

GWAS Data Sharing Policy and Bioinformatics Initiatives at NIH

**US-China Roundtable on Scientific Data Cooperation
March 23, 2009**

James Luo Ph.D.

Program Director

Biomedical Informatics Programs

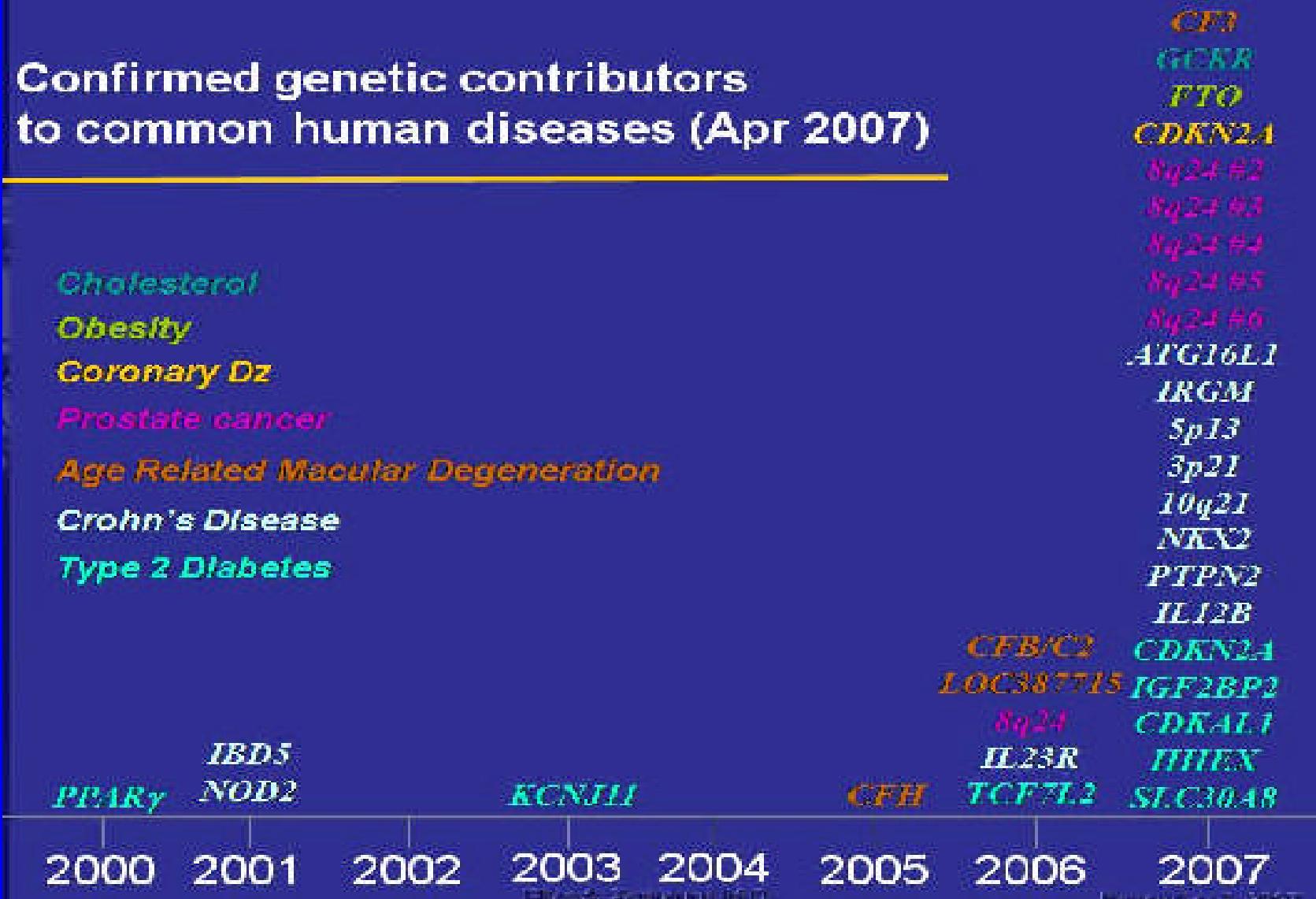
National Institute of Biomedical Imaging and Bioengineering

National Institutes of Health



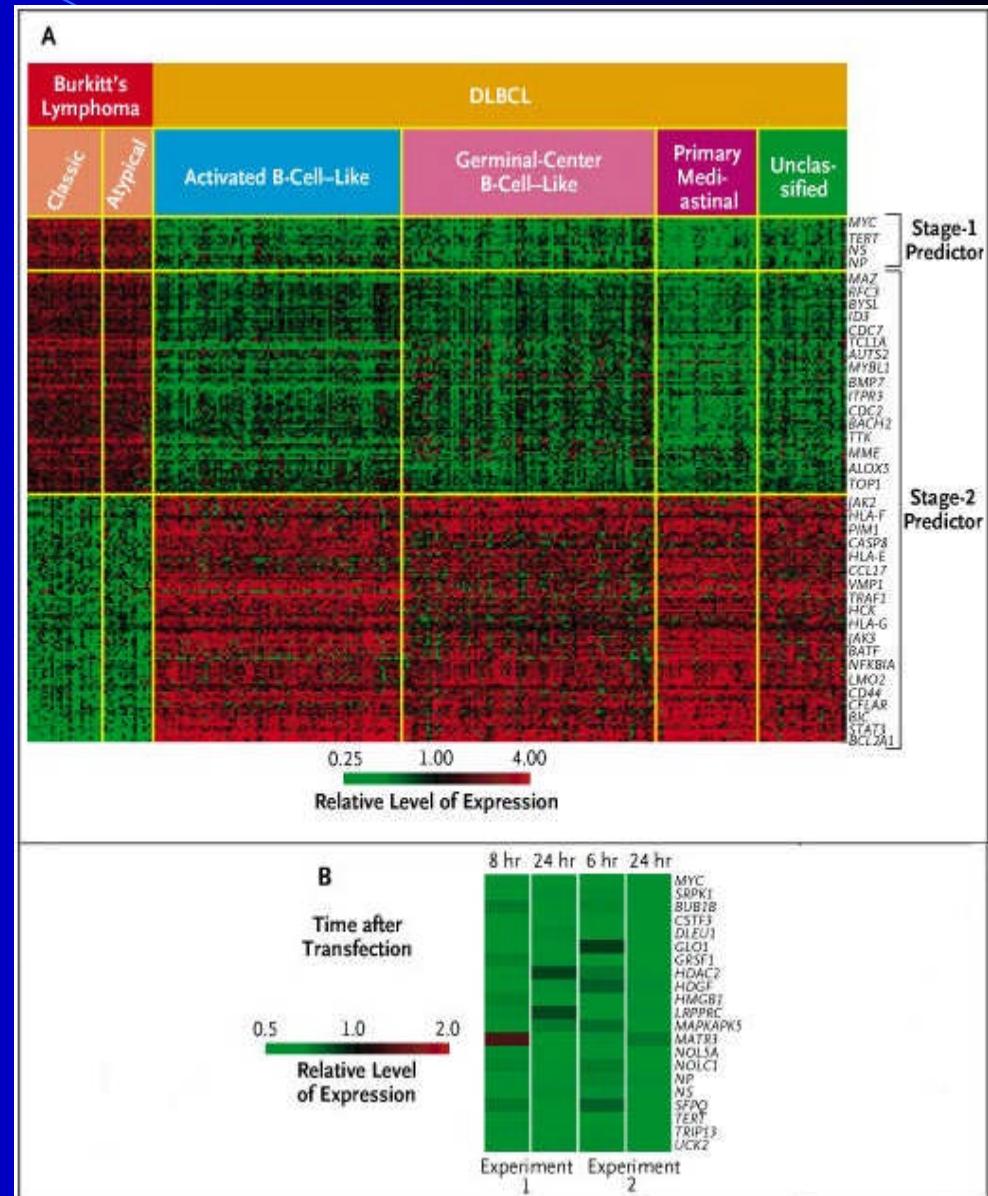
Human Genome and Diseases

Confirmed genetic contributors to common human diseases (Apr 2007)



