



Center
for
Bioinformatics

Biomedical Informatics for Clinical Decision Support – a vision for the 21st Century: **Informatics Infrastructure – NIH Perspective**

Ken Buetow

NCICB/NCI/NIH/DHHS



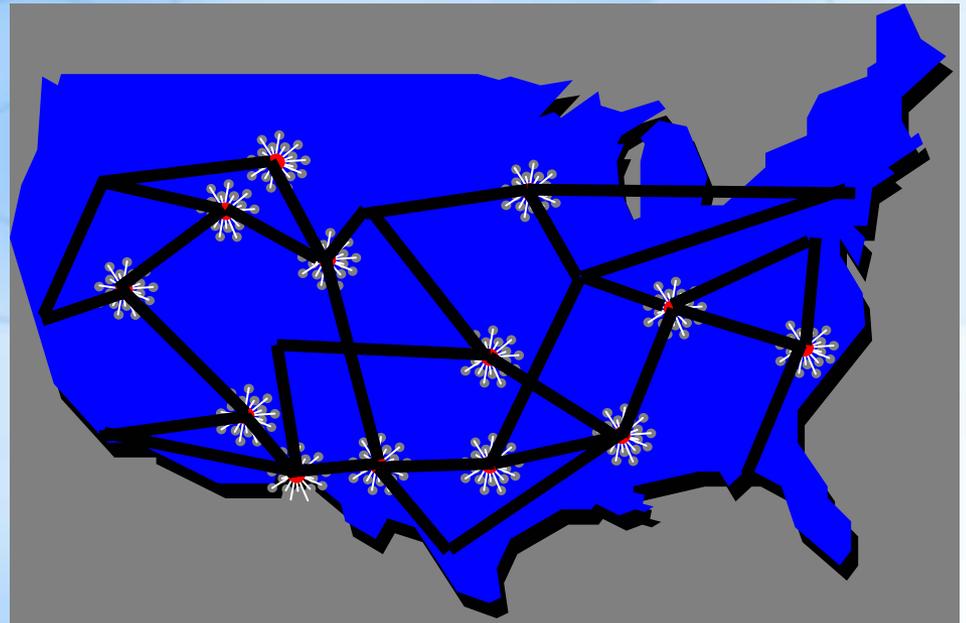
Re-engineering the CR Enterprise

- **Integration of CR networks and CR Informatics**
 - **Pilot NECTAR network**
- **Facilitate translational research**
- **Enhanced training**
- **Clinical Research Corps**
- **Coordination and Support of CR regulatory policies and processes**

Integration of Clinical Research Networks

**Establish
interoperable
networks
where clinical
studies and
trials can be
effectively
conducted**

**Ensure that
patients,
physicians and
scientists form
true
“Communities
of Research”**



NECTAR Network

- A national network of new and extant CR sites, programs, and systems
 - **National Electronic Clinical Trials and Research (NECTAR) Network**
- Linking will synergistically:
 - Expand their utility, functionality
 - Maximize connectivity
 - Provide unprecedented analytical capacity

NECTAR: Three Foci

- **Data Standards**
 - Common exchange standards and vocabularies (e.g., HL-7, LOINC, SNOMED)
- **Tools**
 - Software applications to support clinical research tasks (protocol preparation and review; IRB mgmt; AE reporting; data collection, analysis and reporting)
- **Network Infrastructure**
 - Platforms, architectures, integrating elements

Clinical Research Inventory

- **Estimated Start Date: September 30, 2004**
- **Estimated Completion Date: June 30, 2006**
- **Major Goals:**
 - **Develop inventory and database of existing clinical research networks**
 - **Detailed description of existing practices and assessment of best practices**
 - **National Leadership Forum on the results of inventory and the assessment of studies**

Feasibility of Integrating and Expanding Clinical Research Networks

(BAA-RM-04-23)

- **Foster clinical research networks that are based on common or inter-operable infrastructure elements and that conduct research both in academic and clinical care settings.**
- **Integrate and expand clinical research networks will broaden the kinds of research questions that can be addressed and enhance the efficiency of conducting clinical research.**
- **Solicit feasibility study proposals that will test methods; activities to expand, broaden, and optimize existing approaches and can be generalized to the greater clinical research community.**



