

# Genomic Data Sharing at NIH:

## *The GWAS Policy Example*

---



Laura Lyman Rodriguez, Ph.D.  
National Human Genome  
Research Institute  
National Institutes of Health

SACGHS Genomic Data Sharing Session  
February 4, 2010

# NIH and Data Sharing



*“We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The NIH endorses the sharing of final research data to serve these and other important scientific goals.”*

*- NIH 2003 Data Sharing Policy*

# Why was GWAS different?

- Unprecedented opportunity to advance understanding of common diseases (e.g., diabetes, cancer, heart disease)
- The data generated is far richer than what a single investigator or a collaborative team can fully explore
  - Many different questions may be asked
  - Cross-study analyses are possible, which increases the capacity to address complex questions
- NIH leadership felt that a consistent and robust GWAS policy across the ICs would best serve the public

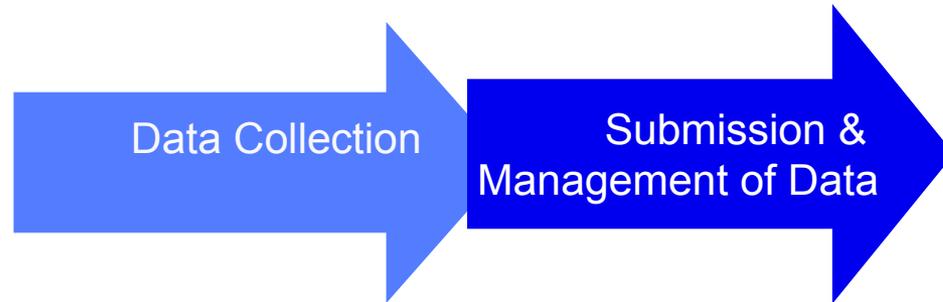
# Guiding Principle

**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.**

# GWAS Policy Elements

- **Data Management**
  - **Data Submission Procedures**
  - **Data Access Principles**
  - **Protection of Research Participants**
- **Scientific Publication**
- **Intellectual Property**

# GWAS Data Management Overview



**Research  
Participants**



→  
**Informed  
Consent**

**Submitting  
Investigators**



→  
**Identifying  
information  
removed,  
replaced with  
random  
unique code**

**GWAS  
Data Repository**



NCBI

All Databases PubMed Nucleotide

Search WGA for all[sb]

