Ethical considerations for investigators proposing samples for inclusion in the 1000 Genomes Project

1. Background

The 1000 Genomes Project aims to provide a resource of genetic variation that includes all but the rarest variants, to support genome-wide association studies and other studies relating genetic variation to health and disease. A major contributor to the rapid progress in human genomics has been the broad use of data made possible by public data release. Thus, to make the resource as useful as possible for future studies, the data on sequence and genetic variants for individual samples will be released publicly on the Internet. More than 1000 samples from unrelated members of 11 populations already exist as part of the extended set of HapMap samples and many of these will be studied by this project; however, additional samples will be needed from some populations.

The Samples/ELSI Group of the 1000 Genomes Project is considering the criteria that will be used to decide which additional populations and samples are needed for the project, and will coordinate their collection. Sample donors should be consented so that the individual sequence and variation data may be released publicly on the Internet, and so that cell lines can be generated and made easily available to researchers without restrictions on their use. Samples must be made available through an institution with the capability to perform the functions of a public repository, such as the Coriell Institute for Medical Research or other appropriate organization.

This document describes the ethical considerations for investigators collecting samples for inclusion in the 1000 Genomes Project, including the elements of informed consent considered essential for the project. The project Steering Committee will decide which sets of samples will be included in the project. All sets of samples used by the project must be collected in the manner described in this document, with input from the 1000 Genomes Project to ensure that the guidelines are interpreted correctly and that the right numbers of samples are collected.

To accompany this document, a template for an informed consent form has been developed for use by investigators who are interested in collecting samples for the 1000 Genomes Project. **The template is intended to be a starting point for investigators** but as described in more detail below, **it may need to be modified , depending on local circumstances**.

The guidelines in this document and the accompanying informed consent template may also be useful, with modifications, for investigators who plan to collect samples for use in other projects that will involve: (1) the generation of extensive, individual-level genomic data (including whole-genome sequence data) from sample donors from identified populations; (2) the deposition of those data in databases available on the Internet; (3) public data sharing; and (4) broad sharing of samples.

2. Appropriate populations and communities for sampling

The 1000 Genomes Project will not include samples from populations or communities that are determined to be particularly vulnerable. For example, small, isolated populations may make participants vulnerable to identification, stigmatization, or invasion of privacy. Other populations may be vulnerable due to factors such as socioeconomic status. Determinations regarding vulnerability will be made by the project Steering Committee in consultation with the Samples/ELSI Group and other relevant experts.

All sample collections should be done with the appropriate approvals of the institutions of the researchers doing the collection and any other institutions that will be directly involved in the sampling. In some cases it may be also necessary to secure other approvals (e.g., licenses from appropriate regulatory agencies to have samples transported outside the country).

3. The informed consent process

Informed consent involves two essential components: a document (consent form) and a process. The consent *form* provides a description of the research project (including the purpose of the research, the study procedures, potential benefits and risks, alternatives to participation, the voluntary nature of participation, etc.) and explains the individual's rights as a research participant. It is just one part of the informed consent *process,* which generally involves a series of conversations between the research team, the participants, and sometimes the community of the participants. The informed consent process provides potential sample donors with extensive information about the project to help them make educated decisions about whether to participate. It frequently also involves ongoing interaction with the sample donors once they have enrolled, and sometimes with communities. Informed consent is thus an ongoing, interactive process-- not a one-time information session followed by the signing of a document.

4. Purpose of the informed consent template

The template for a consent form that accompanies this document is provided as a framework to assist or inform investigators. **It is not intended to be prescriptive,** but should rather be considered to be a starting point, to provide a guide to the basic elements to include. Each investigator will need to customize the document to fit local circumstances. In addition, many institutions have special requirements (for example, regarding the inclusion of particular standardized language on certain topics); these requirements may also necessitate some changes to the structure or content.

The template that accompanies this document is consistent with, and intended to be complementary to, applicable international ethical guidelines for biomedical research and with applicable research regulations in all of the countries currently participating in the project (the U.K., China, Canada, and the U.S.). However, investigators may still need to

modify the template to accord with applicable local laws, regulations, policies, or cultural norms.

