http://grants.nih.gov/grants/gwas/index.htm
http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html
http://grants.nih.gov/grants/gwas/gwas_ptc.pdf

http://www.genome.gov/20019650

National Human Genome Research Institute
Medical Sequencing Policy

Announcement of change in policy for accepting samples for NHGRI Medical Sequencing Projects

February 26, 2008

Summary

Effective February 12, 2008, The National Human Genome Research Institute has changed some of its polices for accepting new sample sets for medical sequencing projects being pursued by the NHGRI Large-Scale sequencing program. These changes relate to the procedures for ensuring that informed consent is adequate to allow samples to be used in the large-scale sequencing program, where sequence and phenotype data will be released into a controlled-access repository.

These consent policies have been changed to make procedures more consistent with the NIH-wide policy for data resulting from genome-wide association studies (see http://grants.nih.gov/grants/gwas/index.htm).

Specifically, for NHGRI to approve the use of a sample set in a medical sequencing project, where the sequence will be linked to phenotype data and/or where the original investigators will retain the link to subject identity, NHGRI will require the institution that the request for sequencing be accompanied by a certification by the responsible Institutional Official(s) of the submitting institution that they approve submission to the NIH data repository (dbGaP).

The certification should assure that:

- The use of samples including data submission to the data repository (as described herein) is consistent with all applicable laws and regulations, as well as institutional policies;
- The appropriate research uses of the data and the uses that are specifically excluded by the informed consent documents are delineated;
- The identities of research participants will not be disclosed to the data repository; and
- An IRB and/or Privacy Board, as applicable, reviewed and verified that:

- The submission of data to the data repository and subsequent sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;
- The investigator's plan for de-identifying datasets is consistent with the standards outlined above;
- It has considered the risks to individuals, their families, and groups or populations associated with data submitted to the data repository; and
- The genotype and phenotype data to be submitted were collected in a manner consistent with 45 C.F.R. Part 46.

NHGRI staff will still evaluate consent forms for proposed projects, and based on that may decide to ask the responsible institution to address specific questions. However, the National Council for Human Genome Research will no longer be consulted about individual consent issues.

NHGRI medical sequencing policies for scientific approval for projects, for data release, for third-party access to data, and other aspects of the program will remain unchanged, as will the underlying rationale for these policies (see http://www.genome.gov/20019650).

These policies were discussed and approved by the National Council on Human Genome research at its February 2008 meeting.

Samples already approved for sequencing under the previous policies--where those projects are assigned to sequencing centers--will not need to seek re-approval. All others may wish to contact the program director listed below (http://www.genome.gov/15014882#8) to discuss whether it will be better to apply the new policy.

Model language for consent forms previously posted by NHGRI (http://www.genome.gov/Pages/Research/SequenceMapsBAC/MedicalSequencing/MSPModelLanguageforConsent.pdf) will still serve as a useful guide for studies going forward.

There is an additional item of note that has arisen due to a technical issue. Some of the newer sequencing platforms produce very short reads. For technical reasons, it is difficult to archive these data in a way that permits efficient release of small, contiguous, and unassociated segments of data -–up to 1 Mb as envisaged by the precious policy---in a fully open repository. Until this technical issue can be addressed, short read data produced on newer sequencing platforms may be deposited solely in the controlled-access repository.

Revised Policy 03142008

THE NATIONAL HUMAN GENOME RESEARCH INSTITUTE
MEDICAL SEQUENCING PROGRAM
Policies and Procedures

This Document contains two parts:

Part 1: General Policies describes a set of policies that NHGRI is adopting for its Medical Sequencing Program.

Part 2: Research Participant Protections: Considerations and Conclusions presents some background considerations that were taken into account in the development of the policies. Some readers may prefer to read Part 2 first, in order to better understand the rationales on which the policies are based.

Part 1: General Policies

Introduction

The policies developed for the Medical Sequencing Program (MSP) at the National Human Genome Research Institute (NHGRI) are designed to balance two important goals: to facilitate the discovery of genetic variants related to health and disease and, at the same time, to respect the research participants whose data and materials have been contributed to the MSP. MSP policies also are intended to promote wide dissemination of the data for use by the biomedical research community. This is imperative for maximum utility of the sequence data, as they will be produced by NHGRI-funded large-scale sequencing centers as a community resource of medically relevant sequence data produced on a scale that cannot be matched in the public sector. Finally, these policies are intended to encourage the development of new prognostic, diagnostic, preventive, and therapeutic products while safeguarding the important and unique contributions made by the scientists who collected the biological samples and associated phenotype data over many years.

The NHGRI is committed to the rapid and complete release of MSP Project Datasets for use by all investigators throughout the global scientific community who, along with their institutions, certify their agreement with MSP policies. All participants in MSP are expected to promote the policies on data access, publication, and intellectual property. Specific terms and conditions for access to and use of MSP Project Datasets by Approved Users can be found in the MSP Data Use Certification (DUC) document (http://www.genome.gov/20019653).

MSP will establish mechanisms to monitor data use in agreement with its policies. Information on these mechanisms can be found within the description of MSP monitoring procedures.

