Responsible Use of Human Genomic Data

An Informational Resource

National Institutes of Health
Turning Discovery Into Health

Purpose of Informational Resource

- To support awareness of the ethical responsibilities associated with responsible use of genomic information (including GSR)
- To promote the NIH intent that use of genomic information be for research or health purposes

(for more information, see NIH Notice Number: NOT-OD-19-023)

Genomic Data Sharing

Potential Benefits

- Enables data to be used to explore a wide range of additional research questions to advance human health
- Increases statistical power and scientific value by enabling data from multiple studies to be combined
- Promotes transparency, reproducibility, and validation of research results
- Facilitates innovation of methods and tools for research

Sharing research data supports the NIH mission

Genomic Data Sharing

Potential Ethical Considerations

- Individual re-identification from de-identified data may be possible as technological capabilities increase and data sources become more readily available
  - Sophisticated statistical methods using GSR could be used to determine if an individual participated in a specific research study (if the researcher had access to the individual's genomic data)
- Individuals, families, or populations may be stigmatized
  - Risks may be higher for identifiable communities, small sample size studies, and underrepresented study populations

NIH Genomic Data Sharing (GDS) Policy

- Premise: Respect for and protection of the interests of research participants are fundamental to the NIH's stewardship of human genomic data
- Purpose: Set forth expectations and responsibilities that ensure the broad and responsible sharing of genomic research data in a timely manner
- Policy: Applies to all NIH-funded research generating large-scale human or non-human genomic data and their use for subsequent research
  - Examples: Genome Wide Association Studies (GWAS), Single Nucleotide Polymorphism (SNP) arrays, genome sequences, expression data (e.g., transcriptomic, epigenomic, genetic)
- Access: Tiered system for human genomic data
  - Unrestricted: accessible to anyone via public website
  - Controlled: accessible upon approval if aligned with consent and any other established criteria for re-use

Genomic Summary Results

- Genomic Summary Results (GSR) are summary genomic data generated from primary analyses of genomic research across many individuals (also referred to as "aggregate genomic data" or "summary data"), includes:
  - Frequency information (genotype counts and frequencies or allele counts and frequencies)
  - Association analysis statistics (effect size estimates, standard errors, p-values, etc.)
- **GSR under NIH Genomic Data Sharing Policy**
  - May be computed by relevant NIH-designated data repository or provided by study investigator (unpublished analyses or final published results)
  - Submitting institutions determine if unrestricted access is appropriate for GSR from their studies
    - For most studies, risks of sharing GSR are low and unrestricted access is appropriate
    - For some studies with sensitivities related to individual privacy or the potential for group harm, GSR may be more appropriately maintained in controlled-access

Conditions for Use of Unrestricted-Access GSR

- NIH expects that GSR will be used responsibly and that users:
  - Follow applicable NIH policies and best practices
  - Review informational resources available on NIH-designated data repositories detailing appropriate uses of genomic data, including GSR
  - Promote scientific research or health through any use of GSR
  - Do not attempt to re-identify or contact any individual or group within a study population, or generate information that could allow participants' identities to be readily ascertained

Human Genomic Data Sharing: Considerations for Data Submission

- Appropriateness of access level (unrestricted vs. controlled for both individual-level data and GSR)
- Informed consent and any exceptions for data submission
- Awareness of and respect for cultural and/or community-based concerns
- Institutional certification and IRB determinations of consent applicability & data protection
- Data repository security framework

Protections and Safeguards for Human Genomic Data - Access

- The data requestor and the requestor's institution sign a Data Use Certification (DUC) agreeing to the NIH terms of the GDS study's Institutional Certification
- NIH Data Access Committees assess whether the proposed research is consistent with the Data Use Limitations afforded by the Institutional Certification
- A Certificate of Confidentiality protects "identifiable, sensitive information" (as defined under section 301(d) of the Public Health Service Act (42 U.S.C 241)) that is gathered or used during applicable research. NIH-funded research that involves the generation or use of individual level, human genomic data from biospecimens, regardless of whether the data is de-identified, is protected by a Certificate. Because the protection provided by the Certificate applies to all copies of the information, any investigator or institution not funded by NIH who receives a copy of identifiable, sensitive information protected by a Certificate, is also subject to the requirements of section 301(d) of the Public Health Service Act.

Human Genomic Data Sharing: Good Practices for Data Access and Use

- Use requested datasets solely in connection with project approved for data use
- Prevent efforts to attempt to identify or contact individual participants from whom these data were collected without appropriate approvals from the relevant IRBs
- Do not distribute data to any entity or individual beyond those specified in a project request
- Adhere to computer security practices that ensure only authorized individuals can gain access to data
- Do not disseminate or submit for publication reports on work prior to the release date listed for underlying datasets
- Acknowledge intellectual property policies as specified
- Report any inadvertent data release, breach of data security, or other data management incidents

Data Security and Protections

- Investigators who are approved to use the data are expected to follow guidance on security practices that outline expected data security protections
  - To ensure that data are kept secure and
  - Not released to any person not permitted to access it
- Examples
  - Physical security measures
  - User training

See dbGaP security Best Practices Requirements:
http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf