NIH Data Sharing Policy for Genome-wide Association Studies ("GWAS")

Steven Hirschfeld, MD PhD
National Institute of Child Health and Human Development
National Institutes of Health
May 10, 2007
Genetic Variation and Common Diseases

- The major genetic risk factors for common diseases such as heart disease, diabetes, some forms of cancer, hypertension, asthma, Alzheimer’s disease, osteoporosis, autism, and mood disorders are likely to be identified in the near future.

- The biology of these genetic risk factors and how to impact the natural history is a key challenge for the coming decades.
Genome-wide association studies (GWAS) rely on research tools and technologies to identify and analyze genetic differences between people with specific observable illnesses or conditions and individuals without the illness or condition.
Special Issues Arising from GWAS

- The amount of data generated by studies of up to 500,000 Single Nucleotide Polymorphisms (SNP) in thousands of participants greatly exceeds the ability of individual investigators to analyze.
- Genome wide studies are resource intensive.
- Sharing of genetic information raises sensitive issues in an area of rapid technological change.
- Genetic information is relevant to family members and community.
Guiding Principle:

The greatest public benefit will be realized if data from GWAS are made available, under terms and conditions consistent with the informed consent provided by individual participants, in a timely manner to the largest possible number of investigators.
Rationale for GWAS Repository

- **Improve Health** – Sharing genome-based research, including both genotypic and phenotypic information, with a broad number of scientists, will enable medical science to better understand the health needs of the public and facilitate development of new technologies and approaches for the prevention, diagnosis and treatment of disease.

- **Maximize Public Investment** – Providing centralized access significantly increases the availability of data for researchers, which is predicted to accelerate the discovery of associations between genetic data and disease while reducing research costs.
Data sharing is common

- Framingham SHARe (NHLBI)
- Many investigators share genomic and other data
- The Broad Institute at MIT shares data widely through a data enclave system
What is the proposed NIH GWAS policy?

The proposed National Institutes of Health (NIH) GWAS Policy calls for investigators funded by the NIH for GWAS

1) to submit de-identified genetic (genotypic and phenotypic) data to a centralized NIH repository; and,

2) to submit documentation that describes how the investigators will protect privacy and confidentiality of research participants.
NIH GWAS Policy applies to:

- **Data Management**
The proposed policy defines how investigators may submit and request access to GWAS data for research purposes. Investigators using GWAS data from the NIH database will continue to ensure the privacy and confidentiality of the individuals that participated in the original genetic association studies.

- **Publication**
The NIH proposes that investigators who submit GWAS data to the repository should be given time to analyze and publish their results before additional investigators would be permitted to submit manuscripts for publication using those data. Secondary investigators will be asked to acknowledge the submitting investigator(s) in all publications utilizing GWAS data.

- **Intellectual Property**
In the proposed policy, NIH would encourage patents for downstream discoveries that would be necessary to develop products to meet public health needs, while discouraging obtaining a patent for early, pre-competitive information that may impede future research.
Goals of the Proposed Policy

- Advance science for the benefit of the public through the creation of a centralized NIH GWAS data repository.

- Facilitate research and medical science to better address the health needs of people based on individual genetic information.
GWAS Data Management Overview

1. Data Collection
   - Research Participants
   - Informed consent

2. Submission & Management of Data
   - Submitting Investigators
   - Identifying information removed, replaced with random unique code

3. Distribution & Secondary Use of Data
   - GWAS data repository
   - Recipient Investigators
   - Data Access Request for Coded data
Data Submission

- All submissions should include an
  - IRB Certification
  - Statement from the institutions that submission is consistent with applicable laws, regulations, institutional policies

- IRB Certification
  - Consistent with informed consent
  - Risks to individuals, families, groups considered
  - Restrictions on data use identified
  - Data collected consistent with Common Rule
  - De-identified per GWAS standard
HIPAA Identifiers

- Names
- All geographic subdivisions smaller than a state, except for the initial three digits of a ZIP Code if certain conditions met.
- All elements of dates (except year), including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates unless aggregated into a single category of age 90 or older.
- Telephone numbers
- Facsimile numbers
- Electronic mail addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers.
- Certificate/license numbers.
- Vehicle identifiers & license plate numbers.
- Device identifiers and serial numbers.
- URLs.
- IP addresses.
- Biometric identifiers.
- Full-face photographic images and any comparable images.
- Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification.
Participant Issues – Data Submission

