Genomewide Association Studies Data Sharing: National Institutes of Health Policy Process

Genomewide association studies (GWAS) are becoming an important part of the human genetics research landscape. As the technological and analytical tools have been developed, many of the National Institutes of Health (NIH) centers and institutes have launched projects to find genes or sets of genes that contribute to a variety of human disorders. Massive data sets are required to perform comprehensive genotype-phenotype analyses. GWAS projects that have been launched by the NIH include the Genetic Association Information Network (GAIN) (http://www.fnih.org/GAIN/GAIN_home.shtml) and the Genes and Environment Initiative (GEI) (http://www.gei.nih.gov). Because the NIH is the steward of the public's investment in biomedical research, the NIH created a mechanism (i.e., a data repository) and related policy to maximize data sharing among the scientific community. The draft policy includes the following major elements: data management (i.e., protection of research subjects, data submission procedures, and data access principles), scientific publication, and intellectual property. The draft policy is available at http://grants2.nih.gov/grants/gwas/.

GWAS Data Management Overview

The proposed federal policy calls for NIH-funded investigators to submit to a data repository GWAS genotype and phenotype data that have been deidentified and uniquely coded by the primary investigators. They would also be required to submit documentation that describes how the confidentiality of research subjects is protected, project protocols, and data variable descriptions. Access to the repository would be granted by an NIH-appointed Data Access Committee, which would review all applicant projects and put data-use agreements in place. Publication exclusivity would last for 9 months, after which subsequent secondary analyses would be required to acknowledge the original contributors of the data and funding organizations. Consistent with other NIH policies, the broad use and sharing of data are encouraged, and the submission of premature intellectual property claims on precompetitive information is discouraged.

The Challenge to the Scientific Community

The draft policy was made available for public comment through October 2006. The leadership of the American Society of Human Genetics (ASHG) had initiated a draft response, and then, at the ASHG Business Meeting in New Orleans, members of the Society raised serious concerns about the draft policy. ASHG immediately requested an extension of the comment deadline, and the NIH responded with an extension to November 30, 2006. President Steve Warren appointed a small working group to draft the Society's comments, and then the leadership reviewed and approved the response that was submitted to the NIH.

ASHG's Response

The Request for Information (RFI) from the NIH was structured as a set of questions, and the complete ASHG response is available on the Web site at http://www.ashg.org/genetics/ashg/news/gwas.shtml. The major points are summarized below.

1. ASHG supports the concept of data sharing and the development of a centralized repository, but this support is contingent on there being changes made to the proposed policy, as suggested below, as well as the addition of clarifying statements and additional scientific consultation.
2. There remain substantial risks to research subject protection with the policy as proposed, since the combination of genotype and significant phenotypic data remains a potentially accurate and unique identifier, even if names and Social Security numbers are removed. We remain concerned that enough information would be available to lead to the identification of specific individuals and their relatives, even with the controls defined. Therefore, the data access certification process must be rigorous, and any misuse of data must have clear and important consequences.
3. Because institutional review boards (IRBs) have great variability in their responses to federal guidelines and policy, more clarification of and specification in the language required to protect subjects is necessary. This specificity and additional IRB training would be helpful for both data depositors and requestors. Informed consent with wide data-sharing provisions may be appropriate for future studies but is not a pragmatic solution for existing studies that may require reconsent.
4. ASHG is assuming that the repository process would be funded by intramural funds and would not adversely impact extramural funding. If this is not the case, then serious concerns remain in the prioritization of research endeavors. In addition, investigators may require additional funding to meet requirements for data preparation and transfer to the repository, as well as additional resources for any process of requisite reconsent.
5. Unlike the more regular features of molecular (sequence and polymorphism) data, the challenges of obtaining and curating phenotypic variate and covariate data, as well as exposures, are not trivial. Variable definitions and protocol differences may require extensive interpretation before meaningful analysis. In addition, the presentation of precomputed simple comparisons may be misleading and produce inaccurate results, and interpretations of secondary analyses may be incomplete or derive inappropriate conclusions.
6. ASHG urged the NIH to engage in more discussion with representatives of the extramural as well as the intramural scientific and advocacy communities before finalizing the policy. The suggestion was made to perform a pilot test of contributed data, to assess the implications before taking the policy public.

Am. J. Hum. Genet. 2007;80:581–582. 0002-9297/2007/8003-0023$15.00
DOI: 10.1086/512166

In December, the NIH held a town hall meeting, presenting an additional overview of the policy and briefly summarizing some of the issues raised in the comments. Earlier the same week, an announcement had been made that the National Center for Biotechnology Information had developed unique databases to make the results of GWAS widely available to the research community, such as the results on macular degeneration and Parkinson disease that can be accessed at the dbGaP Web site (http://www.ncbi.nih.gov/entrez/query.fcgi?db=gap). Until scientists have the opportunity to examine and evaluate the database, further reasoned comments cannot appropriately be made.

Now What?

Although ASHG was pleased to have the opportunity to respond to a formal RFI during a public comment period, major challenges clearly remain with the steps taken in a complicated process that may implicate major changes in the clinical genetics and genetic epidemiology communities. ASHG members should stay informed of changes and possible opportunities to remain engaged in this policy process.

JOANN A. BOUGHMAN, Ph.D.
Executive Vice President