This template is written at a reading level that may be appropriate for some populations but too high for others. **Investigators are encouraged to explore ways to shorten or simplify the language in the consent form, so long as this can be done without omitting essential elements or other important information.**

5. Elements of the consent form

A. Elements required by law

Samples to be considered for inclusion in the 1000 Genomes Project must be collected according to the laws that apply to the researchers who collect the samples, including the laws of the country in which the samples will be collected. In addition, both the consent form and the consent process must satisfy the research regulations of all other countries participating in the project (currently, the U.K., China, Canada, and the U.S.). An international compilation of human research protections, which lists the laws, regulations, and guidelines governing human subjects research in these countries, as well as guidelines from a number of international and regional organizations, can be found at: http://www.hhs.gov/ohrp/international/. However, investigators should also consult with appropriate local experts to ensure that their understanding of the applicable regulatory requirements is accurate and up-to-date.

B. Other essential elements

Consent forms used to collect samples for the 1000 Genomes Project must also include discussion of the following topics:

- **Scope of the research and access to data and samples,** including:
  - the plans to develop cell lines from the samples (so that an unlimited amount of the donor's DNA will become available that may last indefinitely);
  - the plans to generate extensive genomic data about individuals, including whole-genome sequence data and cellular phenotypes such as expression and proteomic studies in the cell lines;
  - the plans to deposit the data in open access (public) scientific databases available on the Internet, so that researchers in universities, hospitals, the government, and companies can use the data;
  - the possibility that the data in the databases will be used to study questions related not only to health and disease but also to various non-medical traits, and that both individual and group comparisons will be made, which may lead to stigmatization risks for both individuals and groups;
  - the possibility that the data in the databases will be used to study questions relating to population history (if this type of research may be of particular

concern in communities where samples are being collected, the associated psychosocial risks should also be described); and
    - the opportunity that will be provided to many investigators around the world to study both the data and the samples for a wide range of studies of genetic variation (without donors being able to "pre-approve" each study).

- **Commercialization,** including the possibility that researchers in companies may study the data and the samples and that if any commercially valuable products result from these studies, donors will not receive any of the profits.

- **Return of results and withdrawal,** including:
    - the inability to return specific research results to individual donors;
    - the mechanisms that will be used to distribute information about general findings from the research or to return information to the communities that provided samples; and
    - the inability of donors to withdraw data from the study of their samples from the databases once the samples have been studied.

6. Other considerations

Depending on the circumstances, discussion of additional topics during the consent process or in the consent form may also be appropriate. For example:

- **New versus existing samples.** The template is designed to be used for the collection of *new samples*. Investigators who propose to obtain re-consent from donors who are already enrolled in existing studies will need to modify the template accordingly.

- **Identifiers and phenotypes.** The template is designed to be used for the collection of samples *without* names, other traditional identifiers, or medical information attached. Investigators who propose to collect samples along with names, other traditional identifiers, medical information, or other phenotypic information (or who propose to obtain re-consent, which means that names must already be available) will need to modify the template accordingly. In particular, investigators in such cases should discuss specifically:
    - How the confidentiality of the identifiers and (if applicable) medical information will be maintained (U.S. investigators should consider applying for a Certificate of Confidentiality from the appropriate NIH institute as an additional way to protect donors' identities from compelled disclosure in a legal proceeding; see http://grants.nih.gov/grants/policy/coc/index.htm);
    - How future re-contact of donors will be handled; and
    - How the right to withdraw samples and data will be handled. If identifiers are maintained so that it would be easy for the investigator to determine which sample came from which person, the opportunity to withdraw

samples should be allowed; however, donors should be told that they cannot withdraw the data from the study of their sample once the data are in the databases.

Note that samples will be used for the 1000 Genomes Project only in cases where either no medical phenotype data exist or where, if they exist, they are available only to the investigators who collected the samples and their close collaborators. Although no names, other traditional identifying information, or medical phenotypes will be included in the database, information about the sex and population group (or, where appropriate, language group) will be included in the database.

- **Samples from legally competent adult donors.** The template is designed to be used in connection with the collection of samples from *legally competent adult donors,* and the 1000 Genomes Project will include samples only from such donors.

- **Samples from unrelated individuals versus trios or family samples.** The template is designed to be used in connection with the collection of samples from *unrelated individuals*. Since sequencing is expensive, and redundant sequence information makes the analyses harder (except where trio samples are studied for validation), it is important that close relatives of donors not be included in the samples studied for the project. Including language in the consent form about how the project will use samples only from unrelated donors may discourage donations from individuals who are biologically related to each other. Collecting samples from mother-father-adult child trios, where feasible, may be valuable for future studies, but investigators who propose to collect samples from such trios instead of from singletons will need to modify the template accordingly.