Molecular Profiling and Treatment Decision

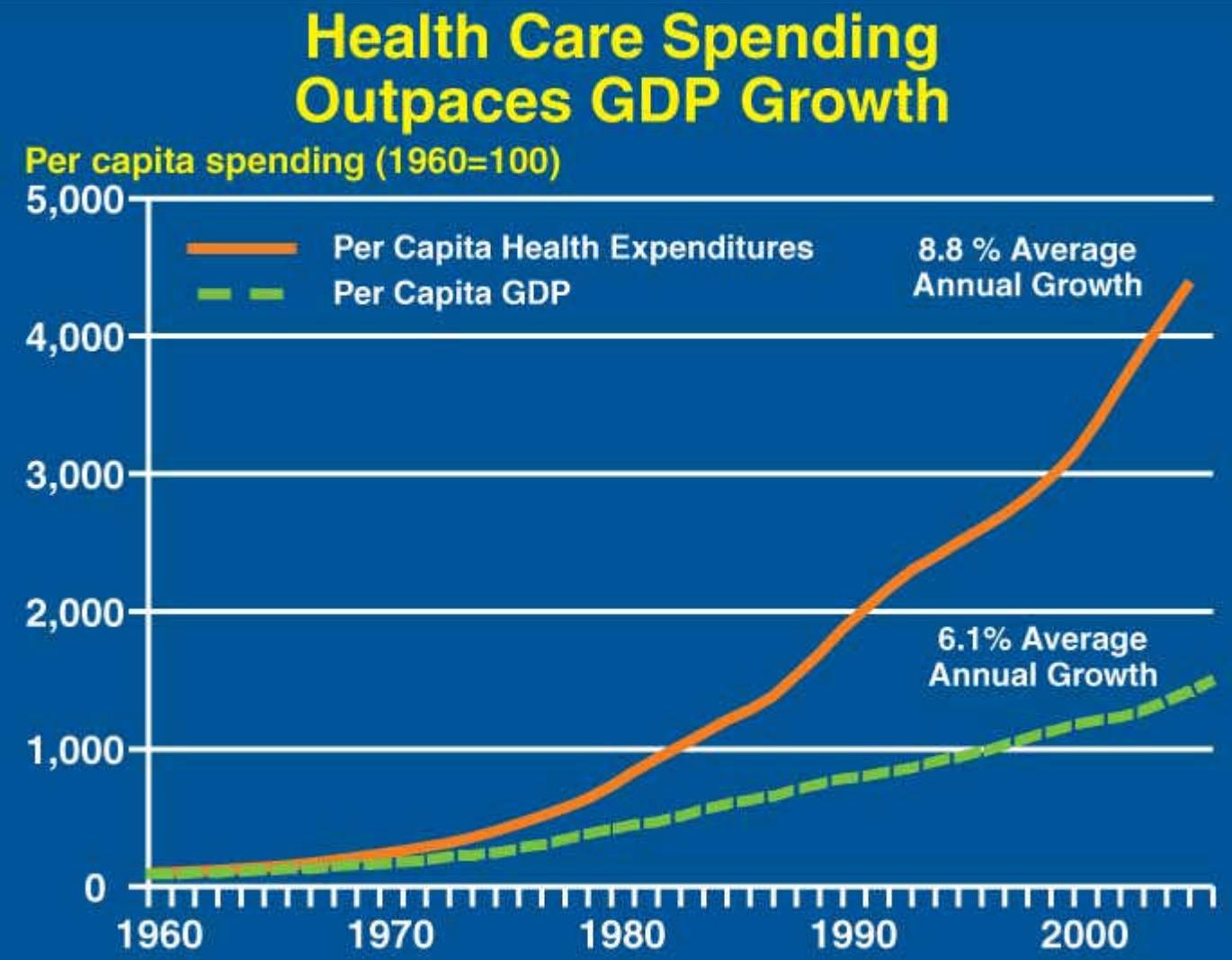
- Molecular profiling was more accurate for B-cell lymphoma leading to better treatment plan



Paradigm of Medicine

20 th Century	21 st Century
Treat disease when symptom appear and normal function is lost	Intervene before symptoms appear and preserve normal function for as long as possible
Did not understand the molecular and cellular event that lead to disease	Understanding preclinical molecular events and ability to detect patients at risk
Expensive in financial and disability cost	Order of magnitude more effective

Paradigm of Medicine

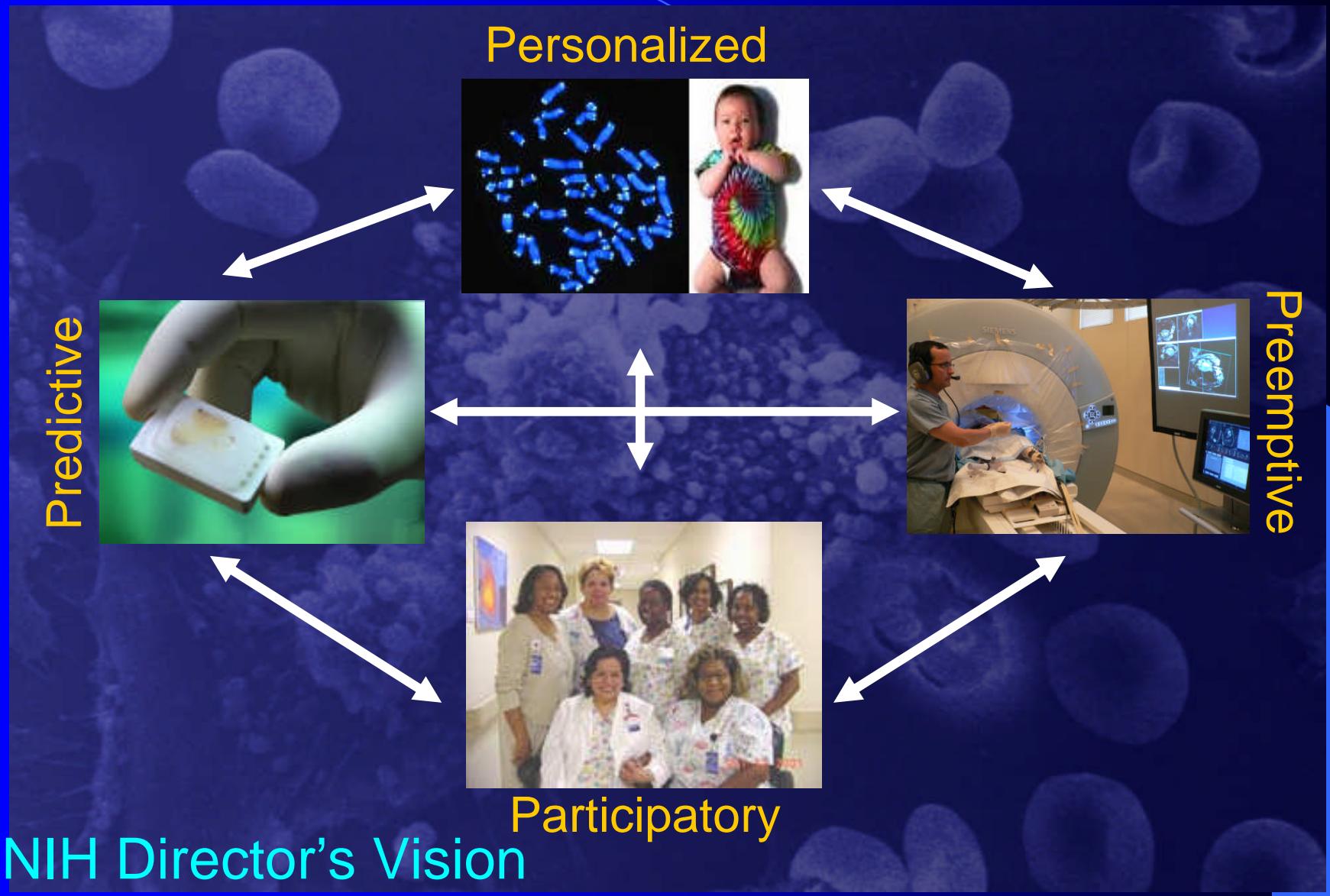


Characteristics of Biological Data

Current	Future
Destructive	Non-Destructive
Qualitative	Quantitative
Uni-Dimensional	Multi-Dimensional
Low Temporal Resolution	High Temporal Resolution
Non-localized	Spatially resolved
Low data density	High data density
Variable standards	Common standards
Non-Cumulative	Cumulative

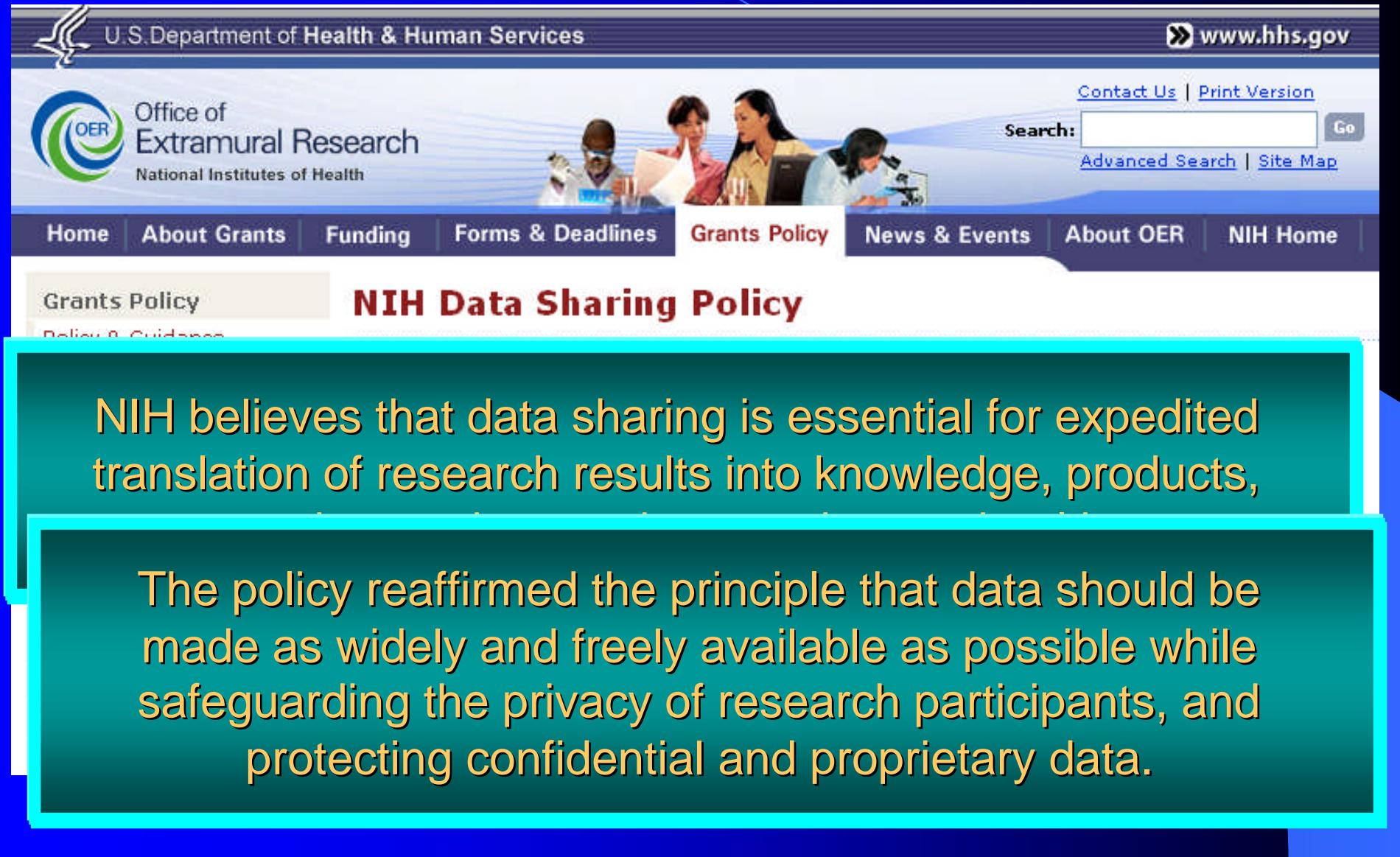
The Future Paradigm: The 4 P's

Transform Medicine from Curative to Preemptive



- Data sharing is essential to conduct biomedical research in the post genomic era.

NIH Data Sharing Policy



The image is a screenshot of the NIH Data Sharing Policy website. The header features the U.S. Department of Health & Human Services logo and the URL www.hhs.gov. Below the header, the Office of Extramural Research (OER) logo is displayed, along with the text "Office of Extramural Research" and "National Institutes of Health". The main navigation menu includes links for Home, About Grants, Funding, Forms & Deadlines, Grants Policy, News & Events, About OER, and NIH Home. The "Grants Policy" link is highlighted in red. A sub-navigation menu under "Grants Policy" includes "Policy & Guidance" and "NIH Data Sharing Policy", with the latter also highlighted in red. The main content area features a large teal box containing the text: "NIH believes that data sharing is essential for expedited translation of research results into knowledge, products, and applications that benefit the public health." Below this, another teal box contains the text: "The policy reaffirmed the principle that data should be made as widely and freely available as possible while safeguarding the privacy of research participants, and protecting confidential and proprietary data." The background of the page features a photograph of four researchers in a laboratory setting.

U.S. Department of Health & Human Services www.hhs.gov

Office of Extramural Research
National Institutes of Health

Contact Us | Print Version

Search: Go

Advanced Search | Site Map

Home | About Grants | Funding | Forms & Deadlines | **Grants Policy** | News & Events | About OER | NIH Home |

Grants Policy | **NIH Data Sharing Policy**

NIH believes that data sharing is essential for expedited translation of research results into knowledge, products, and applications that benefit the public health.

The policy reaffirmed the principle that data should be made as widely and freely available as possible while safeguarding the privacy of research participants, and protecting confidential and proprietary data.

NIH Bioinformatics Initiatives

- NIH GWAS data sharing policy and dbGaP

The goal of these initiatives is to build infrastructure and networks to facilitate data sharing, integration, and interoperability.

Most of software are open source and free to download.

NIH Bioinformatics Initiatives

- NIH GWAS data sharing policy and dbGaP
- caBIG - The Cancer Biomedical Informatics Grid
- BIRN - The Biomedical Informatics Research Network
- CTSA - Clinical and Translational Science Awards
- NIH Blueprint Neuroimaging Informatics
- NCBC - National Centers for Biomedical Computing

NIH GWAS Data Sharing Policy

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 provide powerful research tools for identifying genetic variants that contribute to health and disease.

For these reasons, The NIH believes that the full value of GWAS can be realized only if the resulting genotype and phenotype datasets are made available as rapidly as possible to a wide range of scientific investigators.

NIH dbGaP Database for GWAS

dbGaP - database of Genotype and Phenotype

It is NIH's Genotype and Phenotype association database for sharing GWAS data.

- **Phenotype**
- **Genotype**
- **Genotype X Phenotype Association**

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

The dbGaP Website - Browse Studies

Link back to dbGaP homepage

Instructions Description of dbGaP

Link to study report

List of variables in study

List of documents in study

Automated query to PubMed for genome-wide association study articles

NCBI

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search dbGaP for

Limits Preview/Index History Clipboard Details

Browse dbGaP

By Studies By Diseases Advanced Search

Study	Sub-Studies	Variables	Documents	Participants	Type of Study	Status
NEI Age-Related Eye Disease Study (AREDS)	-	176	37	600	case-control	Completed
NINDS Parkinsonism Study	2	100	5	2573	case-control	Completed

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
[Privacy Statement](#) | [Freedom of Information Act](#) | [Disclaimer](#)

The dbGaP Website - Study Report

Citeable unique stable identifier

NCBI db GaP GENOTYPE and PHENOTYPE

National Eye Institute (NEI) Age-Related Eye Disease Study (AREDS)
ID: phs000001
Version: 1

Description

The Age-Related Eye Disease Study (AREDS) was initially conceived as a long-term multicenter, prospective study of the clinical course of age-related macular degeneration (AMD) and age-related cataract. In addition to collecting natural history data, AREDS included a clinical trial of high-dose vitamin and mineral supplements for AMD and a clinical trial of high-dose vitamin supplements for cataract. Results from these clinical trials have been published. The two clinical trials generally shared 1 pool of participants (Figure 1). The clinical trials were initiated largely because of the widespread public use in the United States of commercially available pharmacologic doses of vitamins and minerals to treat these two eye conditions and the absence of definitive studies on the safety and efficacy of their use.

Eligible AREDS participants were age 55 to 80 years of age at enrollment and had to be free of any illness or condition that would make long-term follow-up or compliance with study medications unlikely or difficult. On the basis of fundus photographs graded by a central reading center, best-corrected visual acuity and ophthalmologic evaluations, participants were enrolled in one of several AMD.

It is hoped that this resource will help researchers better understand two important diseases that affect an aging population. The AREDS Research Group hopes that data from AREDS on progression rates and risk factors for AMD and cataract will further understanding of the clinical course of both conditions, generate hypotheses about etiology and aid in the design of clinical trials of potential interventions.

[AREDS, The National Eye Institute](#)
[AREDS, The EMMES Corporation](#)

- Subjects: 600
- Type: case-control
- Status: Completed

Inclusion/Exclusion Criteria

- [Inclusion/exclusion criteria](#)

History

Publications

Attribution

Access Rules

Genotype x phenotype association or linkage analyses

AREDS

Search Within This Study

Search for: Go

Associated Analyses

+ NEI Age-Related Eye Disease Study (AREDS)

Associated Variables

smk10
smk11
smk12
smk13
specnum
syst00
syst03
syst04
syst05
syst06
syst07
syst08

Associated Documents

+ NEI Age-Related Eye Disease Study (AREDS)

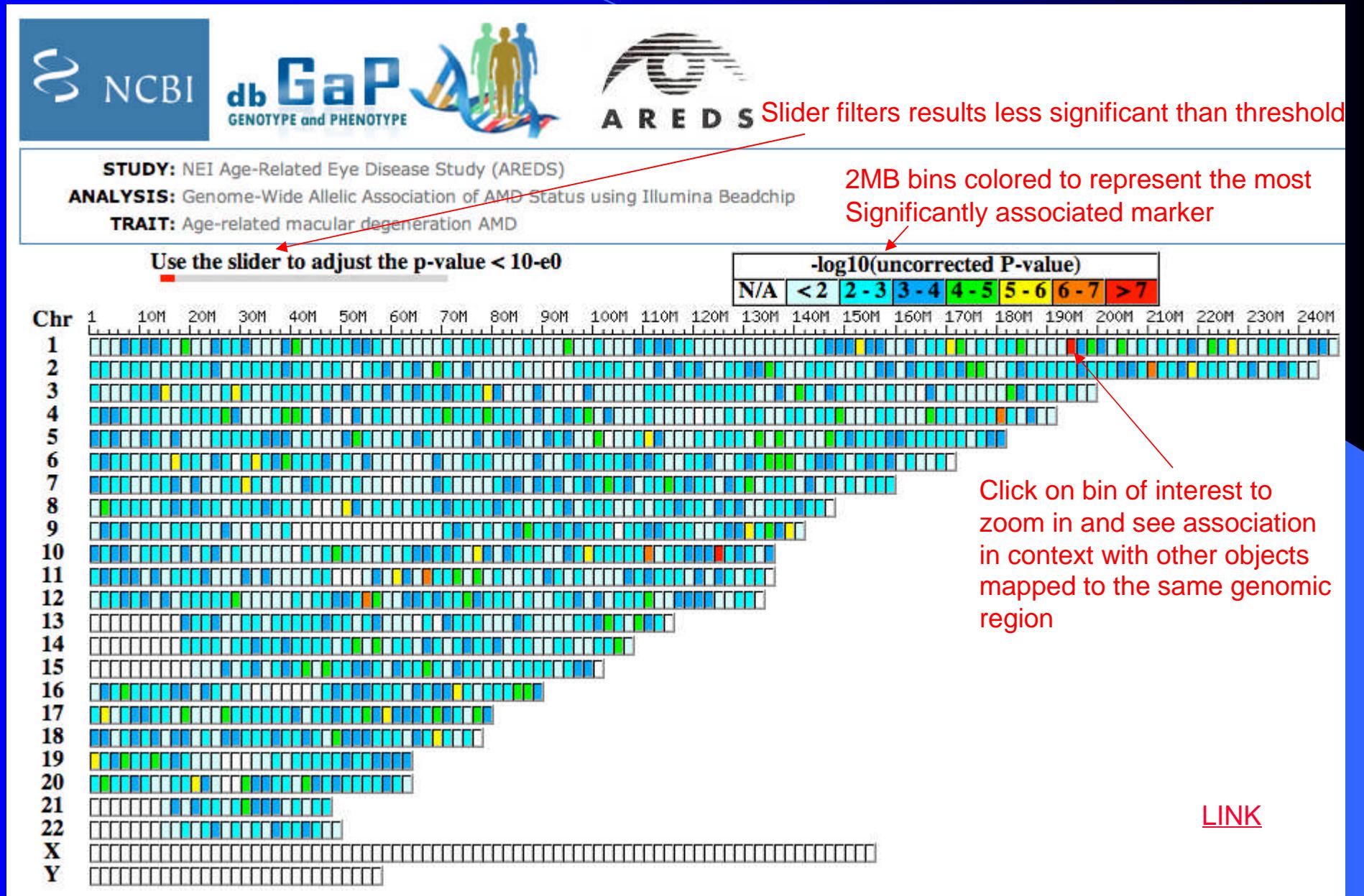
Links back to submitter website

Criteria for inclusion/exclusion

search this study

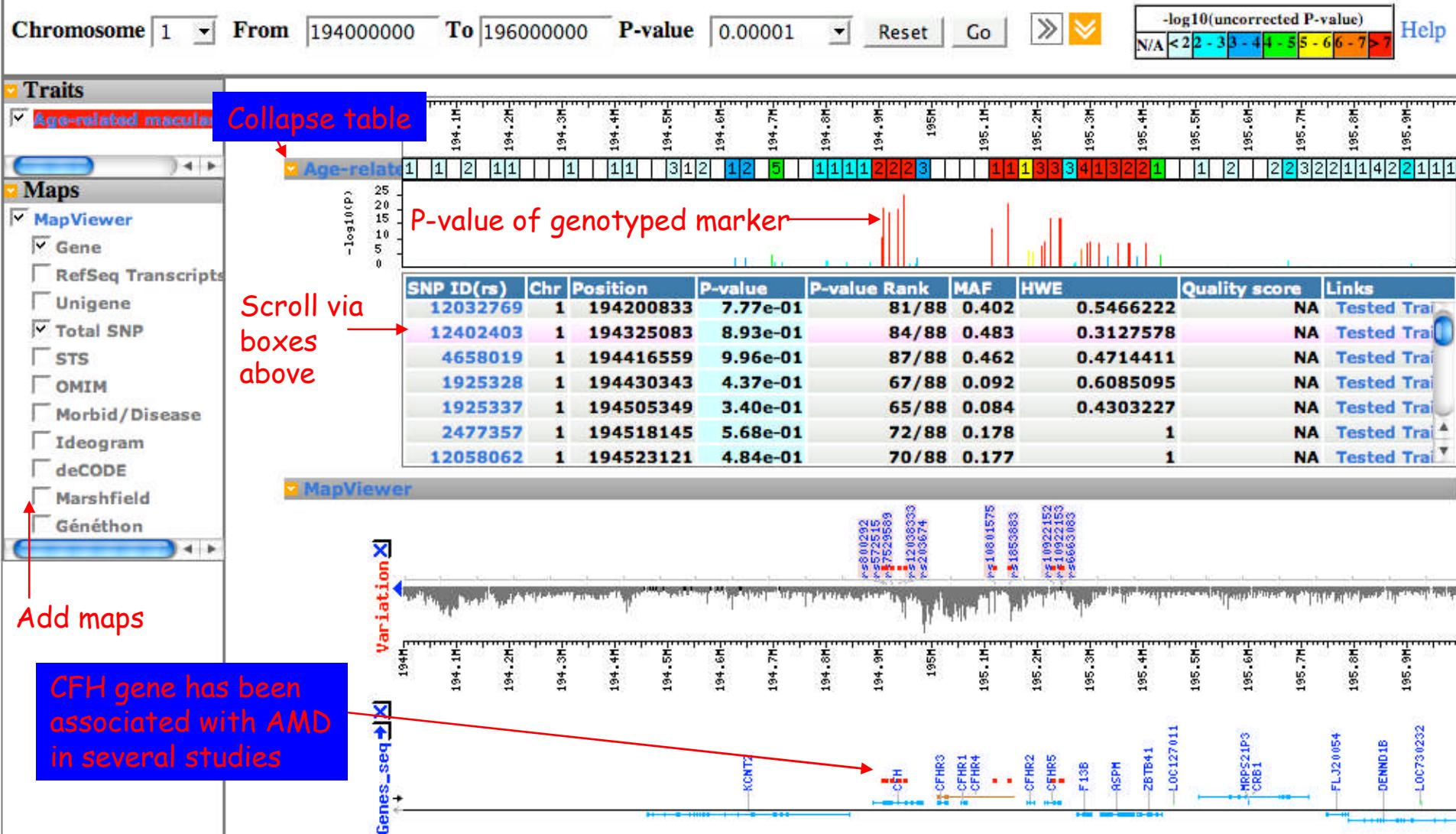
Link to variable report

The dbGaP Website - Analysis Results



The dbGaP Website - Analysis Results

STUDY: NEI Age-Related Eye Disease Study (AREDS)
ANALYSIS: Genome-Wide Allelic Association of AMD Status using Illumina Beadchip



NIH Bioinformatics Initiatives

- NIH GWAS data sharing policy and dbGaP
- caBIG - The Cancer Biomedical Informatics Grid
- BIRN - The Biomedical Informatics Research Network
- CTSA - Clinical and Translational Science Awards
- NIH Blueprint Neuroimaging Informatics
- NCBC - National Centers for Biomedical Computing

The caBIG



National Cancer Institute

U.S. National Institutes of Health | www.cancer.gov



caBIG™
Cancer Biomedical
Informatics Grid™

caBIG™ stands for the cancer Biomedical Informatics Grid™. caBIG™ is an information network enabling all constituencies in the cancer community – researchers, physicians, and patients – to share data and knowledge. The components of caBIG™ are widely applicable beyond cancer as well.

