

NIH Policy for Data Sharing of NIH Supported Genome Wide Association Studies

Elizabeth G. Nabel, M.D.
Director, National Heart, Lung, and
Blood Institute

**Council of Public
Representatives**

April 18, 2008



NIH GWAS Policy

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome- Wide Association Studies (GWAS)

AGENCY: National Institutes of Health,
HHS.

ACTION: Notice.

Background

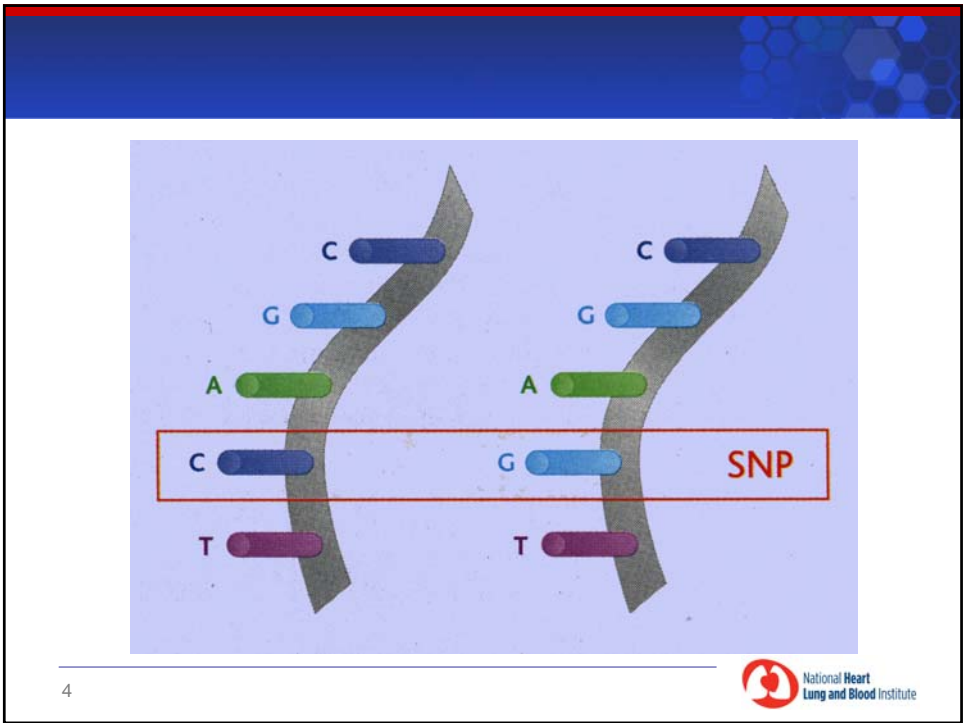
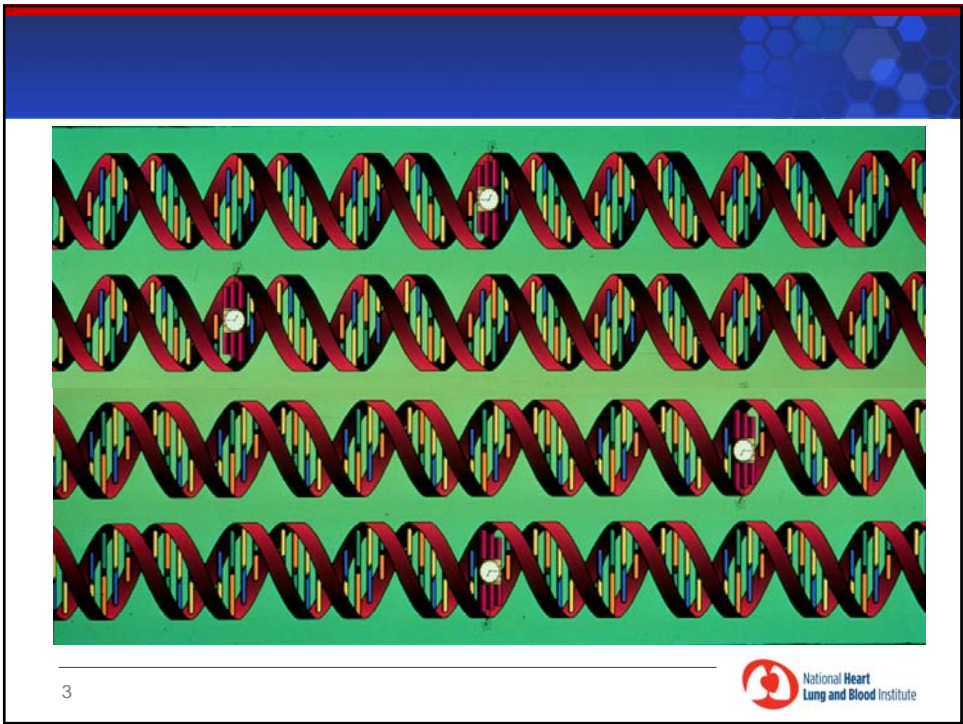
The NIH is interested in advancing genome-wide association studies (GWAS) to identify common genetic factors that influence health and disease. For the purposes of this policy, a genome-wide association study is defined as any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood

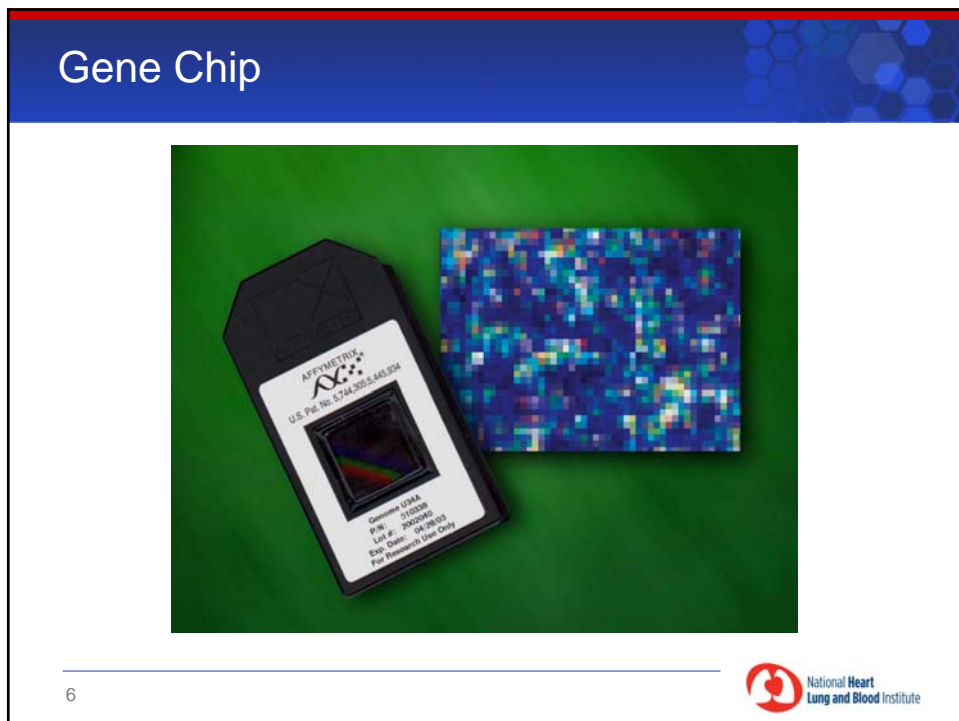
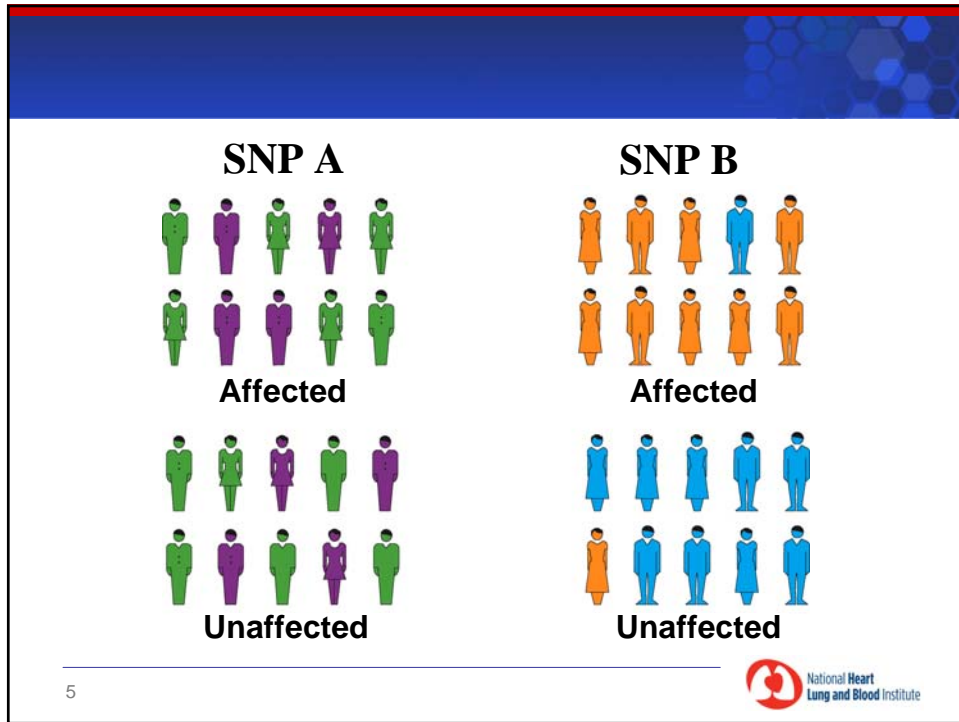
The screenshot shows the NIH Office of Extramural Research (OER) website. The main heading is "Genome-Wide Association Studies (GWAS)". The text states: "The NIH is interested in advancing genome-wide association studies (GWAS) to identify common genetic factors that influence health and disease. For the purposes of this policy, a genome-wide association study is defined as any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood pressure or weight), or the presence or absence of a disease or condition. Whole genome information, when combined with clinical and other phenotype data, offers the potential for increased understanding of basic biological processes affecting human health, improvement in the prediction of disease and patient care, and ultimately the realization of the promise of personalized medicine. In addition, rapid advances in understanding the patterns of human genetic variation and maturing high-throughput, cost-effective methods for genotyping are providing powerful research tools for identifying genetic variants that contribute to health and disease. The purpose of this Website is to support the implementation of the GWAS Policy." Below the text, there is a "Policy Guidance" section with a link to "Governance Structure (PowerPoint - 37 KB)". At the bottom of the page, the "NIH GWAS Policy" link is highlighted.

