capabilities is being negotiated with the Community Advisory Group.

Consistent with international guidelines stating that any knowledge generated from research that is done in a resource-poor country should be made reasonably available for the benefit of its people<sup>33</sup>, the HapMap Project data will be placed in the public domain and will be accessible to researchers around the globe. These researchers will include, among others, investigators who study genetic factors that influence individual risk for infectious diseases, such as malaria and tuberculosis, that remain high-priority health concerns in Nigeria and other resource-poor countries. Although translating the understanding of these genetic factors into improved health outcomes is probably going to be a long, arduous process, the benefits would not be possible in the absence of research.

did not have to be newly collected) before genotyping on the other samples could begin. This created the risk of misperception on the part of the public that higher priority was being given to the CEPH samples when this disparity reflected merely the ethical necessity to take more time for community consultation where new samples were being collected. The delay in the availability of the Yoruba samples (which are expected to show shorter-range associations among SNPs and greater haplotype diversity than those from the other populations) also delayed the development of 'stopping rules' for the Project, which specify when adequate coverage for all the populations will be achieved.

Whether the extra time, effort and expense incurred from the engagement and consultation processes will be counterbalanced by increased public trust remains to be seen. But the experience so far indicates that asking people respectfully about participating in projects of this type, providing complete, balanced and accurate information, giving them a chance to express their views, and (where possible), incorporating their input, need not unduly impede research. Indeed, it can create a climate in which research proceeds in an atmosphere of openness and trust<sup>26</sup>.

**Community advisory groups**

Careful guidelines on sample storage and access were an integral part of the Project, as they have important research and ethical implications (see BOX 3 for a summary).

In anticipation of the possibility that concerns might arise about future uses of the samples despite detailed guidelines in the consent forms, a Community Advisory Group (CAG) was established in each community where new samples were collected. Each CAG will function as a liaison between the community from which the samples were collected and the Coriell Institute. The Coriell Institute will not distribute samples to investigators if the proposed research is inconsistent with the terms of the donors' consent, and will consult the CAGs if any proposed use of the samples raises questions. The CAG in each locality will hold periodic meetings. The Coriell Institute will provide up to US \$1,000 per year per site to defray expenses, and will produce quarterly reports and an annual newsletter describing how the HapMap and the samples are being used.

With this structure in place, it seems improbable that a community would seek to withdraw its samples from the repository except in extraordinary circumstances.

However, pursuant to a written policy of the Coriell Institute, if a CAG wishes to withdraw its community's samples, in response to the views of a substantial portion of the community and after careful consideration, that request will be honoured. The genotype data already in the HapMap database, however, could not be withdrawn (even if removed from the Project web site or the public SNP database, dbSNP), because they would already have been widely distributed.

The purpose of the CAGs is to ensure that the community-engagement or public-consultation processes continue after the samples are collected. The effectiveness of the CAG mechanism, similar to the effectiveness of the community-engagement and public-consultation processes, can only be assessed over time.

**Profits, patents and data release**

It is hoped that the HapMap Project will eventually benefit the health of all people. Most of the benefits, however, will not be immediately apparent, and some might take years to materialize. So, in the short term, the main beneficiaries will not be sample donors, their families or their communities, but researchers, who will gain professional rewards and companies, that will be able to develop drugs, diagnostic tests or other commercial products from research using the HapMap. No commercial products will be developed as part of the Project, however, as the HapMap is merely a resource that catalogues the common patterns of genetic variation. So, although future studies that use information from the HapMap might generate profit, the HapMap Project itself will not do so. In addition, the Coriell Institute does not allow investigators to commercialize the samples.

**Glossary****ADMIXTURE MAPPING STUDY**

Mapping genes that affect a phenotype on the basis of the linkage disequilibrium generated in a population that is formed by admixture between groups that differ in allele frequencies and the frequency of the phenotype.

**ASSOCIATION STUDY**

A set of methods that are used to correlate polymorphisms in genotype to polymorphisms in phenotype in populations.