Limits Preview/Index History Clipt

Display Browse studies Show 20 S

All: 5941

Browse WGA

NCBI

PubMed Search Overview What's New Help FAQ

Other Services Journal Browser MeSH Browser Citation Matcher Clinical Queries

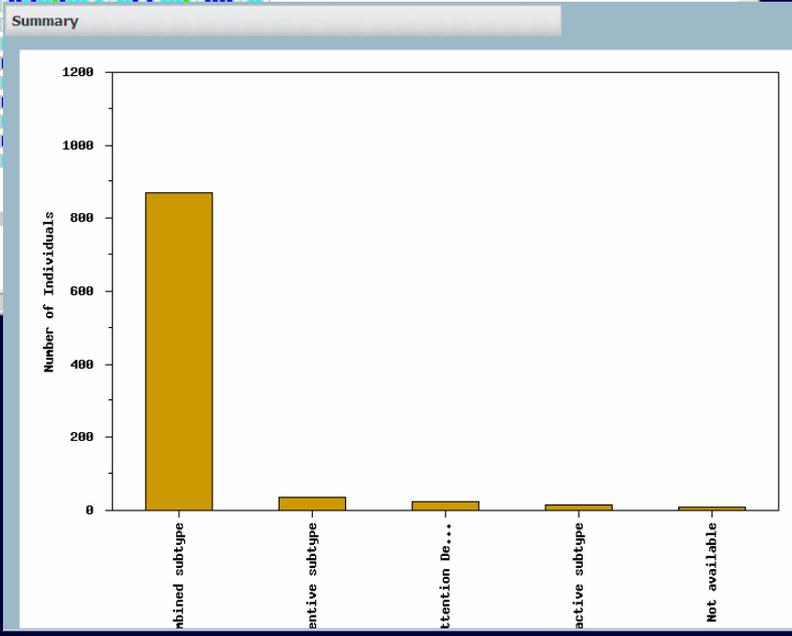
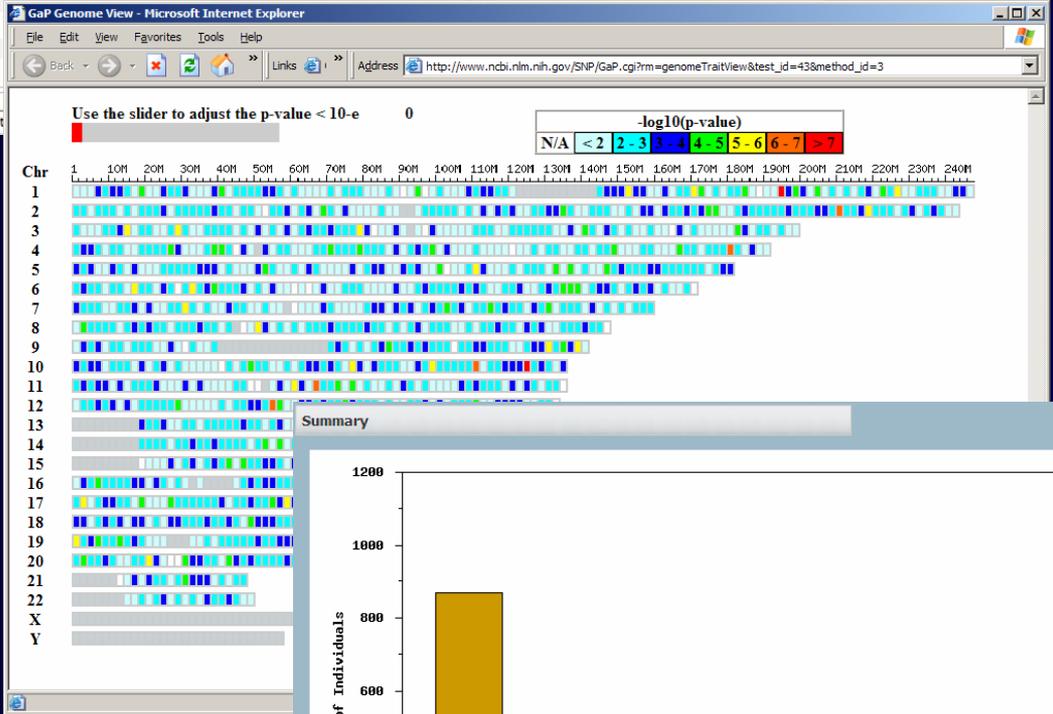


My NCBI

Welcome mmailman. [Sign Out]

PMC Journals Books

Study	Sub-Studies	Variables	Documents	Participants	Type of Study	Status
Age-Related Eye Disease Study (AREDS)	-	182	23	600	case-control	
NINDS Parkinsonism Study		400		250		
NINDS Parkinsonism Study - Cases						



## NCBI WGA Document Age Related Eye Disease Study

### Chapter 7 EXAMINATION PROCEDURES

#### 7.1 INTRODUCTION

The procedures for carrying out the examinations required in the study are described in this chapter. Required ocular examinations include refraction and visual acuity measurements, intraocular pressure measurement, and ophthalmoscopic examination. General characteristic assessments include measurement of height, weight, and blood pressure and determination of past medical history. Risk factor assessments will require the administration of the food frequency and sunlight exposure questionnaires as well as collection of blood specimens. Procedures for participant identification, masking, distribution and management of the supplementation, adherence assessment, and home visit examination are also described. Procedures for taking photographs of the lens and fundus are described in detail in Chapter 8. The schedule and description of participant visits in Chapter 6 outline the examinations required during each visit.

