National Cancer Institute National Human Genome Research Institute

The Cancer Genome Atlas

IOM Workshop

February 9-10, 2010

Joe Vockley, Ph.D.

Director, TCGA Program Office

Bethesda, MD

Precompetitive Collaboration

- Goals of the TCGA Project
- Describe the large scale approach from tissue accrual to data analysis
- TCGA sample data from a central TCGA repository
- Inter-agency and large network lessons

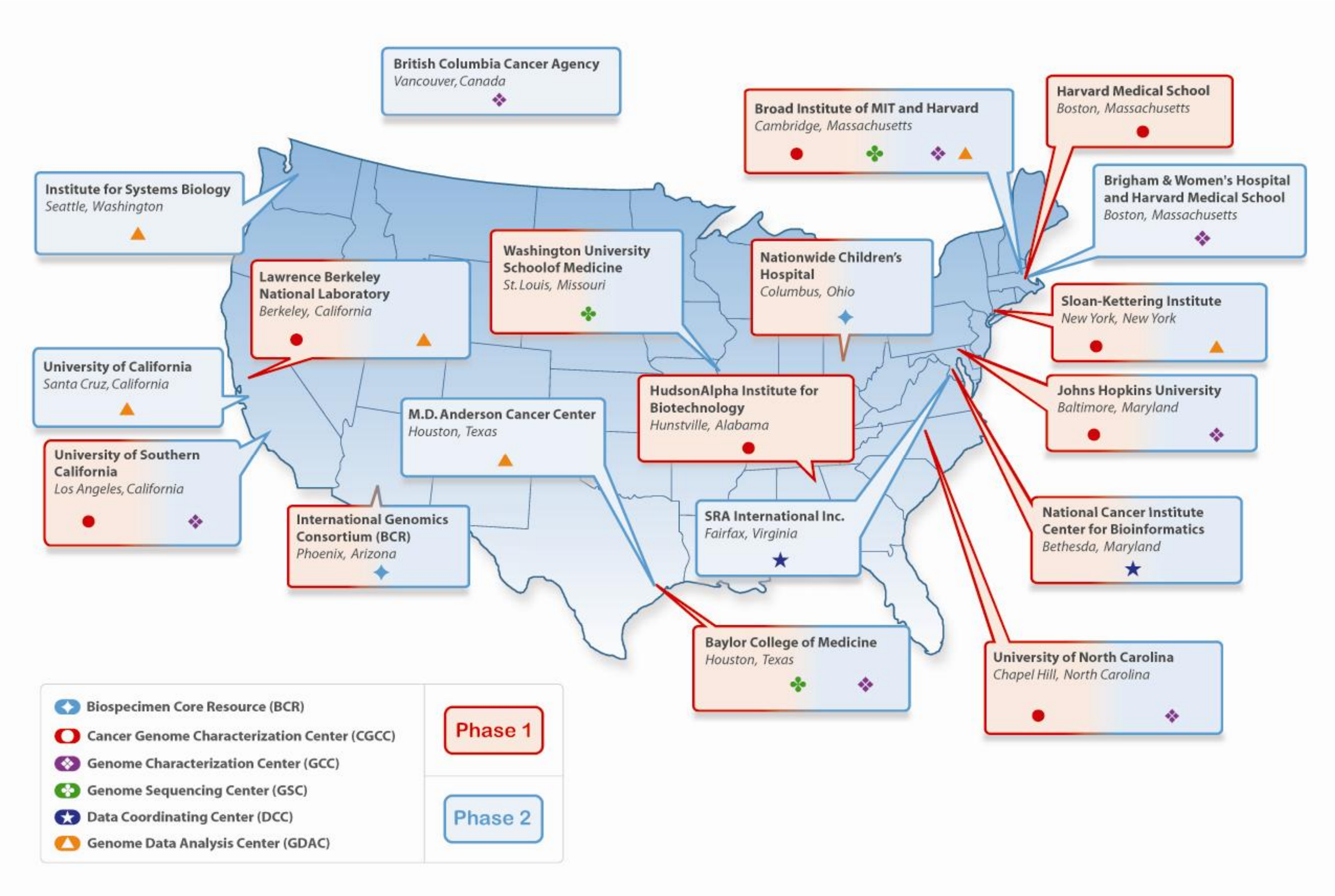
Project Goals

- Generate genome characterization data of 20 tumor types for a statistically significant number of tumors (500 cases/tumor type) over 5 years