**BOTTLENECKS**

A temporary reduction in population size that might cause the loss of genetic variation.

**FAMILY-LINKAGE STUDY**

A study that examines DNA sequence variants in families having multiple members with a disease, to map the genomic location of genes that affect the disease.

**FOUNDER EFFECTS**

A relatively high frequency of an allele in a population because it was founded by a small set of individuals who had the allele at a higher frequency than in the parent population.

The Project has adopted an interim protective strategy to try to ensure that no restrictive patents are filed by researchers who use HapMap Project data. The individual genotype and haplotype data are initially being made available under a click-wrap licence that states that users will agree to not reduce others' access to the data and to share the data only with others who have made the same agreement (see online links box for information on [genotype access registration](#)); when the Project is over (estimated to be at the end of 2005), all of the genotype and haplotype data will be publicly released. Other data, such as SNP allele and genotype frequencies, will be publicly released soon after they are obtained (at the Project website and the public database, dbSNP; see online links box). Project researchers will not seek patents on the data that they generate for which they have not demonstrated a specific use (such as relating a particular haplotype to a disease) and will not use the Project data for other projects in their laboratories before the data are released.

### Benefits and reciprocity

Each genotyping centre has been assigned particular chromosomal regions to genotype in all four sets of samples. Researchers in three of the countries where samples were collected or where re-consent was obtained (that is, Japan, China and the United States) will benefit by participating in the genotyping. Also, the sample collection that took place in each country was supported by funding agencies in that country, except for Nigeria, where that work was funded by the NHGRI; although local investigators engaged the community and collected the samples, no local investigators will be involved in the genotyping. See BOX 4 for a discussion of why, given these circumstances, it was appropriate to collect samples in Nigeria.

Another question is whether spending US \$120 million to create the HapMap is ethically justified when much of the world's population lacks access to basic health care. The hope, as with other research investments, is that the expenditure will eventually prove to have been well justified in terms of its benefits to world health. But understanding patterns of genetic variation and finding health-related genetic variants are merely initial steps; researchers must still discover how genes work and interact with environmental factors in the disease process and then must find a way of translating that knowledge into better health outcomes. It is essential that the HapMap and studies that use it are not 'over-hyped' or used to reinforce mistaken ideas of genetic determinism.

### Conclusions

As medical research scales up case-control and population-based studies that aim to identify genetic and environmental contributions to complex diseases, ethical questions will grow in number and complexity. Those questions will involve privacy, consent and other issues of concern to individual participants, issues related to the populations to which those individuals belong and even potential effects on non-participating populations. In some instances, addressing risks and benefits to populations might be as important as weighing these concerns for individuals.

Although the process we have described here represents what we believe to be a useful starting point for similar future studies, we recognize that it is only a small step towards doing more culturally sensitive genetics research. Those who follow no doubt will improve on these procedures. Ethical standards for the protection of participants in research are continually evolving and require experiences such as the one we have described to explain the philosophical and practical foundations on which those standards are based. Only in this way can bioethics proceed as both an empirical and a philosophical undertaking.

\*Members of the International HapMap Consortium are listed in Box 1.

Correspondence to Morris W. Foster at the University of Oklahoma, Department of Anthropology, 455 W. Lindsey, Room 520, Norman, Oklahoma 73019-0535, USA.  
e-mail: [fost1848@msmailhub.oulan.ou.edu](mailto:fost1848@msmailhub.oulan.ou.edu)

