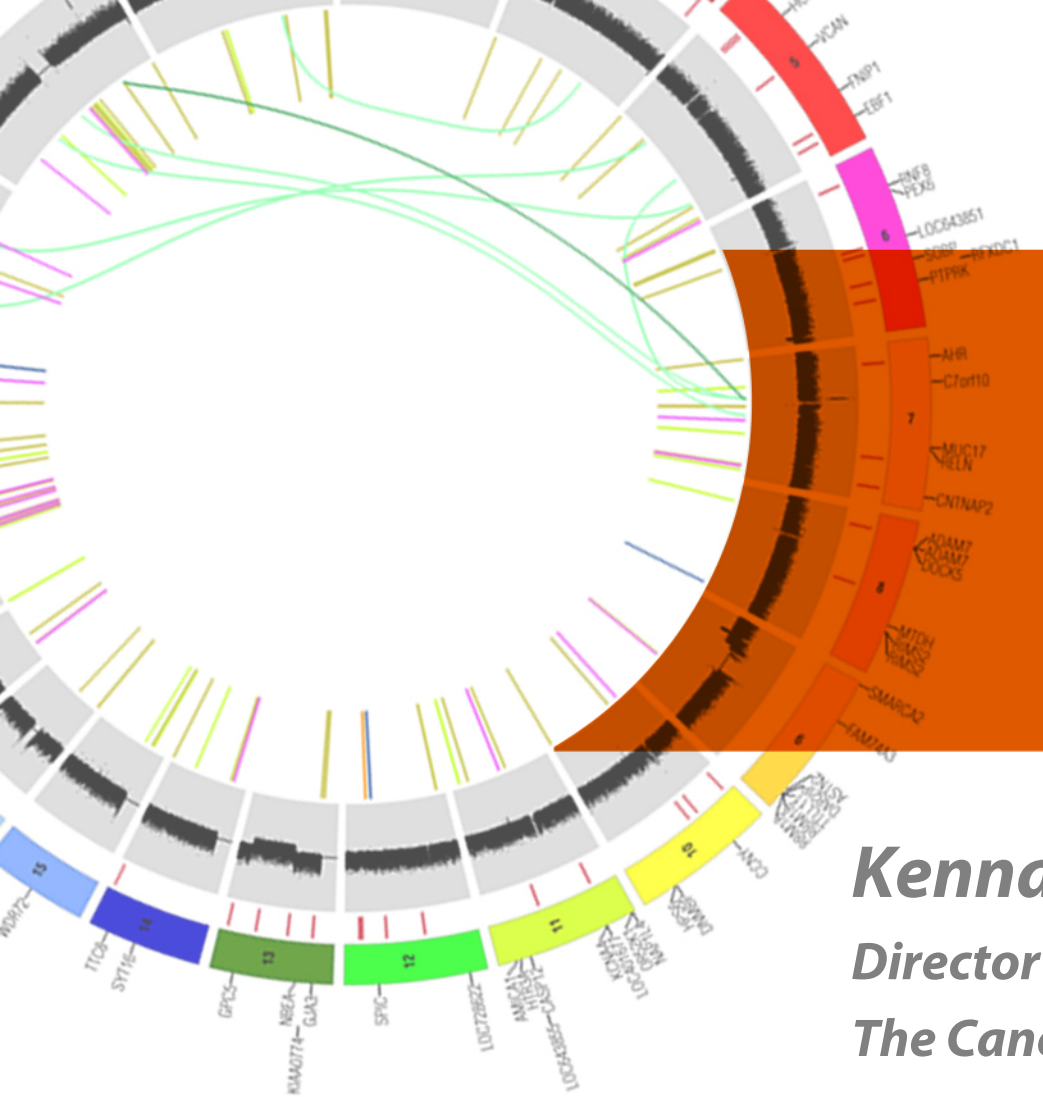


The Cancer Genome Atlas

# TCGA: A Community Resource Looking for a Broader Community


***Kenna Shaw, Ph.D.***

***Director***

***The Cancer Genome Atlas***

# TCGA: Core Objectives

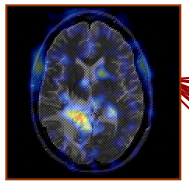
**Launched in 2006 as a pilot and expanded in 2009, the goals of TCGA are to:**