- **Genome Characterization Centers (GCCs)** will perform gene expression, copy number, SNP analysis, DNA methylation and miRNA characterization
- **Genome Sequencing Centers (GSCs)** will generate second generation sequence data for mutation detection
- **Genome Data Analysis Centers (GDACs)** will integrate data from GCCs and GSCs. GDACs will create new models and tools to refine and further add value to data for communities
- **Data Coordinating Center (DCC)** will be make data available on a near real-time basis through a web based data portal.

TCGA Components

THE CANCER GENOME ATLAS



British Columbia Cancer Agency
Vancouver, Canada
❖

Institute for Systems Biology
Seattle, Washington
▲

Lawrence Berkeley National Laboratory
Berkeley, California
● ▲

University of California
Santa Cruz, California
▲

University of Southern California
Los Angeles, California
● ❖

International Genomics Consortium (BCR)
Phoenix, Arizona
✦

M.D. Anderson Cancer Center
Houston, Texas
▲

Washington University Schoolof Medicine
St. Louis, Missouri
✤

HudsonAlpha Institute for Biotechnology
Hunstville, Alabama
●

Baylor College of Medicine
Houston, Texas
✤ ❖

Nationwide Children's Hospital
Columbus, Ohio
✦

SRA International Inc.
Fairfax, Virginia
★

Broad Institute of MIT and Harvard
Cambridge, Massachusetts
● ✤ ❖ ▲

Harvard Medical School
Boston, Massachusetts
●