The NHGRI, in consultation with the National Advisory Council for Human Genome Research, will make all final decisions concerning MSP policies. All MSP policies are subject to change as deemed necessary to sustain program principles and priorities, to ensure the highest standards for responsible research conduct, and to be consistent with comparable policies established by NIH and by NHGRI for other programs.

Access to the MSP data will be managed by the National Center for Biotechnology Information (NCBI), National Library of Medicine, and will be overseen by the National Institutes of Health in accordance with United States Government rules and policies. All changes to policies or procedures will be posted on the NHGRI website.

Definitions of terminology used in these documents are found in the MSP Glossary (http://www.genome.gov/20019649).

An overall description of the goals of the MSP, specific current MSP programs, and a list of the advisors that contributed to identifying sequencing projects or development of policy can be found at http://www.genome.gov/15014882.

MSP Process Workflow
In most cases, the MSP will sequence genomic DNA from samples provided by Contributing Investigators. For projects that are approved (per the policies stated below):

1. Contributing Investigators will provide a letter countersigned by their institutional official certifying that use of the samples is consistent with the informed consent (see Sample Applicant Letter" at http://www.genome.gov/Pages/Research/SequenceMapsBAC/MedicalSequencing/MSPSampleApplicantLetter03.13.2008.pdf. Once this is approved by NHGRI staff, phenotype and (where applicable) exposure data to the NCBI-maintained MSP database, and will also provide samples to the NHGRI Large-Scale Sequencing Centers. Samples will be de-identified and coded, with the code specifying the link between the sample and the phenotype data from an individual research participant. A key to this code, linking the sample/phenotype information to an identified individual, will be maintained only by the Contributing Investigator. All items of readily identifying information will be stripped from the records in the MSP database. Alternatively, in some cases, the samples will be fully anonymous (i.e., the code keys will not be maintained).

2. NCBI will manage the phenotype data as described below.

3. On receipt, the Sequencing Center will log in the coded samples in a secure, automated database that will maintain the link between the coded sample and the sequence data derived from it.

4. As sequence data are generated, they will be deposited by the Sequencing Center to the NCBI open and controlled-access repositories, as described below.

5. NCBI will receive all the sequence data, and the code linking the sample and phenotype data to the sequence data. This latter information will be available only in the controlled-access database (see below).

6. NCBI will re-code the sequence data, and make them available in the Open (trace or equivalent) repository in a way that only a subset of traces can be associated with each other, in an amount that is biologically informative (proposed to be 1 Mb, or roughly the extent of a human gene) while lessening the possibility of providing enough sequence information for someone to be able to uniquely identify the individual from whom the sample was obtained. (*Note that some of the newer sequencing platforms may produce collections of very short reads that are not technically amenable to display in an open repository in a useful way that limits the view to a contiguous small (e.g. 1 Mb) region. Where new platforms are used, data may reside only in a controlled access repository unless this technical issue can be addressed*.)

MSP Data

The MSP will comprise several distinct initiatives. Currently envisaged initiatives include the following:

• *Mendelian Disorders*. In this initiative, sequencing will be done within intervals associated with Mendelian disorders, in situations where the intervals are too large or otherwise too challenging for contributing investigators to sequence without access to large-scale capability.

• *Allelic Spectrum of Common Disease*. In this initiative, sequencing will be done on large numbers of samples, supplied by contributing investigators, from studies of complex disorders. The object is to identify alleles that contribute to these disorders, both to obtain information about the specific variants that lead to disease, and to gain general information about the distribution and frequency of alleles that underlie common diseases.

In addition to the MSP, there are a number of affiliated programs in which NHGRI is participating. These currently are:

• *The Cancer Sequencing Program (CSP).* This program is being carried out by NHGRI and the policies described herein cover the CSP. However, details regarding data access may differ. Please see http://www.genome.gov/19517442 for a description of the CSP program.

• *The Cancer Genome Atlas* (TCGA; http://www.genome.gov/17516564) is a collaboration between NHGRI and NCI. Policies and procedures for data access are separate from those of MSP.

Over time, NHGRI anticipates that the MSP will include additional initiatives. It is anticipated that, in almost all cases, samples will be provided by Contributing Investigators, whose sample collection was funded under other auspices. The policies herein are designed to accommodate both existing and future initiatives.

Sequence and associated phenotype data will be placed into databases maintained at the National Center for Biotechnology Information (NCBI), as follows:

Open-access data will be available in public databases, e.g. the NCBI Trace repository, the MSP web site, dbSNP, etc. These data types include:

• Short stretches of DNA sequence that cannot be associated with each other beyond the extent that would constitute a single gene locus, or 1000 kb maximum. In some cases, the fragment size may be made smaller if there are concerns about the level of risk. (Also see "Controlled-access data" below.)

• Traces (short stretches of DNA sequence) that cannot be associated with each other beyond the extent that would constitute a single gene locus, or 1000 kb maximum. Certain technical issues attendant to short read technologies may complicate this to the extent that these type of data are not in an open repository.

• Minimal annotation, including: name of study, study authors, disease affected status, sex, basic population information, age range.

• Study protocols.

• Data summaries such as genotype frequencies and phenotype means.

• Newly discovered variants (in dbSNP, germ-line and somatic).

