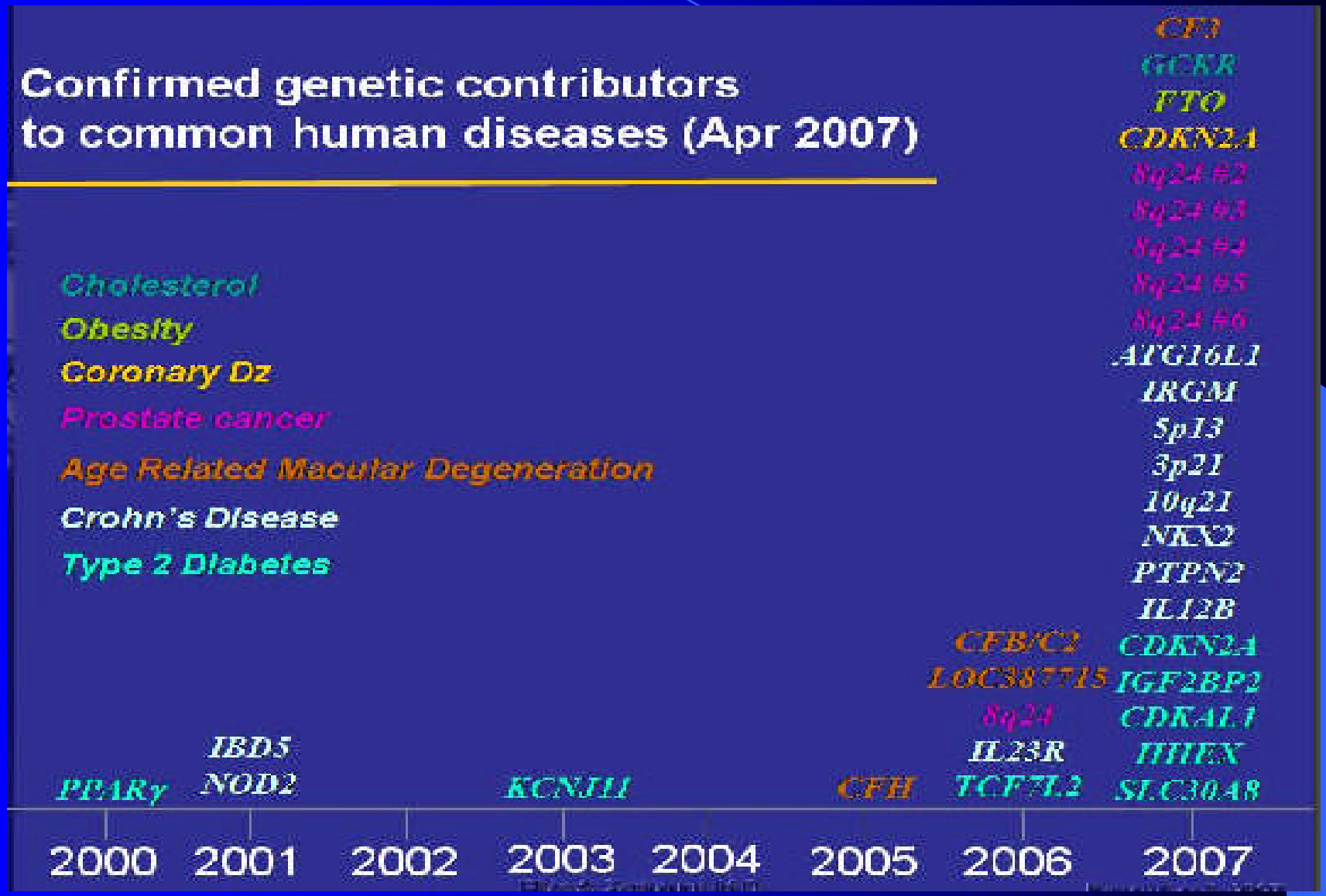
GWAS Data Sharing Policy and Bioinformatics Initiatives at NIH

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

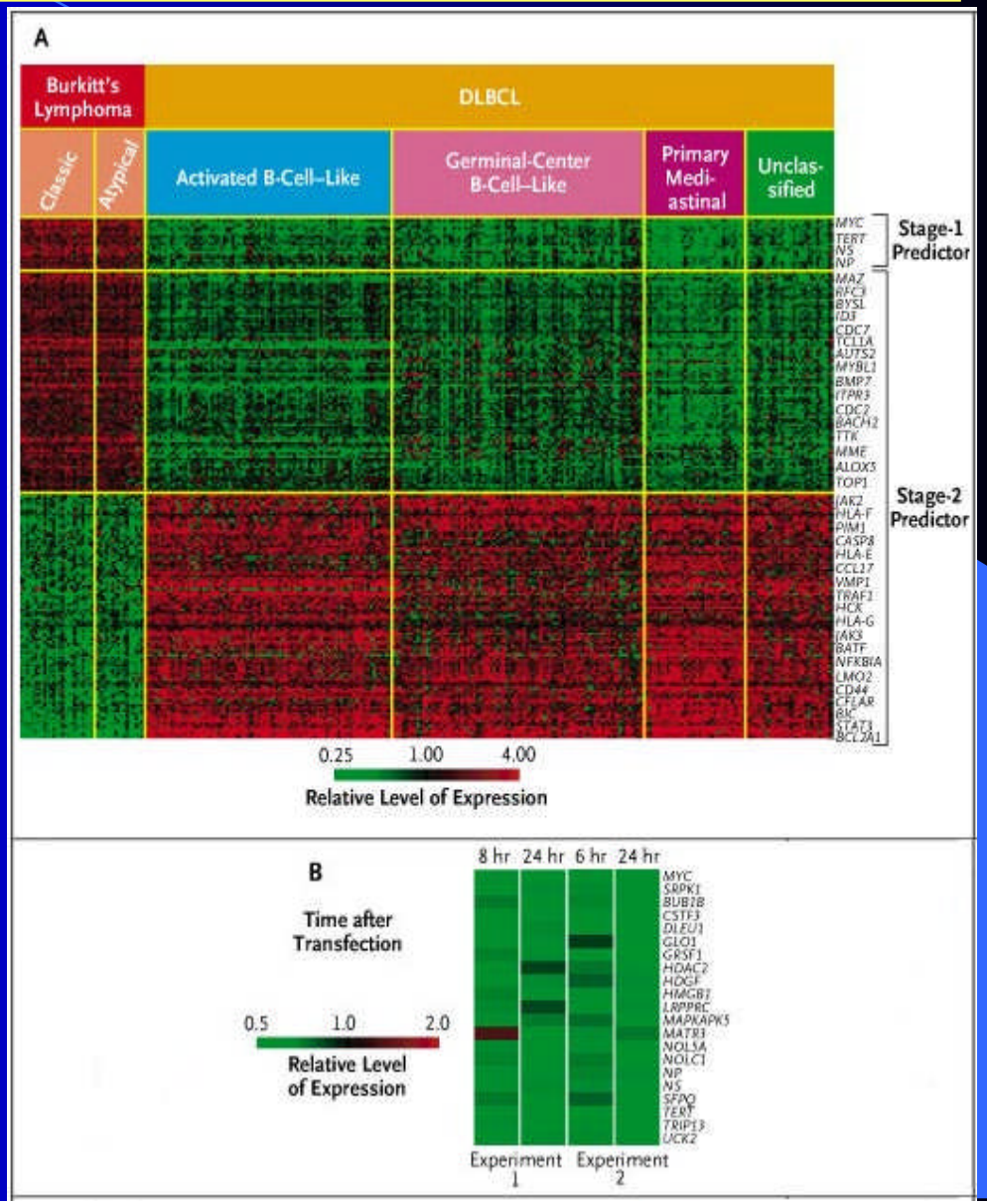
James Luo Ph.D.
Program Director
Biomedical Informatics Programs
National Institute of Biomedical Imaging and Bioengineering
National Institutes of Health