#### 7.2 REFRACTION AND VISUAL ACUITY

A manifest refraction and visual acuity measurement according to the detailed study protocol must be performed during (a) the Qualifying Visit when the visual acuity score using Chart R is 73 letters or less in at least one eye, (b) the Randomization Visit, (c) Annual Visits, and (d) any Nonannual Visit when the visual acuity score using Chart R has dropped by 10 letters or more compared to the Randomization Visit score for the first time. Participants' pupils should not be dilated at the time of visual acuity testing at any study visit, except they may be dilated during the Qualifying Visit. Pinhole acuity will not be tested as part of AREDS. At the Qualifying Visit, visual acuity may be initially assessed utilizing the participant's current distance glasses. At the Nonannual Visits, visual acuity is initially assessed utilizing the previously obtained manifest refraction. Participants will be asked to read the letters on Chart R only (not Charts 1 or 2), using the equipment described in Section 7.2.1. They will start reading from the top left-most letters--first with the right eye and then with the left eye. A visual acuity score will be calculated as described in Section 7.2.3.3. If at the Qualifying Visit the visual acuity is 74 letters or more in each eye or if at a Nonannual Visit the visual acuity is within nine letters of the Randomization Visit score in each eye, or a vision drop has already been documented in each eye, the visual acuities measured will be entered on the study form. For these participants, a manifest refraction and measurement of best-corrected visual acuity, using the detailed protocol (Sections 7.2.1 - 7.2.3), will not be required.

##### 7.2.1 Visual Acuity Equipment and Facilities

**7.2.1.1 Introduction.**—The visual acuity of participants will be measured according to the standard procedure developed for the Early Treatment diabetic Retinopathy Study (ETDRS) and adapted for AREDS. The procedure is described in this section. The following equipment is used in AREDS: a set of three Lighthouse Distance Visual Acuity Test charts (second edition), which are modified ETDRS Charts 1, 2, and R, 1 and a retroilluminated box providing standardized chart illumination, as modified from the design by Ferris and Sperduto. 2 The charts and boxes are manufactured by:

Lighthouse Low Vision Products  
36-02 Northern Boulevard  
Long Island, New York 11101

