



# The Cancer Biomedical Informatics Grid™ (caBIG™)

*2006 CODATA Conference  
Beijing, China*

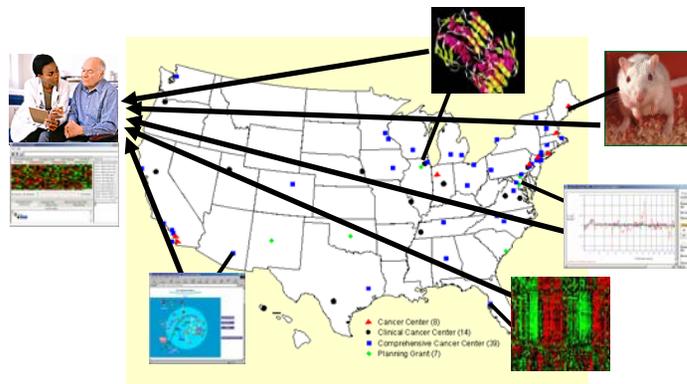
**Mary Jo Deering , Ph.D.**  
**Director, Informatics Dissemination**  
**NCI Center for Bioinformatics**





## Cancer Biomedical Informatics Grid™ (caBIG™)

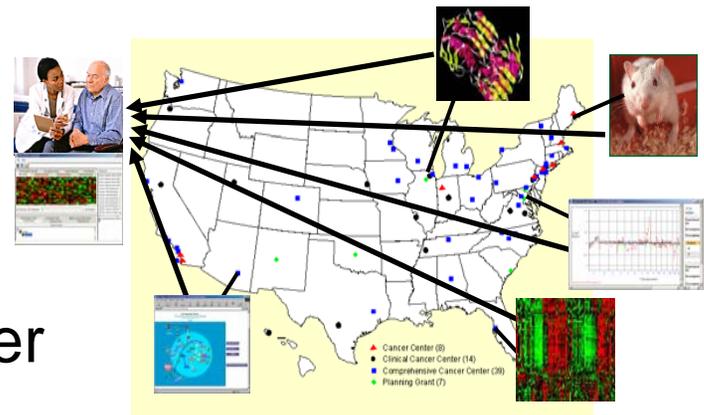
- Common, widely distributed infrastructure permits research community to focus on innovation
- Shared vocabulary, data elements, data models facilitate information exchange
- Collection of interoperable applications developed to common standards
- Raw published cancer research data is available for mining and integration