Brigham & Women's Hospital and Harvard Medical School
Boston, Massachusetts
❖

Sloan-Kettering Institute
New York, New York
● ▲

Johns Hopkins University
Baltimore, Maryland
● ❖

National Cancer Institute Center for Bioinformatics
Bethesda, Maryland
★

University of North Carolina
Chapel Hill, North Carolina
● ❖

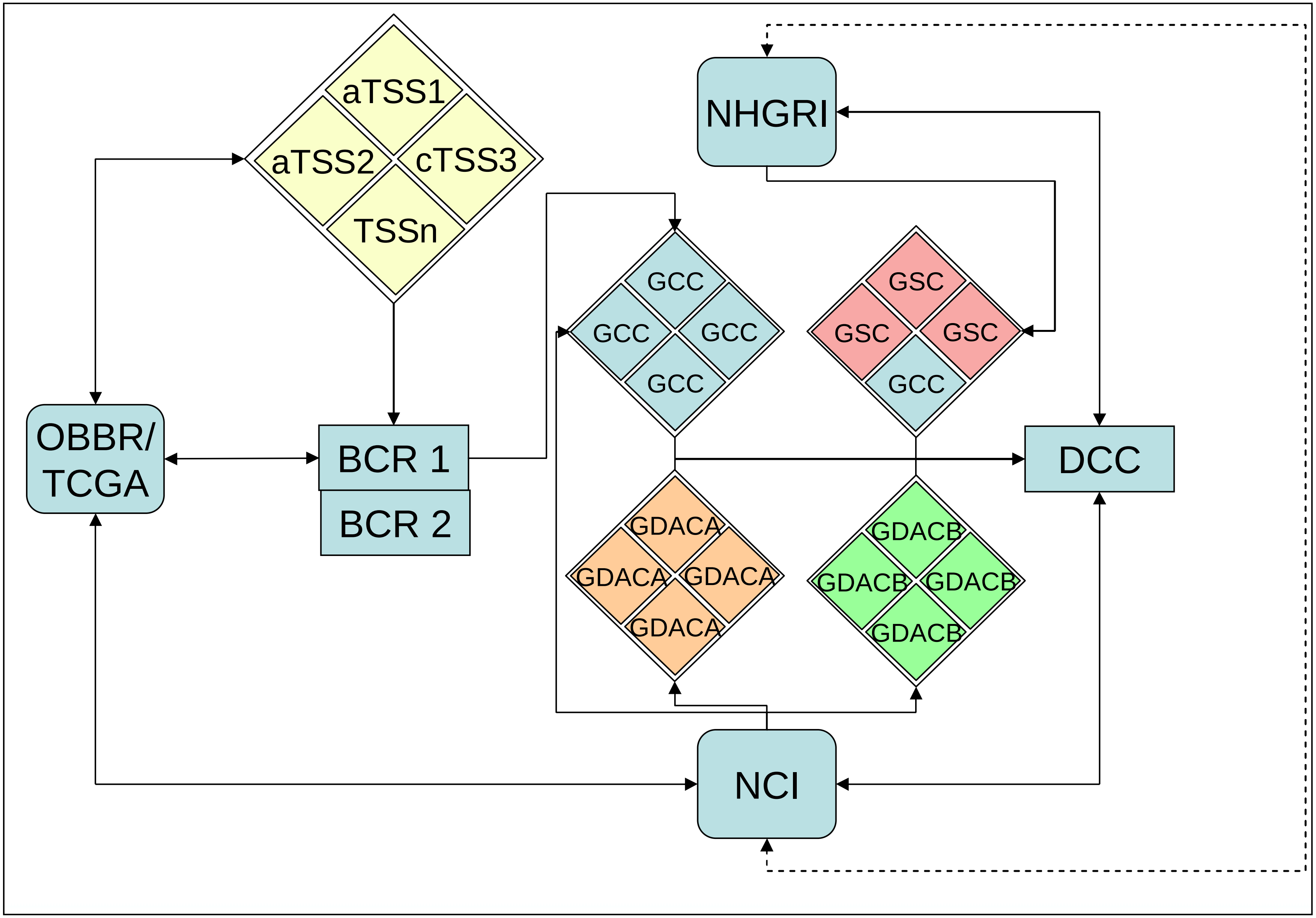
- ✦ Biospecimen Core Resource (BCR)
- ● Cancer Genome Characterization Center (CGCC)
- ❖ Genome Characterization Center (GCC)
- ✤ Genome Sequencing Center (GSC)
- ★ Data Coordinating Center (DCC)
- ▲ Genome Data Analysis Center (GDAC)

Phase 1

Phase 2

TCGA Phase II Network

THE CANCER GENOME ATLAS



Genome Characterization Centers

- **Characterization performed on 500 cases per tumor type**
- **Gene expression**: RNA-seq
- **miRNA**: 2nd generation sequencing
- **SNP and copy number variation:** Affymetrix 6.0 SNP array
- **Epigenomics**: Illumina Infinium DNA methylation27
- BeadArray

Genome Sequencing Centers

- **Discovery phase sequencing**
  - 200 cases sequenced for each of 20 different tumor types
  - 10% whole genomes, 90% whole exomes

- **Validation phase sequencing**
  - 300 cases sequenced for regions of interest identified during discovery phase

Genome Data Analysis Centers

- **Data standardization and integration**
  - Standard data formats and fields
  - Methods to integrate data within and across platforms