Human Genome and Diseases



Molecular Profiling and Treatment Decision

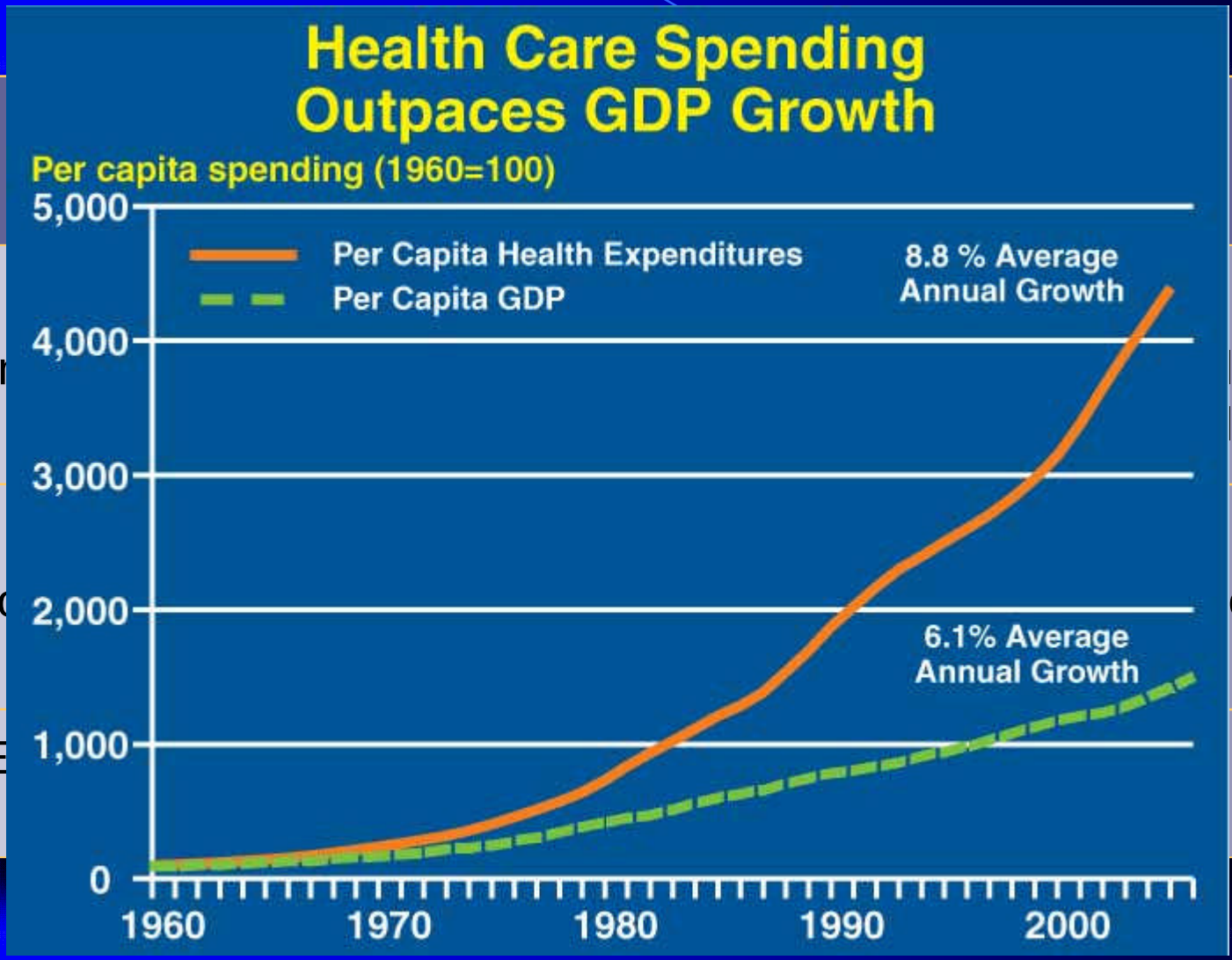
- Molecular profiling was more accurate for B-cell lymphoma leading to better treatment plan



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

Personalized

Predictive

Preemptive

Participatory

NIH Director's Vision

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

NIH Data Sharing Policy

NIH believes that data sharing is essential for expedited translation of research results into knowledge, products,

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

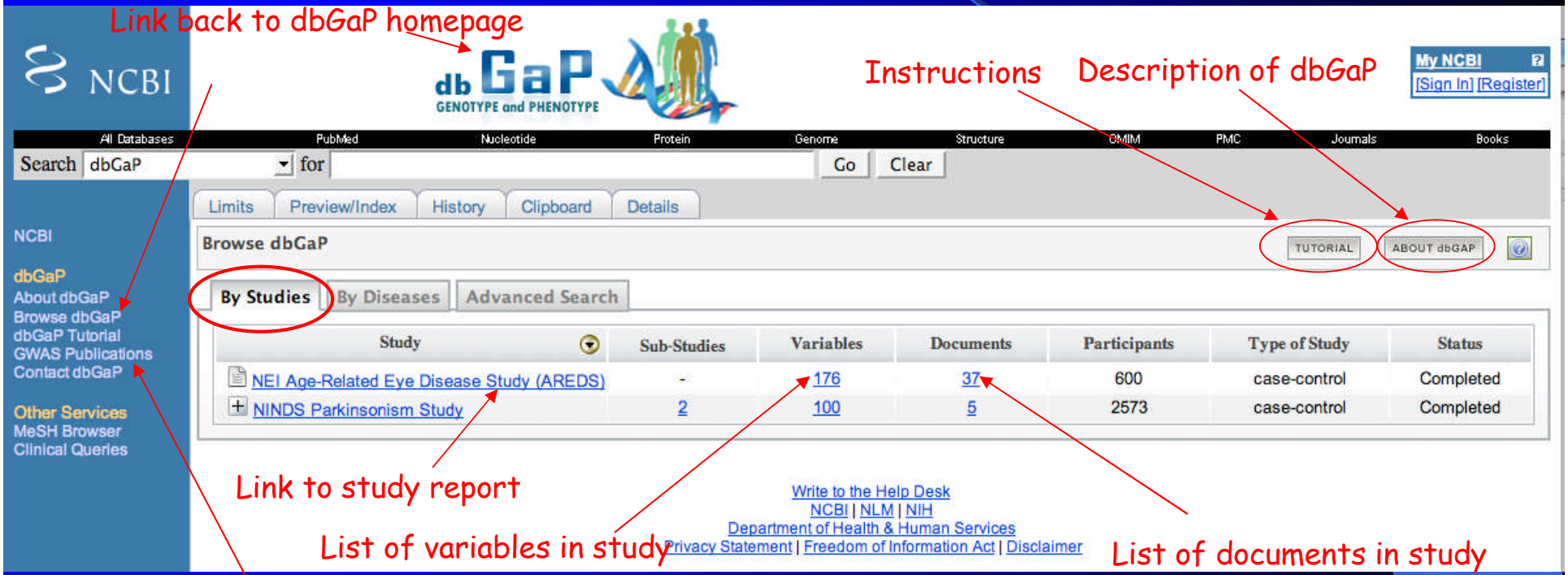
dbGaP - database of Genotype and Phenotype

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



The dbGaP Website - Study Report

Citeable unique stable identifier

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

National Eye Institute (NEI) Age-Related Eye Disease Study (AREDS)

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

Links back to submitter website

Inclusion/Exclusion Criteria

- Inclusion/exclusion criteria

Criteria for inclusion/exclusion

History

History
Publications
Attribution
Access Rules

Genotype x phenotype association or linkage analyses

search this study

Search Within This Study

Search for: Go

Associated Analyses

NEI Age-Related Eye Disease Study (AREDS)