## Cancer Biomedical Informatics Grid™ (caBIG™)

- caBIG™ infrastructure and tools are widely applicable outside cancer



- caBIG™ components may be used by anyone



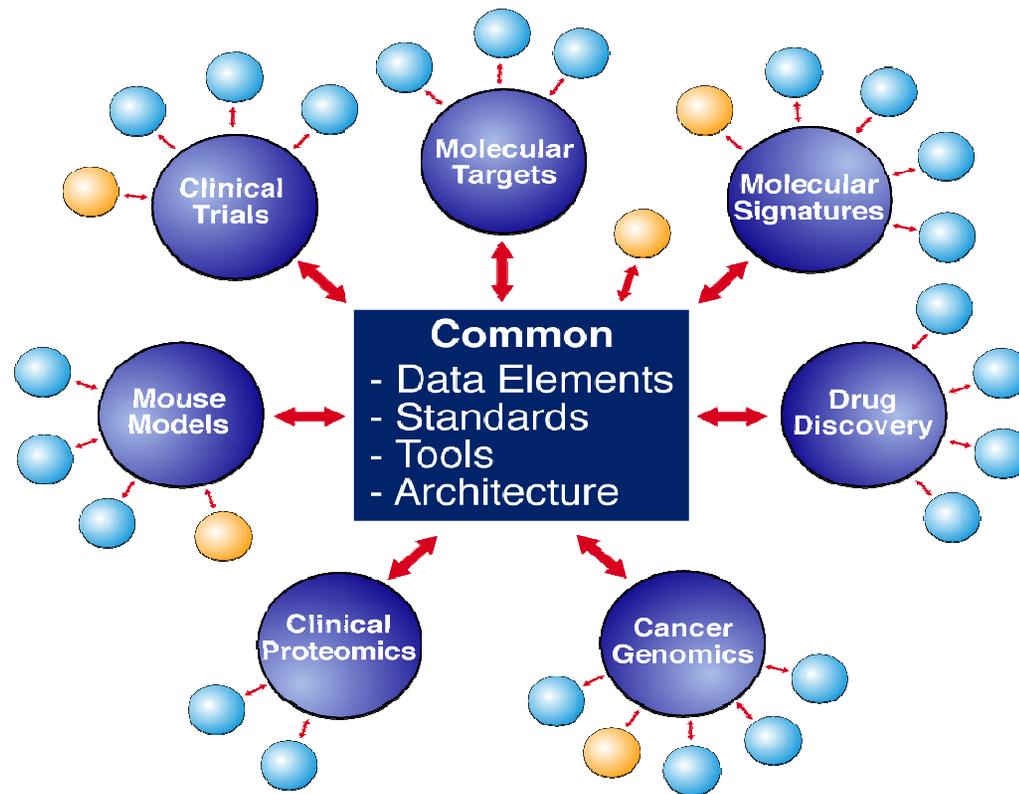
# caBIG™ principles

- Open source
- Open access
- Open development
- Federated



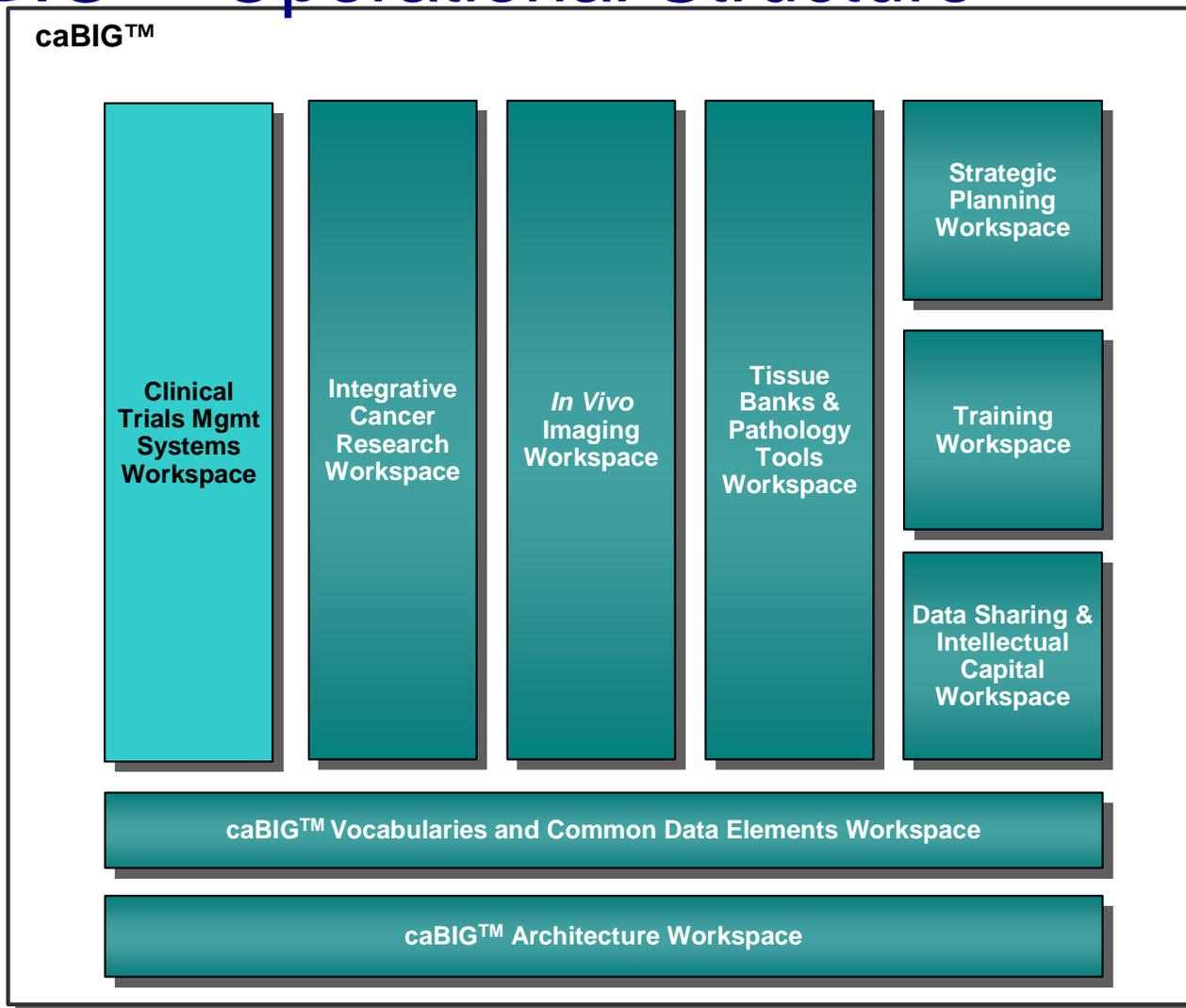


# caBIG™'s Informatics Core





# caBIG™ Operational Structure





## 2006 Clinical Trial Tools Development Activities

- caAERS
- Patient Study Calendar
- Lab Data Hub
- Making other CTMS systems caBIG compatible





# Clinical Research IT Infrastructure

Clinical Systems

Cancer Institute

Labs, EMR, Tissue, etc.