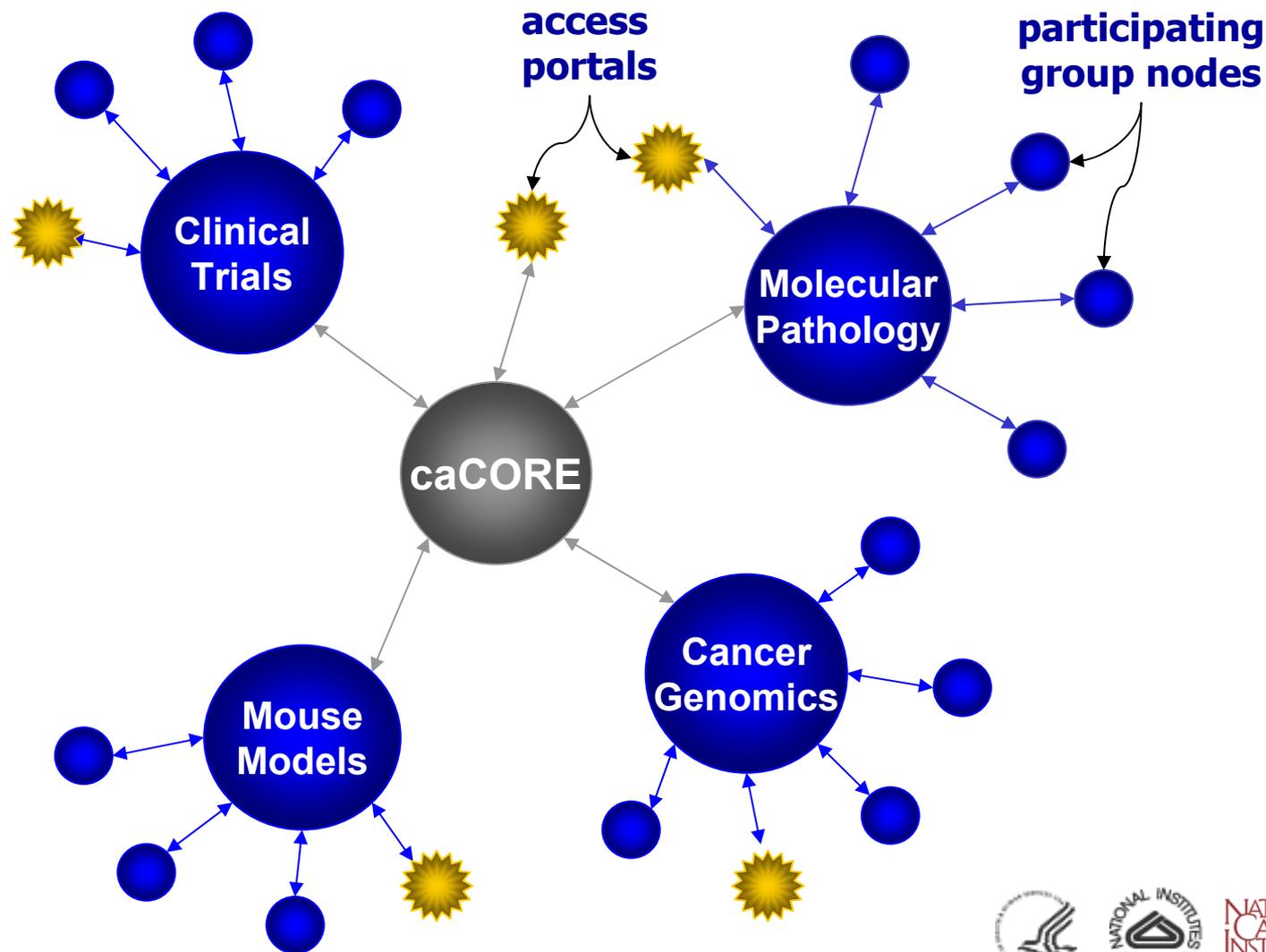
NCI biomedical informatics

- ***Goal:*** A virtual web of interconnected data, individuals, and organizations redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise





building common architecture, common tools, and common standards



Interoperability

■ in·ter·op·er·a·bil·i·ty

- ability of a system...to use the parts or equipment of another system

Source: Merriam-Webster web site

■ interoperability

- ability of two or more systems or components to exchange information and to use the information that has been exchanged.

Source: IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries, IEEE, 1990]

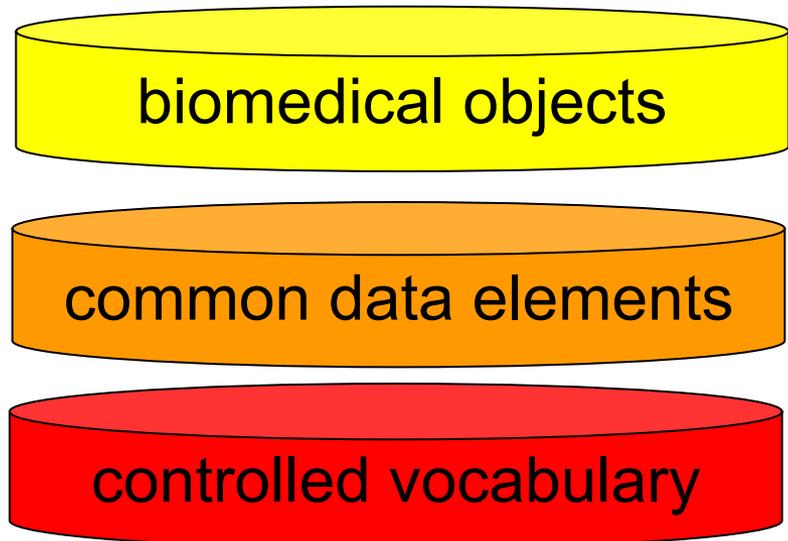
Syntactic
interoperability

Semantic
interoperability



caCORE – common ontologic representation environment

- Information integration
- Cross-discipline reasoning





Enterprise Vocabulary

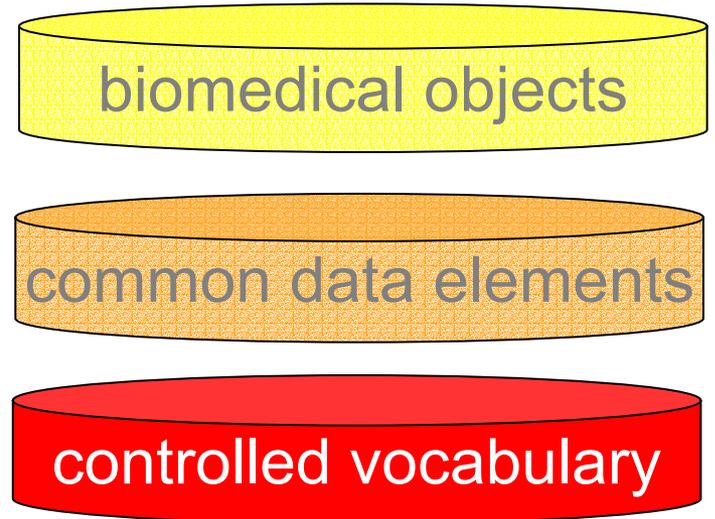
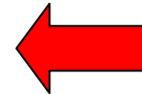
■ NCI Meta-Thesaurus

(Cross-map standard vocabularies/ontologies, e.g. SNOMED, MEDRA, ICD)

- Semantic integration, inter-vocabulary mapping
- UMLS Metathesaurus extended with cancer-oriented vocabularies
 - 800,000 Concepts, 2,000,000 terms and phrases
 - Mappings among over 50 vocabularies

■ NCI Thesaurus

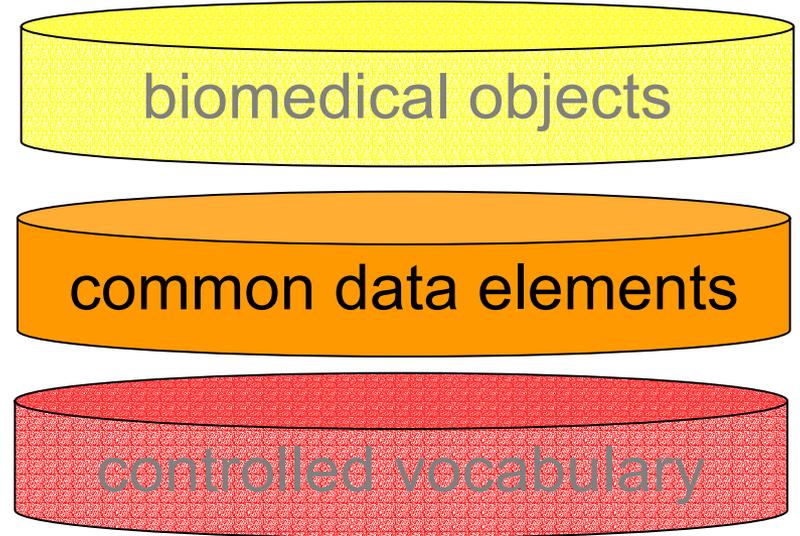
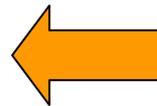
- Description logic-based
- 18,000 "Concepts"
 - Concept is the semantic unit
 - One or more terms describe a Concept – synonymy
 - Semantic relationships between Concepts