• Pre-computed analyses, including associations among the variants and phenotypes, and variants in LD with those variants.

Specific information that is judged to be potentially identifying, for example geographical location of participants with rare and phenotypically distinctive Mendelian disorders, will not be posted in the Open-access database.

Controlled-access data will not be available to the public, but will be made available to any researcher for biomedical research, once the investigator seeking data access, along with his/her institution, has certified agreement to the statements within the Data Use Certification (DUC) and acknowledged the intent of the NHGRI that users of MSP Datasets follow the NHGRI/MSP policies on data access and intellectual property. These data types will include:

• Phenotype data.

• The information linking together all sequence traces that come from a single (de-identified) individual. For short read technologies, such linked short read sequence data will be in the controlled-access repository.

• The information linking sequence and phenotype data from a single (de-identified) individual.

• In some cases, fragmentary sequence data will be made available only though the controlled- access repository, for example when it is judged by NHGRI to be too risky to deposit the sequence data based on what was in the original participant consent form.

In no case will readily identifiable information, such as name, social security number, etc. be put in any MSP database. All such information will be removed from records before samples are transferred to the NHGRI Large Scale Sequencing Centers and before data are submitted to NCBI. In most cases, only the Contributing Investigator will hold a coded key. In those cases, NHGRI will adhere to the information in the consent form under which the samples were originally obtained. In some cases, all linking information will be severed.

In general, the MSP will implement projects where the data can be used for any biomedical research problem by Approved Users of the MSP data, in which case Approved Users will have access to all MSP data. However, there may be specific MSP initiatives or projects where samples were consented for only limited research use (for example, only for cardiovascular disease research). This is expected to be rare, but when it occurs, such data will be provided as a separate MSP data set, for which Requestors will need to make a separate data access request.

NHGRI and its advisors will continually evaluate the risks and benefits associated with deposition of all MSP data and will modify its policies accordingly when appropriate. See Part 2: Research Participant Protections: Considerations and Conclusions.
Links to the data maintained at NCBI will be made available through the NHGRI Web Site.

Contributing Investigators
Contributing Investigators must provide a letter to NHGRI program staff (See http://www.genome.gov/Pages/Research/SequenceMapsBAC/MedicalSequencing/MSPSampleApplicantLetter03.13.2008.pdf), which certifies their understanding of the following points:

• It is the responsibility of the Primary Contributing Investigator and Major Co-investigators to obtain approval for participation in MSP from the appropriate institutional officials and committees at all sites at which data and samples being submitted for sequencing were collected according to applicable federal, state, and local laws and regulations and any relevant institutional policies. This will require an explicit institutional certification.

The certification should assure that:

- The use of samples including data submission to the data repository (as described herein) is consistent with all applicable laws and regulations, as well as institutional policies;
- The appropriate research uses of the data and the uses that are specifically excluded by the informed consent documents are delineated;
- The identities of research participants will not be disclosed to the data repository;

and

- An IRB and/or Privacy Board, as applicable, reviewed and verified that:

- The submission of data to the data repository and subsequent sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;
- The investigator's plan for de-identifying datasets is consistent with the standards outlined above;
- It has considered the risks to individuals, their families, and groups or populations associated with data submitted to the data repository;
- It has considered specific questions raised by the NHGRI staff, if any; and
- The genotype and phenotype data to be submitted were collected in a manner consistent with 45 C.F.R. Part 46.

Please see the "Points to Consider" at http://www.genome.gov/Pages/Research/SequenceMapsBAC/MedicalSequencing/MSPPtstoConsider03.12.08.pdf for institutions and IRBs that are considering this certification. A certification is sufficient for use of the samples to go forward, where the scientific aspects of the project plan have been approved, except in rare cases where NHGRI staff has identified a specific contradiction between the consent form and the intended use, that cannot be addressed by the IRB in the certification.

• NHGRI will, in cases where the Contributing Investigator retains a coded key to the samples, request a copy of the informed consent under which participants were enlisted to the study. The purpose of this is to frame specific questions, if any, to be addressed in the institutional certification beyond what is already outlined in the certification.

• Coded phenotype and exposure data (if any) associated with the DNA samples that are proposed for sequencing will be submitted to the MSP Database managers according to the requirements in the MSP Dataset Submission instructions. These data must be submitted before sequencing can begin, unless a waiver is provided from NHGRI program staff. The MSP has a range of distinct initiatives, and conditions for submitting data will vary between those programs, depending on how each is implemented. In essentially all cases, submission of data must occur before sequencing will begin. In some cases, NHGRI will ask that data be submitted at the time of application to allow the review process to assess the range of phenotypic measures, the amount of effort required to put these data into the web resource database, and the completeness of the data submitted. In rare cases, submission of phenotype data may be waived. It is the responsibility of the Contributing Investigator and his/her institution to ensure that all data are submitted in accordance with applicable federal, state, and local laws and regulations. In the event that an application is not approved or samples are not sequenced, MSP will destroy all data submitted.

• Investigators contributing data and DNA samples to the MSP will access MSP Datasets through the same procedures and data access request documents as other investigators.

Depending on the specific initiative (or even project) within MSP, Contributing Investigators may or may not retain a coded key that links the data back to the individual participant. NHGRI will consider whether or not this link is maintained (i.e., whether or not samples are completely anonymous) in applying these policies.

Informed Consent

(See Part 2 for more detailed information.)
For MSP projects, NHGRI believes that the informed consent under which samples were collected should be consistent with the aims of the MSP, to provide participant protections while also distributing data to the biomedical research community with minimal obstacles or restrictions. NHGRI recognizes, however, that many existing consents pre-date or may not have anticipated widespread data sharing, yet the principles of data sharing as proposed under the MSP may be consistent with participants' wishes.

NHGRI believes that the ideal informed consents for medical sequencing projects should contain:

A. Allowance of wide and indefinite sharing of genomic and health data, preferably with reference to placing data in databases that will be available on the Internet.

B. Description of risks associated with wide sharing of genomic data.

C. Unrestricted use with respect to what disorders can be studied with the data.

D. Realistic discussion of return of results. Specifically, only investigators with access to information linking the data to the research participant can return results. Third-parties (NHGRI sequencing centers, those with access to the medical sequencing repository) cannot.

E. Realistic discussion of the ability of the research participant to withdraw. Once data are deposited in the data repository, withdrawal of data from the repository becomes difficult.

NHGRI has developed a model language for consent forms containing these basic elements that IRBs or investigators can modify to meet their needs. It is available at http://www.genome.gov/Pages/Research/SequenceMapsBAC/MedicalSequencing/MSPModelLanguageforConsent.pdf.

NHGRI recognizes that many samples will have been collected before the existence of the MSP or similar programs that seek to broadly disseminate data to the scientific community. These samples are (in our experience) unlikely to have been collected under consents that fully anticipated wide data release. The certification discussed above will, in almost all cases, be sufficient to allow such studies to go forward.

In cases where the samples are particularly rare or scientifically compelling, the responsible institution is unable to certify their use as envisaged by the NHGRI program, and re-consent is impracticable, NHGRI may be able to consider two alternatives.

Samples may be acceptable if they can be fully anonymized (i.e., no code linking data to personal identifiers will be maintained, even by the Contributing Investigator). NHGRI staff will review the original consent forms to ensure that MSP procedures and data release policies are not inconsistent with terms explicit in those consents, particularly terms that make commitments to participants that cannot be kept if data are broadly released. [See also Anonymous Samples in Part 2.]

If anonymization is not feasible, and the study is small, NHGRI may consider waiving data deposition requirements, and the study would then go forward as a normal collaboration

between a Contributing Investigator and an NHGRI-funded large-scale sequencing center. (See http://www.genome.gov/Pages/Research/SequenceMapsBAC/MedicalSequencing/MSPExemptionsfromDataReleaseRequirement.pdf)

Due to the complex nature of these issues, and because they must be resolved before work can begin, NHGRI encourages prospective Contributing Investigators to contact NHGRI staff (http://www.genome.gov/15014882#8) to discuss them as early as possible.

In no case is the NHGRI review of consent forms intended to substitute for the opinion of the Contributing Investigator's local institution and IRB which, in any event, are responsible for complying with all applicable federal, state, and local laws and regulations relevant to the submission of samples for the MSP. Specifically, NHGRI will not proceed with samples in cases where a local IRB believes it is not appropriate.

Access to MSP data

All MSP data sets will be available via NCBI, through dbGaP (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap).

Some "open access" data will be available in other venues (for example, summary or derived SNP data may be available through dbSNP).

Investigators seeking access to controlled-access MSP data via dbGaP will be asked to complete a Data Access Request (DAR). The DAR entails that investigators, along with their institutions, have agreed to the requirements and terms of access. Further, access to controlled-access MSP data will be granted with the understanding that the data will be used in accord with the conditions to be described on dbGaP for the appropriate research uses, including any limitations on such use, if any, of a given dataset. In most cases, NHGRI anticipates that MSP data will not include any limitations on use to a particular disorder. NHGRI aims to make all MSP data from all MSP studies available via a single request. However, NHGRI may accept some studies where data are restricted to specific uses—those will require separate access requests.

DARs will be evaluated by a Data Access Committee constituted by NHGRI. NHGRI anticipates that most DARs will be evaluated within two weeks of receipt. Applicants that are approved will become Approved Users for one year (subject to adherence to MSP policies).

All Approved Users will certify through the DAR process that they will not distribute individual MSP Controlled-access data in any form to any third parties, other than those of their own research staff who have agreed to the terms of the DAR. Approved Users who are not Contributing Investigators shall also certify that they will not attempt to identify the individual participants. For collaborative projects, any independent collaborating investigator from a separate institution involved in the use of the MSP data is required to submit a separate DAR. All Approved Users and their institutions will be required to acknowledge responsibility for ensuring that all uses of the data are consistent with federal, state, and local laws and regulations, and any relevant institutional policies.

Customary scientific use of results derived from Controlled-access data (for example, publication results, Web posting of summary or aggregate results that does not in effect disclose individual genomic or phenotypic data that are in the Controlled access repository) is not restricted in any way. NHGRI encourages such publication with appropriate attribution (see MSP Publication Policy, below). Investigators that have questions about what may constitute individual genomic data should contact NHGRI program staff (LINK TO PROGRAM CONTACT).