Patient Health Record

## Translation Service

HL7-v2.x, other

HL7/CAM  
SDK

HL7-v3

HL7 transactional database

## Clinical Trials

etc.  
Lifecycle Management  
Adverse Events  
Participant Registry  
EDC  
Clinical Data Mgmt

HL7-v3, Janus

Clinical Research Information Exchange

HL7-v3, Janus

## External Reporting

FDA  
SPONSOR  
NCI  
other

De-identification Services

Research Data Warehouse





## Integrated Cancer Research

- Microarray Repositories
- Data Analysis & Statistics
- Informatics for Proteomics
- Genome Annotation
- Pathways Tools
- Translational Tools
- Population Sciences and Cancer Control





Cancer Array Informatics - Microsoft Internet Explorer

Address: <http://caarray.nci.nih.gov/>

Links: FXndi, FX, NCICB, SAIC, caIMAGE, caMOD, caARRAY, T, WP, SD, PubMed, GMOD, MW, Google, LL, NIHcsc, WM

**cancer.gov**

**NATIONAL CANCER INSTITUTE**

**CARRAY**

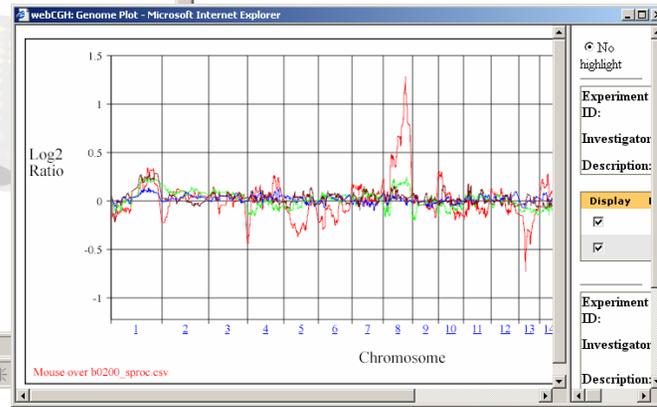
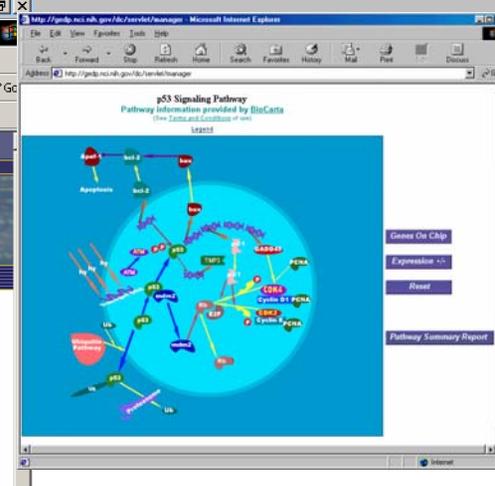
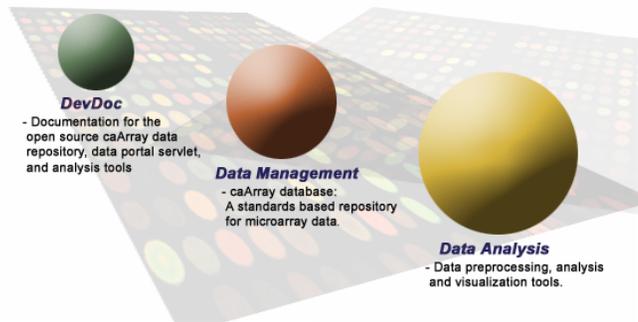
Cancer Array Informatics

NCICB Home | Home | DevDoc | Data Management | Data Analysis | Support

- HOME**
- [caArray Project Plan](#)
- NCI Supported Programs**
- [Director's Challenge](#)
  - [MMHCC](#)
  - [SPORes](#)
- What's New**
- [caARRAY release notes](#)
  - [XpressionWay](#)
  - [RiOgopher](#)
- caArray Team**
- [Staff](#)
  - [Contributors](#)
- Technical Documentation**
- [Latest technical docs](#)
- Related Links**
- [NCICB Home](#)
  - [MGED](#)
  - [BioGRID](#)

## caArray Data Portal

The NCICB's cancer array informatics project, caArray, consists of a microarray database and microarray data analysis and visualization tools. CaArray is an open source project, and the source code and APIs are available in the [caArray Informatics](#) page. The goals of the project are to make microarray data publicly available, and to develop and bring together open source tools to analyze these data.