- Establish infrastructure for effective team science
- Develop a scalable "pipeline" beginning with highest quality samples
- Determine the feasibility of a large-scale, high throughput approach to identifying the molecular 'parts-list'
- Evaluate using statistically-robust sample sets
- Make the data publicly and broadly available to the cancer community while protecting patient privacy

# TCGA: "No Platform Left Behind"

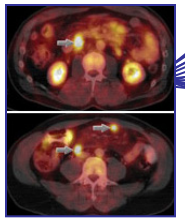
## 25* forms of cancer

glioblastoma multiforme (brain)



squamous carcinoma (lung)

serous cystadenocarcinoma (ovarian)



Etc. Etc. Etc.

Biospecimen Core Resource with more than 150 Tissue Source Sites

6 Cancer Genomic Characterization Centers

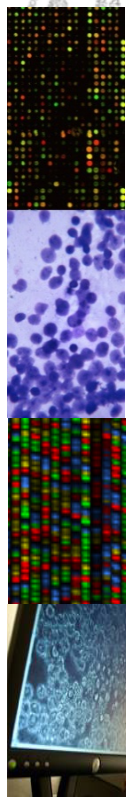
3 Genome Sequencing Centers

7 Genome Data Analysis Centers

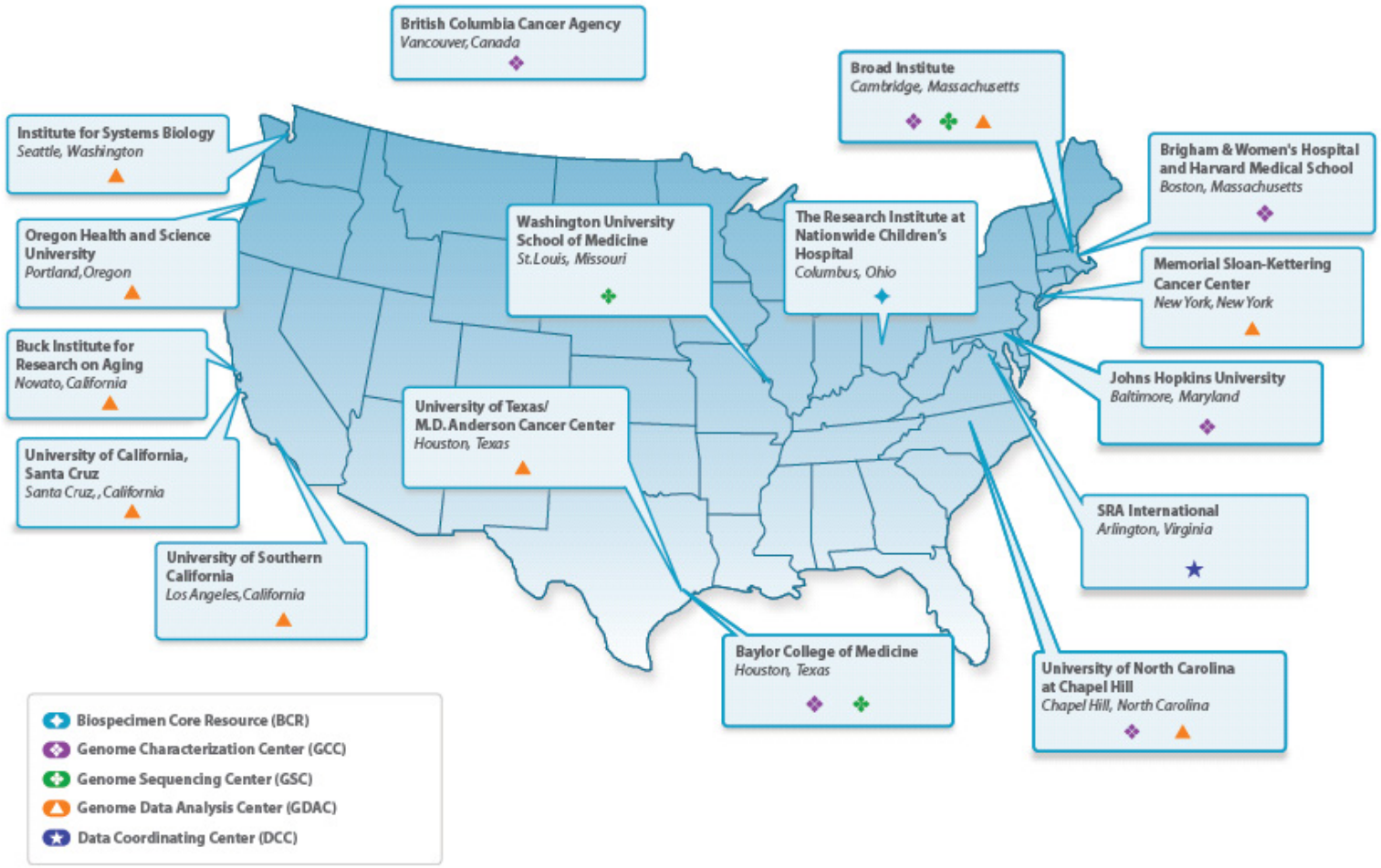
Data Coordinating Center

## Multiple data types

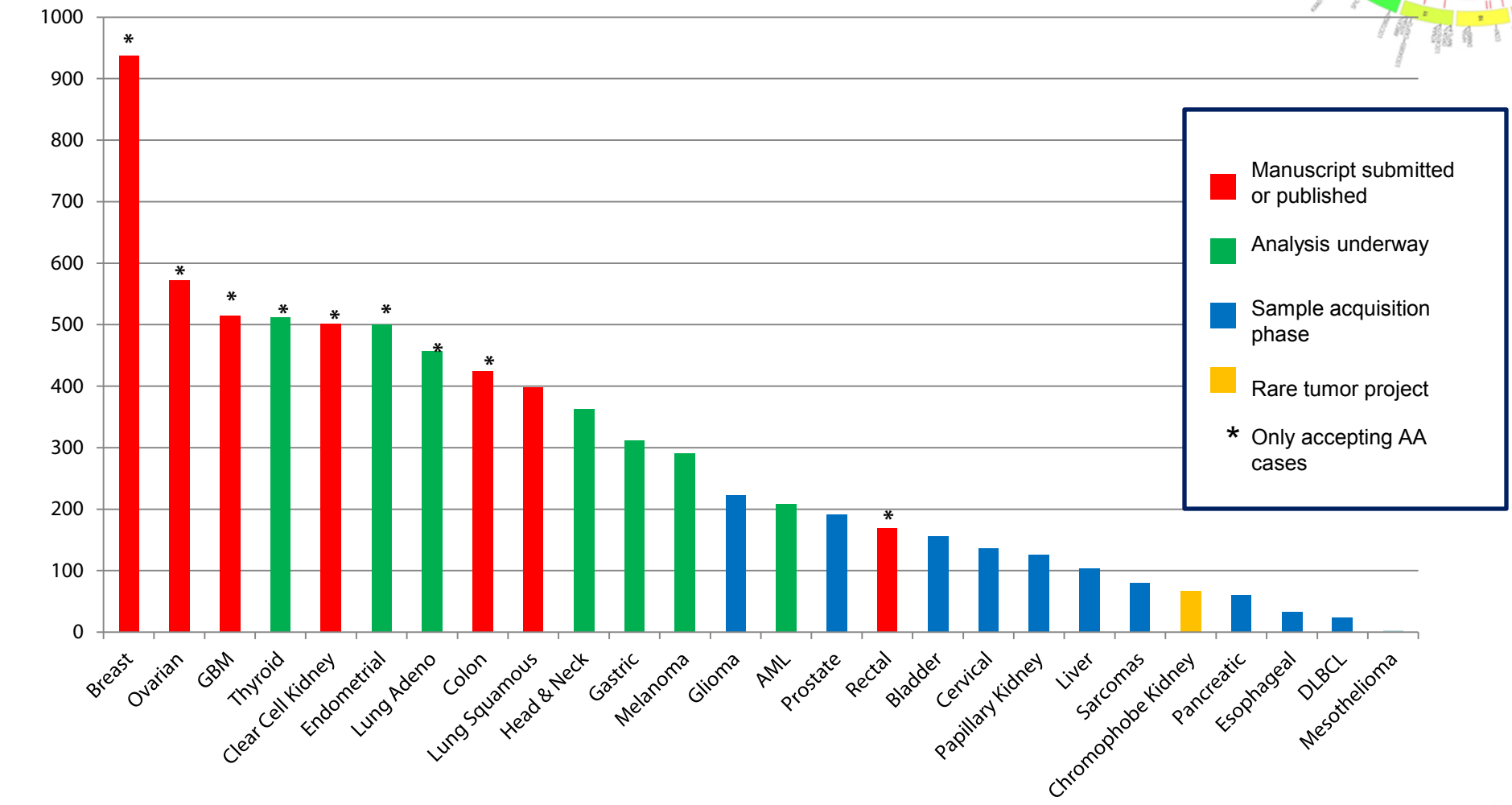
- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic report/images
- Tissue anatomic site
- Surgical history
- Gene expression/RNA sequence
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence
- RPPA (protein)
- Subset for Mass Spec