Federal Register: August 28, 2007
Policy Effective: January 25, 2008

NIH OER Website:
<http://grants.nih.gov/grants/gwas/index.htm>

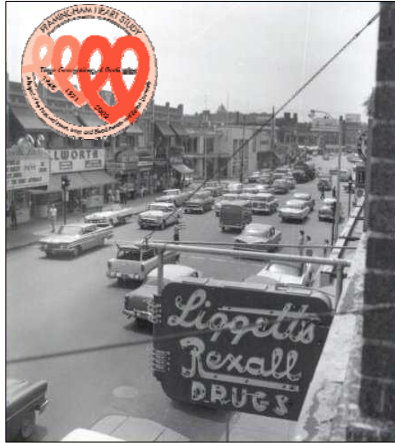






Framingham Heart Study

Downtown Framingham, MA
(circa 1960)



Risk Factors for Heart Attack and Stroke

- High blood pressure
- High cholesterol
- Cigarette smoking
- Diabetes mellitus
- Parental or sibling history
- Obesity

7



Framingham Heart Study: Population-Based Family Study

1948 → 1958 → 1968 → 1978 → 1988 → 1998 → 2008

1948 → → → → → → 2008

Original cohort: n=5209 men and women (ages 28-62)
1644 spouse pairs, 596 extended families

1972 → → → 2008

Offspring study: n=5124 men and women (ages 5-70)
1576 spouse pairs, 3514 biological offspring