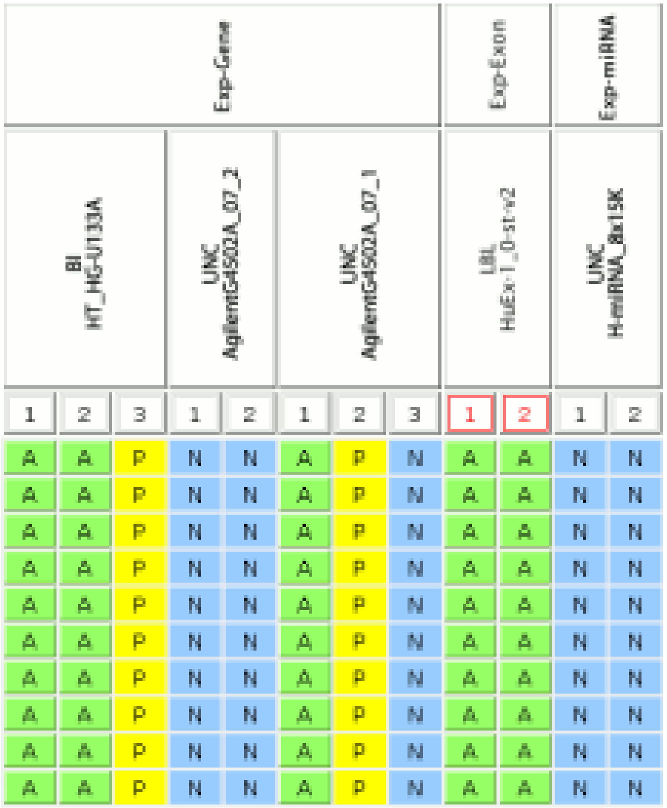
- **Support data analysis needs of the project**
  - ARRA phase represents a challenge with a deadline
  - Post-ARRA phase (last 3 years of project)

- **Data analysis pipeline**
  - Multi-faceted pipeline for target discovery
  - Centralized early access tool with downstream expansion

- **Visualization tools**
  - Web-based visualization tools for TCGA data
  - Integration of non-o

TCGA Data Portal

The Cancer Genome Atlas



Exp-Gene								Exp-Exon		Exp-miRNA	
BI HT_HG-U133A			UNC AgilentG4502A_07_2		UNC AgilentG4502A_07_1			LBL HuEx-1_0-st-v2		UNC H-miRNA_8x15K	
1	2	3	1	2	1	2	3	1	2	1	2
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N
A	A	P	N	N	A	P	N	A	A	N	N

- Browse and download TCGA data
- Data analysis software tools
- Access to controlled data
- No Intra-network data access

http://cancergenome.nih.gov/dataportal/data/about/

Potential Tumors for TCGA

Tissue	Tumor Type	Tissue	Tumor Type
Lung	Adenocarcinoma	Thyroid	Follicular and Papillarycarcinoma
Breast	Invasive ductal carcinoma	Soft Tissue	Sarcoma
Breast	Invasive lobular carcinoma	Prostate	Ductal adenocarcinoma
Kidney	Clear cell renal cell carcinoma	Bladder	Urothelial carcinoma, invasive, nonpapillary
Kidney	Papillary renal cell carcinoma	Bladder	Urothelial carcinoma, invasive, papillary
Uterus	Endometrial adenocarcenoma	Pancreas	Ductal adenocarcinoma
Lymphoma	Non Hodgkin's Lymphoma	Lymphoma	Diffuse large B-cell lymphoma
Cervix	Cervical Cancer	Leukemia	Acute myeloid leukemia
Colon	Adenocarcinoma	Leukemia	Chronic lymphocytic leukemia
Stomach	Adenocarcinoma (intestinal or diffuse)	Oral cavity	Squamous cell carcinoma
Liver	Hepatocellular carcinoma	Skin	Cutaneous melanoma w/ lymph node metastasis?
Brain	Astrocytoma, Mid-grade	Myeloma	Multiple Myeloma

OVARIAN

Coverage(T/N)	Callable	Purity	Ploidy
31x / 30x	**81%**	**90%**	**2.8**

Name	**TCGA-13-0751**
Alias	**OV-0751**
Issued By	**Broad Institute**
Issue Date	**July 8, 2009**

Point Mutations

Rate/Mb	**0.75**
Total	**1786**
Coding	**9**

MIS 5
STOP 1
INDEL ---

HIGHLIGHTS

GENE	MUTATION	FUNCTION
TP53	Insertion	Tumor suppressor
EXOC6B	Missense	protein transport, exocytosis
ANKRD6	Missense	ankyrin
AHNAK	Missense	CNS development
C11orf52	Nonsense	?
GABRB3	Missense	GABA receptor