Associated Variables

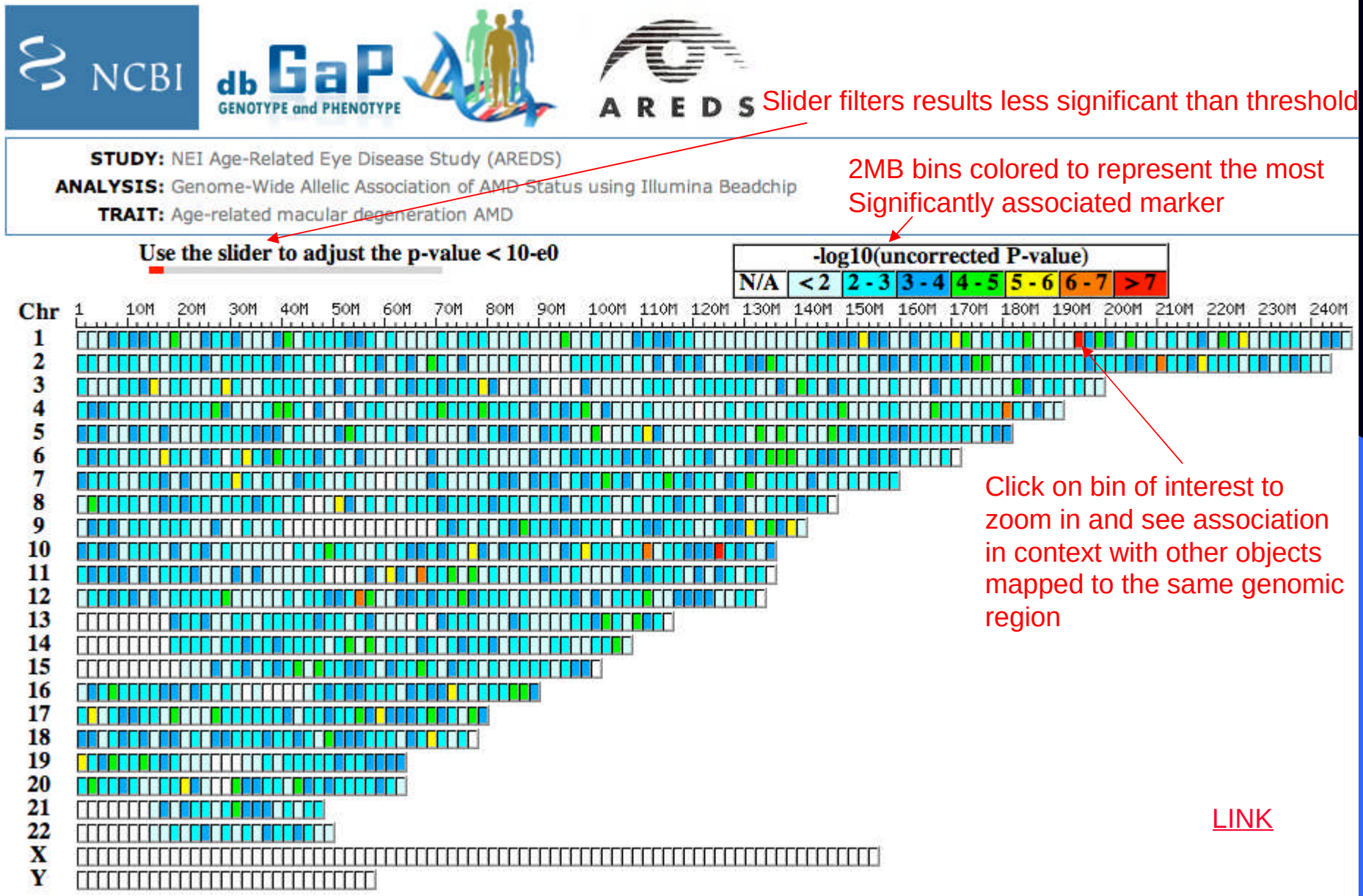
smk10
smk11
smk12
smk13
specnum
syst00
syst03
syst04
syst05
syst06
syst07
syst08

Link to variable report

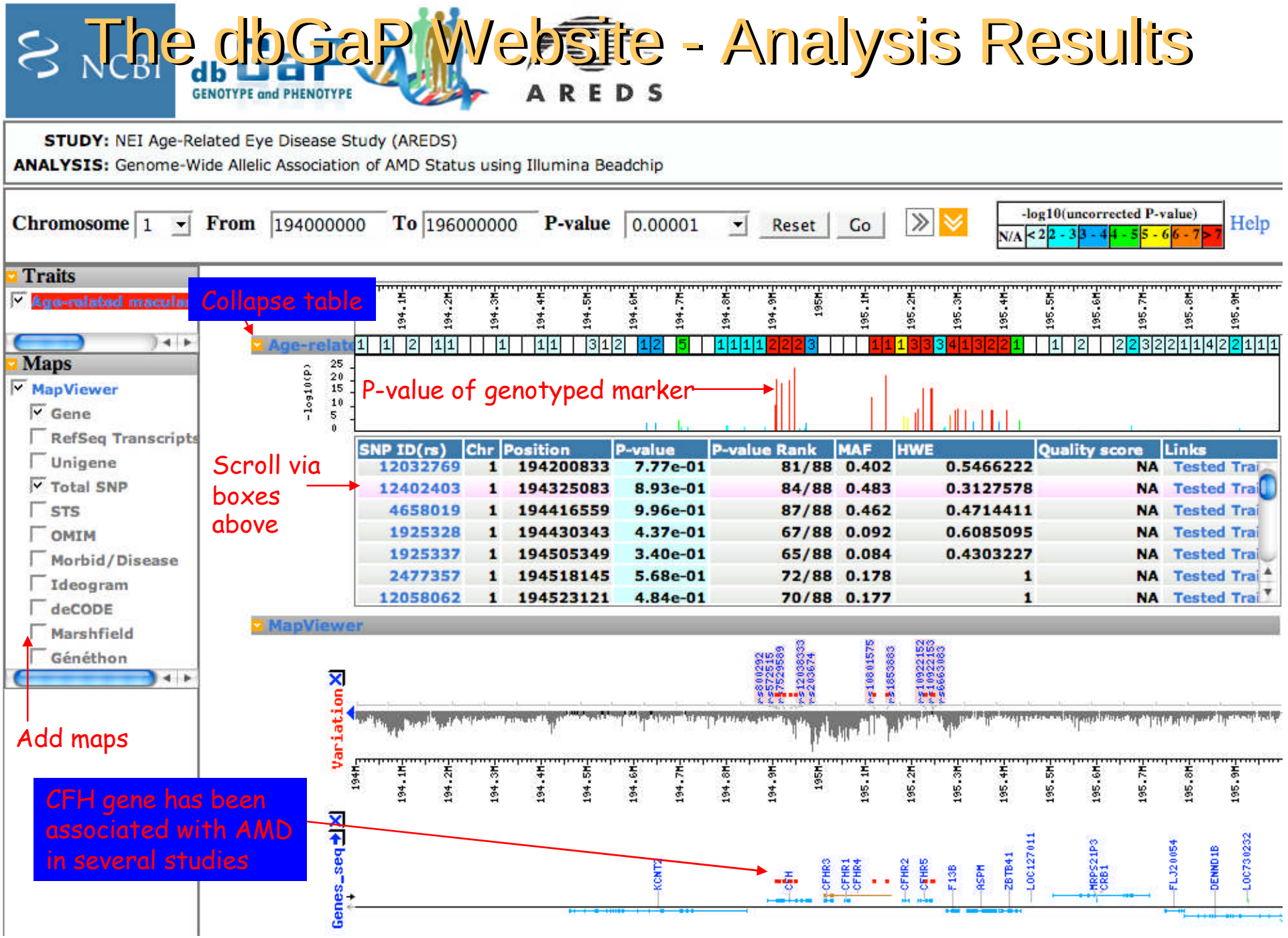
Associated Documents

NEI Age-Related Eye Disease Study (AREDS)