Third Generation study:
n=3500 men and women 2002 →

8



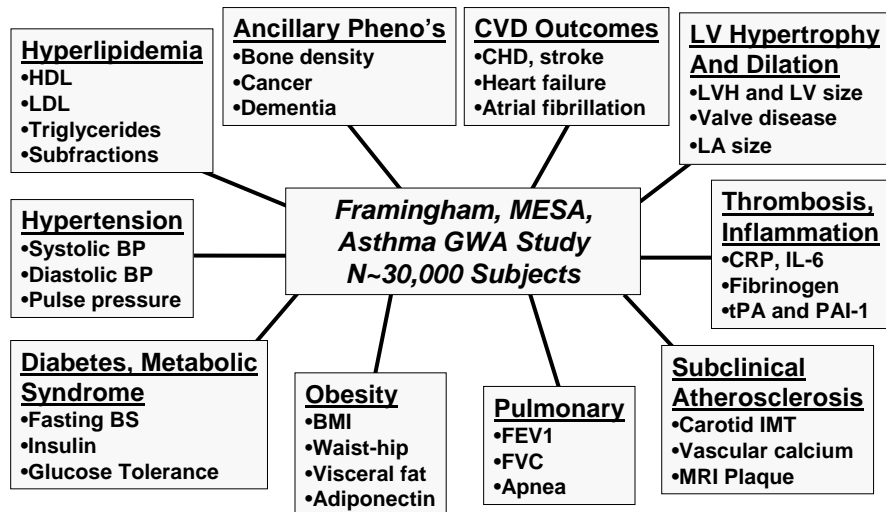
Framingham SNP Health Association Resource (SHARe)

- Genotypes: ~10,000 Caucasians from 3 generations - Affymetrix 500,000 SNP chip
- Phenotypes: >1,000 risk factor, subclinical and clinical CV phenotypes, from 60 years of exams
- Genotypes and phenotypes placed in a web-based dataset, called dbGaP, maintained at the NIH
- Framingham SHARe dataset contains 5.5 billion genotypes, >5.5 trillion association tests
- Available to biomedical researchers through dbGaP October 1, 2007

9



SHARe Phenotypes in Framingham, MESA and Asthma



10



NIH GWAS Data Sharing Policy: 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.

11



Rationale: Sharing Data

- The opportunities for scientific advances in our understanding of complex, common diseases are now extraordinary.
- The richness of the data generated is far greater than what a single investigator or a group of collaborators can explore
 - Many different scientific questions may be addressed through a single dataset and multiple types of analyses may be needed
 - The opportunity for cross-study analyses increases the capacity to address complex biological questions and can enhance data quality by increasing power
- Information which is not shared represents a lost opportunity to improve the health of the public.
- NIH has been encouraging wide sharing of information for several years.

12



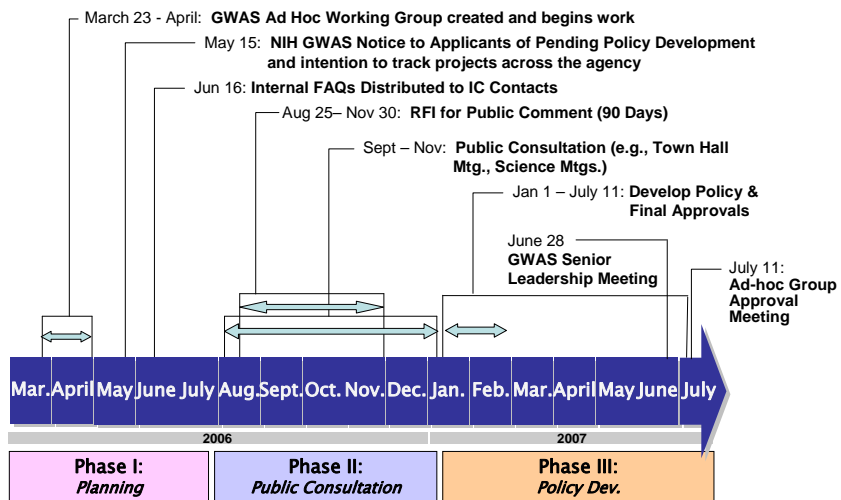
Why Now?