Common Data Elements

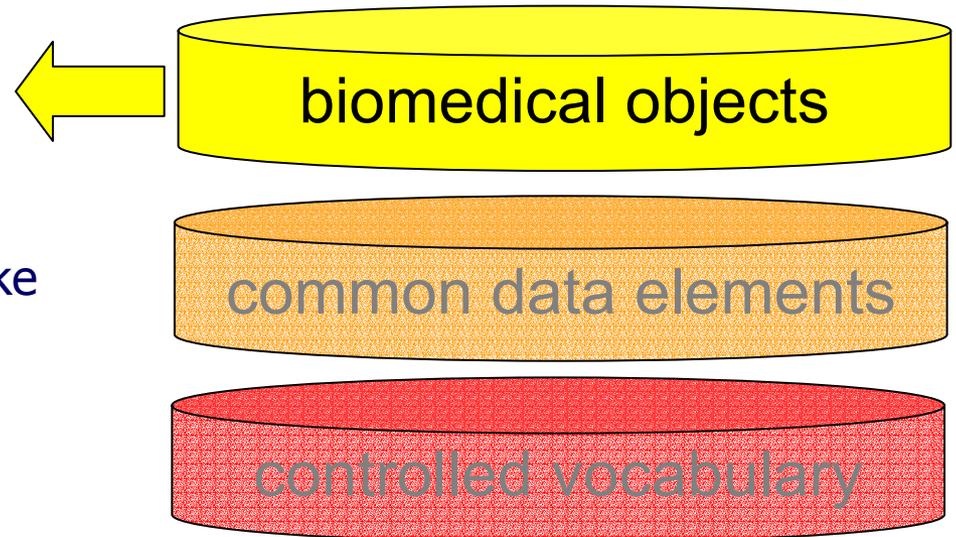
- Structured data reporting elements (e.g. LOINC)
- ISO11179 compliant





Biomedical Information Objects

- Computer model of a biomedical object – “Plato’s Forms”
 - capture properties of object
 - can be joined together to make complex systems
 - isolate data from data source
 - isolate applications from data
- Examples:
 - HL7-RIM
 - MAGE-OM





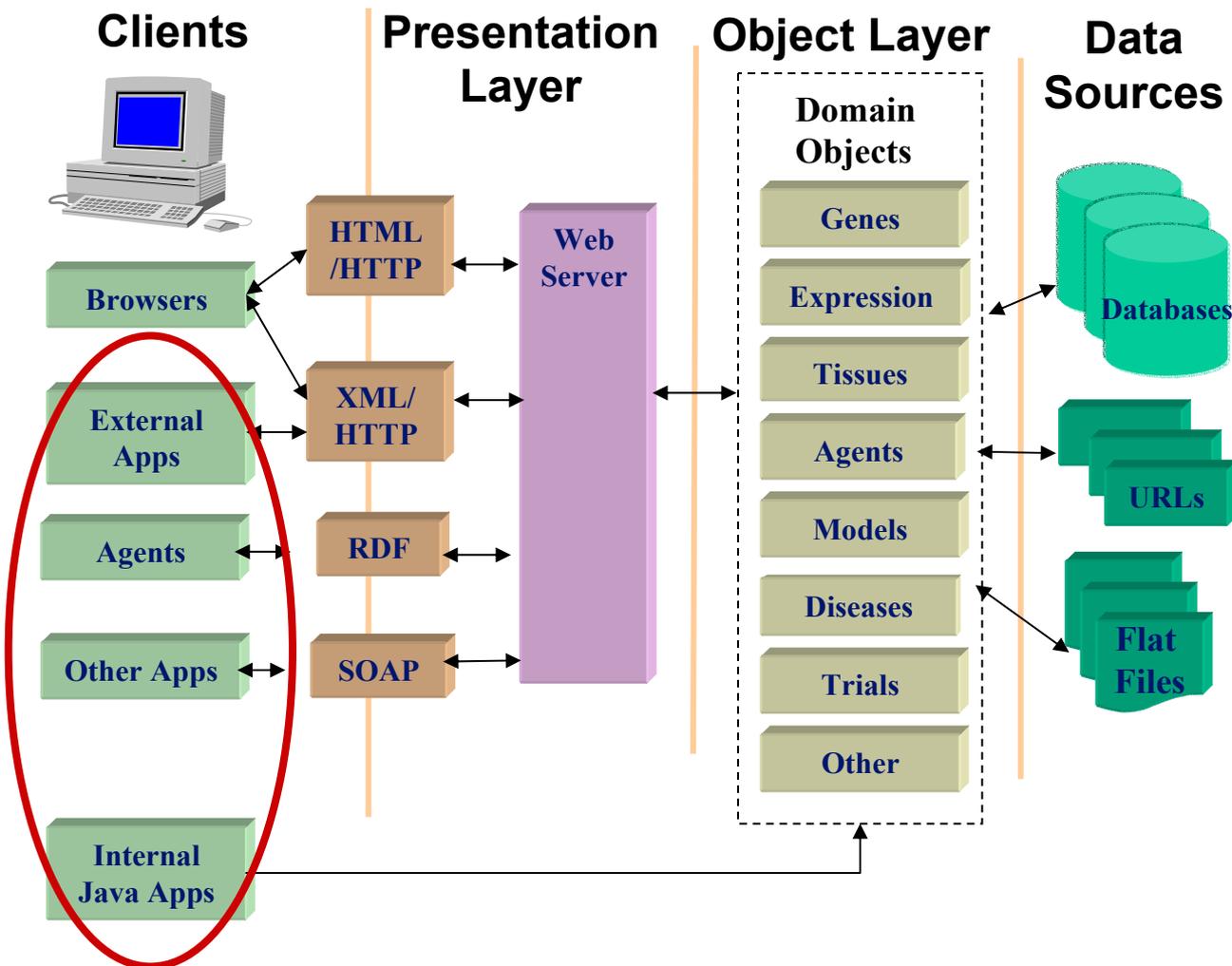
Standards Supporting Infrastructure

- Enterprise Vocabulary Services (EVS)
- cancer Data Standards Repository (caDSR)
- cancer Bioinformatics Infrastructure Objects (caBIO)





...an open, sharing architecture



- open source
- open development
- open access
- federated





caBIG

*cancer Biomedical
Informatics Grid*



Cancer Biomedical Informatics Grid (caBIG)

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





caBIG

*cancer Biomedical
Informatics Grid*



caBIG action plan

- Establish pilot network of Cancer Centers
 - Groups agreeing to caBIG principles
 - Mixture of capabilities
 - Mixture of contributions
- Expanding collection of participants
- Establish consortium development process
 - Collecting and sharing expertise
 - Identifying and prioritizing community needs
 - Expanding development efforts
- Moving at the speed of the internet...





caBIG

cancer Biomedical
Informatics Grid



Three Domain Workspaces and two Cross Cutting Workspaces will be launched during the Pilot phase

DOMAIN WORKSPACE 1 Clinical Trial Management Systems

Will address the need for consistent, open and comprehensive tools for clinical trials management.

DOMAIN WORKSPACE 2 Integrative Cancer Research

Will provide tools and systems to enable integration and sharing of information.

DOMAIN WORKSPACE 3 Tissue Banks & Pathology Tools

Will provide for the integration, development, and implementation of tissue and pathology tools.

Will be responsible for evaluating, developing, and integrating systems for vocabulary and ontology content, standards, and software systems for content delivery

CROSS CUTTING WORKSPACE 1 Vocabularies & Common Data Elements

Will develop architectural standards and provide architectural assistance as necessary to other workspaces.

CROSS CUTTING WORKSPACE 2 Architecture





caBIG

*cancer Biomedical
Informatics Grid*



caBIG deliverables

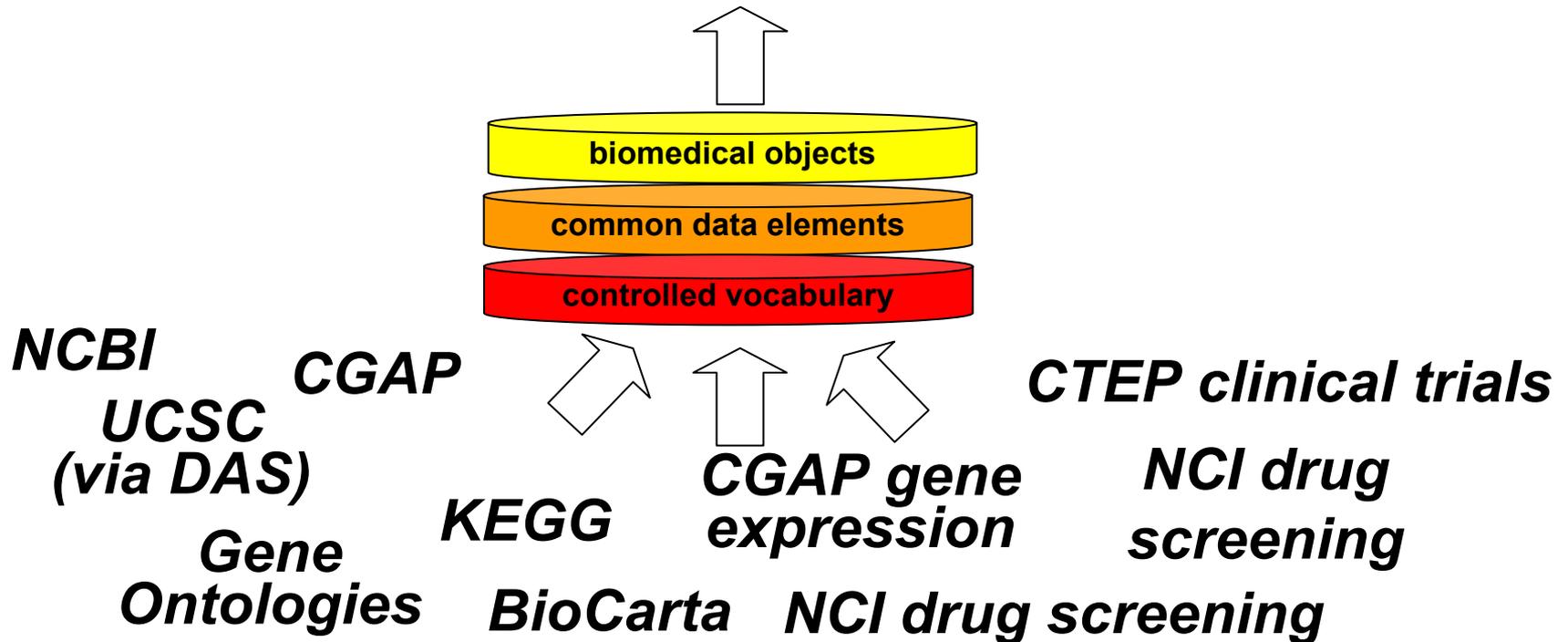
- **Componentized, standards-based Clinical Trials Management System**
 - e-IND filing/regulatory reporting with FDA
 - Electronic management of trials
 - Integration of diverse trials
- **Tissue Management System**
 - Systematic description and characterization of tissue resources
 - Ability to link tissue resources to clinical and molecular correlative descriptions
- **“Plug and Play” analytic tool set**
- **Diverse library of raw, structured data**





Cancer Molecular Analysis Project (CMAP) - a prototypic biomedical data integration effort

Profiles, Targets, Agents, Clinical Trials



cancer.gov



Cancer Molecular Analysis Project

Molecular Profiles

Molecular Targets

Molecular Targeted Agents

Trials

Current Context

Tissue:

[All Tissues](#)

Histology Type:

[All Types](#)

Histology Subtype:

[All Subtypes](#)

[Change Context](#)

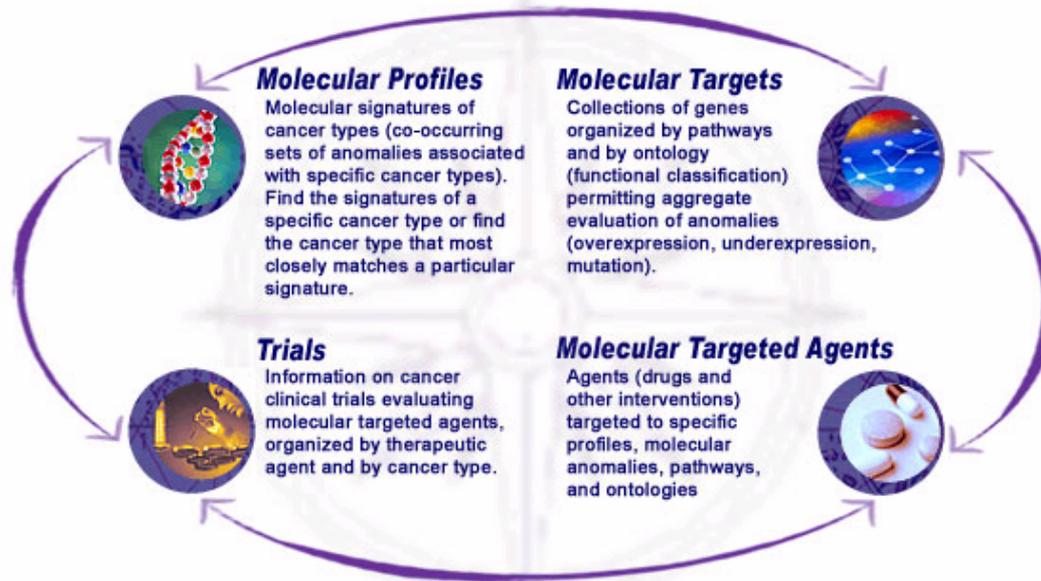
[About CMAP](#)

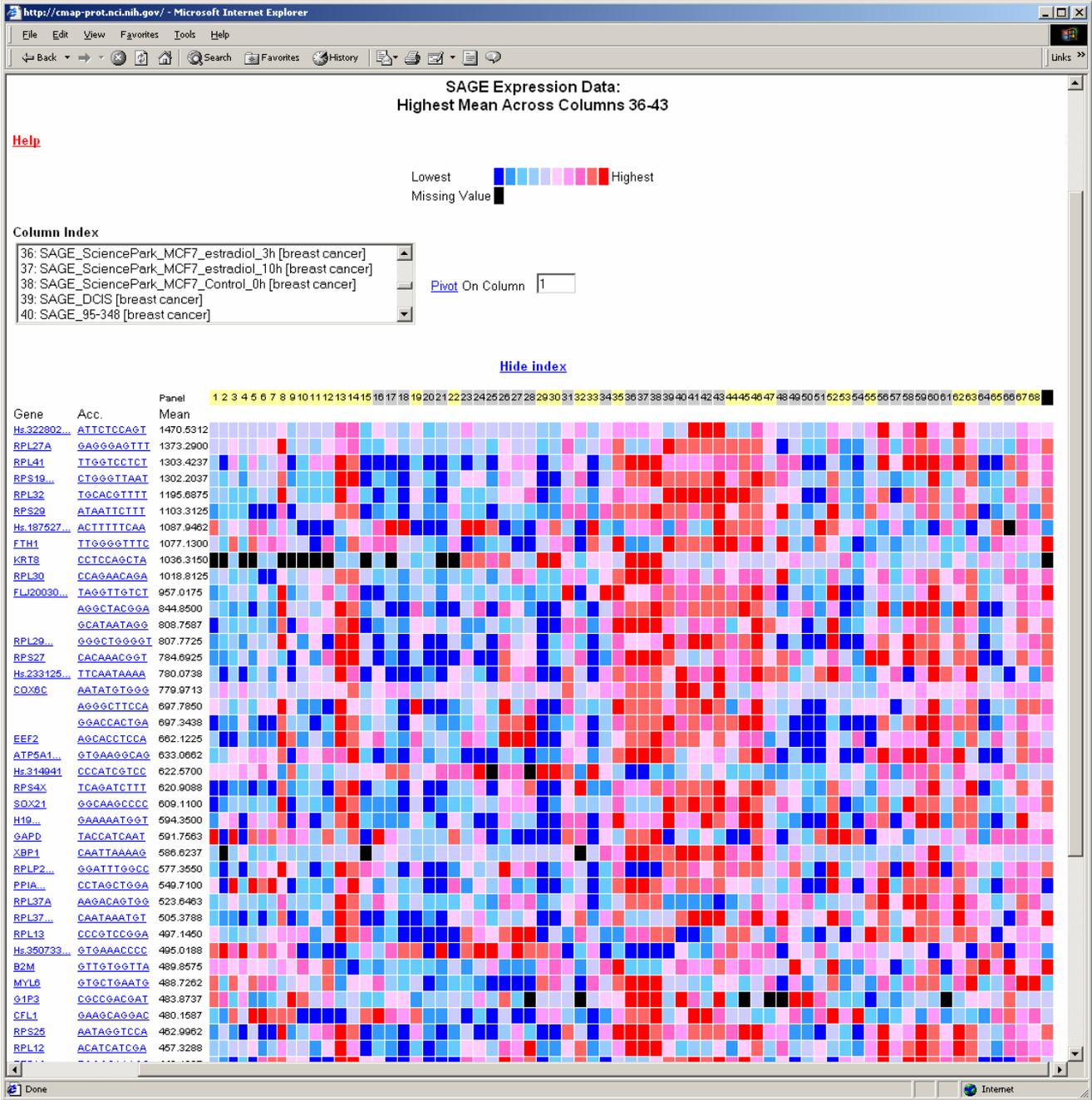
Quick Links:

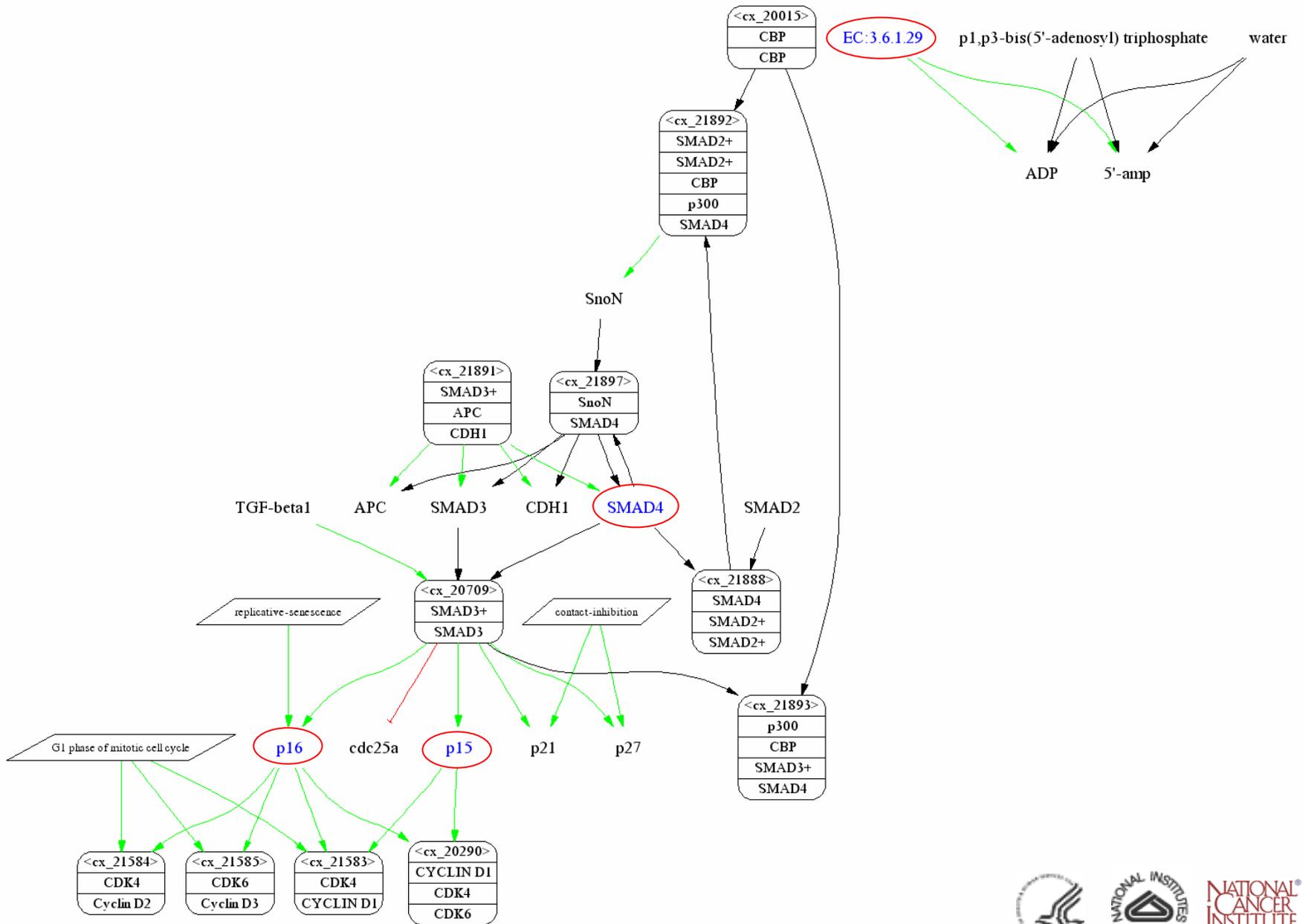
- [NCI Home](#)
- [NCICB Home](#)