The dbGaP Website - Analysis Results



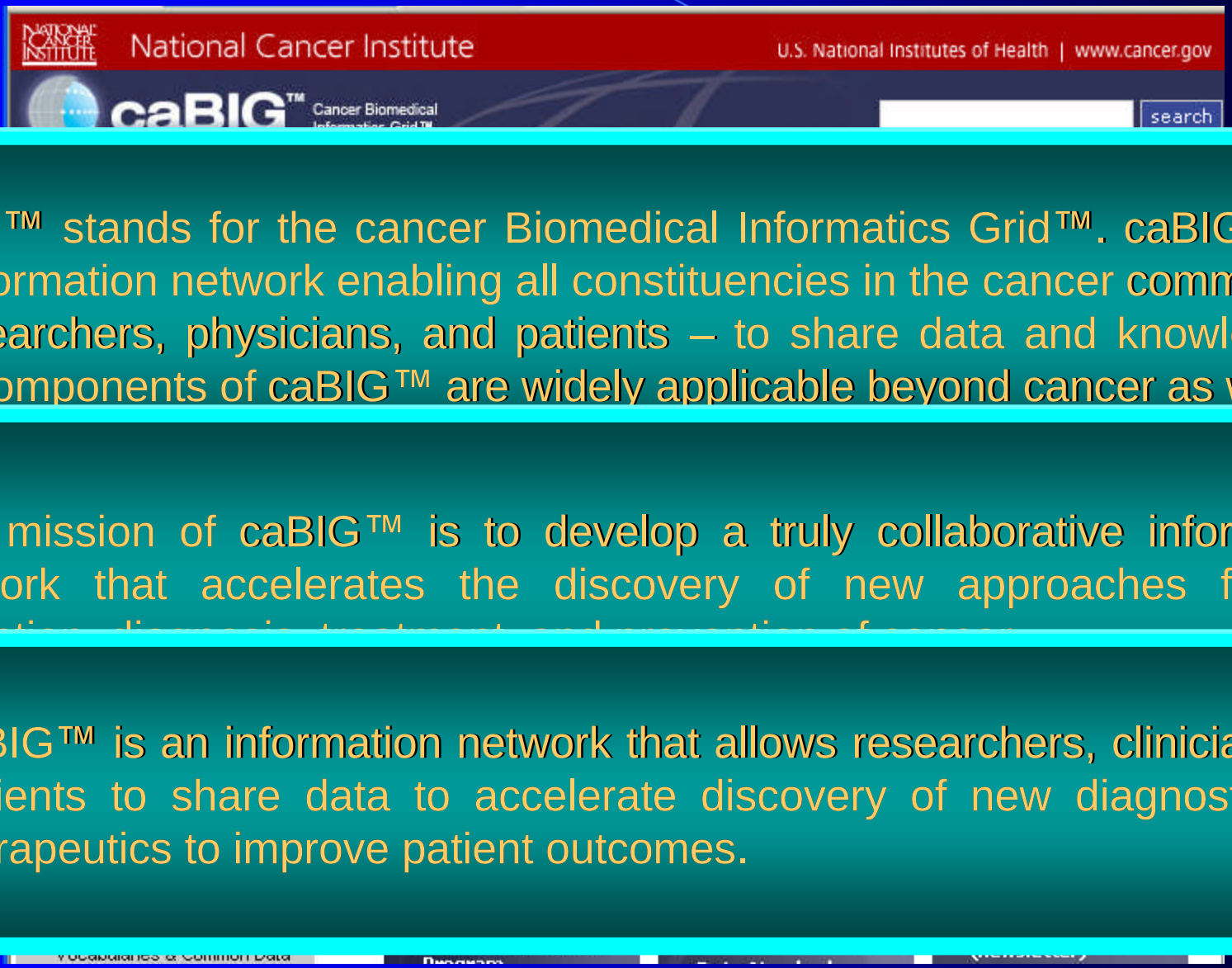
The dbGaP Website - Analysis Results



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

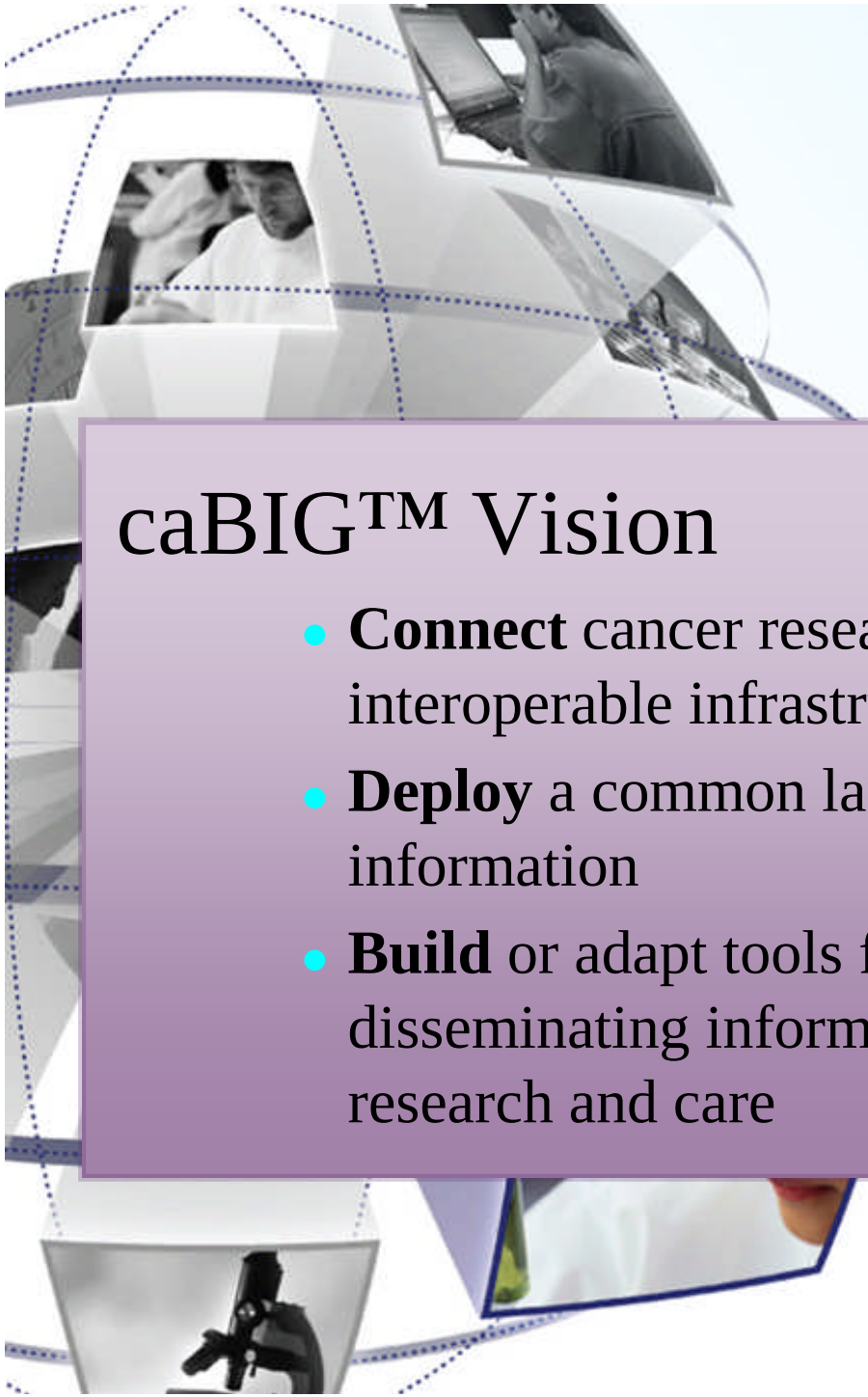
The caBIG



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

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.



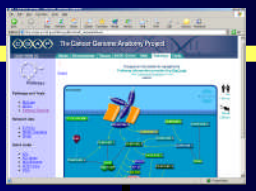
caBIG™
cancer Biomedical Informatics Grid™

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

Pathway Analysis



Protein Classification

Genomic Annotations

Detection

Etiology/ Genetics

Clinical Protocols

Translational Research

Laboratory Management

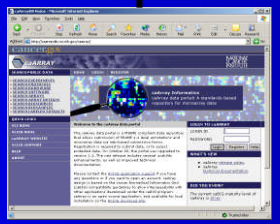


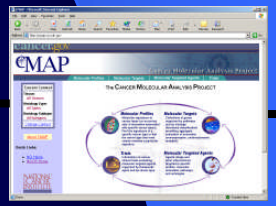
Image Archive

Gene/Protein Expression Studies

Animal Models

Prevention

Treatment



Molecular Analysis

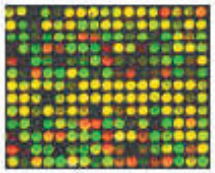
- *Interconnecting over 60 cancer centers*
- *Enabling novel biomarker discoveries towards personalized medicine*
- *Received 2005 Service to America Award for contributions to brain cancer research*

- *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



Expression array data

SNPArray data

Clinical data

Pathology

Radiology

Proteomics data

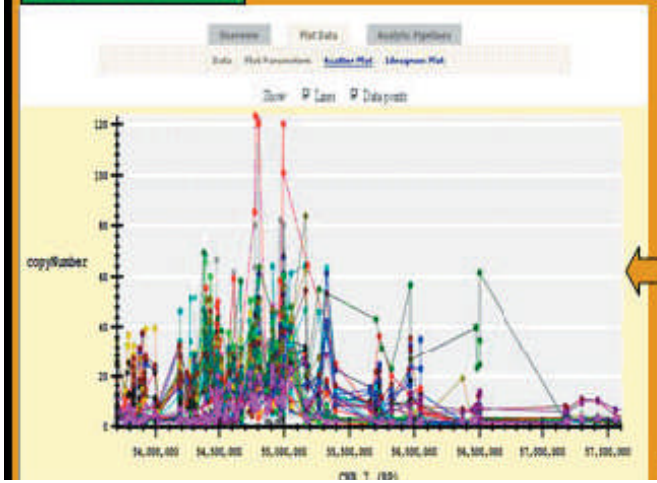
Datawarehouse

Query Engine

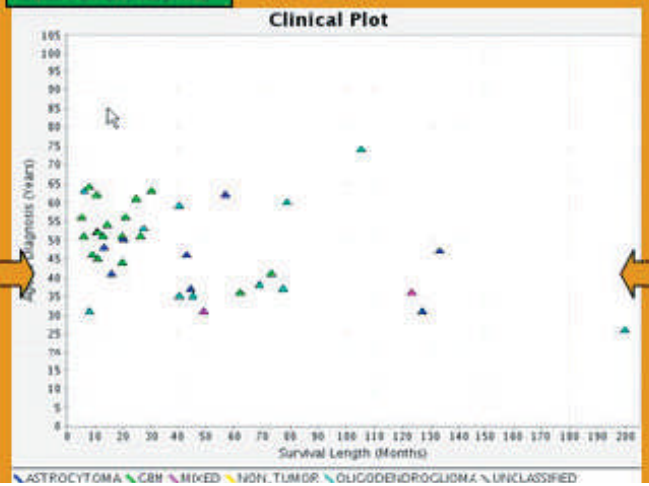
Comparison Tools	Visualization Tools	Developer Tools	Intuitive User Interface
-Access and Integrate clinical and genomic data - ntegrate copy number data with Gene expression data -Compare molecular signatures from different disease groups	-Copy number frequency charts -Ideograms -PCA graph -Kaplan-Meier survival charts -user-friendly reports	-Service APIs to access Clinical genomic objects and Biomarker Findings -Analysis APIs -DD_ and data load scripts -Java Docs	-Navigation based on user selection -Context-sensitive help throughout the application -Minimal training required to use the system