- The cost of extensive genotyping has fallen rapidly, and continues to fall, making studies feasible which would not have been possible even 4 years ago.
- NIH is receiving many applications for GWAS, representing many millions of dollars of research investment.
- NIH leadership believes that a consistent and robust GWAS policy across the ICs best serves the research community and the public (i.e., research participants)

13



GWAS Timeline



14



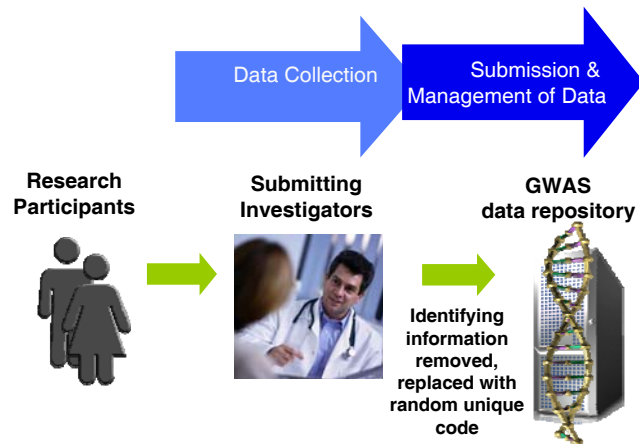
NIH Policy for GWAS

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

15



GWAS Data Management Overview



16



Data Submission

- Local institution will certify approval of submission to GWAS data repository.
- Certification will include institutional statement that data are provided in accord with all applicable laws and regulations and that an IRB or Privacy Board has reviewed the submission plans.
- 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 will not be engaging in human subjects research, according to OHRP.
 - Data will be coded by submitting investigators (no HIPAA identifiers)
 - Agreements will be signed stipulating that the identities of research participants will not be disclosed to the NIH GWAS data repository

17



Data Submission - Points to Consider for IRBs

- Intent is to provide investigators and IRBs reviewing datasets for GWAS submission with information on important participant protection considerations
- Topics include:
 - Background on the scientific opportunities presented by GWAS
 - Elements of the GWAS policy
 - Discussion of the ethical issues relevant to the review of submission plans for GWAS datasets
 - Benefits, risks and safeguards that will be used to protect the data
 - Specific points to consider in the evaluation of informed consent documents