doi:1038/nrg1351

- International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
- The International SNP Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
- King, R. A., Rotter, J. I. & Motulsky, A. G. *The Genetic Basis of Common Diseases (Oxford monographs on medical genetics No. 20)* (eds Motulsky, A. G., Harper, P. S., Scriver, C. & Bobrow, M.) (Oxford Univ. Press, Oxford, 1992).
- Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
- 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).
- Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
- Dawson, E. *et al.* A first-generation linkage disequilibrium map of human chromosome 22. *Nature* **418**, 544–548 (2002).
- 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).
- Lonjou, C. *et al.* Linkage disequilibrium in human populations. *Proc. Natl Acad. Sci. USA* **100**, 6069–6074 (2003).
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, New Jersey, 1994).
- 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).
- Rosenberg, N. A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
- Jorde, L. B. *et al.* Microsatellite diversity and the demographic history of modern humans. *Proc. Natl Acad. Sci. USA* **94**, 3100–3103 (1997).
- Cardon, L. R. & Palmer, L. J. Population stratification and spurious allelic association. *Lancet* **361**, 598–604 (2003).
- Kidd, J. R. *et al.* Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. *Am. J. Hum. Genet.* **66**, 1882–1899 (2000).
- Dausset, J. *et al.* Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**, 575–577 (1990).
- Sharp, R. R. & Foster, M. W. An analysis of research guidelines on the collection and use of human biological materials from American Indian and Alaskan Native communities. *Jurimetrics* **42**, 165–186 (2002).
- King, P. A. *Gene Mapping: Using Law and Ethics As Guides* (eds Annas, G. J. & Elias, S.) 94–111 (Oxford Univ. Press, Oxford, New York, 1992).
- 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).
- Greeley, H. T. The control of genetic research: involving the 'groups between'. *Houst. L. Rev.* **33**, 1397–1430 (1996–1997).
- Reilly, P. R. Rethinking risks to human subjects in genetic research. *Am. J. Hum. Genet.* **63**, 682–685 (1998).
- Foster, M. W. *et al.* The role of community review in evaluating the risks of human genetic variation research. *Am. J. Hum. Genet.* **64**, 1719–1727 (1999).
- Juengst, E. T. Commentary: what 'community review' can and cannot do. *J. Law Med. Ethics* **28**, 52–54 (2000).
- Sharp, R. R. & Foster, M. W. Involving study populations in the review of genetic research. *J. Law Med. Ethics* **28**, 41–51 (2000).
- Weijer, C. & Emanuel, E. J. Protecting communities in biomedical research. *Science* **289**, 1142–1144 (2000).
- Marshall, P. A. & Rotimi, C. Ethical challenges in community-based research. *Am. J. Med. Sci.* **322**, 241–245 (2001).
- Knoppers, B. M., Hirtle, M. & Lormeau, S. Ethical issues in international collaborative research on the human genome: the HGP and the HGDP. *Genomics* **34**, 272–282 (1996).
- Macer, D. R. Ethical opportunities offered by the Human Genome Diversity Project. *Politics Life Sciences* **18**, 325–327 (1999).
- Reardon, J. The human genome diversity project: a case study in coproduction. *Soc. Stud. Sci.* **31**, 357–388 (2001).
- Committee on Human Genome Diversity, National Research Council. *Evaluating human genetic diversity* (National Academy of Sciences, Washington DC, 1997).
- Knoppers, B. M. (ed.) *Populations and Genetics: Legal and Socio-Ethical Perspectives* (Martinus Nijhoff, Leiden, Boston, 2003).
- Human Genome Organisation (HUGO). *Statement on the Principled Conduct of Genetics Research* [online]. <<http://www.gene.ucl.ac.uk/hugo/conduct.htm>> (1996).
- Council for International Organizations of Medical Sciences (CIOMS). *International Ethical Guidelines for Biomedical Research Involving Human Subjects* [online]. <[http://www.cioms.ch/frame\\_guidelines\\_nov\\_2002.htm](http://www.cioms.ch/frame_guidelines_nov_2002.htm)> (2002).
- Quebec Network of Applied Genetic Medicine (RMGA), Deschênes M., Gardinal G., Knoppers B. M. & Laberge C. Statement of Principles on the Ethical Conduct of Human Genetic Research Involving Populations. *Res. Health* **30**, 1–4 (2003).
- United Nations Educational, Scientific and Cultural Organization (UNESCO). Universal Declaration on the Human Genome and Human Rights. *J. Med. Philos.* **23**, 334–341 (1998).
- United Nations Educational, Scientific and Cultural Organization (UNESCO). *International Declaration on Human Genetic Data* [online]. <[http://portal.unesco.org/en/ev.php?URL\\_ID=17720&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/en/ev.php?URL_ID=17720&URL_DO=DO_TOPIC&URL_SECTION=201.html)> (2003).
- Juengst, E. T. Groups as gatekeepers to genomic research: conceptually confusing, morally hazardous, and practically useless. *Kennedy Inst. Ethics J.* **8**, 183–200 (1998).
- Bowekaty, M. B. & David, D. S. Cultural issues in genetic research with American Indian and Alaskan Native people. *IRB* **25**, 12–15 (2003).
- World Medical Association (WMA). in *Ethics and Research on Human Subjects, International Guidelines, Proceedings for the XXVth CIOMS Conference* (eds Bankowski, Z. & Levine, R. J.) 278–281 (CIOMS, Geneva, 1993).
- People's Republic of China, Ministry of Science and Technology and the Ministry of Public Health. *Interim Measures for the Administration of Human Genetic Resources* [online]. <<http://www.ebnic.org/interim.htm>> (1998).
- Human Genome Organisation (HUGO). *Statement on DNA Sampling: Control and Access* [online]. <<http://www.gene.ucl.ac.uk/hugo/sampling.html>> (1998).



42. Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) *Council for Science and Technology. Fundamental Principles of Research on the Human Genome* [online], <[http://www.mext.go.jp/a\\_menu/shinkou/shisaku/principles.htm](http://www.mext.go.jp/a_menu/shinkou/shisaku/principles.htm)> (2000).
43. The Japanese Ministry of Education, Culture, Sports, Science and Technology, the Japanese Ministry of Health, Labour and Welfare, and the Japanese Ministry of Economy, Trade and Industry *Ethical Guidelines for Analytical Research on the Human Genome/Genes* [online], <<http://www.biol.tsukuba.ac.jp/~macer/eghgr.htm>> (2001).
44. Juengst, E. T. Group identity and human diversity: keeping biology straight from culture. *Am. J. Hum. Genet.* **63**, 673–677 (1998).
45. Macer, D. R. Ethical considerations in the HapMap Project: an insider's personal view. *Eubios J. Asian Int. Bioeth.* **13**, 125–127 (2003).
46. Human Genome Organisation (HUGO). *Statement on Benefit Sharing* [online], <<http://www.gene.ucl.ac.uk/hugo/benefit.html>> (2000).

#### Acknowledgements

Full details of acknowledgements are given in the online [supplementary information S2](#) (box). We thank many people who contributed to addressing the ethical, social and cultural issues in this project: J. Greenberg, R. Anderson, J. Beck and the staff of the Coriell Institute, M. Inaba, H. Zhao, Y. Wang, W. Hu, H. Zhao, Y. Gao, Q. Zhang, Y. Zheng, D. Guan, W. Jiang, J. Li, Z. Li, W. Luo, K. Shen, X. Zhou, Y. Li, X. Feng, J. Ren, M. Deschênes, B. Godard, S. Adeniyi-Jones, D. Burgess, W. Burke, T. Citrin, D. Cowhig, P. Epps, K. Hofman, A. Holt, E. Juengst, J. Levin, A. Obuoforibo, F. Romero, C. Tamura, Y. Wang, S. Olson, A. Peck, J. Witonsky, E. DeHaut-Combs, S. Saylor, M. Gray, 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, Génom

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, NCRR, NEI, 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.