AREDS: Age Related Eye Disease Study

# GWAS Data Management Overview



**Research Participants**

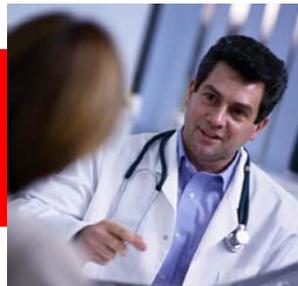
**Submitting Investigators**

**GWAS Data Repository**

**Recipient Investigators**



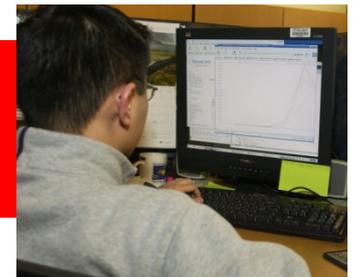
**Informed Consent**



**Identifying information removed, replaced with random unique code**



**Data Access Request for Coded data**



**Data Use Limitations**

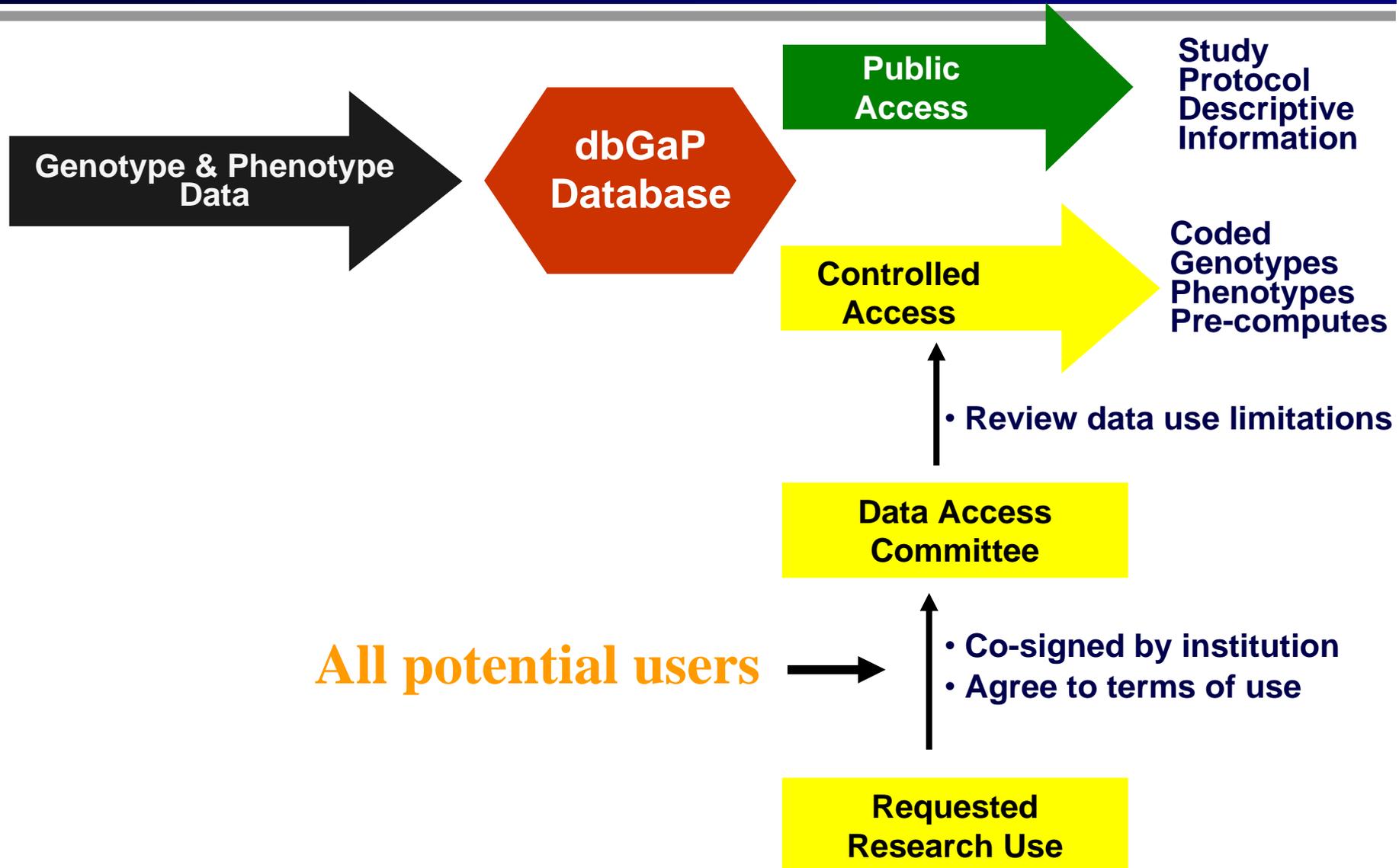
# Data Submission

- **Local institution** will certify approval of submission to GWAS data repository, including statements that:
  - data are provided in accord with applicable laws and regulations
  - an **IRB** or Privacy Board has reviewed the submission plans
- The **PI** will remove HIPAA identifiers and retain the keycode to the data
- Any limitations on data use are requested at time of application (e.g., limitations imposed by existing informed consent).