User Interface

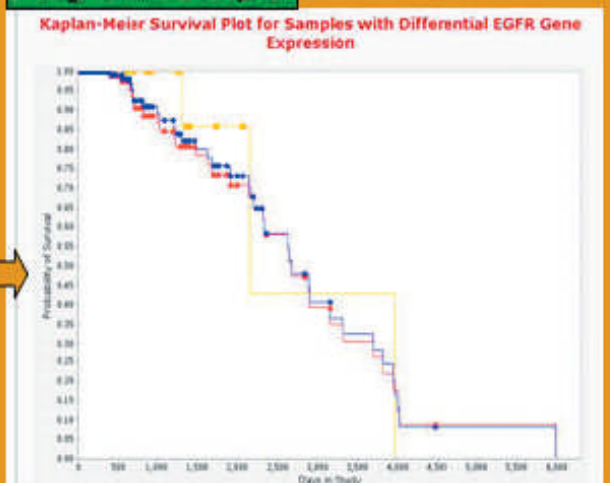
Genomic Data



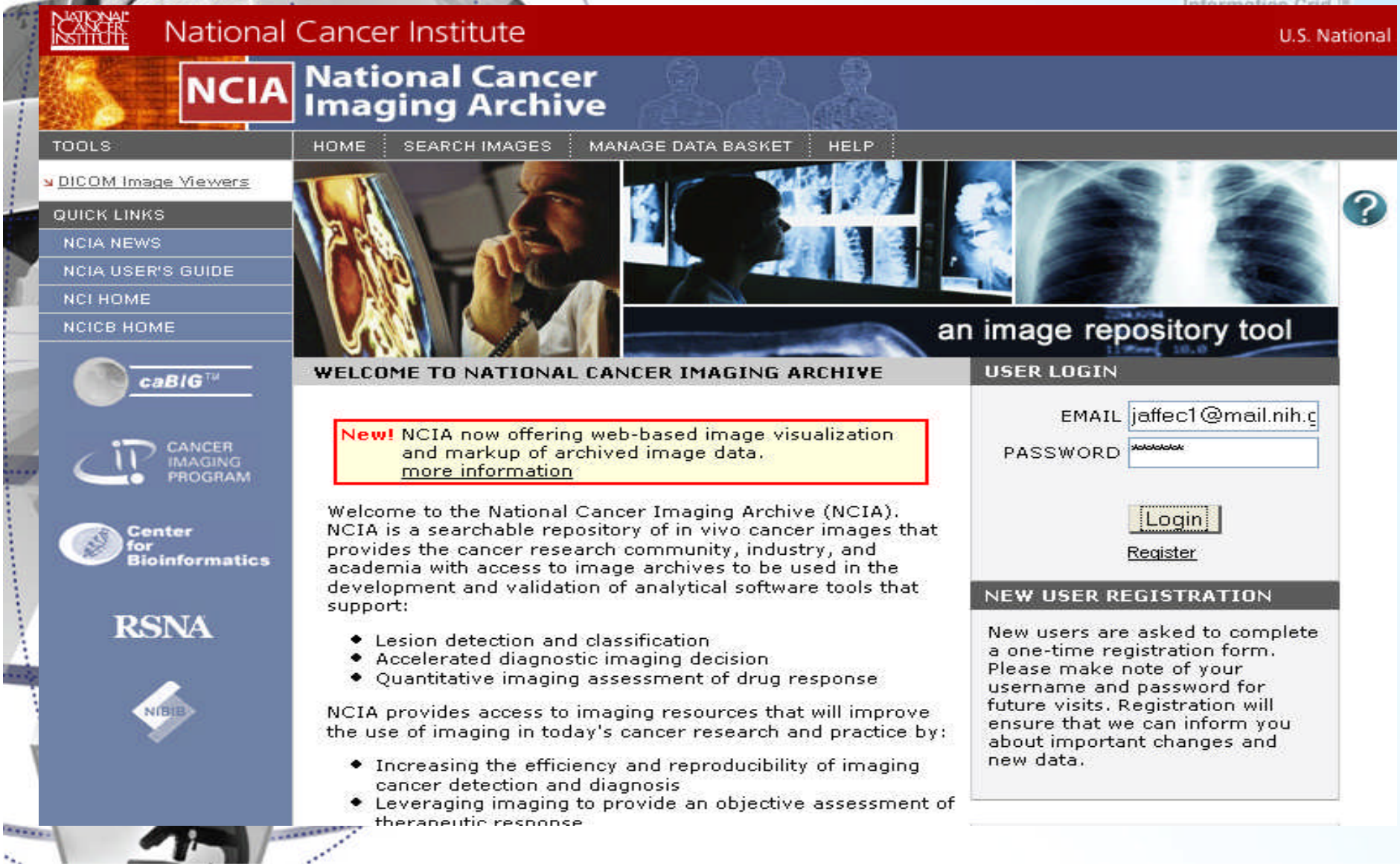
Clinical Data



Integrative Analysis



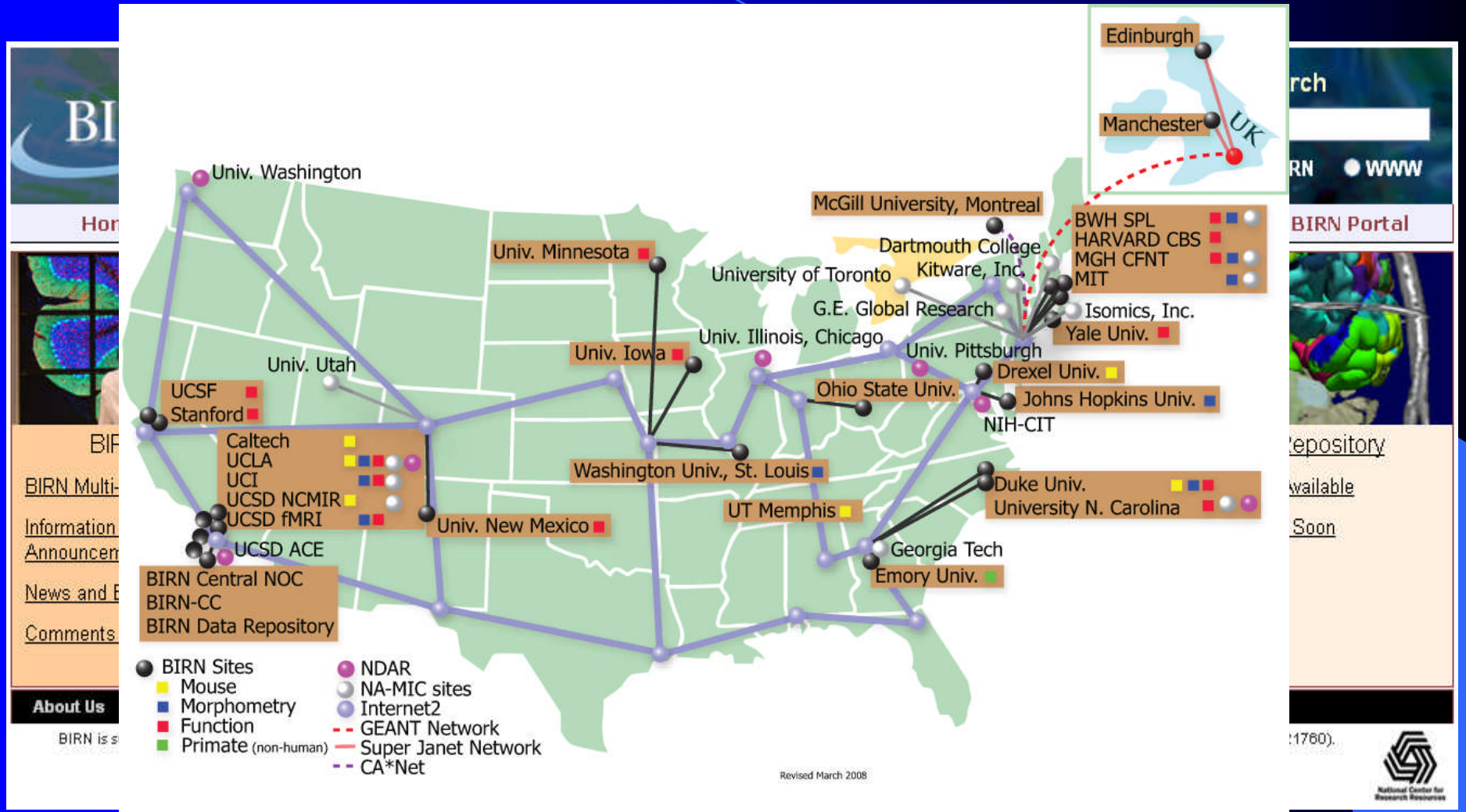
National Cancer Imaging Archive - NCIA



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


Edinburgh
Manchester
UK
Univ. Washington
McGill University, Montreal
BWH SPL
HARVARD CBS
MGH CFNT
MIT
Dartmouth College
Kitware, Inc.
Univ. Minnesota
University of Toronto
G.E. Global Research
Isomics, Inc.
Yale Univ.
Univ. Illinois, Chicago
Univ. Iowa
Univ. Pittsburgh
Drexel Univ.
Univ. Utah
Ohio State Univ.
Johns Hopkins Univ.
UCSF
Stanford
NIH-CIT
Caltech
UCLA
UCI
UCSD NCMIR
UCSD fMRI
Washington Univ., St. Louis