Contributing Investigators will not be provided advance research access to MSP data that are in the controlled-access database. The terms and conditions governing data access for research use of Contributing Investigators will be identical to those for any other member of the scientific community seeking to become an Approved User. All submitted samples provided by Contributing Investigators for MSP use will be returned or destroyed following the completion of the specific MSP project according to the procedures set by the contributing study site and the sequencing centers.

MSP Publication Policy

MSP publication policy is intended to balance two factors. On the one hand, it seeks to recognize the substantial long-term commitment that Contributing Investigators have made in the collection, phenotypic characterization, and analysis of the study samples. On the other, NHGRI believes that the MSP is a "community resource projects" and as such, there is maximum scientific benefit to making the data available to the scientific community as soon as possible.

NHGRI intends to release MSP sequence data as rapidly as possible after they are produced, with no restrictions on use. Phenotype and other data associated with the sequence data will be released (in the controlled-access database) to all Approved Users as soon as it is clear that the sequence data from submitted samples will be produced in full, that is, that the project is determined to be technically feasible and NHGRI has made a firm commitment to perform the sequencing as put forward in any particular project description.

Approved Users will agree not to submit for publication any results or analyses derived from the use of any MSP data without specifically acknowledging the Contributing Investigators, the funding organization that supported the Contributing Investigators, the Sequencing Centers, the MSP database, and NHGRI.

Further, NHGRI considers that the MSP data in the databases are unpublished data until the Contributing Investigator publishes a paper describing the results of a particular MSP project, specifically in this case, results describing the association of genomic variation with a phenotype. Approved Users are asked to apply the normal rules of scientific etiquette when deciding to publish results (association of a variation with phenotype) based substantially on unpublished data, which may be unvalidated or of otherwise undocumented quality. Contributing Investigators are asked to keep in mind that they have a reciprocal responsibility to publish significant results rapidly.

Based on experience, NHGRI believes that one highly productive outcome that recognizes the contributions of both Contributing Investigators, Sequencing Centers, and data users is for separate parties with an interest in the data to engage in collaborations.

MSP Intellectual Property Policy

The goal of the MSP Intellectual Property (IP) Policy is to maximize the public benefit of research fostered by NHGRI to identify the genes and gene variants that contribute to diseases, as well as molecular targets useful in the prevention, diagnosis, and treatment of these diseases. It is the intent of NHGRI to promote broad freedom-to-operate for all users of MSP data by rapidly placing data in the public domain. Further, it is also the intent of NHGRI that the genotype-phenotype associations identified through the MSP remain in the public domain unencumbered by intellectual property claims. The NHGRI believes that this policy will avoid premature claims on pre-competitive information, while promoting opportunities to develop IP and file claims on downstream discoveries, which will be necessary to support full investment in products that the public needs.

To facilitate the goals for this IP Policy, the MSP database will provide rapid, no-cost, and complete release of all data for access by Approved Users. It is expected that MSP-supported data and conclusions derived therefrom will remain freely available, without requirement for licensing, for applications such as, but not necessarily limited to, the use of markers in developing assays and as guides toward identification of new drug targets, therapeutics and diagnostics. NHGRI encourages broad use of MSP data coupled with a responsible approach to management of intellectual property derived from downstream discoveries that are consistent both with the recommendations cited in NIH's Best Practices for the Licensing of Genomic Inventions and the NIH Research Tools Policy.
(http://www.genome.gov/15014882#8)

The filing of patent applications in a manner that might restrict use of MSP data could substantially diminish the value and public benefit provided by these community resources. Approved Users, including Contributing Investigators and their affiliated organizations, must acknowledge the MSP IP Policy, the goal of which is to sustain the public benefit of MSP by not pursuing intellectual property protections that would prevent or block access to, or use of, any element of MSP data, or conclusions drawn directly from those data.

Part 2: Research Participant Protections: Considerations and Conclusions

Introduction

The information below summarizes our considerations and key conclusions in four areas related to the protection of research participants: *Consent, Database Structure, Consideration of Human Subjects, and Return of Clinically Relevant Results.* This information is specifically intended to provide investigators and institutions with detailed background information as to how NHGRI made decisions regarding these important policy issues. NHGRI believes this background information is particularly important because of the significance of the topic, because the conclusions were the result of a deliberative process that revealed a range of well-considered opinions rather than absolute consensus, and because the conclusions and policies are open to change due to the nature of the program, related programs, and the state of the science.

The most difficult aspect of establishing a sound MSP policy was properly balancing the requirement to protect research participants with the importance to biomedical research of making MSP data available to a wide research community. During this process, NHGRI received advice from multiple sources, including advisors to NHGRI's Ethical, Legal and Social Implications (ELSI) research program, the Medical Sequencing Working Group, and the National Advisory Council for Human Genome Research (rosters for these groups are available at http://www.genome.gov/10000905). As may be expected, there was not unanimity of views with regard to many of the specific issues involved.