# TCGA Research Network

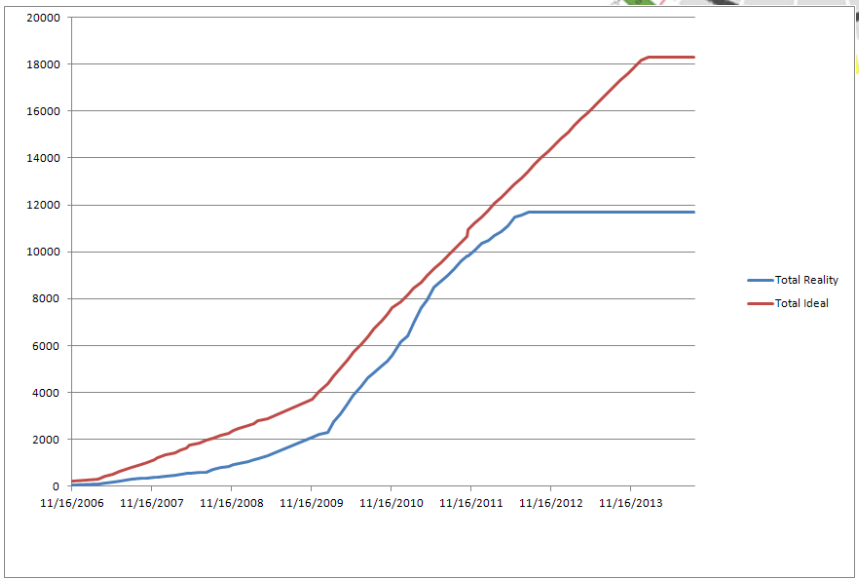


# Tumor Project Progress



# **Rare Tumor Project** *(Initiated March 2012)*

- Adrenocortical Carcinoma
- Adult ALL (B-cell and T-Cell)
- Anaplastic Thyroid
- Cholangiocarcinoma
- Chromophobe kidney
- High Risk MDS (del 5q- cases)
- Mesothelioma
- Paraganglioma/Pheochromocytoma
- Testicular Germ Cell
- Thymoma
- Uterine Carcinosarcoma
- Sarcomas
- Others??



# TCGA: Platforms- Then and Now

| Platform | Pilot | Expansion |
|---|---|---|
| SNP/CNV | Affy SNP 6.0<br>Agilent CGH Array<br>Illumina 1M Duo | Affy SNP 6.0<br>Low Pass Sequencing* |
| Methylation | Infinium Array | Infinium Array |
| mRNA | Agilent 244K Array<br>Affy Human Exon Array<br>Affy U133 Array | RNAseq |
| miRNA | Agilent 8 x 15K Array | RNAseq |
| Mutation | 600-1000 genes | DNAseq:<br>100% whole exomes<br>10% whole genomes |
| Proteomics | None | Reverse Phase Protein Arrays |
| Clinical Data | Minimum Enrollment & Follow-Up*<br>H&E from Frozen Section Images<br>Treatment Data | Minimum Enrollment & Follow-Up*<br>Pathology Reports<br>H&E from Frozen & Diagnostic Images |

