The American Society of Human Genetics
Policy Statement

November 30, 2006

ASHG Response to NIH on Genome-Wide Association Studies

The comments shown below have been submitted to the NIH on behalf of the American Society of Human Genetics (ASHG) in response to Request for Information (RFI): Proposed Policy for Sharing of Data obtained in NIH supported or conducted Genome-Wide Association Studies (GWAS), as published in the NIH Guide Notice NOT-OD-06-094 and the Federal Register.

Do you support broad access of phenotypic and genotypic data to advance medical research?
Support

Do you support the proposed policy?
Support with changes

Please explain?
Yes, as a matter of scientific principle, the American Society of Human genetics supports broad access to all data and research materials in the peer-reviewed and published literature, particularly that funded by US governmental agencies.

Yes, the American Society of Human Genetics supports the proposed policy but this support is contingent on requiring significant changes and appropriate consultations with the scientific community and members of the public before a final policy is established.

Yes, we support the proposed centralized data repository at the NIH but this support is contingent on requiring significant changes and appropriate consultations with the scientific community and members of the public before the nature of this repository is finalized.

Assuming personal identifying information is removed, do you support the proposal for a centralized NIH data repository of phenotypic and genotypic data for Genome-Wide Association Studies (GWAS)?
Support

1. What are the potential benefits and risks associated with wide sharing of phenotypic and genotypic data where identifying information has been removed?

Benefits: We concede that wide sharing of NIH-supported genotypic and phenotypic data has not been ideal to date, that there are roadblocks and impediments to gaining access even after publication, and, that with the advent of GWAS this can be greatly exacerbated. Ready access by the ASHG membership to large phenotype-genotype data sets representing a variety of disorders and traits, will doubtless achieve our mutual goal of advancement of science and the enhancement of translation of research findings to better address health needs. First, this will increase the efficiency of existing research programs. Second, this will lead to greater robustness

of research findings since associations will be based on data from individual investigators but combined with publicly-accessible data. Third, this will rapidly spur development of new analytical tools based on large-scale data.

Risks: There are significant risks to the policy as proposed, primarily related to the ill-defined term "identifying information". Does "identifying information" refer only to name and social security number or possibly to the many additional identifiers referenced in the HIPAA? The ASHG is acutely aware that the most accurate individual identifier is the DNA sequence itself or its surrogate here, genotypes across the genome. It is clear that these available genotypes alone, available on tens to hundreds of thousands of individuals in the repository, are more accurate identifiers than demographic variables alone; the combination is an accurate and unique identifier. We presume that many GWAS will use pedigree information either to select probands or to implement family-based association studies. Thus, the potential to identify both individuals and families exist. We believe that highly specific definitions of which identifying information will be available through the repository is needed. The NIH has proposed that a Data Use Certification process be used to theoretically screen those who would define potential use inappropriately. We laud this provision, but note that this approach is ineffective unless there are clear penalties for misuse of the information by unscrupulous players. We also need a clearer definition of who the potential users can and cannot be, and delineation of a transparent process for the decision-making process. In other words, the benefits of wide sharing have to be balanced against the real possibility and risk of inadvertent identification of research subjects. We are willing to help in this endeavor.

2. In addition to removing personal identifying information, what protections are needed to minimize risks to research participants whose phenotypic and genotypic data are included in a centralized NIH data repository and shared with qualified investigators for research purposes?

Many, if not most, of the phenotypic data will arise from existing studies across multiple institutions with informed consent language that represented the best judgment of the investigators involved but rarely included the open dissemination of genetic data. Prospectively, informed consent documents will require significant change to reflect the potential wide sharing of both biological samples and data and the degree of risk involved based on the extent of sharing. The response of individual IRBs to this change is unpredictable but clarified guidelines for IRB's may be helpful to reduce the variability of IRB expertise and policy implementation. Paradoxically, the suggested NIH policy may exclude many studies from this dissemination model based on their current consent documents. It is sometimes suggested that in this new genetics environment investigators should reconsent their subjects. Informed consent with wide genome-wide data sharing may be appropriate for some proposed future studies, but is not a pragmatic solution for existing ones: reconsent will lead to a smaller sample size, may lead to study bias based on research attitudes and associated demographic variables and is hardly the best use of limited funding budgets at the NIH.

We propose that the NIH convene a working group of extramural scientists, experts In research ethics, and members of the public to examine which specific types of identifying data and

informed consent forms might be relevant to open data dissemination. In fact, finding out the status of existing studies and creating a path that investigators could traverse to attain the wide data dissemination ideal will be more educational to the scientific community and beneficial to the American public than the rapid institution of an untested policy that will exclude more investigators than it will include. Importantly, one can use one model study as a test example to assess both the problems and the creative solutions it can engender. We are supportive of this working group reporting in a short period (6 months); we, the leadership and membership of ASHG, are willing to work with the NIH on accomplishing this goal.