The policies on Research Participant Protections take into account all the advice accrued by NHGRI from its advisors, and all the policies established for the related programs mentioned above. The MSP policies, in some cases, place somewhat more emphasis on research participant protections based on the advice we received about the specific MSP programs. One result is that the MSP policies may be considered more stringent than current practice at many institutions, or than strictly called for in OHRP guidelines.

One very significant piece of advice that shaped these policies was that that information about research participants can not truly be considered to be anonymous if it includes significant amounts of genomic data, simply because those data have an increasing potential to be identifying as more human genomic (and other) information becomes widely available in the near future. As a corollary to that, these policies were developed with the understanding that the standards for appropriate use of genomic data are likely to evolve rapidly.

In addition, these policies were written from the point of view that, initially, most MSP projects will sequence existing sets of samples where consent did not anticipate broad data distribution. NHGRI expects that, over time, consent forms will begin to anticipate wide distribution of data, perhaps driven by MSP and other similar programs (e.g. GAIN). Accordingly, over time, as more samples become available where consent is consistent with broad distribution of data, it is likely that NHGRI will place more emphasis on those samples.

In addition, NIH is embarking on a number of related programs, including GAIN (http://www.fnih.org/GAIN/GAIN_home.shtml), the Genes and Environment Initiative (GEI, http://www.gei.nih.gov/index.asp), The Cancer Genome Atlas (http://cancergenome.nih.gov/index.asp), and the ongoing establishment of policy regarding Genome-Wide Association studies, or GWAS (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-06-071.html). Although related policies are still being developed for each of these programs, in developing the MSP policies, NHGRI attempted to take these other discussions into account in order to present as consistent a set of policies to the research community as possible.

As stated above, in no case is the NHGRI review of consent forms intended to substitute for the opinion of the Contributing Investigator's local institution and IRB which, in any event, are responsible for complying with all applicable federal, state, and local laws and regulations relevant to the submission of samples for the MSP. Specifically, NHGRI will not proceed with samples in cases where a local IRB believes it is not appropriate. In addition, NHGRI will request that the contributing investigator provide documentation that the local IRB is aware of the proposed use of the samples, by providing a certification from the institution, as described above. Such a certification will, in all but rare cases, be sufficient for samples to be used.

Informed Consent

In the process of establishing MSP policies, NHGRI staff and advisors reviewed over 25 consent forms used by Contributing Investigators to obtain samples and data. (These Contributing Investigators were participating in a set of pilot efforts for the MSP.) This review led to a critical observation: consent forms for existing studies rarely contain adequate descriptions about the idea that genetic or genomic data could be shared widely, particularly on the internet, nor do they generally describe the risks associated with wide data sharing. In cases where consent forms do allow wide sharing, they often stipulate that the data can only be used for a specific research purpose (e.g., heart disease). Neither of these observations was consistent with NHGRI's desire to make MSP data widely available while maintaining sufficient participant protections.

Based on this, NHGRI believes that re-consent will always be preferable, if the original consent does not clearly allow wide data sharing, and for a variety of disorders. In cases where the consent is very clearly not consistent with wide data sharing, NHGRI will continue to recommend that the Contributing Investigator re-consent samples for MSP projects, and will over time encourage newly consented samples to include explicit consideration of wide data sharing and broad use.

However, with the benefit of nearly two years of experience with this policy, NHGRI together with its National Advisory Council has decided to moderate its approach somewhat. First, in order to be more consistent with NIH-wide GWAS policy, NHGRI will leave it almost entirely to the responsible institution and its IRB to determine whether the sharing of data as outlined herein is appropriate or not. In addition, NHGRI is able to accommodate use of samples that have restrictions on use, but in all cases we will prefer samples that have allow broad use.

Model language for new consents was developed in conjunction with staff and outside ELSI advice, and is available at http://www.genome.gov/Pages/Research/SequenceMapsBAC/MedicalSequencing/MSPModelLanguageforConsent.pdf. The main concepts that NHGRI has concluded should be included in any consent form/process (and agreed to by participants) for the MSP are:

• Voluntary agreement by the participant to donate a blood or other tissue sample to be used for this and other research projects. (The sample specifically would be used by the sequencing center to produce data.)

• Voluntary agreement by the participant to allow release of information from her/his medical records for this and other research projects. (Phenotype data would be included in the controlled access database.)

• Voluntary agreement by the participant to have his/her coded genetic information and coded medical information shared among researchers, ideally with specific reference to data being placed in databases on the Internet, as described in the *Storage and Release of Samples and Medical Information* section of this document. Understanding on the part of the participant that her/his coded genetic information and coded medical information in the Internet databases will be used in this and in other research projects.

• Understanding on the part of the participant that there is a risk that someone in the future might be able to use information in these databases to identify him/her or possibly his/her blood relative(s).

• Understanding on the part of the participant that data, once in the MSP database, cannot be withdrawn. (However, the participant can withdraw from the study to the extent that the Contributing Investigator can still control, e.g, samples, key codes, and local records can be destroyed.)

The exact wording to be used in the consent process will left up to investigators and their IRBs as long as the consent form contains all of the above concepts.

