

How Does the NHGRI Ensure Open Access to Genomics Databases and Unrestricted Use of Data?

Open Access to Public Sector Information
Brisbane, Australia
July 13, 2007

Claire T. Driscoll

Director, Technology Transfer Office

National Human Genome Research Institute (NHGRI)

National Institutes of Health (NIH)

U.S. Department of Health & Human Services (DHHS)



National Institutes of Health



- Largest financial backer of basic biomedical research in the world
- Mission: Uncover new knowledge that will lead to improved public health by conducting and supporting research



NIH's Gifts to the World: Public Sector Information "Treasures" including....

■ GenBank

- Contains human sequence data collected by the Human Genome Project and many other related NIH-funded genomics projects
- Many analysis tools & other features
- **Cost-free, full, open access to all with no restrictions on use**

■ MedLine & PubMed

- Database of Biomedical Literature (1000s of journals)
- Many full text & downloadable electronic files available for FREE
- Can be used and accessed by anyone w/ no restrictions on use imposed by the NIH



Example NIH Research Efforts—All Involve Collection of Genomic Info & Scientific Data

- Human Genome Project (HGP) & The International Genomics Consortium
- Single Nucleotide Polymorphism (SNP) Consortium
- The Pharmacogenetics Research Network
- International Haplotype Project (HapMap)
- Mammalian Gene Collection (MGC)
- Encyclopedia of DNA Elements (ENCODE)
- Knock-out Mouse Project (KOMP)
- Genetic Associations Information Network (GAIN)
- Molecular Libraries project (part of NIH Roadmap)
- Genes and Environment Initiative (GEI)
- Cancer Genome Atlas Project (CGAP)





Human Genome Project...Early, Key Data Release & Use Principles

- **"..access to all sequences and materials generated by these publicly funded projects should and even must be made freely available"** National Research Council on Mapping and Sequencing the Human Genome, 1988
- **"...the data must be in the public domain, and the redistribution of data should remain free of royalties."** NIH Ad Hoc Program Advisory Committee on Complex Genomes, 1988





A Vision for the Future of Genomics Research

“An important lesson of the Human Genome Project has been the benefit of immediately releasing data from large-scale sequencing projects....*Scientific progress and public benefit will be maximized by early, open and continuing access to large data sets...*The development of effective systems for achieving the rapid release of data without restrictions and providing continuing access...should be an integral component of the planning and development of new [scientific] community resources.

Collins, F. *et al Nature* 242: 835 (2002)



NIH Statement on Data Sharing (for Grantees) 2003

- “Data sharing is essential for expedited translation of research results into knowledge, products and procedures to improve human health.”
- Proposed data-sharing plans are now a required part of large NIH grant applications (e.g., all with >\$500K/yr in funding)
- Policy makes concessions for exceptions (e.g., there are legitimate reasons for not always being able to share info-notably patient info)
 - DEFAULT POSITION = YOU MUST SHARE YOUR DATA!!
- http://grants2.nih.gov/grants/policy/data_sharing/





“Typical” Data Release, Use & Re-Use Terms for a NHGRI Project

- Data available in the public domain for any and all uses (research uses and commercial uses) without restriction
- Rapid deposition of data (often within 24 hours) of generation or at latest after data validation
- Original/Primary Data generators given the exclusive opportunity to publish first (within set timelines depending on the specific data set)
 - Data can still be used by others during the time period the original data generators are preparing manuscripts
 - Data Users must not publish during the agreed upon “waiting period”



NHGRI Guiding Principle for Studies Involving Human Subjects (Driver for All Policy Decisions)



The greatest public benefit will be realized if data 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 researchers.



Benefits of the NHGRI's Data Policies



- **Improve Health:** Better understand the health needs of the public and facilitate the development of improved diagnostic tools and design of new, safe and highly effective treatments
- **Maximize Public Investment:** Accelerate the discovery of associations between genetic data and disease, while minimizing research costs





Summary of Key Tenets

- **“Precompetitive” genomic info should be available in public databases without any restrictions on future use**
- **Many such data and reagent sets can be produced most effectively by Government entities and their grantees & also via public-private partnerships such as multi-party consortia**
- **Univ. and companies have little incentive to “give away” potentially licensable data and materials—government needs to lead the way**
- **Companies come and go (or change business strategies) but NIH will always be there...**



What Should We All be Doing?

For Genetic, Genomic & Other types of scientific Data:

- Adopt data release, data sharing & IP policies that foster unrestricted, cost-free and rapid access for all purposes (research and commercial) for pre-competitive/fundamental data
- Do not extend copyright/IP protection to databases (e.g., *sui generis* protection as in the EU)
- Encourage and provide funding for research collaborations & public-private partnerships that are committed to placing pre-competitive data in the public domain
- Funding agencies must provide additional resources to grantees so that they can set up the needed cyber infrastructure



For more information:



- Technology Transfer Office, NHGRI
 - **Building 12A Room 1033**
 - **tel: (301) 402-2537/ (301) 594-2235**
 - **e-mail: cdriscol@mail.nih.gov**
- National Human Genome Research Institute,
National Institutes of Health
<http://www.genome.gov>