# Points to Consider for IRBs

- Provides investigators & IRBs with information on important participant protection considerations related to submission of data
- Not intended to serve as a checklist
- Topics include:
  - Background on the scientific opportunities presented by GWAS
  - Discussion of the ethical issues relevant to the review of submission plans for GWAS datasets
  - Specific points to consider in the evaluation of informed consent documents
- Available at: [http://grants.nih.gov/grants/gwas/gwas\\_ptc.pdf](http://grants.nih.gov/grants/gwas/gwas_ptc.pdf)

# Data Access is Two-Tiered



# Data Access

- **Investigators and home institutions** responsible for compliance with federal, state, and local policies
  - Local institutional review – HIPAA – 45 CFR 46
- Secondary data users not conducting human subjects research
- Data Access Committees (DACs) will review requests for consistency with data use limitations
  - Federal staff with appropriate expertise
  - Also responsible for tracking and reviewing Approved User Annual Reports

# Data Use Certification Agreement

- There is a common framework for all NIH Data Use Certifications (DUCs)
- Terms and conditions include that requesters will:
  - be responsible for compliance with federal, state, and local policies
  - only use the data for the specified research use
  - not identify study participants
  - not transfer data beyond approved users
  - immediately notify the DAC if a security breach occurs
  - submit brief annual updates on research and publications
  - be identified as an Approved User within dbGaP
  - acknowledge other GWAS policies

# GWAS Policy Elements

- **Data Management**
  - Data Submission Procedures
  - Data Access Principles
  - Protection of Research Participants
- **Scientific Publication**
- **Intellectual Property**