Anonymous samples

In cases where samples have significant scientific importance or address compelling public health needs and re-consent is not feasible, it may be possible to proceed with samples that are fully anonymized, that is, not even the Contributing Investigator holds a key linking the samples to an individual participant. Each sample set represents a unique situation, so it is not possible to state all criteria under which anonymization would be considered. However, the cost of re-consent alone will not be considered as a sufficient reason. Rather, NHGRI will consider issues such as the practical ability of the participants to be re-contacted, and the likely rate of success in obtaining re-consent.

In situations where anonymization is being considered, NHGRI staff will still evaluate the consent forms for language that constitutes a direct commitment to the participants that cannot be maintained even with complete anonymization. For example, some consent forms promise the return of results, which is precluded by anonymization. Others state that the

results will only be used for a single disease study or type of disease, or that results will only be available to the Contributing Investigator, which is not possible with wide data distribution. However, samples with a consent that made no such direct commitments, but, for example, was mute on the idea that data would be distributed on the Web, may proceed with anonymization.

Samples from deceased individuals
Samples from deceased individuals may be used in MSP projects without re-consent unless the original consent made a direct commitment not to use such samples in this way, e.g., a statement saying samples will not be used after the participant's death. NHGRI will review the original consent.

Database Structure
To help ensure the protection of research participants in a manner consistent with consent forms, NHGRI decided that the MSP database must take steps to ensure that the deposited data cannot readily be used to identify research participants. Deposited data could include genome sequence, phenotype, demographic, and other data. In no case would MSP data include readily identifiable data such as name, address, social security numbers, contact information, or other HIPAA identifiers.

In addition, NHGRI decided that all samples would have to be coded as to how they were linked to a particular identified individual, and that only the Contributing Investigator could hold a key to that code. The Contributing Investigator is presumed to have had full IRB approval for the study under which samples were originally gathered (and must certify that approval to participate in an MSP program). Thus, neither the Sequencing Centers nor the Databases would know the identities of research participants.

However, it is possible that genomic information (DNA sequence) can potentially be identifying should a second sample from a research participant (or blood relative) be obtained and analyzed (as might happen in a forensic analysis). Although the risk of this occurring was judged to be slight at present, NHGRI and its advisors decided to apply a stricter procedure than is currently required by the NIH Office of Research Protections (OHRP) to deal with even this slight risk. See the section on Human Subjects below for more discussion. Thus, the MSP established a policy of requiring that MSP data be deposited in a two-tiered database, as described in Part 1 above.

"Human subjects" or not
After considering the above issues, some have concluded that research using this dataset, from the point of view of a third party other than the Contributing Investigator accessing MSP data, does not involve human subjects. This conclusion is based on the Office of Human Research Protections' (OHRP) "Guidance on Research Involving Coded Private Information or Biological Specimens" published on August 10, 2004 which can be found at: http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.pdf

This guidance states:
Under the definition of human subject at 45 CFR 46.102(f), obtaining identifiable private information or identifiable specimens for research purposes constitutes human subjects research. Obtaining means receiving or accessing identifiable private information or

identifiable specimens for research purposes. OHRP interprets obtaining to include an investigator's use, study, or analysis for research purposes of identifiable private information or identifiable specimens already in the possession of the investigator.

In general, OHRP considers private information or specimens to be individually identifiable as defined at 45 CFR 46.102(f) when they can be linked to specific individuals by the investigator(s) either directly or indirectly through coding systems.

Conversely, OHRP considers private information or specimens not to be individually identifiable when they cannot be linked to specific individuals by the investigator(s) either directly or indirectly through coding systems. For example, OHRP does not consider research involving only coded private information or specimens to involve human subjects as defined under 45CFR46.102(f) if the following conditions are both met:

(1) the private information or specimens were not collected specifically for the currently proposed research project through an interaction or intervention with living individuals; and (2) the investigator(s) cannot readily ascertain the identity of the individual(s) to whom the coded private information or specimens pertain because, for example:

> *(a) the key to decipher the code is destroyed before the research begins;*
> *(b) the investigators and the holder of the key enter into an agreement prohibiting the release of the key to the investigators under any circumstances, until the individuals are deceased (note that the HHS regulations do not require the IRB to review and approve this agreement);*
> *(c) there are IRB-approved written policies and operating procedures for a repository or data management center that prohibit the release of the key to the investigators under any circumstances, until the individuals are deceased; or*
> *(d) there are other legal requirements prohibiting the release of the key to the investigators, until the individuals are deceased.*

Others thought that the MSP policies should use a more stringent procedure for protection of participants than that called for in the OHRP guidance for one or more reasons, for example a belief that participants should be specifically consented for this type of research, the long-standing precedent that human subjects are involved when there is coded, but linkable, private information being made available, and unease as to whether sequence data can truly be considered to be not readily identifying data. Their conclusions were based on their interpretation of other OHRP guidances, including: Issues to Consider in the Research Use of Stored Data or Tissues published November 7, 1997 which can be found at:
http://www.hhs.gov/ohrp/humansubjects/guidance/reposit.htm
and OHRP Decision Charts of September 24, 2004, which can be found at:
http://www.hhs.gov/ohrp/humansubjects/guidance/decisioncharts.htm

Because there was no consensus on this issue, we decided on a somewhat stricter policy than called for by the August 10 2004 OHRP guidance. Specifically, we require Institutions to make a determination, based on their own standards of research practice, whether research involving the de-identified, coded and potentially linkable information in the MSP dataset by Approved Users involves human subjects or not. We presume that this institutional determination will be made in consultation with the local Institutional Review Board (though that may not be absolutely required in all cases). If the conclusion is that there are human subjects, then the next question is whether the proposed research qualifies for Exemption #4

(Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects) or whether a full IRB review is necessary.