The CANCER MOLECULAR ANALYSIS PROJECT







cancer.gov

Cancer Molecular Analysis Project

Molecular Profiles

Molecular Targets

Molecular Targeted Agents

Trials

Current Context

Tissue:

mammary_gland

Histology Type:

All Types

Histology Subtype:

All Subtypes

[Change Context](#)[About CMAP](#)

Quick Links:

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

Pathway information provided by [BioCarta](#)(See [Terms and Conditions](#) of use)For information on sources of Pathway diagrams, see [BioCarta Pathways HowTo](#)Current Context:mammary_gland; all; all [Switch to a Wider Context](#)

A

- Acetylation and Deacetylation of RelA in Nucleus
- Actions of Nitric Oxide in the Heart
- Activation of cAMP-dependent Protein Kinase, PKA
- Activation of CSK Inhibits Signaling through the T Cell Receptor
- Activation of PKC through G-Protein Coupled Receptors
- Acute Myocardial Infarction
- Adhesion and Diapedesis of Granulocytes
- Adhesion and Diapedesis of Lymphocytes
- Adhesion Molecules on Lymphocyte
- ADP-Ribosylation Factor
- Agrin in Postsynaptic Differentiation
- AhR Signal Transduction Pathway
- AKAP95 Role in Mitosis and Chromosome Dynamics
- AKT Signaling Pathway
- ALK in Cardiac Myocytes
- Alternative Complement Pathway
- Angiotensin II Mediated Activation of JNK Pathway via Pyk2 Dependent Signaling
- Angiotensin-converting Enzyme 2 Regulates Heart Function
- Anthrax Toxin Mechanism of Action
- Antigen Dependent B Cell Activation
- Antigen Processing and Presentation
- Antisense Pathway
- Apoptotic DNA Fragmentation and Tissue Homeostasis
- Apoptotic Signaling in Response to DNA Damage
- Aspirin Blocks Signaling Pathway Involved in Platelet Activation
- ATM Signaling Pathway
- Attenuation of GPCR Signaling

Current Context

Tissue:
mammary_gland

Histology Type:
All Types

Histology Subtype:
All Subtypes

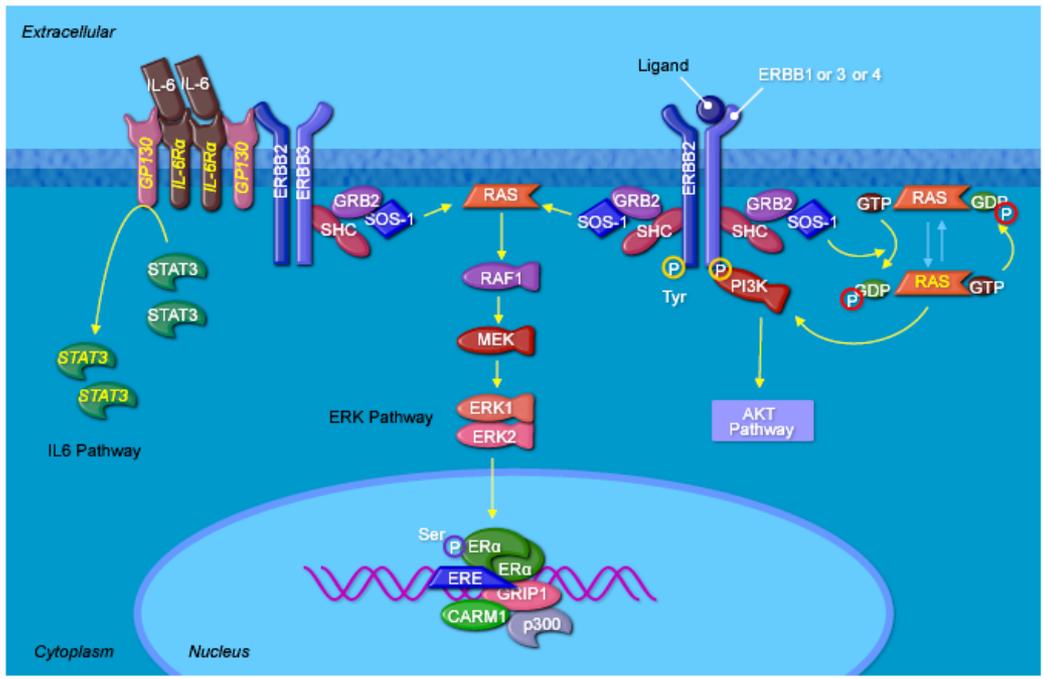
[Change Context](#)

[About CMAP](#)

- Quick Links:
- [NCI Home](#)
 - [NCICB Home](#)



Role of ERBB2 in Signal Transduction and Oncology
 Pathway information provided by [BioCarta](#)
 (See [Terms and Conditions](#) of use)
[Legend](#)



- [All Anomalies](#)
- [Mutated](#)
- [Expression +/-](#)
- [Reset](#)
- [Pathway Summary Report](#)

Current Context
Tissue: **mammary_gland**
Histology Type: **All Types**
Histology Subtype: **All Subtypes**
[Change Context](#)