# Scientific Publication

Browse dbGaP

By Studies

By Diseases

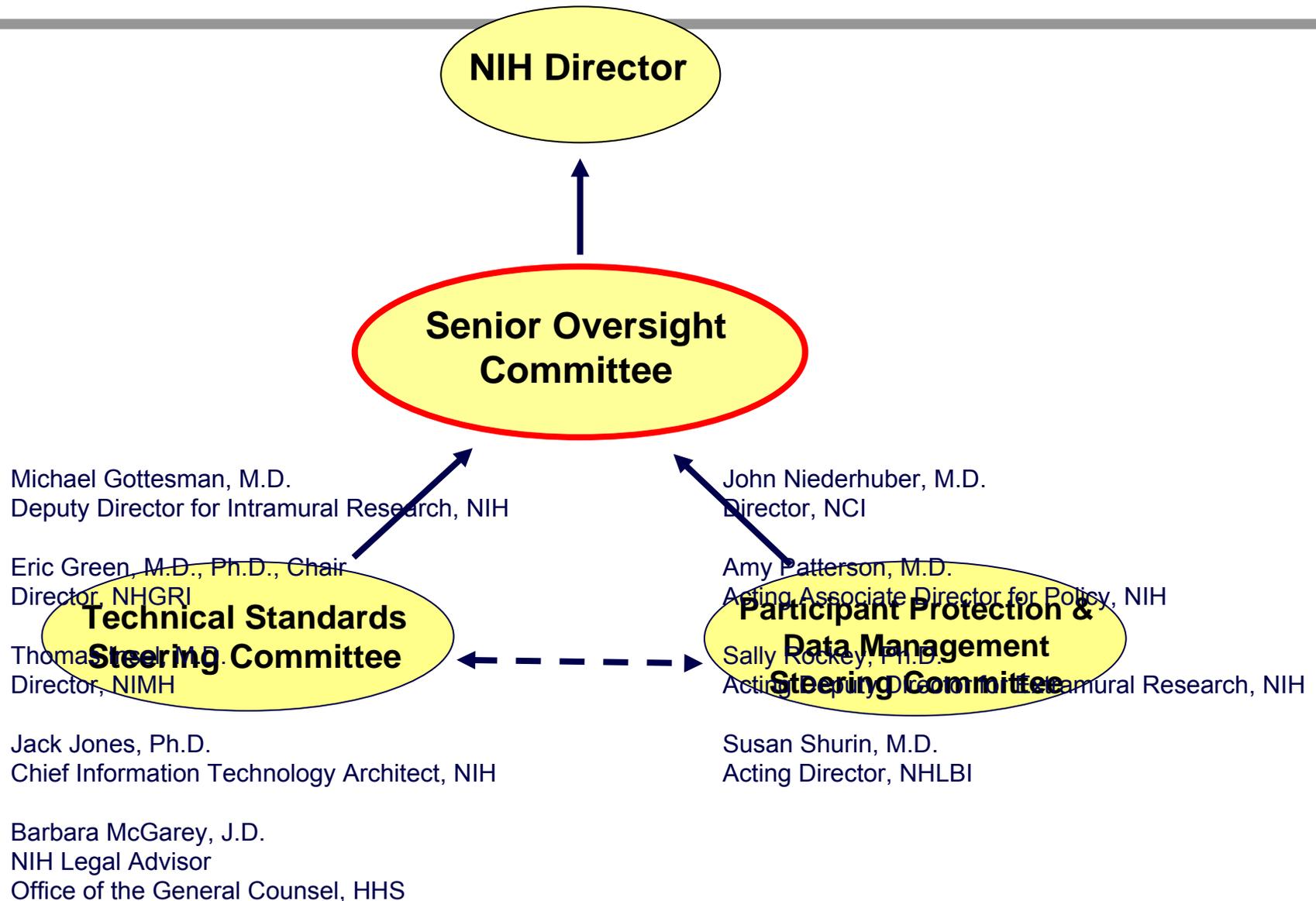
Advanced Search

Study	Embargo Release	Details
 <a href="#">Framingham SHARe</a>	Version 1: Oct 19, 2008. Version 2: Feb 01, 2009.	
 <a href="#">GAIN: Collaborative Association Study of Psoriasis</a>	Aug 13, 2008	
 <a href="#">GAIN: Genotyping the 270 HapMap samples for GAIN by Broad</a>		
 <a href="#">GAIN: Genotyping the 270 HapMap samples for GAIN by Perlegen</a>		
 <a href="#">GAIN: International Multi-Center ADHD Genetics Project</a>	Mar 26, 2008	
 <a href="#">GAIN: Linking Genome-Wide Association Study of Schizophrenia</a>	Version 1: Nov 07, 2008. Version 2: Dec 11, 2008.	

# Intellectual Property

- Consensus is that GWAS data should be pre-competitive
  - Automated calculations to identify first round genetic associations will be made available through dbGaP
- NIH urges that associations remain available to all investigators & discourages premature claims
- Users & data submitters must “acknowledge” this position
- NIH encourages broad use of GWAS data consistent with NIH’s Best Practices for Licensing with Genomic Inventions.

# NIH GWAS Governance & Oversight



# For More Information...

- Watch [gwas.nih.gov](http://gwas.nih.gov)...update underway
- Email [gwas@mail.nih.gov](mailto:gwas@mail.nih.gov)

U.S. Department of Health & Human Services [www.hhs.gov](http://www.hhs.gov)  
[www.nih.gov](http://www.nih.gov)

Genome-Wide Association Studies (GWAS)

Google™ Custom Search Search

Home  
GWAS  
Data Repository  
Policy  
Policy Oversight  
Related Resources  
Researchers  
Participants (General Public)  
Institutions & IRBs

GWAS Data Repository Policy

### Introduction

Genome-wide association studies (GWAS) are used to identify common genetic factors that influence health and disease. In January 2008, the NIH implemented a new policy for the sharing of data obtained in NIH-supported or conducted GWAS. The purpose of the policy is to foster science for the benefit of the public through the creation a centralized NIH GWAS data repository. This Website supports the GWAS policy's implementation.

The NIH will continue to release additional information about the NIH GWAS policy on this site. Please e-mail questions about the policy to [GWAS@mail.nih.gov](mailto:GWAS@mail.nih.gov).