The mission of caBIG™ is to develop a truly collaborative information network that accelerates the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer.

caBIG™ is an information network that allows researchers, clinicians, and patients to share data to accelerate discovery of new diagnostics and therapeutics to improve patient outcomes.

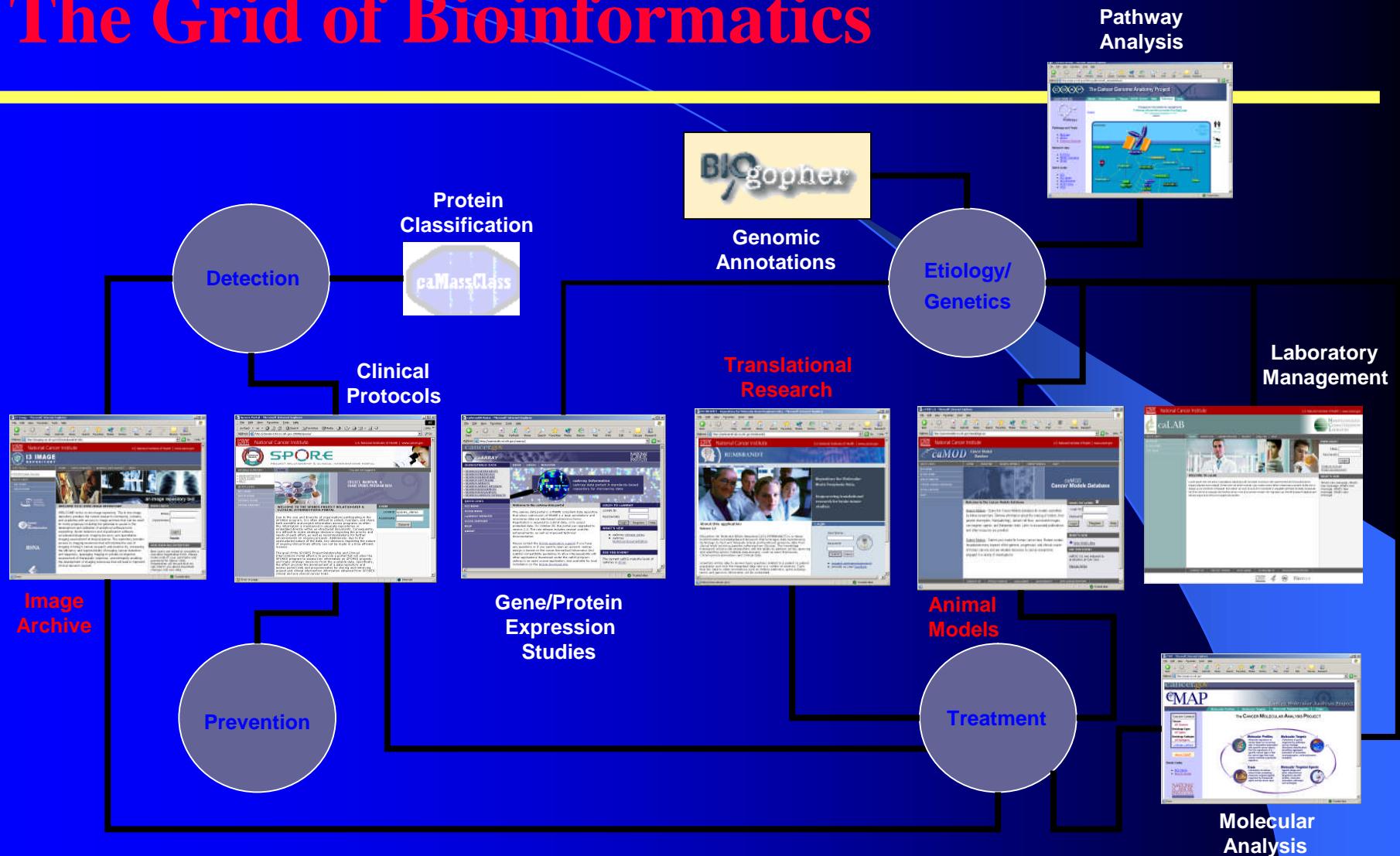


caBIG™
cancer Biomedical
Informatics Grid™

caBIG™ Vision

- **Connect** cancer researchers by a shareable, interoperable infrastructure
- **Deploy** a common language to more easily share information
- **Build** or adapt tools for collecting, analyzing, and disseminating information associated with cancer research and care

The Grid of Bioinformatics



- Interconnecting over 60 cancer centers
- Enabling novel biomarker discoveries towards personalized medicine
- Received 2005 Service to America Award for contributions to brain cancer research



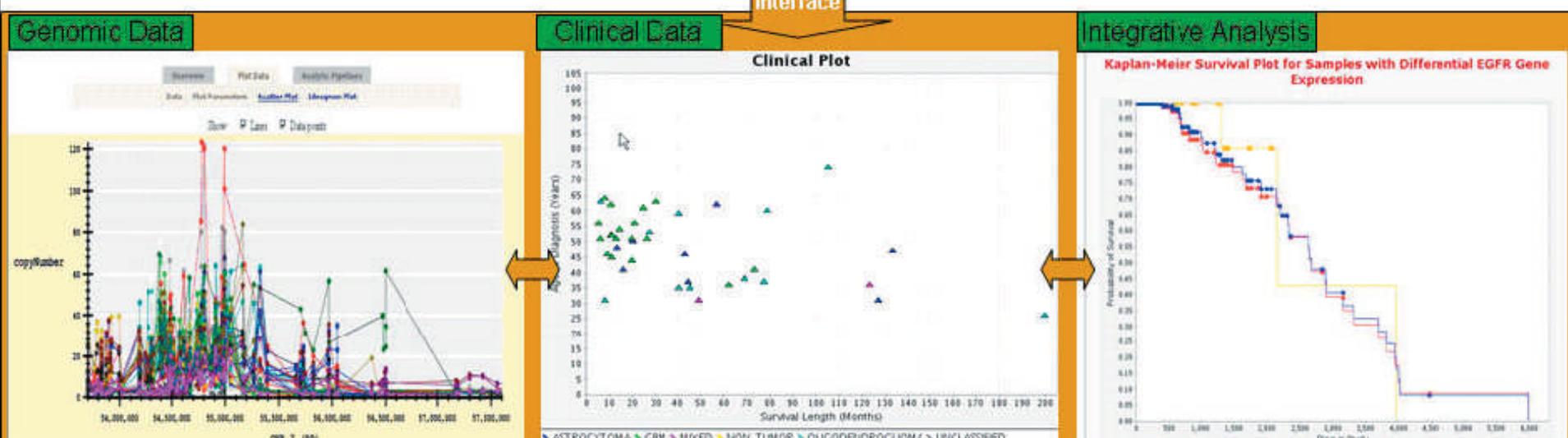
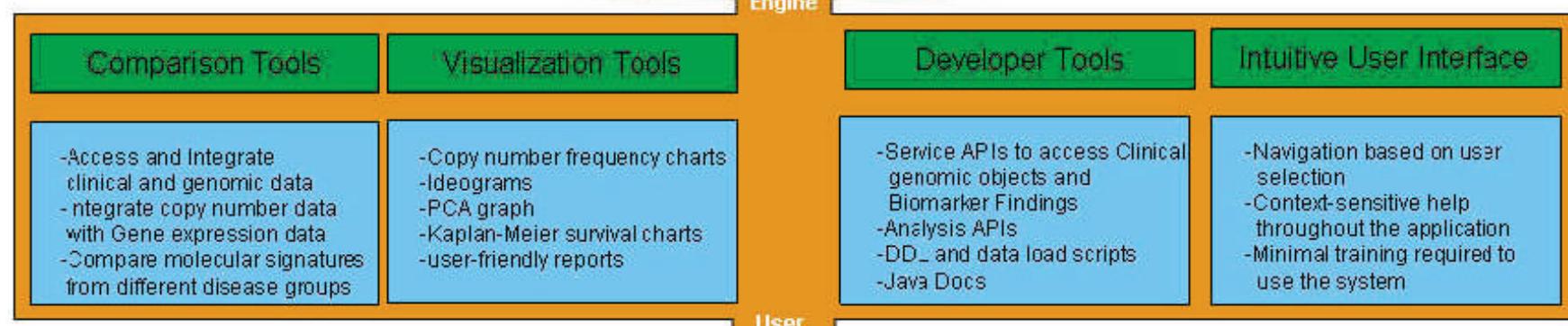
- Promoting early detection and therapeutic response evaluation via cancer imaging
- Defining the infrastructure and standards supporting pre-clinical and clinical oncology

caBIG in the Research Continuum

cancer Biomedical
Informatics Grid™



Disparate Data Sources



National Cancer Imaging Archive - NCIA



U.S. National

NATIONAL CANCER INSTITUTE National Cancer Institute

NCIA National Cancer Imaging Archive

TOOLS

- DICOM Image Viewers

QUICK LINKS

- NCIA NEWS
- NCIA USER'S GUIDE
- NCI HOME
- NCICB HOME

caBIG™

CANCER IMAGING PROGRAM

Center for Bioinformatics

RSNA

NIBIB

HOME SEARCH IMAGES MANAGE DATA BASKET HELP

an image repository tool

WELCOME TO NATIONAL CANCER IMAGING ARCHIVE

New! NCIA now offering web-based image visualization and markup of archived image data. [more information](#)