Duke Univ.
University N. Carolina
UT Memphis
Univ. New Mexico
Georgia Tech
UCSD ACE
Emory Univ.
BIRN Central NOC
BIRN-CC
BIRN Data Repository
BIRN Sites
Mouse
Morphometry
Function
Primate (non-human)
NDAR
NA-MIC sites
Internet2
GEANT Network
Super Janet Network
CA*Net
Revised March 2008

CTSA Clinical & Translational Science Awards

Translating Discoveries to Medical Practice

Home | About CTSA | Participating Institutions | Events | CTSAs in the News | NIH CTSA Information | Consortium Login

CTSA Committees & Activities

Oversight Committees

- CTSA Consortium Oversight
- Pediatrics Oversight Committee

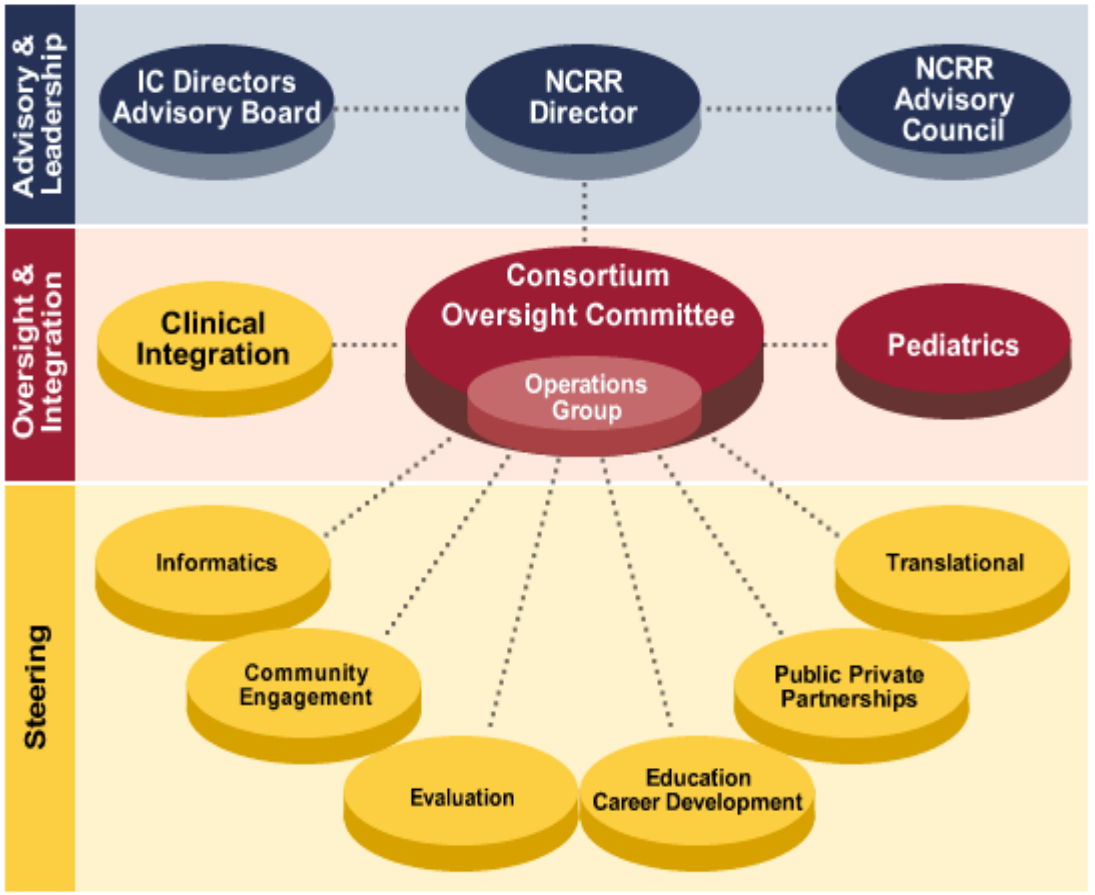
Clinical Integration Committee Workgroups:

- 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

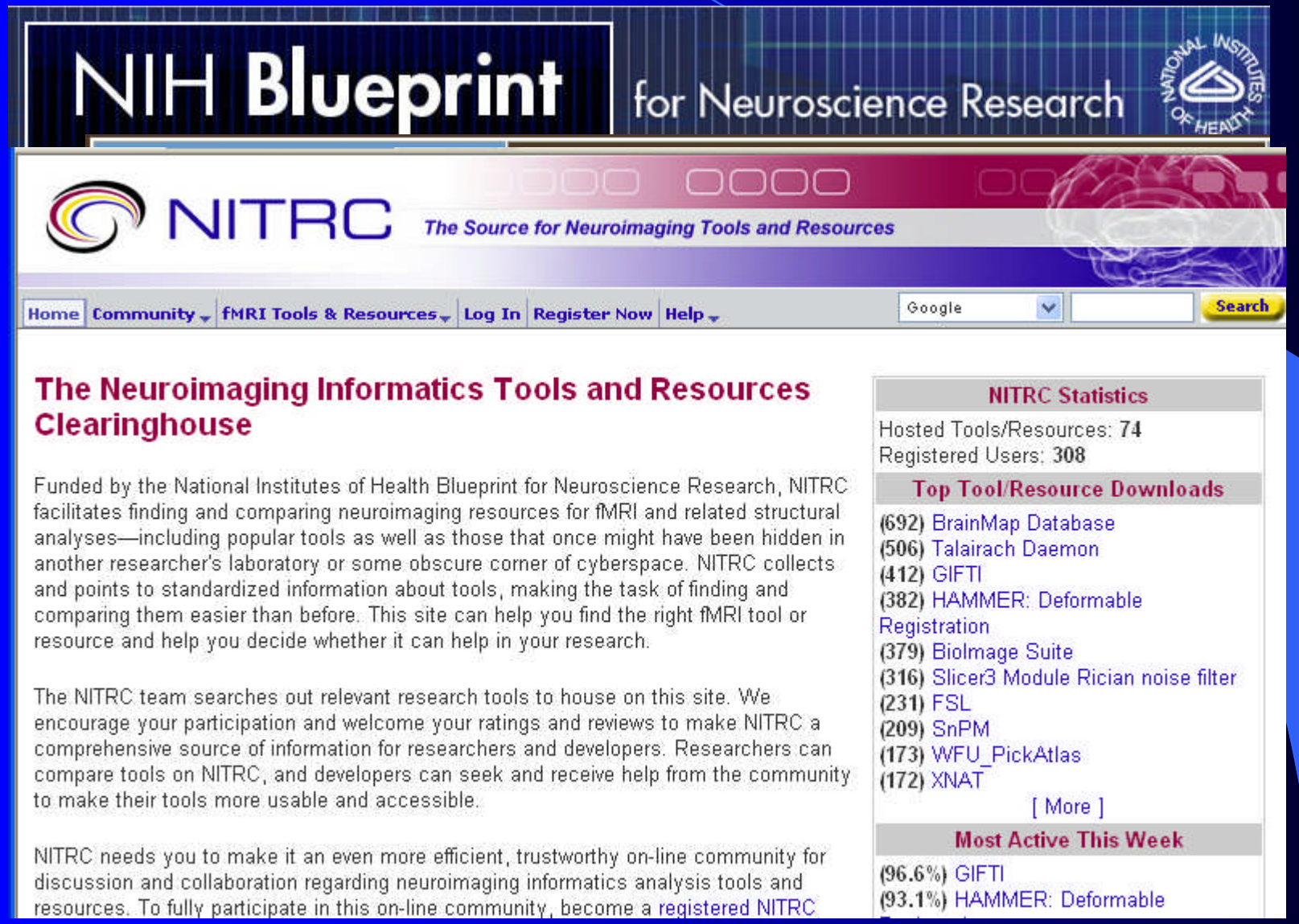
Quick Links

- 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

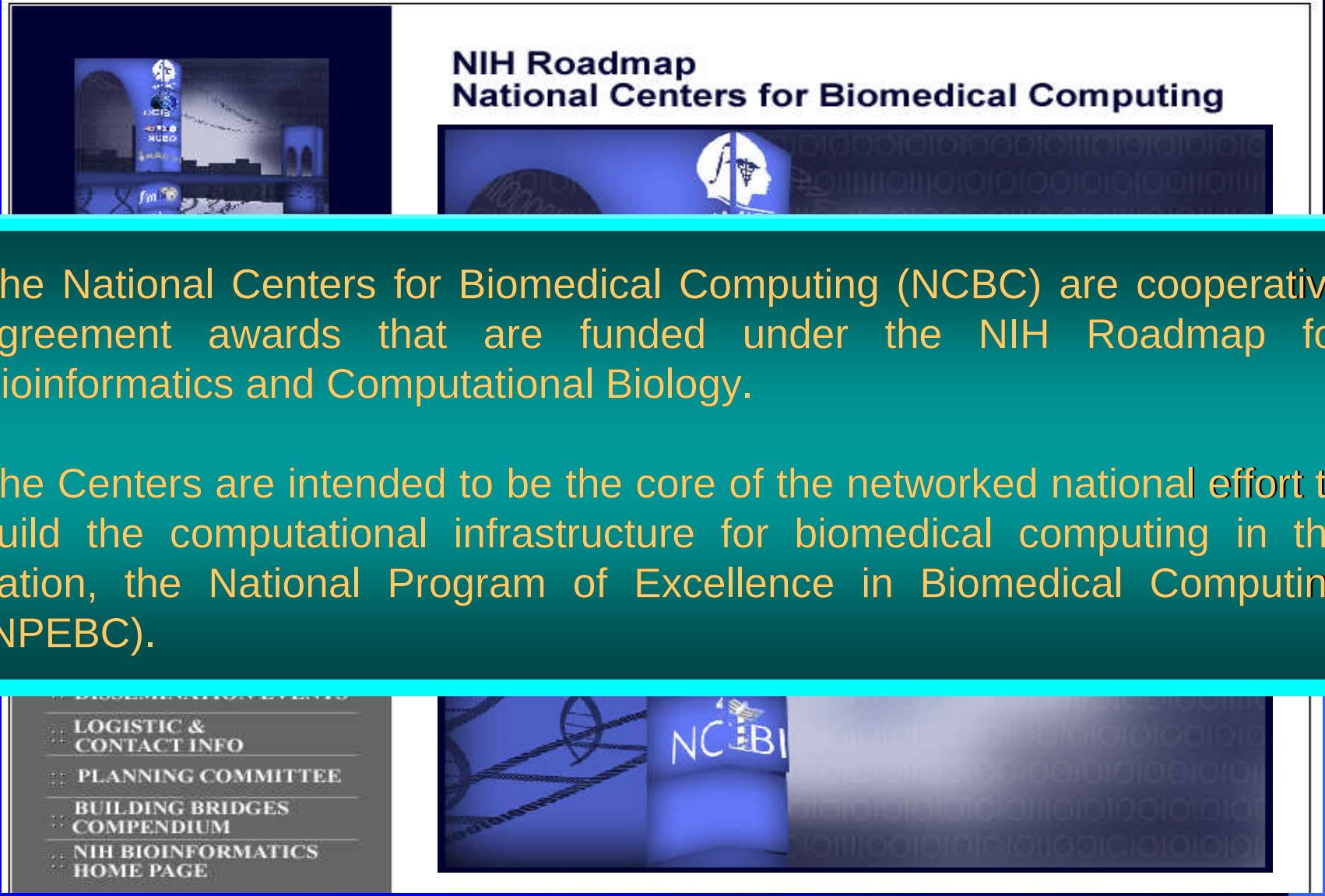
NIH Blueprint



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

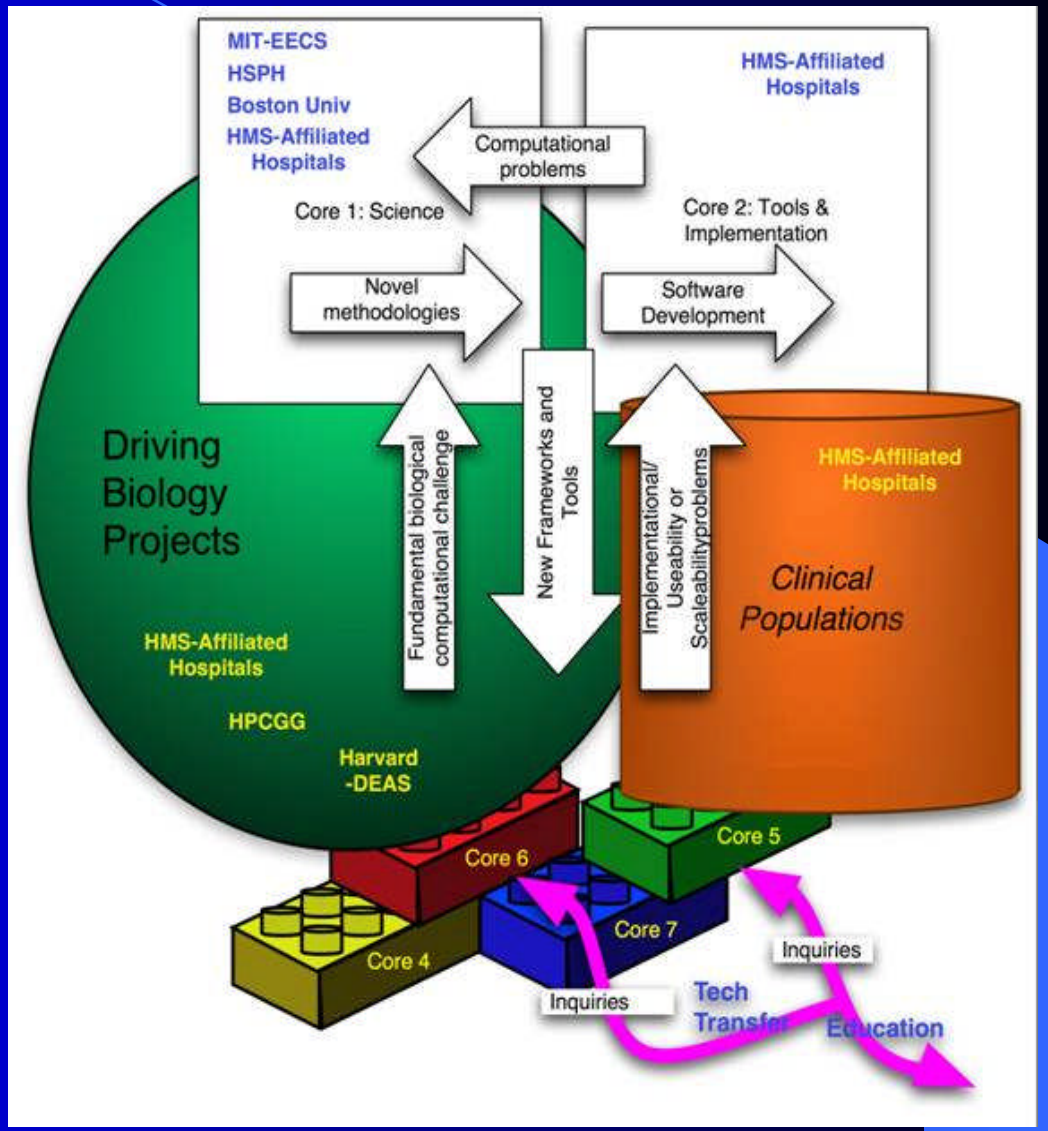


The National Centers for Biomedical Computing (NCBC) are cooperative agreement awards that are funded under the NIH Roadmap for Bioinformatics and Computational Biology.