3. What are the advantages and disadvantages of the proposed:

i. Centralized NIH data repository?

A centralized repository that would provide ready access to data to our researchers is a boon but there are many unanswered questions. We assume that this repository will be funded from within the NIH intramural program and will not impact extramural funding; we require assurance on this matter. If not, an open peer-reviewed competition for the design and maintenance of the repository is desirable. More importantly, this repository requires the active collaboration of geneticists, epidemiologists and computer scientists; we are not aware that this has happened or is planned. Finally, the contents of the repository require continual peer review and input from the scientific community. This plan is sadly lacking. It is understood that a Data Access Group will review the applications for access to the repository data, but this is not the same as scientific input and buy-in from the scientific community or the community at large.. However, a community-curated expert database (by hypertension, colon cancer, etc. experts) supported by the NIH with contemporary data standards will be a great boon to the research of our members.

ii. Approach to data submission?

The approach to data submission is necessarily broad and is difficult to comment on with specifics. The plan to transfer and provide large data sets to the NIH is not adequately defined. Unlike the regular features of molecular (sequence, polymorphism) data the challenges of obtaining and curating phenotypic and covariate data are not trivial. The scientific community will need to be assured with a plan of what the data protocols will be so that we can be assured that the data are of high-quality and accurate. Once again, it might be helpful for the repository to undertake a pilot example to understand the challenges. It would have been helpful if the NIH had requested the help of the ASHG membership, who have deep expertise in this area, to help examine and provide solutions to these issues.

iii. Approach to scientific publication?

It is tempting to invoke the principles of DNA sequence data release to patient-oriented data but there are significant differences. From the publication viewpoint, a 9 month handicap appears reasonable, but the NIH must realize that this is an "experiment." Why 9 months and not 6 or 12 months? What happens to agreements in existing collaborations in large studies? Whose agreements wins and why and who adjudicates them on what authority? However, we support

the concept of uniform principles and standards under which data from publicly-funded research gets widely disseminated after providing the primary researchers a safe period for analysis.

iv. Approach to intellectual property?

The application of the meaning of intellectual property is narrow in the proposed policy, focusing on the patenting of technologies and pharmaceuticals in the more traditional sense. Even in this sense, the policy is not attached to any mechanism that would really protect well-meaning investigators from very aggressive players that would continue to push the patenting and licensing systems to its outer legal limits. As we have seen in the testing arena, patenting and aggressive licensing situations have limited access to the testing, and it would follow that the very reason for developing a repository, namely data access, could also be greatly compromised with no consequential terms for these activities.

In its current form the proposed policy does not address the copyrighting of questionnaires, data collection instruments or algorithms used in the definition of phenotypes?

4. What specific resources may investigators and institutions need to meet the goals of this proposed policy?

Institution of the proposed policies will affect individual investigators in a variety of ways as outlined and require the indicated resources:

1) NIH guidelines for standard consent forms which include data and biological sample sharing for certain types of studies. The aim is not to reduce the autonomy of IRBs but to obtain a greater uniformity of response to some generic types of studies.

2) Investigators will require specific budgetary support for data preparation and transfer to the NIH repository. This needs to be done uniformly for disease-oriented studies using modular budgets since this is one area where uniformly grant proposal budgets are routinely decreased.

3) Investigators will require specific budgetary support for data analysis during the grace period since otherwise the proposed 9-month period is of little use to the investigator conducting the study.

4) Investigators may require resources to reconsent study participants.

5. Other comments you would like to submit associated with draft policy

Institution of the proposed policies will affect individual investigators in a variety of ways as outlined and require the indicated resources:

1) NIH guidelines for standard consent forms which include data and biological sample sharing for certain types of studies. The aim is not to reduce the autonomy of IRBs but to obtain a greater uniformity of response to some generic types of studies.

2) Investigators will require specific budgetary support for data preparation and transfer to the NIH repository. This needs to be done uniformly for disease-oriented studies using modular budgets since this is one area where uniformly grant proposal budgets are routinely decreased.
3) Investigators will require specific budgetary support for data analysis during the grace period since otherwise the proposed 9-month period is of little use to the investigator conducting the study.
4) Investigators may require resources to reconsent study participants.

First Name: Joann
Last Name: Boughman
Email: jboughman@ashg.org
Title: Executive Vice President of ASHG
Affiliation: American Society of Human Genetics (My comments represent The Collective Position of American Society of Human Genetics)

I am responding from the perspective of:
Professional Organization (ASHG)

Organizational affiliation:
Professional Society (ASHG)