### In the Spotlight

RECENT NEWS



# Trends in Access Requests

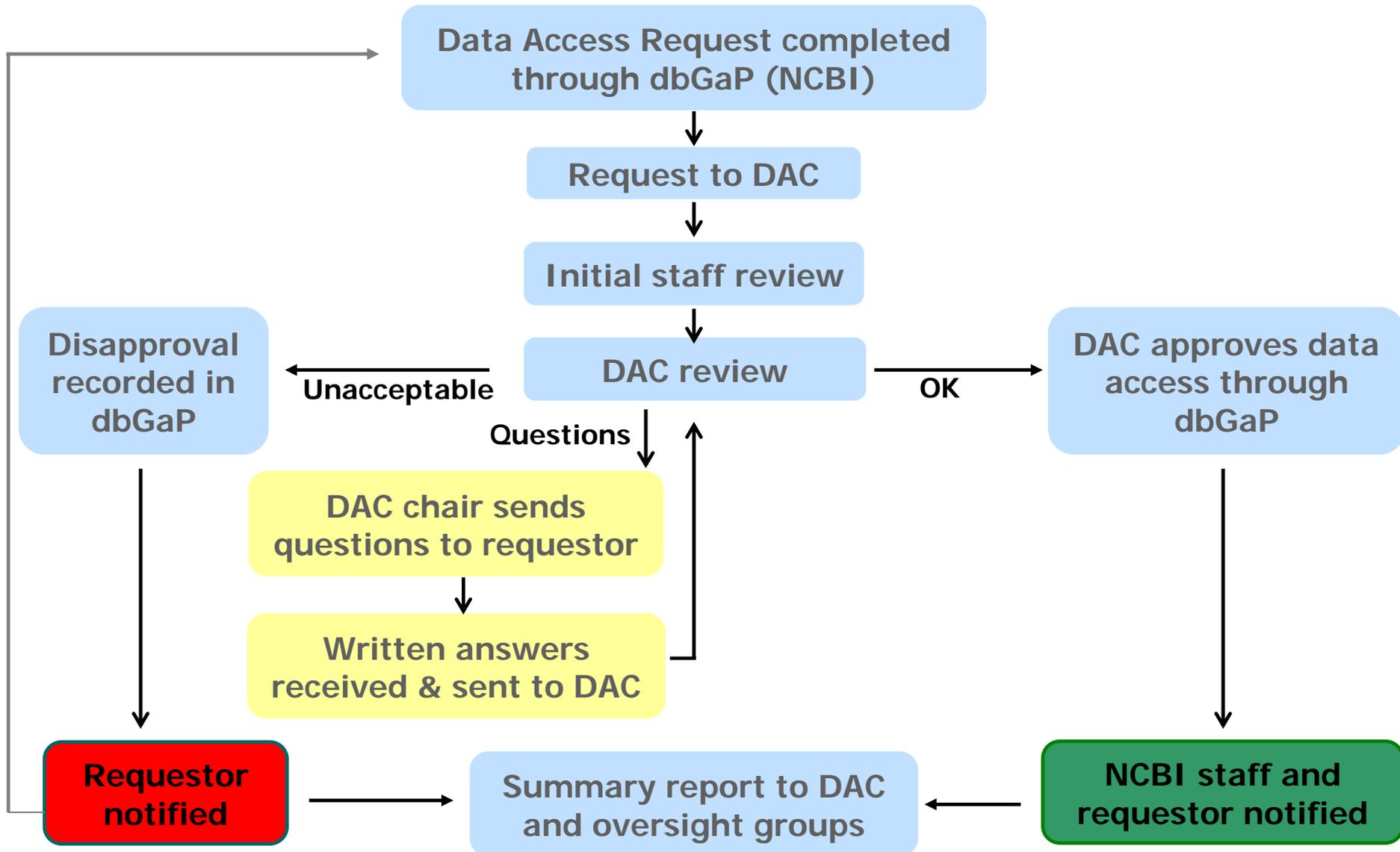
- Requestors come from across the research community, with most residing in academic institutions
- No sanctions have been identified for PIs or their collaborators
- Proposed research uses include:
  - understanding the etiology of the target disease or related conditions/traits
  - testing new statistical methods to identify disease susceptibility genes or gene-environment interactions