Welcome to the National Cancer Imaging Archive (NCIA). NCIA is a searchable repository of *in vivo* cancer images that provides the cancer research community, industry, and academia with access to image archives to be used in the development and validation of analytical software tools that support:

- Lesion detection and classification
- Accelerated diagnostic imaging decision
- Quantitative imaging assessment of drug response

NCIA provides access to imaging resources that will improve the use of imaging in today's cancer research and practice by:

- Increasing the efficiency and reproducibility of imaging cancer detection and diagnosis
- Leveraging imaging to provide an objective assessment of therapeutic response

USER LOGIN

EMAIL

PASSWORD

[Register](#)

NEW USER REGISTRATION

New users are asked to complete a one-time registration form. Please make note of your username and password for future visits. Registration will ensure that we can inform you about important changes and new data.

NIH Bioinformatics Initiatives

- NIH GWAS data sharing policy and dbGaP
- caBIG - The Cancer Biomedical Informatics Grid
- BIRN - The Biomedical Informatics Research Network
- CTSA - Clinical and Translational Science Awards
- NIH Blueprint Neuroimaging Informatics
- NCBC - National Centers for Biomedical Computing

BIRN



CTSA Committees & Activities

Oversight Committees

CTSA Consortium Oversight

Pediatrics Oversight Committee

Clinical Integration Committee Workgroups:

Administration

Biostatistics / Epidemiology / Research Design

Clinical Research Ethics

Communications

Participant and Clinical Interactions Resources

Regulatory Knowledge

Steering Committees

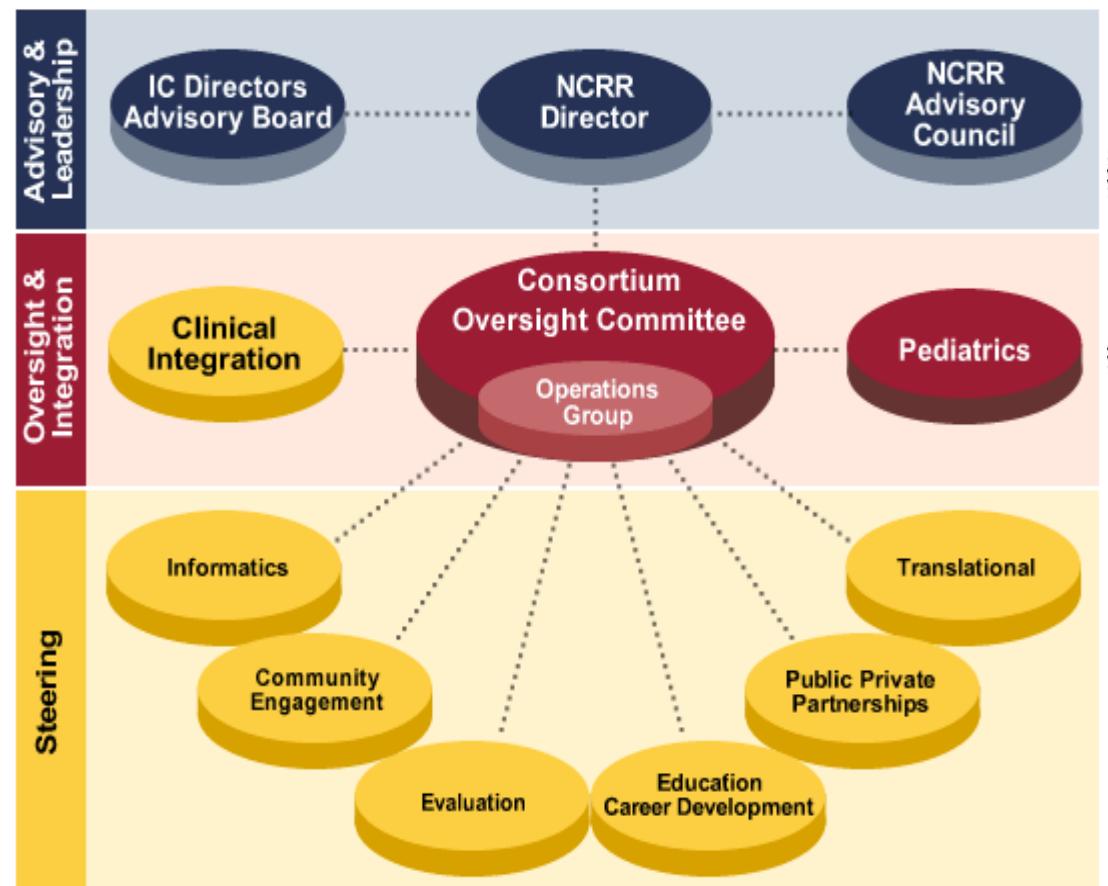
Community Engagement

Education / Career Development

Evaluation

Informatics

Clinical and Translational Science Awards



- **Advisory:** providing guidance and input to the NCRR Director on the CTSA Consortium
- **Oversight:** identifying and selecting collaborative opportunities to facilitate research throughout the CTSA program, coordinating Consortium-wide approaches to research and overseeing topic-specific efforts across the Consortium
- **Steering:** coordinating institutional topic-specific efforts with the national CTSA Consortium; each Steering Committee has an Operations subgroup that takes timely action on emergent topic issues

Quick Links

[Calendar](#)

[Current CTSA e-Newsletter](#)

[CTSA Committee Representatives](#)

[Communication Toolkit](#)

[Resources for Researchers](#)

[Governance Manual](#)

[Institution Search](#)

NIH CTSA Information

[NEW - NCRR CTSA Administrative Supplements](#)

[NCRR CTSA Information](#)

[CTSA Solicitations](#)

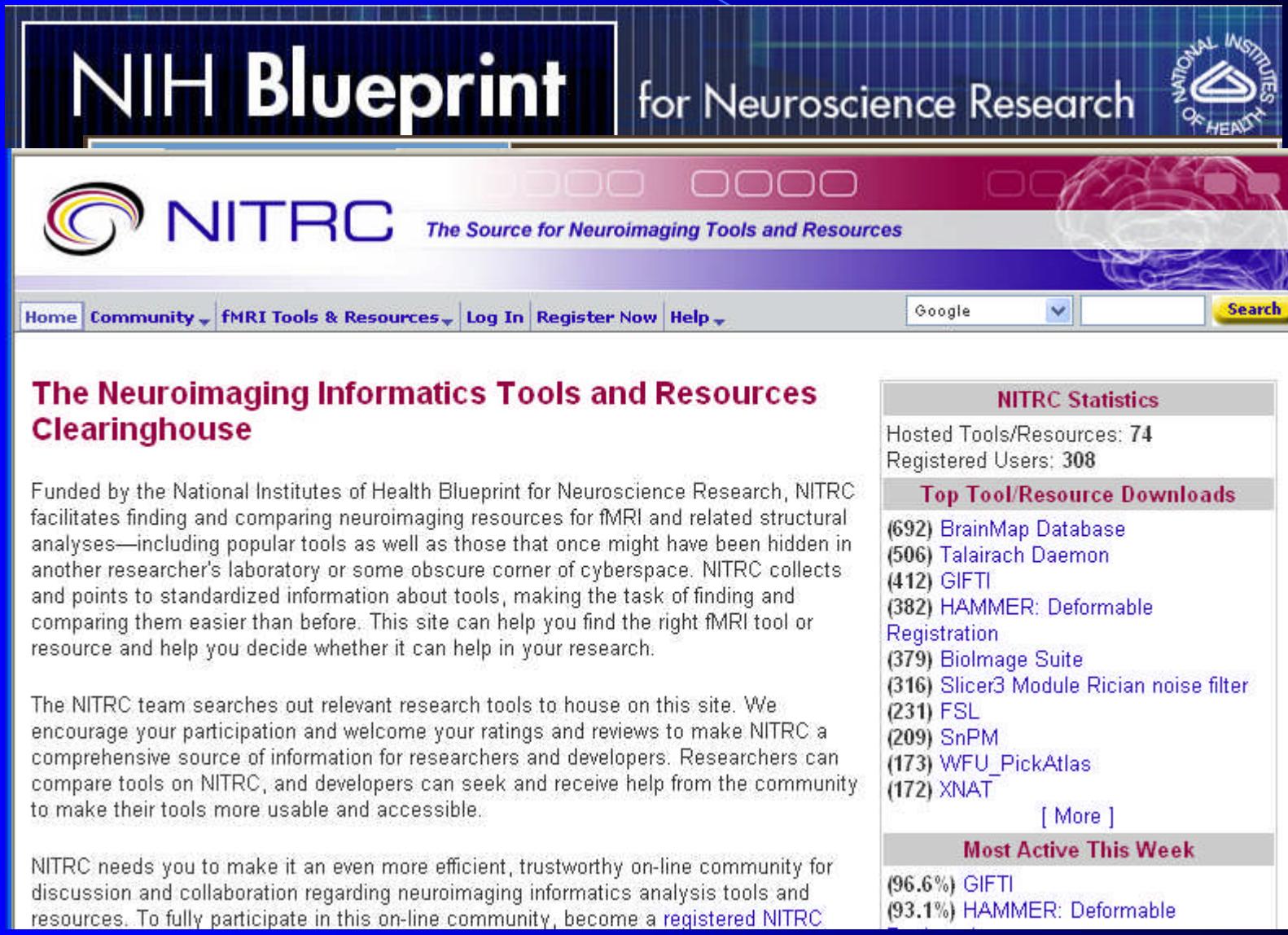
[Current RFA](#)

[CTSA Funding Guidelines](#)

[Questions and Answers](#)

[Fact Sheet](#)

NIH Blueprint



The screenshot shows the NIH Blueprint for Neuroscience Research website. At the top, the NIH logo is visible. Below it, the NITRC logo and the tagline "The Source for Neuroimaging Tools and Resources" are displayed. The main navigation menu includes Home, Community, fMRI Tools & Resources, Log In, Register Now, and Help. A search bar is also present. The main content area features a section titled "The Neuroimaging Informatics Tools and Resources Clearinghouse" with a description of its purpose. Below this, there is a section about the NITRC team and their mission. At the bottom, there is a call to action for users to participate in the online community. On the right side, there are three boxes: "NITRC Statistics" (Hosted Tools/Resources: 74, Registered Users: 308), "Top Tool/Resource Downloads" (listing BrainMap Database, Talairach Daemon, GIFTI, HAMMER: Deformable Registration, Biolmage Suite, Slicer3 Module Rician noise filter, FSL, SnPM, WFU_PickAtlas, and XNAT), and "Most Active This Week" (listing GIFTI and HAMMER: Deformable).

NIH Blueprint for Neuroscience Research

NITRC *The Source for Neuroimaging Tools and Resources*

Home | Community | fMRI Tools & Resources | Log In | Register Now | Help | Google | Search

The Neuroimaging Informatics Tools and Resources Clearinghouse

Funded by the National Institutes of Health Blueprint for Neuroscience Research, NITRC facilitates finding and comparing neuroimaging resources for fMRI and related structural analyses—including popular tools as well as those that once might have been hidden in another researcher's laboratory or some obscure corner of cyberspace. NITRC collects and points to standardized information about tools, making the task of finding and comparing them easier than before. This site can help you find the right fMRI tool or resource and help you decide whether it can help in your research.

The NITRC team searches out relevant research tools to house on this site. We encourage your participation and welcome your ratings and reviews to make NITRC a comprehensive source of information for researchers and developers. Researchers can compare tools on NITRC, and developers can seek and receive help from the community to make their tools more usable and accessible.

NITRC needs you to make it an even more efficient, trustworthy on-line community for discussion and collaboration regarding neuroimaging informatics analysis tools and resources. To fully participate in this on-line community, become a registered NITRC

NITRC Statistics

Hosted Tools/Resources: 74
Registered Users: 308

Top Tool/Resource Downloads

- (692) BrainMap Database
- (506) Talairach Daemon
- (412) GIFTI
- (382) HAMMER: Deformable Registration
- (379) Biolmage Suite
- (316) Slicer3 Module Rician noise filter
- (231) FSL
- (209) SnPM
- (173) WFU_PickAtlas
- (172) XNAT

[More]

Most Active This Week

- (96.6%) GIFTI
- (93.1%) HAMMER: Deformable

NIH Bioinformatics Initiatives

- NIH GWAS data sharing policy and dbGaP
- caBIG - The Cancer Biomedical Informatics Grid
- BIRN - The Biomedical Informatics Research Network
- CTSA - Clinical and Translational Science Awards
- NIH Blueprint Neuroimaging Informatics
- NCBC - National Centers for Biomedical Computing

NIH Roadmap - NCBC



NIH Roadmap National Centers for Biomedical Computing



The National Centers for Biomedical Computing (NCBC) are cooperative agreement awards that are funded under the NIH Roadmap for Bioinformatics and Computational Biology.

The Centers are intended to be the core of the networked national effort to build the computational infrastructure for biomedical computing in the nation, the National Program of Excellence in Biomedical Computing (NPEBC).