- **Sample repository.** The template assumes that samples will be retained and that they will be deposited in a public repository, such as the Coriell Institute or other appropriate repository. (Investigators who wish to send their samples to the Coriell Institute should contact us to discuss the possibility of arranging for the shipping, transformation, and distribution of cell lines and DNA, and to discuss the handling of associated costs.) Regardless of where the samples will be stored, the consent form should specify where the samples will be kept, how they will be distributed, and what mechanisms will be used to provide feedback to donors (or donor communities) about how they are being used. (At the Coriell Institute, this can in some cases be done through Community Advisory Groups; however, different models may be appropriate for different communities.)

- **Number of samples.** The template states that more samples will be collected than will be used. This is intended as an additional protection of privacy, so that neither the sample donors nor the researchers collecting the samples know for sure whether any individual's sample is among the final set of samples

studied from that population. For the 1000 Genomes Project, the number of samples that will be needed from each population will be decided by the Steering Committee after discussion with the Samples and ELSI Group and the Analysis Group.

- **Inclusion of additional topics.** Depending on local circumstances, it may in some cases be advisable to address additional topics in the consent form. For example, depending on the cultural characteristics of certain communities, it may be appropriate to include additional language about what types of research the samples may or may not be used for (e.g., that the samples may be used for studies of population history, but will not be used for human cloning).

7. Beyond individual informed consent: Considering the appropriateness of community engagement/consultation

The 1000 Genomes Project, like other research on human genetic variation, may have implications not only for the individuals who donate samples, but also for the broader communities and populations of which they are a part. This is because the research involves the potential for comparing allele frequencies among groups whose ancestors came from different geographic regions, sometimes in a context where societal, racial, or ethnic discrimination exists.

Thus, in some, but not all, situations, in addition to obtaining informed consent from individual sample donors, it may be appropriate to conduct a process of community engagement or community consultation. The goal of community engagement or consultation is to give a broad range of members of the communities approached for the donation of samples an opportunity to:

- obtain information about the project so that the decisions of individuals whether to donate samples will be better informed;
- share their views about the ethical, social, and cultural issues the project raises for them, their immediate communities, and the broader communities and populations of which they are a part;
- provide input into such matters as how the samples from their locality will be collected and described;
- remain informed about how the data and the samples are being used and about findings from future studies based on the data and the samples.

In some cases it may also be appropriate to establish a Community Advisory Group (CAG) or similar body to provide ongoing feedback about the project and about how the community's samples are being used (see above). For more information on the CAG approach (as developed by NHGRI and the Coriell Institute for the International HapMap Project), see http://ccr.coriell.org/Sections/Support/NHGRI/nhgriImpInfo.aspx?PgId=389&coll=HG.

Final decisions regarding the need (if any) for community consultation or engagement in a particular situation, and the form it should take, will be made jointly by the investigator and the 1000 Genomes Project Steering Committee, with input from the Samples/ELSI Group and in consultation with experts in the ethical, legal, and social implications (ELSI) of human genetic research in the area where sampling is being proposed. The appropriateness of establishing a Community Advisory Group will be considered on a case- by-case basis.

8. Contact Us

Investigators who are considering collecting samples or seeking reconsent from donors of existing samples for use in the 1000 Genomes Project should contact us early in the planning process. Investigators will be asked to submit their proposed sample collection protocols, consent forms, and any materials relating to community engagement or consultation, for review and approval by the 1000 Genomes Project Steering Committee with input from the Samples/ELSI Group before submitting them to their institutional review boards or ethics committees for final approval. We are happy to work with investigators and IRBs to help them with such matters as determining whether the population is appropriate for inclusion in the project; developing or implementing consent processes and consent forms; determining whether some form of community engagement or consultation is appropriate; exploring the range of possible methodologies for conducting community engagement or consultation; and (where appropriate) setting up a Community Advisory Group or similar entity.

For more information contact:

Jean E. McEwen, J.D., Ph.D.
Program Director
Ethical, Legal, and Social Implications Program
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, Suite 4076
Bethesda, MD 20892
jm522n@nih.gov
Phone: (301) 496-7531
Fax: (301) 480-2770