Plug and Play Framework

File Help

Normalization Results | Annotations Panel | CGAP Pathways | Color Mosaic

Microarray | Analysis Results | Microarray Table | Filtering Results

All Arrays  All Markers  Enabled

Normalization Pane | Comments Pane | Marker Linear Graph Panel

Project Panel | Analysis Pane | Filtering Pane

Project Folders

Gene List

- Homo sapiens transcript
- Homo sapiens transcript
- Homo sapiens transcript
- Homo sapiens transcript
- Human 18S rRNA gene
- Human glyceraldehyde-3
- Human glyceraldehyde-3
- Human glyceraldehyde-3
- X00351 Human mRNA fr
- X00351 Human mRNA fr
- X00351 Human mRNA fr

DF2.tes

### Current Context

**Tissue:**

[All Tissues](#)

**Histology Type:**

[All Types](#)

**Histology Subtype:**

[All Subtypes](#)

[Change Context](#)

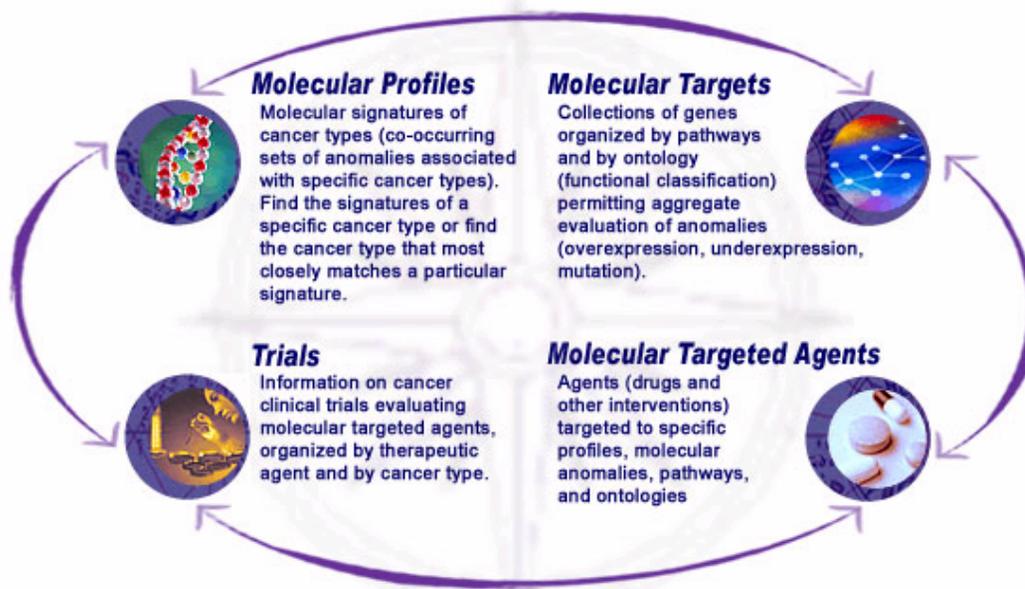
[About CMAP](#)

### Quick Links:

- [NCI Home](#)
- [NCICB Home](#)



## The CANCER MOLECULAR ANALYSIS PROJECT





NATIONAL CANCER INSTITUTE
U.S. National Institutes of Health | www.cancer.gov

Division of Cancer Epidemiology and Genetics

[Welcome](#) | [Browse Data](#) | [Bulk Data Download](#) | [Support](#) | [Feedback](#) | [Login/Register](#)

CGEMS Data Portal Version 1.0

This is the home page of the Cancer Genetic Markers of Susceptibility (CGEMS) data access. The following links provide information on the [project](#) and [background](#). The CGEMS study design uses cases and controls drawn from well designed epidemiological studies of prostate and breast cancer. DNA from these subjects is being used to generate genotypes to perform a Genome-Wide Association Study (GWAS) on over 500,000 genetic variants to determine their role in cancer susceptibility.

**CGEMS Prostate Scan Phase 1**

---

A GWAS has been conducted in a large, national study in the U.S.A., the Prostate, Lung, Colorectal, and Ovary study (PLCO). The analysis

[Browse Data](#)
[Bulk Data Download](#)

**For more information on:**

- [About CGEMS Study](#)
- [How to use the CGEMS data portal](#)
- [Register to access raw data](#)

? Click the question mark icon for context sensitive help throughout the application.

**CGEMS updates:**

# National Cancer Institute



caBIG<sup>™</sup> cancer Biomedical  
Interoperable Research