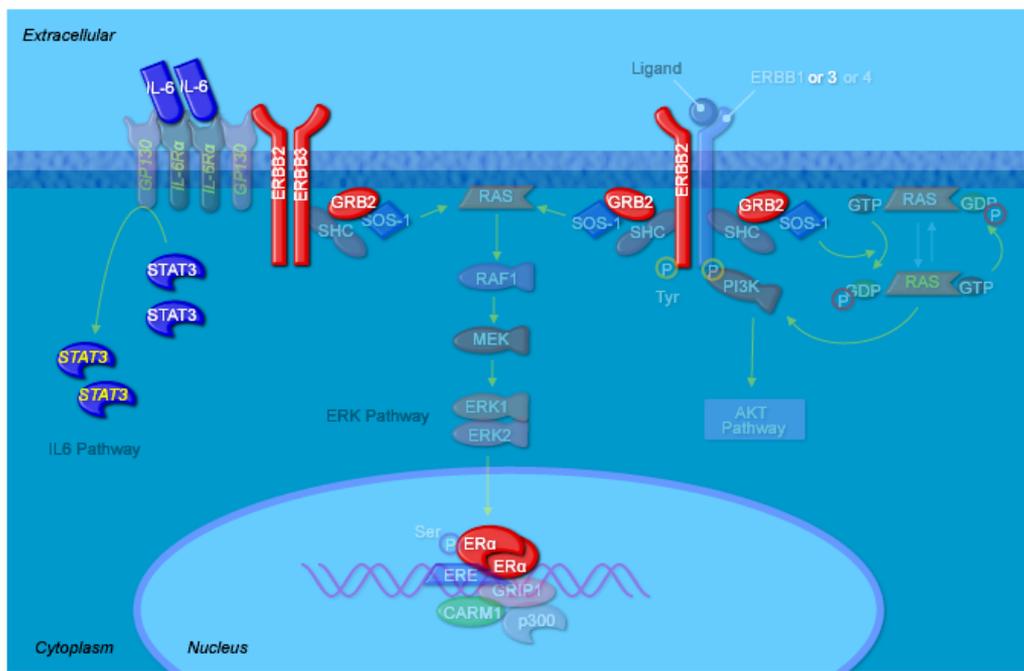
[About CMAP](#)

Quick Links:

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



Role of ERBB2 in Signal Transduction and Oncology
Pathway information provided by [BioCarta](#)
(See [Terms and Conditions](#) of use)
[Legend](#)



- [All Anomalies](#)
- [Mutated](#)
- [Expression +/-](#)
- [Reset](#)

[Pathway Summary Report](#)

CMAP DotBlot - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media Print Mail News RSS Links

Google Search Web Search Site Options

CMAP Report

[Wider Context](#)

Current Context

Tissue:
mammary_gland

Histology Type:
All Types

Histology Subtype:
All Subtypes

[Change Context](#)

About CMAP

Quick Links:

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



Target	SAGE Data			EST Data			Anomaly	Agent
	Normal	Cancer	P	Normal	Cancer	P		
MAPK3			0.20			0.29		
ESR1			0.00			0.20		Anastrozole (Arimidex) ERA-923 Exemestane (Aromasin) Faslodex (Fulvestrant, ICI182780) LY353381 Idoxifene
HRAS			0.26			0.42		
MAP2K1			0.15			0.22		
LRPPRC			0.05			0.02		
RAF1			0.38			0.41		BAY 43-9006 L-779,450 Isoxazole urea ISIS 5132
EP300			0.35			0.44		
ERBB3			0.00			0.45		
MAPK1			0.20			0.05		
CARM1			0.22			0.31		
GRB2			0.00			0.37	Activated	
STAT3			0.00			0.14		
SHC1			0.31			0.11		
ERBB2			0.00			0.00		Trastuzumab (Herceptin(R))
SOS1						0.04		
IL6R			0.36			0.20		
ERBB4						0.32		
EGFR						0.00	Amplification/Mutation	ABX-EGF BIBX 1382 C225/Cetuximab CGP 75166/PK1166 CI-1033/PD183805 E7.6.3 EKB-569 EMD 55 900 GW2016 ICR62 MDX-447 OSI-774 PD 168 393 Thera-CIM-hr3 ZD1839 (IRESSA)
PIK3CG								

Virtual Northern - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Home Search Favorites Media Print Copy Paste Options

Google Search Web Search Site

Current Context

Tissue: **mammary_gland**

Histology Type: **All Types**

Histology Subtype: **All Subtypes**

[Change Context](#)

Virtual Northern

Expression Pattern for cluster **Hs.446352** [Text Legend](#)

Tissue	EST Data		SAGE Data		EST Data			SAGE Data		
	Normal	Cancer	Normal	Cancer	Normal	Cancer	P	Normal	Cancer	P
ALL TISSUES					126 / 2076088	230 / 2025142	0.00	85 / 3799195	705 / 8471612	0.00
adipose					1 / 10228	0 / 767	0.48	--	--	--
adrenal cortex					--	0 / 6427	--	--	--	--
adrenal medulla					--	0 / 321	--	--	--	--
bone					0 / 13882	2 / 43514	0.36	--	--	--
bone marrow					0 / 15343	1 / 21661	0.38	0 / 204563	0 /	--
brain					2 / 216857	6 / 155717	0.09	4 / 658515	97 / 3814713	0.00
cartilage					0 / 4526	4 / 34253	0.37	--	17 / 676933	--
cerebellum					0 / 4413	0 / 0	--	1 / 90885	2 / 252749	0.46
cerebrum					4 / 67147	--	--	--	--	--
cervix					0 / 1029	1 / 40518	0.49	--	--	--
colon					7 / 17813	60 / 149060	0.45	5 / 98089	8 / 635770	0.05
ear					0 / 12587	--	--	--	--	--
embryonic tissue					0 / 711	--	--	--	--	--
endocrine					0 / 7403	0 / 3054	--	--	--	--
esophagus					0 / 84	0 / 2665	--	--	--	--
eye					12 / 69445	0 / 44426	0.01	--	--	--
gastrointestinal tract					0 / 664	5 / 12612	0.42	--	--	--
genitourinary					0 / 1371	5 / 29155	0.43	--	--	--
germ cell					--	2 / 49078	--	--	--	--
head and neck					1 / 44279	11 / 59793	0.02	--	--	--
heart					3 / 54913	--	--	0 / 83063	--	--
kidney					0 / 61078	4 / 73670	0.13	4 / 106467	0 /	--
limb					--	--	--	--	--	--
liver					1 / 59298	4 / 72824	0.23	1 / 144205	0 /	--
lung					11 / 101837	3 / 165354	0.01	3 / 159917	5 / 122803	0.22
lymph node					3 / 81735	0 / 47408	0.24	1 / 99426	--	--
lymphoreticular					0 / 31611	1 / 77136	0.42	--	--	--
mammary gland					0 / 40987	51 / 82938	0.00	5 / 538122	496 / 1662198	0.00
muscle					11 / 71277	3 / 36706	0.20	0 / 107836	--	--
nervous					5 / 10465	1 / 57863	0.01	--	--	--

Quick Links:

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



Done Internet

Gene Viewer.cgi - displays SNPs in mRNA sequences - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS Chat