- Local IRB approval required prior to submission to GWAS data repository
- Submission accompanied by institutional statement that data is provided in accord with all applicable laws and regulations
- Information regarding any limitations on data use is requested at time of application (e.g., limitations imposed by existing informed consent)
- The GWAS Database itself would not be engaging in human subjects research
  - Data will be coded by submitting investigators
  - Agreements will be signed stipulating that key codes will not be shared
Informed Consent Comments

Points for IRBs to Consider

- Description of GWAS activities
- Appropriate research use
- Data distribution
- Risks: to individuals, families, groups and populations
  - Privacy and use of genetic information
- Follow up such as return of results, participant withdrawal, possible commercial use
Data Repository Issues

- Integrity of the data and the process
- Security procedures
  - Variable strategies based on “risk” within data and potential use
  - Policy language non-restrictive for flexibility going forward
  - Potential vulnerability to Freedom of Information Act requests and law enforcement requests
- Operating Policies and Procedures
  - Transparency and assurance regarding on-going attention to technical and policy issues
  - Security and access procedures
  - Provide source content for GWAS oversight bodies (internal and external) to review and maintain (internal)
Data Access – DACs

Committee structure
- Institute/Center/Program specific
- Federal expertise from relevant scientific areas and bioethics/human subjects issues
- Coordinate practices and develop collective experience through RPSC and SOC within the GWAS framework

Request review function
- Scientifically and ethically appropriate proposed use
- Consistency with pre-defined “appropriate use” for a given dataset
- May consult with other experts as needed
  - Scientific, privacy, or potential community/group harm concerns
Participant Issues – Data Access

- OHRP has confirmed that secondary data users will not be conducting human subjects research under 45 CFR 46.
- Access requests may ask for proposed research use of data.
  - Specified research use parameters should respect original informed consent provisions.
- Access requests will stipulate that requestors will NOT:
  - Attempt to identify individuals within the study.
  - Share the data with third parties.
- Investigators and home institutions will be responsible for compliance with federal, state, and local policies, such as:
  - HIPAA.
  - 45 CFR 46.
  - Local institutional review.
Data Access – Secondary User

■ Request application to include Data Use Certification agreement co-signed with home institution
  – Use of the data for approved purpose
  – Protect data confidentiality
  – Follow security protections
  – Follow applicable laws, etc.
  – No attempt to identify individuals
  – Non-transferability
  – Summary of proposed use and name/organization posted to GWAS database
  – Provide annual progress reports
Data Access – Secondary User

Extended Security Model

- Concern that assurance model insufficient for the nature and potential sensitivity of the data
- PI to develop a security plan and implement appropriate security practices
- Specifics evolving through interaction with expert advisors to NCBI - recommendations back to GWAS
- Policy language proposed to reflect core areas to be addressed:
  - Physical security
  - Information technology security
  - User training
Scientific Publication

- Period of exclusivity for Primary Investigators
  - Proposed period 12 months

- Acknowledgement of contributing investigators and funding organization
Intellectual Property

- NIH urges that genotype-phenotype associations remain available to all investigators, unencumbered by IP claims
- NIH discourages premature claims on pre-competitive information
- NIH encourages broad use of NIH supported genotype-phenotype data consistent with NIH’s Best Practices for Licensing with Genomic Inventions
Legal and regulatory considerations

- **De-identification Standard**
  - Common Rule: identity must be “readily ascertainable”
  - HIPAA Privacy Rule:
    - No inclusion of the 18 identifiers
    - No knowledge of identifiability within data
  - Genomic data is the ultimate identifier: what is not identifiable now is plausible to be so in the future

- Data held by the government is subject to FOIA unless protected by a waiver for privacy

- Control of access to data by law enforcement

- Certificate of Confidentiality
Shared responsibility and ownership

- NIH policy aims at maximizing public benefit of public investment, but NIH is one of several components in a complex process
- Participants, investigators and sponsors all have claims to data
- Primary study/investigators & submitting institutions own the consent process
For further information

- NIH GWAS web site

- GWAS Proposed Policy