18



dbGaP: Database of Genotype and Phenotype

dbGaP
Genotype and Phenotype

My NCBI

Published Nucleotide Protein Genome Structure OMIM PMC Journals Books

Limits Preview/Index History Clipboard Details

Browse dbGaP

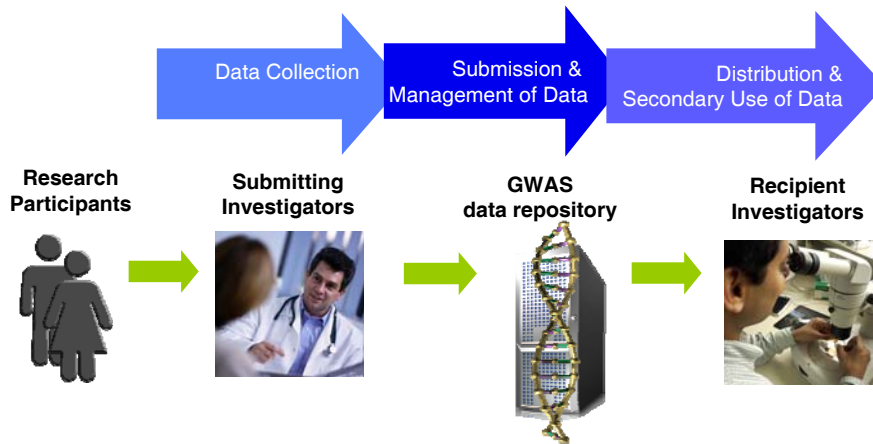
By Studies By Diseases Advanced Search

Study	Embargo Release	Details	Participants	Type of Study	Project
# Erasmus SHARE	phs000007.v1.p1: Oct 19, 2008 phs000007.v2.p1: Jan 31, 2009		14261	Community-based, longitudinal, family-based cohort	
GAIN: Collaborative Association Study of Psoriasis	Dec 31, 2008		2875	Case-control	
GAIN: Genotyping the 270k HapMap samples for GAIN by Broad			-	Parent-offspring trios	
GAIN: Genotyping the 270 HapMap samples for GAIN by Perlegen			-	Parent-offspring trios	
GAIN: International Multi-Center ADHD Genetics Project	Mar 26, 2008		2835	Parent-offspring trios	
GAIN: Linking Genome-Wide Association Study of Schizophrenia	phs000021.v1.p1: Nov 07, 2008 phs000021.v2.p1: Dec 11, 2008		5066	Case-control	
GAIN: Major Depression: Stage 1 Genomewide Association in Population-Based Samples	Jul 15, 2008		3741	Case-control	
GAIN: Search for Susceptibility Genes for Diabetic Neuropathy in Type 1 Diabetes	Jul 09, 2008		1825	Case-control	
GAIN: Whole Genome Association Study of Bipolar Disorder	Dec 30, 2008		2160	Case-control	
Ischemic Stroke Genetics Study (ISGS)			-	Case-control	
Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson's Solutions) Collaboration	Mar 03, 2008		1550	Case-control	
NEI Age-Related Eye Disease Study (AREDS)	Jun 11, 2007		600	Case-control	
NINDS Parkinson's Disease	Oct 12, 2007		535	Case-control	
NINDS Parkinsonism Study	Oct 12, 2007		1283	Case-set	
NINDS Repository: Cardiovascular Disease/Stroke Study			-	Case-set	
NINDS Repository: Neurologically Normal Control Collection			-	Control-set	
Study of Irish Amyotrophic Lateral Sclerosis (SIALS)	Oct 12, 2007		-	Case-control	
The Finland-United States Investigation of NIDDM Genetics (FUSION) study			2335	Case-control	
Whole Genome Association Study of Systemic Lupus Erythematosus			-	Case-control	

19 <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>



GWAS Data Management Overview



20



Data Access: Data Access Committees (DACs)

- Applications for datasets will be submitted to dbGaP, which then will triage applications to the responsible NIH Institute.
- Each NIH Institute has constituted a DAC for administrative review of applications. DACs are constituted by Institute staff.
- DACs review the application and relay the approval back to dbGaP who releases the dataset.
- DACs consist of Federal staff with expertise in science, bioethics, and privacy/confidentiality issues.

21



Data Access: Data Access Committees (DACs)

- DACs will determine which investigators will have access to the datasets in the NIH GWAS data repository based on the proposed research use of the data.
- DACs will also play a critical role in protecting the research participants.
- Annual reports from recipient investigators will be required as a condition for continued use of the datasets.
- DACs will review the annual reports for new issues and to confirm that research use matches approved request.

22



Data Use Certification

- Investigators and their institutions who request GWAS data must provide the following through a Data Use Certification agreement:
 - Description of proposed research projects
 - Agreement to use the data only for the approved research
 - Agreement to follow appropriate data security protections
 - Agreement to follow all applicable laws, regulations, and local institutional policies and procedures for handling GWAS data
 - Agreement not to attempt to identify individual participants within a dataset
 - Agreement not to sell any of the data elements from datasets obtained from the NIH GWAS data repository
 - Agreement not to share with individuals, other than those listed on the request, any of the data elements from datasets obtained from the NIH GWAS data repository
 - Agreement to follow the GWAS policy with regard to publication and intellectual property

23



dbGaP: Database of Genotype and Phenotype

Study	Embargo Release	Details	Participants	Type of Study	Project
# Framingham SHARE	phs000007.v1.p1: Oct 19, 2008 phs000007.v2.p1: Jan 31, 2009	[D]	14261	Community-based, longitudinal, family-based cohort	
# GAIN_Collaborative Association Study of Psoriasis	Dec 31, 2008	[D]	2875	Case-control	
# GAIN_Genotyping the 270_HapMap samples for GAIN by Broad		[D]	-	Parent-offspring trios	
# GAIN_Genotyping the 270_HapMap samples for GAIN by Perlegen		[D]	-	Parent-offspring trios	
# GAIN_International Multi-Center ADHD Genetics Project	Mar 26, 2008	[D]	2635	Parent-offspring trios	
# GAIN_Linkage Genome-Wide Association Study of Schizophrenia	phs000021.v1.p1: Nov 07, 2008 phs000021.v2.p1: Dec 11, 2008	[D]	5066	Case-control	
# GAIN_Major Depression: Stage 1 Genome-wide Association in Population-Based Samples	Jul 15, 2008	[D]	3741	Case-control	
# GAIN_Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes	Jul 09, 2008	[D]	1825	Case-control	
# GAIN_Whole Genome Association Study of Bipolar Disorder	Dec 30, 2008	[D]	2160	Case-control	
# Ischemic Stroke Genetics Study (ISOGS)		[D]	-	Case-control	
# Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson's Solutions) Collaboration	Mar 03, 2008	[D]	1550	Case-control	
# NEI Age-Related Eye Disease Study (AREDS)	Jun 11, 2007	[D]	600	Case-control	
# NINDS Parkinson's Disease	Oct 12, 2007	[D]	835	Case-control	
# NINDS Parkinsonism Study	Oct 12, 2007	[D]	1283	Case-set	
# NINDS Repository_Cerebrovascular Disease/Stroke Study		[D]	-	Case-set	
# NINDS Repository Neurologically Normal Control Collection	Oct 12, 2007	[D]	2723	Control-set	
# Study of Irish Amyotrophic Lateral Sclerosis (SIALS)		[D]	-	Case-control	
# The Finland-United States Investigation of NIDDM Genetics (FUSION) study		[D]	2335	Case-control	
# Whole Genome Association Study of Systemic Lupus Erythematosus		[D]	-	Case-control	