Google Search Web Search Site Options Address

To view protein sequence alignments, statistics about how SNPs affect the fit of protein domains to motif models, and three-dimensional structures of motifs (if known), click [here](#).

mRNA: NM_004448.1
UniGene: Hs.446352 (build 166) [[CGAP Gene Info](#)] [[CGAP SNP Summary](#)]
Description: v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)
Gene symbol: ERBB2
LocusLink ID: [2064](#)

motifs: ■ Receptor L domain
■ Furin-like cysteine rich region
■ Protein kinase domain

SNP ID	CGAP SNP?	Nucleotide	cds SNP?	Status	Primer Set
ss2421434	no	2110	Ile654Val	-	-
ss1335874	no	2113	Val655Ile	-	-
ss1540630	no	2113	Val655Ile	-	-
ss2421433	no	2113	Val655Ile	-	-
ss1540633	no	3658	Ala1170Pro	-	-

Done Internet

pdb1igr Chain A - Microsoft Internet Explorer

NM_004448.1		LPG id: R7972		
Domain #1				
PF01030	IG1R_HUMAN	52	to	184
pdb1igr	Chain: A	51	to	171

Molecule-View **Backbone-View**

Drag mouse to rotate;
Drag mouse while holding the Shift key to zoom in or out.

Agent Info - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media Print Mail News RSS Feeds

Google Search Web Search Site Options

cancer.gov

CMAP Cancer Molecular Analysis Project

Molecular Profiles Molecular Targets Molecular Targeted Agents Trials

Agent Info

Trastuzumab [Herceptin(R)]

Database Links

[DTP](#) [Cancer Models](#)

CMAP Targets

[ERBB2](#)

Protocols

[Narrower Context](#)

Current Context

Tissue: [mammary_gland](#)

Histology Type: [All Types](#)

Histology Subtype: [All Subtypes](#)

[Change Context](#)

[About CMAP](#)

Quick Links:

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

NATIONAL CANCER INSTITUTE

Document Number	Protocol Title
195	Phase II Trial of Anti-HER-2 Monoclonal Antibody Trastuzumab (Herceptin) in Combination with Low Dose Interleukin-2 (Proleukin) in Metastatic Breast Cancer Patients who have Previously Failed Trastuzumab
5330	A Phase II Evaluation of the Efficacy and Safety of R115777 (NSC 702818), a Non-Peptidomimetic Farnesyl Transferase Inhibitor, and Trastuzumab in Patients with Advanced Breast Cancer
5867	A Phase I Study of Herceptin/Flavopiridol in HER-2 Positive Metastatic Breast Cancer
CALGB-49808	A 2 X 2 X 2 Factorial Randomized Phase III Trial of Multimodality Therapy Comparing 4 Cycles of Doxorubicin and Cyclophosphamide with or without Dexrazoxane (AC +/- Z) Followed by 12 Weeks of Weekly Paclitaxel with or without Trastuzumab (T +/- H) Followed by Local Therapy Followed by 40 Weeks of Weekly Trastuzumab or None in Women with Her-2+ Stage IIIA, IIIB or Regional Stage IV Breast Cancer
CALGB-9840	A Phase III Study of Paclitaxel Via Weekly 1 Hour Infusion Versus Standard 3 Hour Infusion Every 3 Weeks with Herceptin (Trastuzumab) (NSC # 688097) in the Treatment of Patients with/without HER-2/Neu-Overexpressing Breast Cancer
E1100	A Phase III Trial of Herceptin and ZD1839 (Iressa, NSC #715055, IND #61187) in Patients with Metastatic Breast Cancer that Overexpresses HER2/neu (erbB-2)
E2103	A Phase II Trial of Trastuzumab Plus Weekly Ixabepilone (BMS-247550) and Carboplatin in Patients with HER2/Neu-Positive Metastatic Breast Cancer
E2198	Pilot Trial of Paclitaxel-Herceptin Adjuvant Therapy for Early Stage Breast Cancer
E3198	A Safety and Efficacy Study of Doxil and Taxotere + Herceptin in Advanced Breast Cancer
N9831	Phase III Trial of Doxorubicin and Cyclophosphamide (AC) Followed by Weekly Paclitaxel with or without Trastuzumab as Adjuvant Treatment for Women with HER-2 Over-Expressing or Amplified Node Positive or High Risk Node Negative Breast Cancer
NCCTG-98-3257	Phase II Trial of Paclitaxel, Carboplatin and Trastuzumab (Herceptin?) as First-Line Chemotherapy in Patients with Metastatic Breast Cancer which Overexpresses or Amplifies HER 2



caBIG

*cancer Biomedical
Informatics Grid*



caBIG community contributions

- Infrastructure
 - Ontologies
 - Databases
- Applications
 - Clinical trials support
 - Analytic tools
 - Data mining
- Data
 - Trials
 - Experimental outcomes
 - Genomic
 - Microarray
 - Proteomic



News/Events

- News
 - **February 19,20 2004 caBIG Workspace and Working Group Kickoff meeting**
 - **caBIG featured in NCI Cancer Bulletin**
 - **caBIG Interactive Overview**
 - **NEW caBIG Online Forum**
 - **caBIG Program Updates**
 - **caBIG FAQs**
 - **The NCI Director's Corner**
 - **Download caBIG Seminar PowerPoint Presentations**

- Related Links
- [NCICB Site](#)

[About caBIG](#)

To expedite the cancer research communities' access to key bioinformatics platforms the NCI plans to deploy an integrating biomedical informatics infrastructure: the cancer biomedical informatics grid (caBIG). In partnership with the cancer research community, the NCI is creating a common, extensible informatics platform that integrates diverse data types and supports interoperable analytic tools. This platform will allow research groups to tap into the rich collection of emerging cancer research data while supporting their individual investigations.



Overview
caBIG at a GLANCE - Overview of Activities and Accomplishments to Date.



Progress
Selecting Pilot Center Participants.



Pilot Structure
Proposed Pilot Structure and Management Approach.



Workspaces
Areas or virtual environments where caBIG pilot project activities will be grouped.



Inventory
Infrastructure, Applications, and Datasets used to support the caBIG initiative.

Please send comments and suggestions to ncicb@pop.nci.nih.gov • [Privacy Notice](#) • [Accessibility information](#)



acknowledgements

- NCICB
 - Peter Covitz
 - Sue Dubman
 - Carl Schaefer
 - Mervi Heiskanen
 - Denise Hise
 - Kotien Wu
 - Fei Xu
 - Ulli Wagner
 - Frank Hartel
 - Sheri De Coronado
 - Gilberto Fragoso
- LPG/CCR
 - Michael Edmundson
 - Bob Clifford
 - Cu Nguyen



Booz | Allen | Hamilton

<http://ncicb.nci.nih.gov>

<http://cmap.nci.nih.gov>

<http://caBIG.nci.nih.gov>