LOGISTIC &
CONTACT INFO

PLANNING COMMITTEE

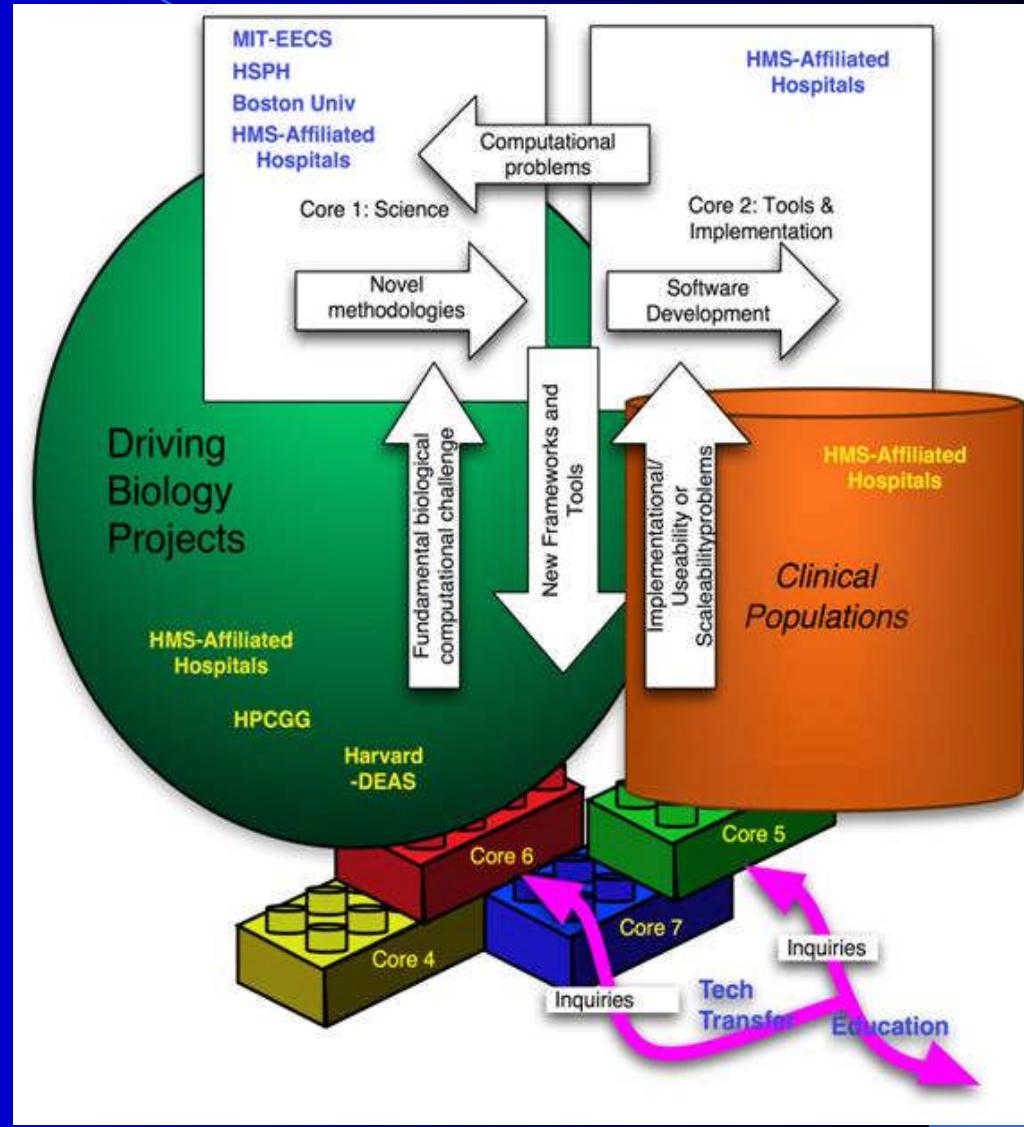
BUILDING BRIDGES
COMPENDIUM

NIH BIOINFORMATICS
HOME PAGE



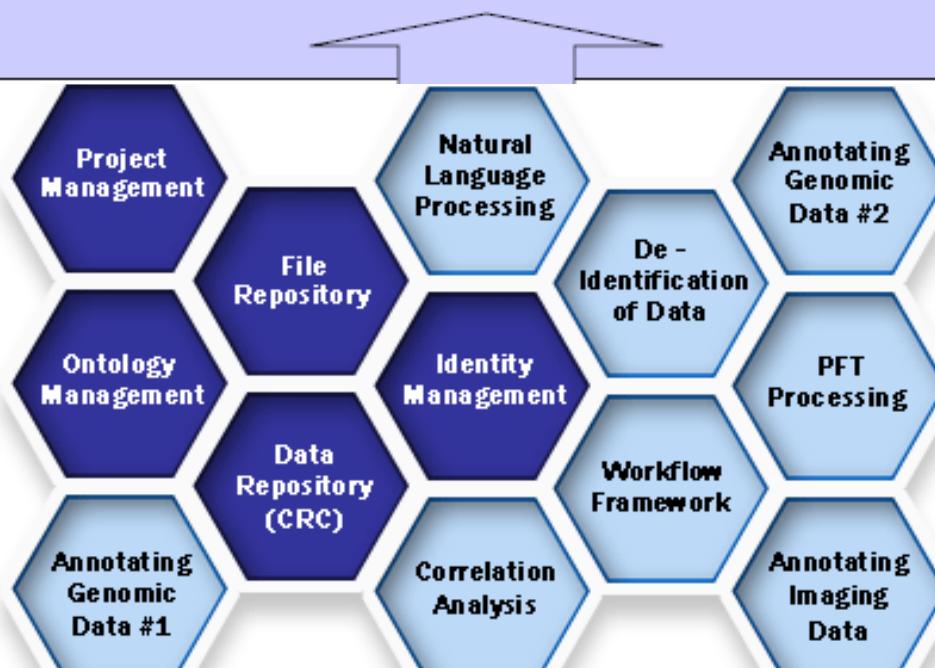
NCBC - i2b2

- The i2b2 is designed to address is that of creating a comprehensive software and methodological framework to enable clinical researchers to accelerate the translation of genomic and “traditional” clinical findings into novel diagnostics, prognostics, and therapeutics.





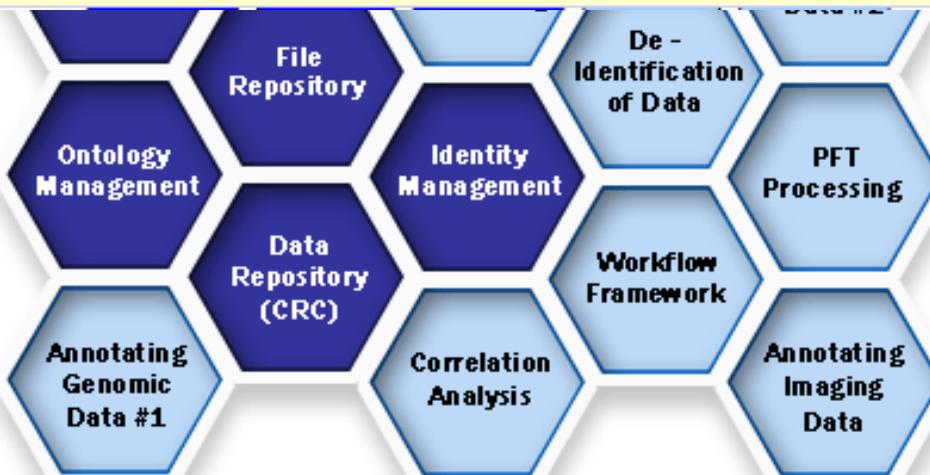
Enterprise-wide repurposing and distribution of medical record data for research



- Enable high performance collection of medical record data for querying and distribution
 - Enterprise web client
- Enable discovery within data on enterprise wide scale
 - Relationship networks
 - Pharmacovigilance

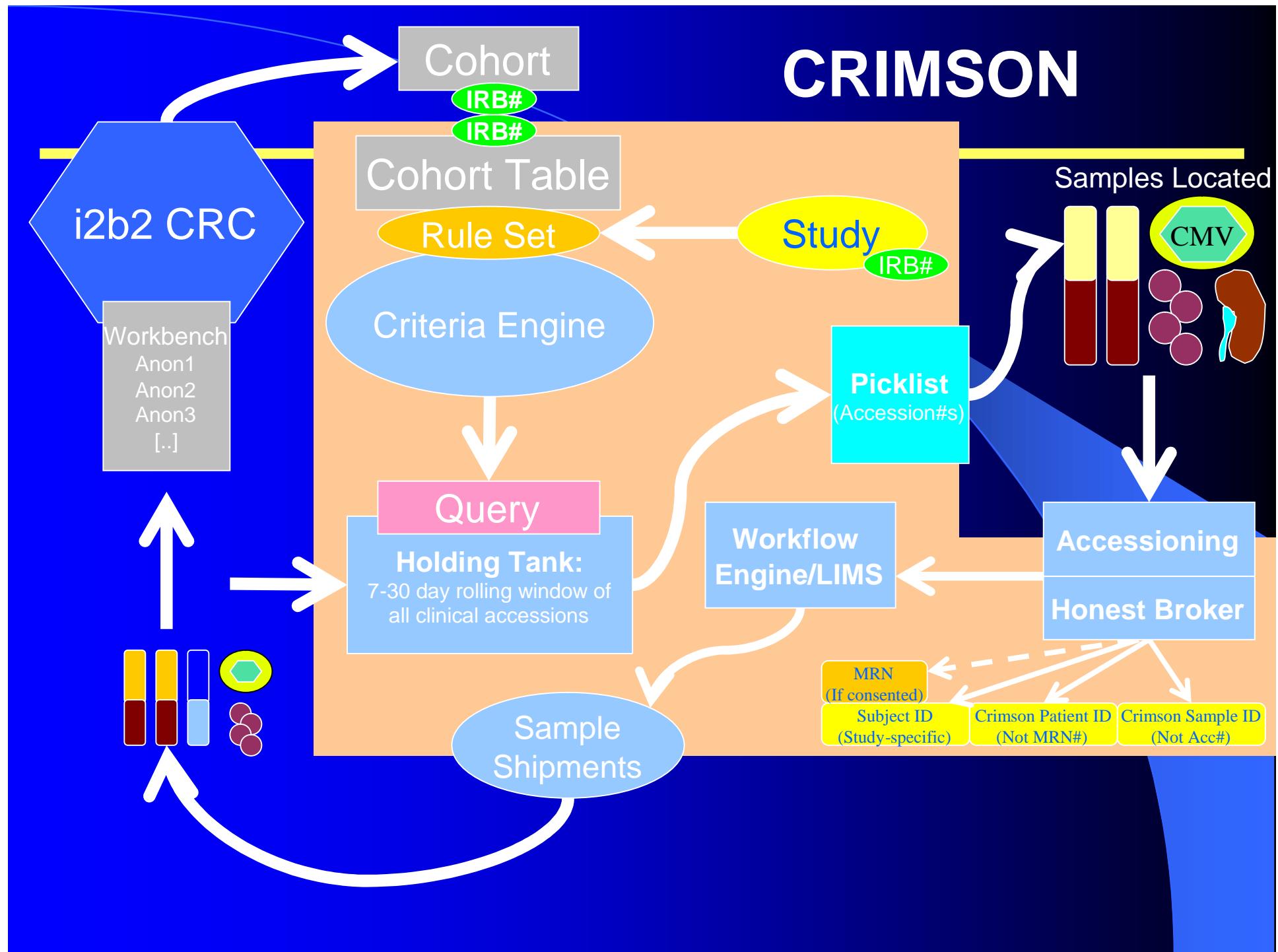


- Repurpose medical record information for research studies
 - I2b2 Workbench
 - Natural language processing
- Enable genomic studies
 - Tissue/blood selection
 - Data integration



Use of medical record data in clinical studies
focused upon genomics and pharmacology

CRIMSON



Cost and Throughput Comparison

- Study desires 10,000 samples for epidemiologic analyses
- Avg. cost/sample for the study: \$1,200
 - *\$12,000,000 to collect 10K samples*
- Throughput of 5-10 samples/month
 - *120 years to collect 10K with current process.*
- Forwarded cohorts via i2b2
- Avg cost for collection: \$8-9/sample
 - Costs for collection of 10K samples: \$85,000
- Avg throughput:
 - 4-600 samples/month (1 Crimson node)
 - 1000+ with 2 Crimson nodes operational.
 - Collection of controls in <1 year
 - Experimental samples in 1.5 - 4 years.

Pharmacovigilance