24

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>



dbGaP: Public Site

- Search for studies, protocols, questionnaires
- View phenotype summary data
- View genotype summary data
- View pre-computed or published genetic associations
- Identify studies of interest, learn about their consent conditions, and learn how to apply for data access
- Locate potential collaborators for follow up studies
- No individual data can be viewed

25



dbGaP: Controlled Access Site – Security Measures

- Login/password required to download data on local computer after approval granted by DAC.
- On local computer, each data set is encrypted with a unique password for each Approved User and file.
- PI must use unique password to decrypt file on local machine.
- Consent and terms of use for each data set are included with downloaded files.

26



Publication

- Period of exclusivity for Primary Investigators
 - Proposed period of 12 months for PIs to publish
 - Exclusivity to apply to any public dissemination of the data or analyses
 - Institutes may elect to shorten this time period
- Acknowledgement of contributing investigators and funding organization

27



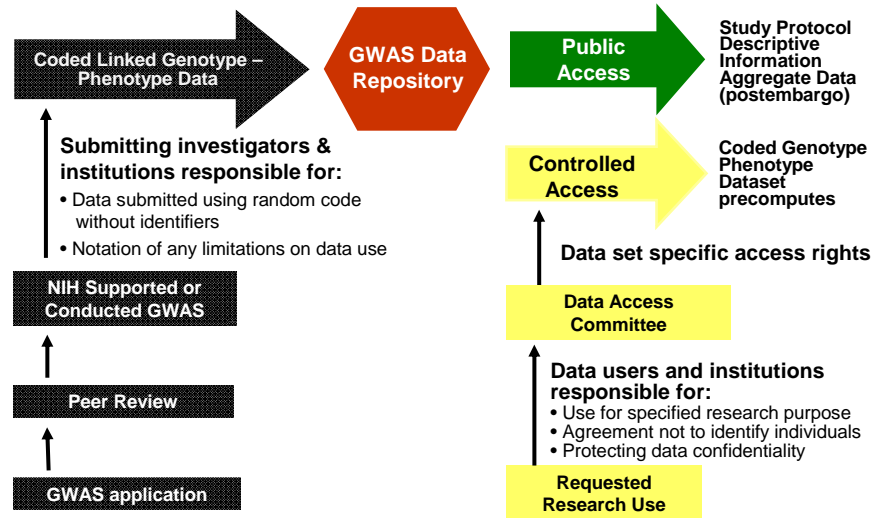
Intellectual Property

- 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.