# dbGaP by the Numbers

## As of Fall 2009:

- 39 deposited studies involving 79 institutions
- 57,612 phenotypes measured
- Over 500 approved users with at least 1 project
- Users from 196 institutions in 25 countries
- 48 additional studies in process

# Data Request Review Procedures



# Annual Report Elements

- Summary of research progress
- Proposed plans for further research utilizing currently approved NIH GWAS datasets
- List of all completed or accepted scientific presentations that include (or will include) findings made with the individual-level NIH GWAS data accessed through dbGaP.
- List of manuscripts submitted
- Description of any intellectual property generated as a result of using the NIH GWAS individual-level data
- Summary information on any inappropriate data release incidents or other data security issues
- General comments on process & Suggestions for improving dbGaP, NIH GWAS, study-specific data access, or NIH GWAS policy or procedures in general

# Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays

Nils Homer<sup>1,2</sup>, Szabolcs Szelinger<sup>1</sup>, Margot Redman<sup>1</sup>, David Duggan<sup>1</sup>, Waibhav Tembe<sup>1</sup>, Jill Muehling<sup>1</sup>, John V. Pearson<sup>1</sup>, Dietrich A. Stephan<sup>1</sup>, Stanley F. Nelson<sup>2</sup>, David W. Craig<sup>1\*</sup>

<sup>1</sup> Translational Genomics Research Institute (TGen), Phoenix, Arizona, United States of America, <sup>2</sup> University of California Los Angeles, Los Angeles, California, United States of America

## Abstract

We use high-density single nucleotide polymorphism (SNP) genotyping microarrays to demonstrate the ability to accurately and robustly determine whether individuals are in a complex genomic DNA mixture. We first develop a theoretical framework for detecting an individual's presence within a mixture, then show, through simulations, the limits associated with our method, and finally demonstrate experimentally the identification of the presence of genomic DNA of specific individuals within a series of highly complex genomic mixtures, including mixtures where an individual contributes less than 0.1% of the total genomic DNA. These findings shift the perceived utility of SNPs for identifying individual trace contributors within a forensics mixture, and suggest future research efforts into assessing the viability of previously sub-optimal DNA sources due to sample contamination. These findings also suggest that composite statistics across cohorts, such as allele frequency or genotype counts, do not mask identity within genome-wide association studies. The implications of these findings are discussed.

# Early Headlines...

*Los Angeles Times*

## **DNA databases blocked from the public**

**The National Institutes of Health removes patients' genetic profiles from its website after a study reveals that a new type of analysis could confirm identities.**

By Jason Felch

Los Angeles Times Staff Writer

## **Good for Cops, Bad for NIH**

By Jennifer Couzin

*ScienceNOW* Daily News

29 August 2008

## **Forensic Breakthrough Stirs NIH to Close GWAS Data from Public View**

August 29, 2008

*By Matt Jones,*

*a GenomeWeb staff reporter*

# Inferring Placement from Allele Frequencies

Snp	Allele Frequency ( $Y_{ij}$ )	Interpretation at the given SNP
	0.0      0.25      0.50      0.75      1.0	
j		
j+1		
j+2		

Y = Person of Interest; Pop = Reference Population; M = Mixture

Homer N et al, *PLoS Genet* 2008 Aug 29;4(8):e1000167.

# dbGaP by the Numbers

As of Fall 2009:

- 39 deposited studies involving 79 institutions
- 57,612 phenotypes measured
- Over 500 approved users with at least 1 project
  - Investigators span research sectors, but primarily reside in academic-based institutions
- Users from 196 institutions in 25 countries
- 48 additional studies in process