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).

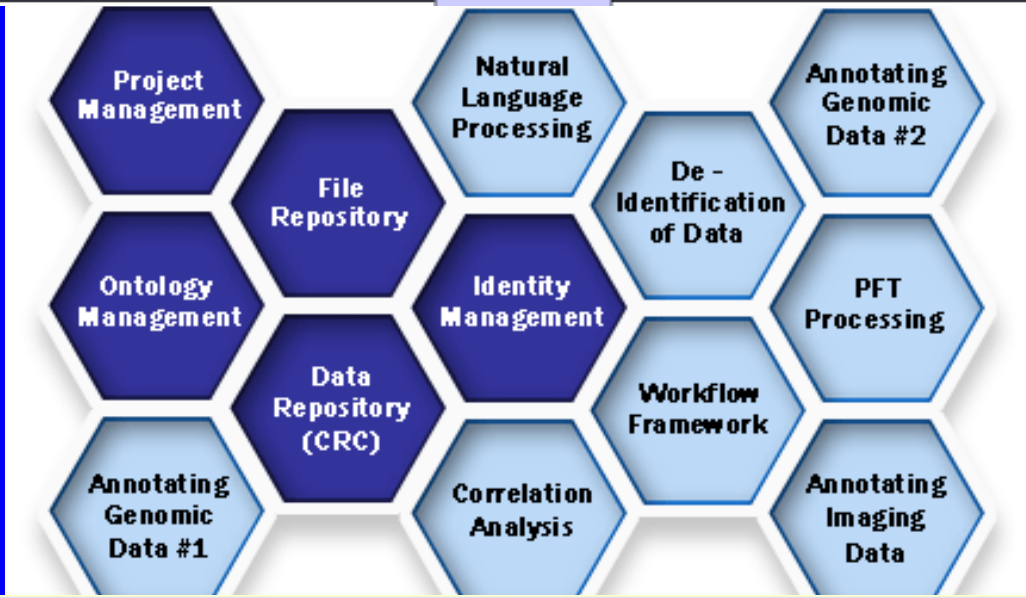
NCBC - i2b2

- 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.



i2b2 Hive

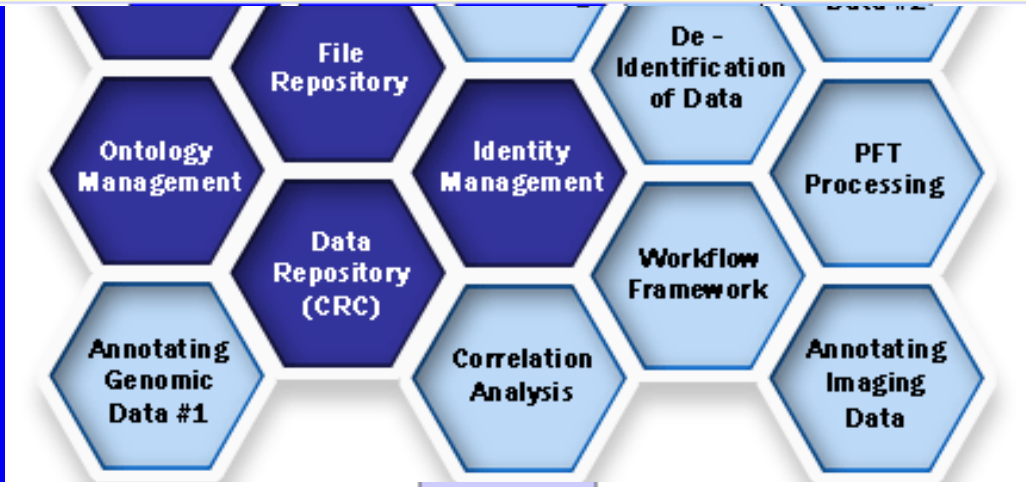
Enterprise-wide repurposing and distribution of medical record data for research



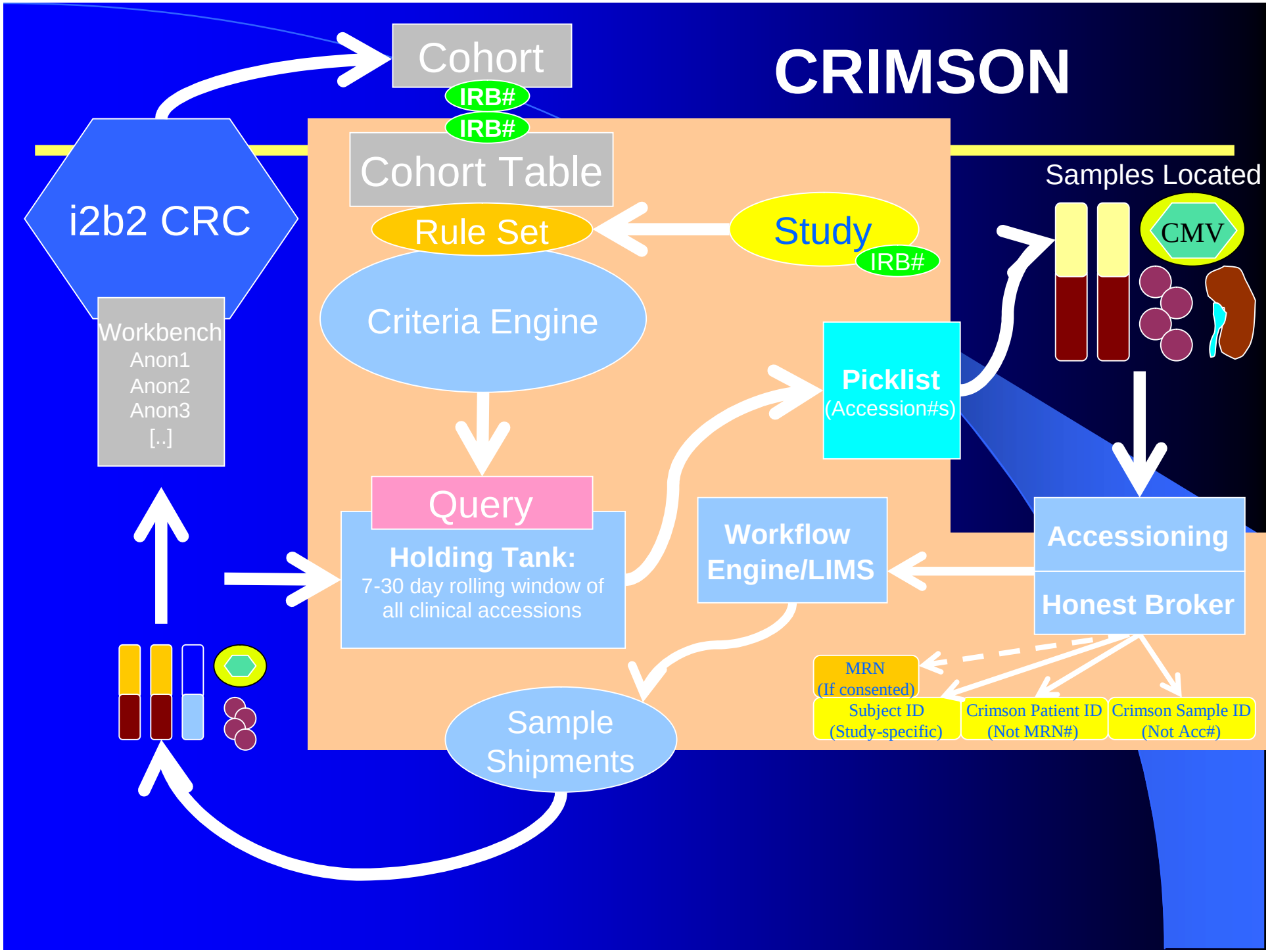
- Enable high performance collection of medical record data for querying and distribution
  - Enterprise web client
- Enable discovery within data on enterprise wide scale
  - Relationship networks
  - Pharmacovigilance

i2b2 Hive

- 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
Cohort
IRB#
IRB#
Cohort Table
i2b2 CRC
Rule Set
Study
IRB#
Criteria Engine
Workbench
Anon1
Anon2
Anon3
[..]
Samples Located
CMV
Picklist
(Accession#s)
Query
Holding Tank:
7-30 day rolling window of
all clinical accessions
Workflow
Engine/LIMS
Accessioning
Honest Broker
MRN
(If consented)
Subject ID
(Study-specific)
Crimson Patient ID
(Not MRN#)
Crimson Sample ID
(Not Acc#)
Sample
Shipments

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