28



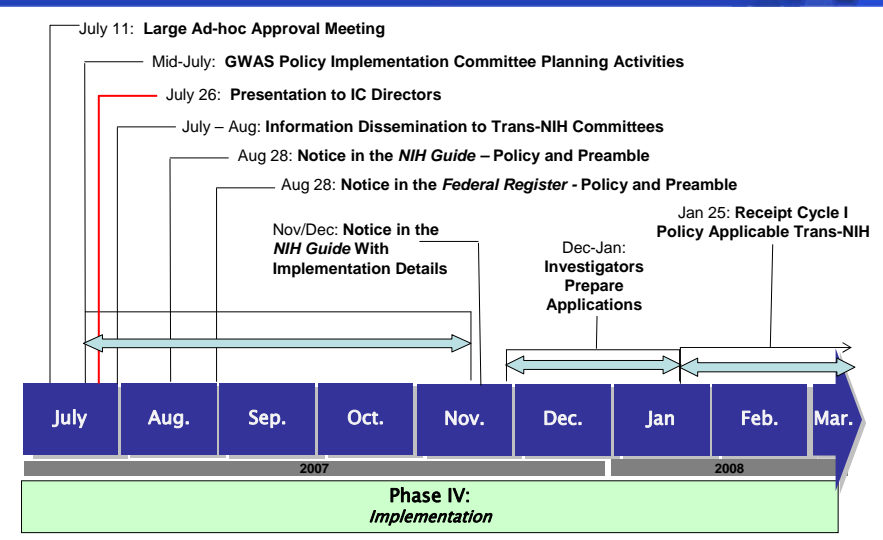
GWAS Information Flow



29



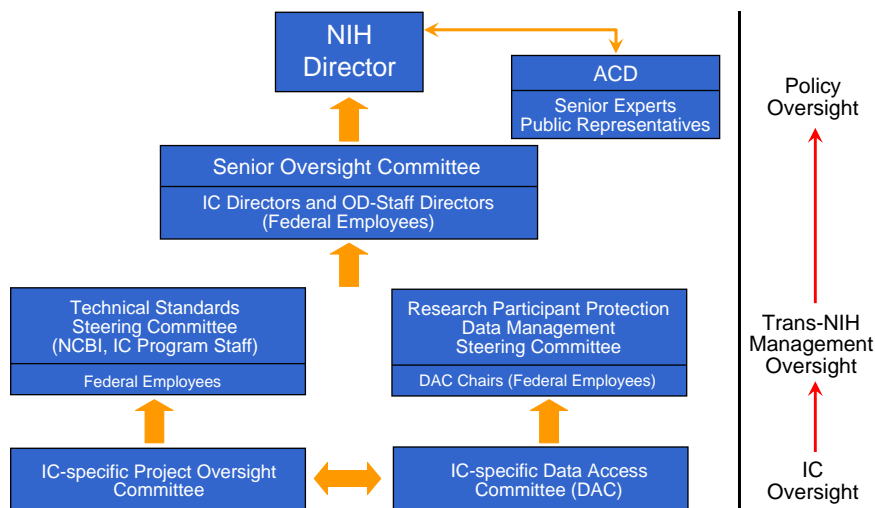
GWAS Implementation Timeline



30



Proposed GWAS Oversight Structure



31



Participating Institutes, Centers, & Offices

- NHLBI
- NHGRI
- NLM
- NCI
- NCMHD
- NCRR
- NEI
- NIA
- NIAAA
- NIAID
- NIAMS
- NICHD
- NIDA
- NIDCD
- NIDCR
- NIDDK
- NIEHS
- NIMH
- NINDS
- NINR
- OER
- OIR
- OSP
- OGC
- CSR
- OBSSR
- OCPL
- OLPA

32



Acknowledgements

Elizabeth Nabel - Chair
Susan Shurin
Carl Roth
Christopher O'Donnell
Richard Fabsitz
Barbara McGarey
Valerie Bonham
Annette Levey
Lauren Higgins
Don Schneider
Stephen Chanock
Deborah Winn
Daniela Gerhard
Robert Hoover
Daniela Seminara
Maria Giovanni
William Sharrock
John Ilekis
Jeff Evans
Catherine McKeon

Francis Collins
Laura Rodriguez
Mark Guyer
Teri Manolio
Lisa Brooks
Jean McEwen
Elizabeth Thomson
Jerome Wilson
Katrina Gwinn
Anthony Hayward
Elaine Collier
Hemin Chin
James Battey
Jerome Wilson
Katrina Gwinn-Hardy
Anthony Hayward
Elaine Collier
Zhaoxia Ren
Thomas Hart
Vivian Ota Wang

Lana Skirboll
Marianna Bledsoe
Amy Patterson
Sarah Carr
Norka Ruiz-Bravo
Valery Gordon
JP Kim
Sam Shekar
John Burklow
Marin Allen
David Lipman
James Ostell
Steve Sherry
Alan Graeff
Michael Gottesman
Jerry Menikoff
Charlotte Holden
Jonathan Pollock
Steven Kleeberger
Charlene Cho
Tom Lehner
Melinda Tinkle