#### Online links

##### FURTHER INFORMATION

###### Coriell Cell Repositories:

<http://arginine.umdnj.edu/ccr/corsumm.html>

###### Coriell Cell Repositories NIGMS Human Genetic Cell

###### Repository submission information:

<http://locus.umdnj.edu/nigms/comm/submit/collpolicy.html>

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

###### HapMap — about the project:

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

###### HapMap — genotype access registration:

<http://www.hapmap.org/cgi-perl/registration>

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

###### National Human Genome Research Institute — past

###### events 2001 HapMap agenda and participants:

<http://www.genome.gov/10001685>

###### National Human Genome Research Institute — past

###### events 2001 HapMap working groups roster:

<http://www.genome.gov/10001685>

##### SUPPLEMENTARY INFORMATION

See online article: S1 (box) | S2 (box)

Access to this links box is available online.

during the experiment, which ended not long before the Community itself was dissolved in 1880. The Community's business enterprises became a joint stock company, which still exists as Oneida Ltd, an important multinational manufacturer of cutlery and flatware<sup>7</sup>. Even today, some descendants of the Community members live around the site of their original home, some continuing to work for the company.

The children of the breeding experiment — the stirpicults — were carefully followed up and were deemed to be superior in intellect, educational and professional achievement and physique<sup>8</sup>. The experiment and its outcome were presented by two Community members at the Second International Congress of Eugenics in 1921 (REF. 9). This large international gathering represented a high watermark for eugenics movements. However, despite the report on the Oneida experiment at such an important conference, it had little or no influence on subsequent eugenic policies and practices. I suggest that this was because of the difficulties in generalizing from the special context in which the experiment took place.

#### Oneida, a utopian community

Since its first colonial days, the United States has had many utopian communities — communities in which people came together to live and work, cooperatively bound together by shared beliefs and commitments<sup>10</sup>. The first communities were religious and were founded by groups who set out to live their chosen life in freedom in the New World, fleeing religious persecution in Europe (BOX 1).

In contrast to earlier communities, Oneida was a home-grown religious group, the members of which were mainly at least second-generation European-Americans. Its inspiration was religious, but several of the ideas for their social organization came from secular roots in other communities, especially from the phalanxes of Fourier (BOX 1).

The founder of Oneida, John Humphrey Noyes (1811–1886), came from an educated background in Vermont, USA. Initially destined for a career in law, he was converted at a revival meeting and went to Yale Theological Seminary to train as a preacher. Having left the seminary, he began to espouse the heretical doctrine of PERFECTIONISM, and his licence to preach was revoked. Perfectionists believed that through striving and communion with Christ, a sinless perfection was attainable in this life<sup>11</sup>. This allowed a heaven on Earth to be established, and John Humphrey Noyes became the prophet of that unfolding kingdom. As the Bible says, “My

#### TIMELINE

## A nineteenth-century experiment in human selective breeding

Martin Richards

A human selective breeding programme took place in a North American bible communist community, Oneida, between 1869 and 1879. It was probably the first such breeding experiment of the modern era, and for this reason, we might expect it to have been influential for the subsequent eugenics movements. Although it attracted much attention in its day, its longer-term influence seems to have been surprisingly slight, largely because its specific context meant that it was not a model that provided an acceptable way to reach eugenics goals.

Francis Galton, generally regarded as the father of eugenics (literally meaning ‘well born’), coined the term in 1883 and defined it as “the study of the agencies under social control that improve or impair the racial qualities of future generations either physically or mentally” according to a later Eugenics Society poster<sup>1–3</sup>. He contributed substantially to that

study through his own research on the families of men who were outstanding in intellect and in public life, and argued that intellectual abilities were mainly a matter of inheritance<sup>4,5</sup>. But, the idea that intellectual ability and other mental characteristics were inherited and that there could be dangers of transmitting hereditary ‘taints’ already concerned medical men and others much earlier in the nineteenth century<sup>6</sup>. It is these beliefs and the example of the effectiveness of selective breeding in animals that provided the backdrop for the Oneida selective breeding experiment.

Oneida, an upper New York State Christian utopian community, was founded in 1848. In the early years, the Community grew largely through recruitment, but once it was well established and financially secure, the decision was taken to have more children and to breed them scientifically in a ‘stirpiculture’ experiment, the Community's term for selective breeding. Some 58 children were born