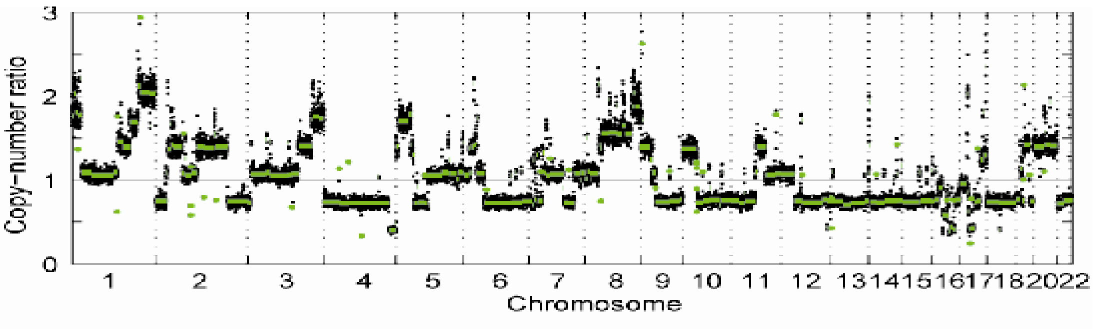
Lost BRCA1 germline indel

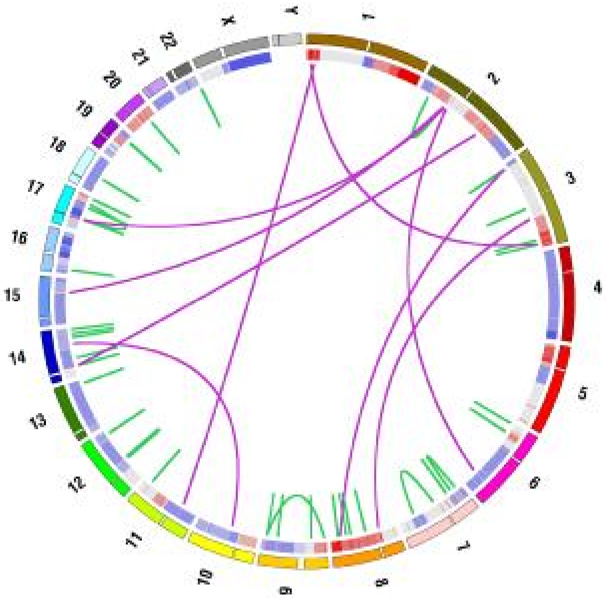
Chr. Aberrations

CNA Breaks	**---**
TX-Inter	**9**
TX-Intra	**15**

HIGHLIGHTS

NF1-EFCAB5 fusion gene probably inactivating validated by RNA-seq

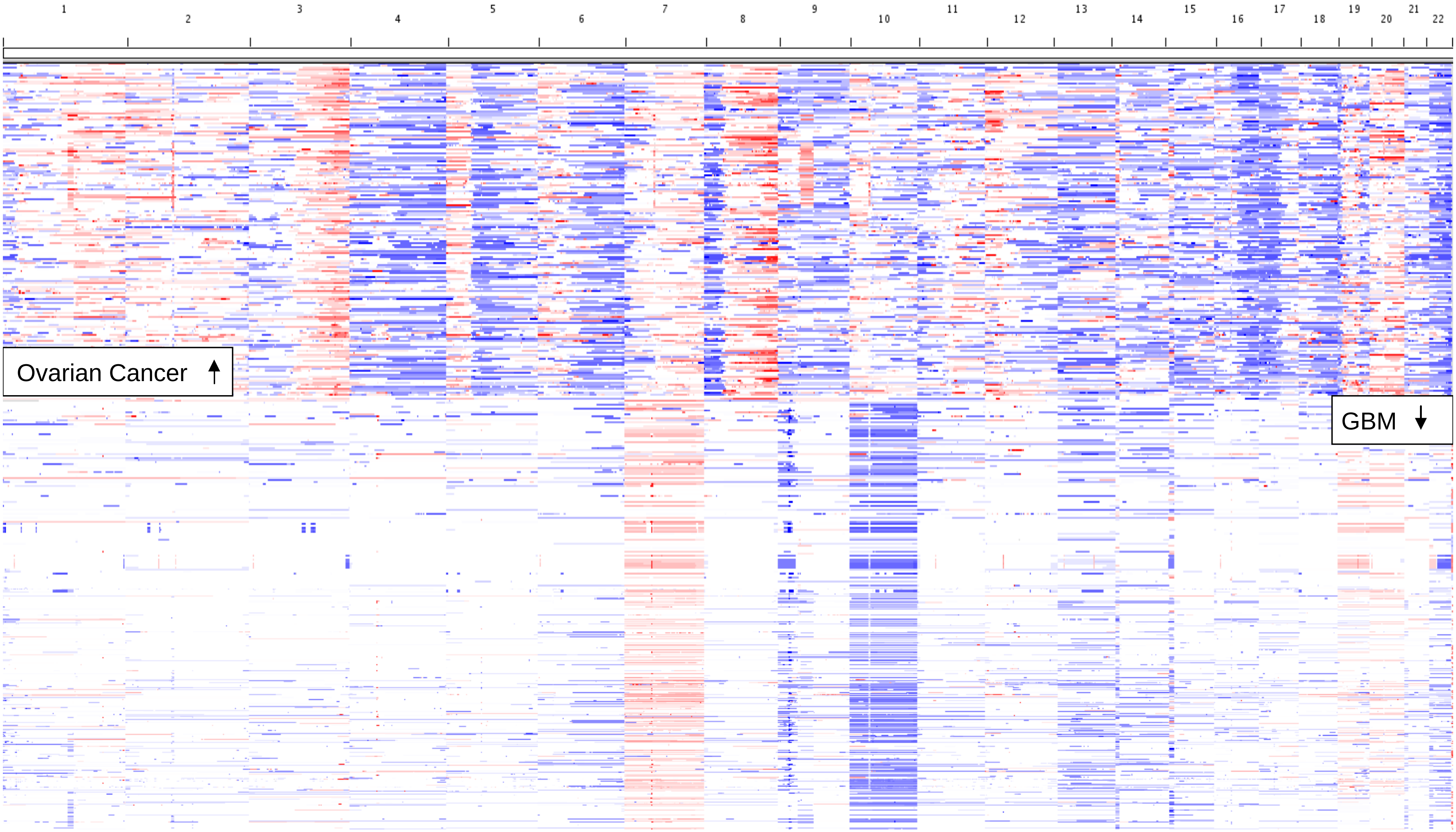




Courtesy of Gaddy Getz – Unpublished, Analysis in Process

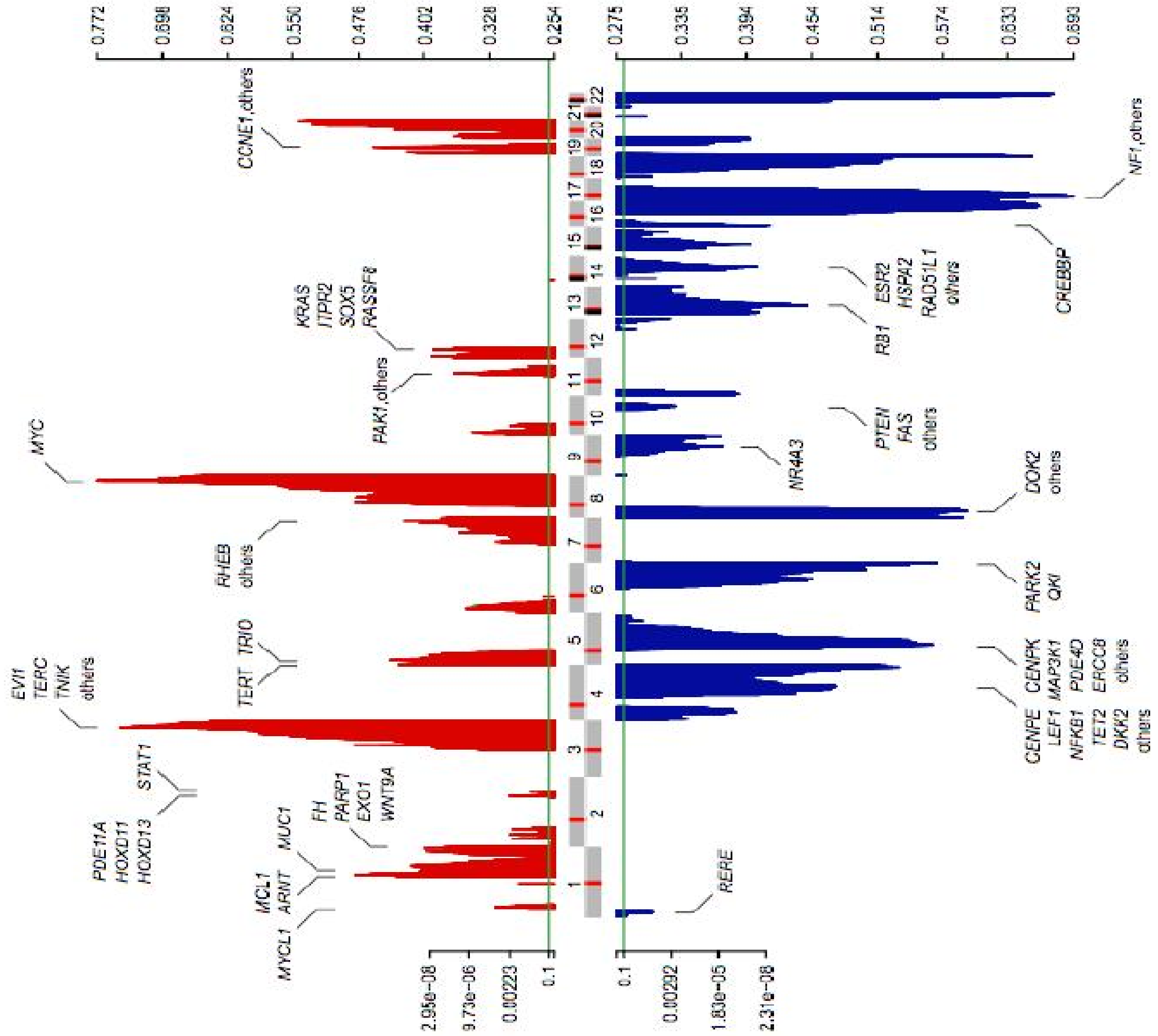
CNA in GBM and Ovarian Cancers

THE CANCER GENOME ATLAS



Courtesy of The Broad Institute and Memorial Sloan Kettering Cancer Center

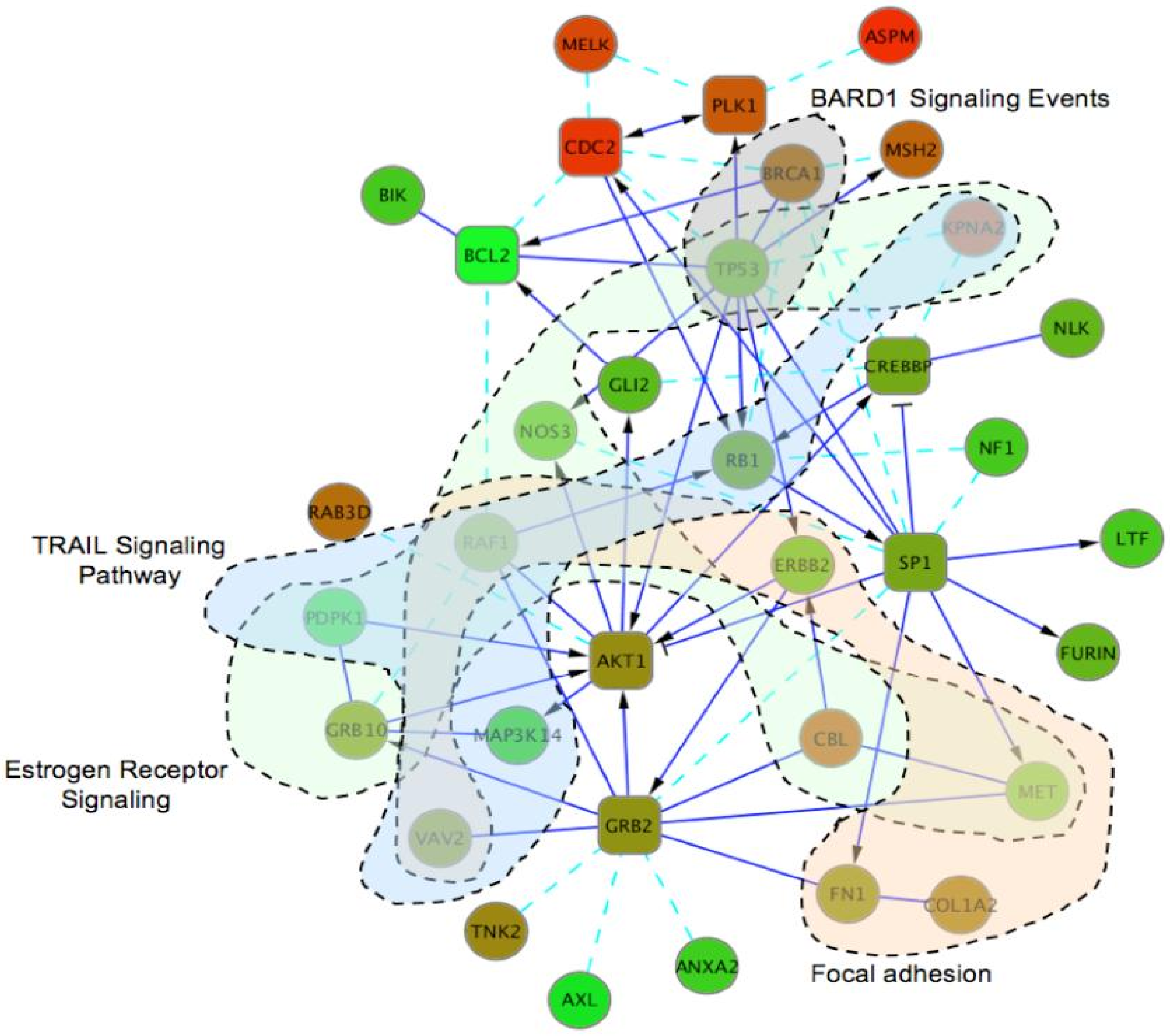
Ovarian Cancer
Chromosomal Amplifications and Deletions