Return of Clinically Relevant Research Results to Participants

It is likely that some Contributing Investigators will want to disclose research results to participants; some have actually promised in their consent forms to do so. In other cases, whether results should be disclosed will be less clear. In still others, it may be more harmful to disclose than to remain silent.

Within the MSP, the decision about whether research results are returned to participants will be made by the Contributing Investigator in consultation with his/her IRB, taking into consideration what was promised in the consent process. The contributing investigator has sole ability to return results, and sole responsibility for delivering information about results to research participants, if she or he chooses to do so. In some MSP projects, data may be completely anonymized, precluding the ability to return results. The Contributing Investigator must consider that MSP data were not generated in CLIA-certified laboratories, and thus the Contributing Investigator will be responsible for ensuring that results are validated in a CLIA-certified facility before results are returned.

In cases where the Sequencing Centers or Approved Users of data find results that they believe have the potential to be clinically significant for an individual, they are encouraged to alert the Contributing Investigator to the results. However, they are not required to do so.

NHGRI has an interest in knowing what kinds of research results are coming out of the MSP and what kinds of decisions are being made by investigators and their IRBs about the return of results. Thus, the NHGRI will establish a Data Use Review Board (DURB) that will have expertise to analyze the results and make recommendations to NHGRI about whether the joint goals of participant protection and wide data usage are being achieved. Contributing Investigators (who will be required to become Approved Users to access controlled-access data) will be asked to provide NHGRI with information about the return of results in annual renewals of data access permissions.

Considerations
The policy on whether to return potentially clinically significant results was challenging to formulate because of our basic lack of knowledge in many cases regarding the correlation of genetic variation with a particular disease, and variation between specific MSP projects and the study designs of Contributing Investigators. In developing this policy, potential Contributing Investigators and NHGRI advisors differed in their opinions and their desire to return results in studies that they personally lead. The consensus was that MSP policy should not make return of results mandatory, but neither should it be precluded.

In general, the duty to report findings increases in cases where the results of an MSP project find variations in an individual which have a high and/or well-defined probability to predict a

serious disease, or if reported could lead to effective preventative measures or easily avoided risk factors.

There are multiple complicating factors in deciding whether to return results. There may be findings in which the implications of a variation are not clear, or not particularly serious. The evidence for clinical utility may be weak, but there may be some who believe participants have the right to know. Finally, there may be research results that should not be disclosed to research participants. This may include research findings which, while they may be of interest to researchers, have no relevance to research participants, such as information for which there is no analytic or clinical validity, no clear explanation of the meaning of the research results, and no evidence of harm if the information is not disclosed. Although family studies may not often be conducted as a part of medical sequencing, it would be our recommendation not to disclose the identification of misattributed paternity.

The Contributing Investigator is in the best, perhaps only, position to make the judgment on whether results should be returned. He or she will have the appropriate Human Subjects protections and IRB approvals in place, and will be most likely to be in a position to contact research participants, both because only she/he holds the key linking data to participant identity and because he/she may actually be in continuing contact with research participants. As noted above, the Contributing Investigator must ensure that results are validated in a CLIA-certified facility before reporting them to participants. NHGRI-funded sequencing centers are not CLIA-certified. The Sequencing Centers are engaged in basic discovery research. Thus neither they, nor NHGRI, can be held responsible for the quality and/or reliability of any individual variation that is discovered, or its ability to predict disease or any other phenotype.

Withdrawal

Informed consent usually informs the participant that he/she has a right to withdraw from studies. The responsibility to comply with a request for withdrawal lies with the investigator(s) that collected the samples, and who maintains the list of links between sample data and individual participant. However, once data are published, this limits the practical ability for an investigator to withdraw data from a study.

Similarly, for NHGRI MSP, there are practical limits to the extent that a request for withdrawal can be complied with. Once data are in the MSP database and have been distributed to approved users, it may not be possible to fully ensure withdrawal of data. NHGRI will honor requests from Contributing Investigators to withdraw data sets from individuals from the next regularly updated version of the MSP database.

Data Use Review Board

NHGRI will establish a Data Use and Review Board (DURB) to help provide guidance about the four major issues discussed above. In particular, the DURB will provide comment to NHGRI regarding the appropriateness and effectiveness of its policies on protection of research participants on an ongoing basis. The DURB will have access to all data about reportable events as communicated by Approved Users of MSP data in regular reports. Duties of the DURB will include:

• Evaluating, on an ongoing basis, the risks to research participants that may be entailed by having their samples used in the MSP.

• Evaluate the effectiveness of MSP policies in protecting research participants from risks.

• Evaluate the operations of the DAC.

• Recommend changes to MSP policies, ether in cases where a policy may be inadequate to protect research subjects, or cases where policies may be too stringent.

• Become a resource to provide advice regarding any questions about return of results.