Data and figure generated by TCGA Network

Ovarian Cancer
Pathways and Altered Expression

THE CANCER GENOME ATLAS



Data and figure generated by TCGA Network

TCGA as an inter-agency project

- NCI and NHGRI co-fund and co-manage TCGA
- Advantages
  - Learn new ways to accomplish goals
  - More diverse set of expertise in management staff
  - Exposure to new technologies
- Challenges
  - Different agency culture
  - Different administrative policies
  - Finding common ground takes time
  - Perception of “inter-agency disagreements” by funded community

Collaboration in a Large Network

- Focuses the expertise of many researchers with different resources on a single goal
  - Accelerates discovery
  - Facilitates new collaborations

- Challenges in a large network
  - The larger the decision making group, the harder it is to make a consensus decision
  - Make sure everyone is on the same page
    - Different understanding of goals results in different outcomes
    - Especially if it is your staff that is not on the same page
  - Too many meetings
    - Illusion of control by the project team
    - People stop participating in a meaningful way

The Right Pre-Project Plan is Important

- Start planning early
  - The best thought out plan always misses something
- Example: TCGA tissue accrual
  - Start collecting samples now or two years ago
  - Sequential vs simultaneous accrual of tissues
  - Know impact of sample criteria on project timeline
  - Don’t make the criteria so rigid that the goals can’t be achieved
  - Be flexible as long as it doesn’t destroy the science

Advantage: Extremely high quality data

Challenge: Too few samples to do the project

Impact of TCGA

- Promote novel findings of cancer genomics and biology
- Accelerate discoveries of diagnostic and therapeutic targets
- Stimulate and potentiate future funding opportunities
- Enable personalized medicine