*- Not a core platform; Not all samples will have data file for this platform

More information on platforms and data available at: http:/tcga-data.nci.nih.gov/tcga/tcgaPlatformDesign.jsp

# TCGA: Data Availability

- 7,136 cases across 20+ tumor types

- 5865 with minimum clinical data set
- 3893 with at least 1 year follow-up; ~50% with treatment data

- 105,000 samples of RNA/DNA/protein shipped between 2006 & 6/2012

- All but 13,000 samples have data returned:
  - ~87% of all samples have Level 1 data publicly available
  - TCGA Program Office to complete accounting with Batelle/QMS by end of 2012

# TCGA: The Pipeline for Comprehensive Characterization

Tissue Sample

Pathology QC

DNA & RNA Isolation, QC

Sequencing

Expression, CNA & LOH, Epigenetics

Data Storage at DCC & CGHub

GDAC

Integrative Analysis

Comprehensive Characterization of a Cancer Genome

*3 months – 2 years*

*~90d*

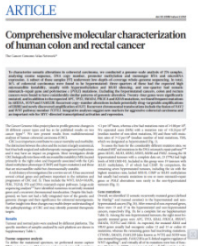
SNP 6.0 ~45d

Methylation ~60d

miRNAseq ~105d

mRNAseq ~120d

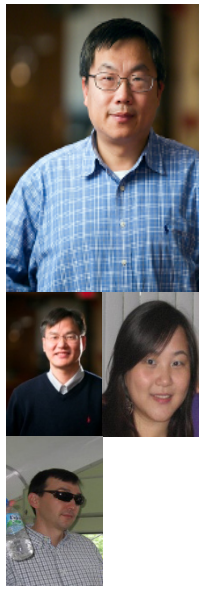
DNAseq Exome ~180d


ARTICLE

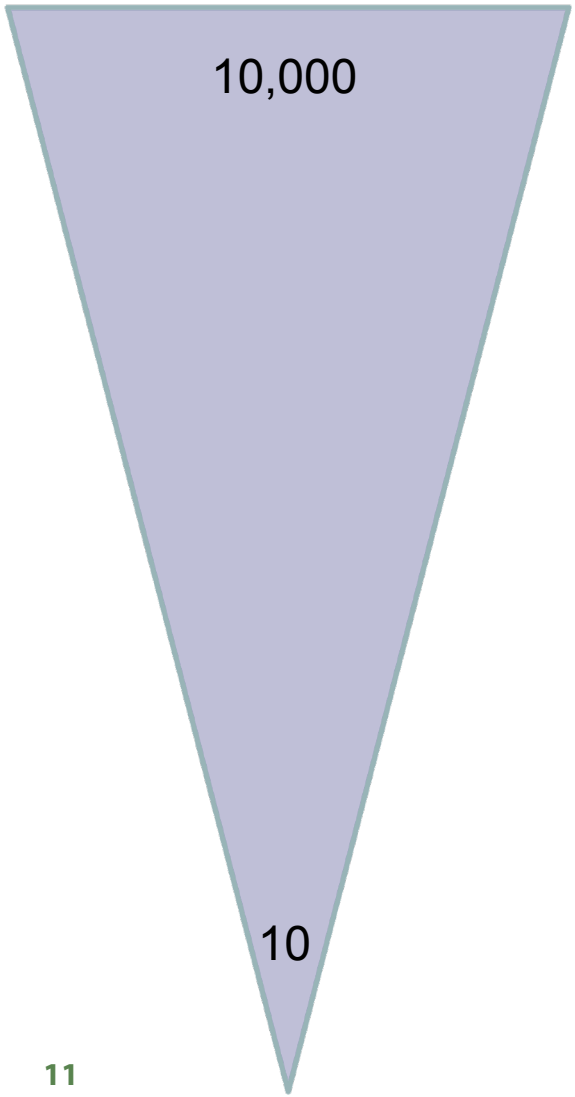
Comprehensive molecular characterization of human colon and rectal cancer

*~12-24 months*

# TCGA: The Pipeline for Comprehensive Characterization

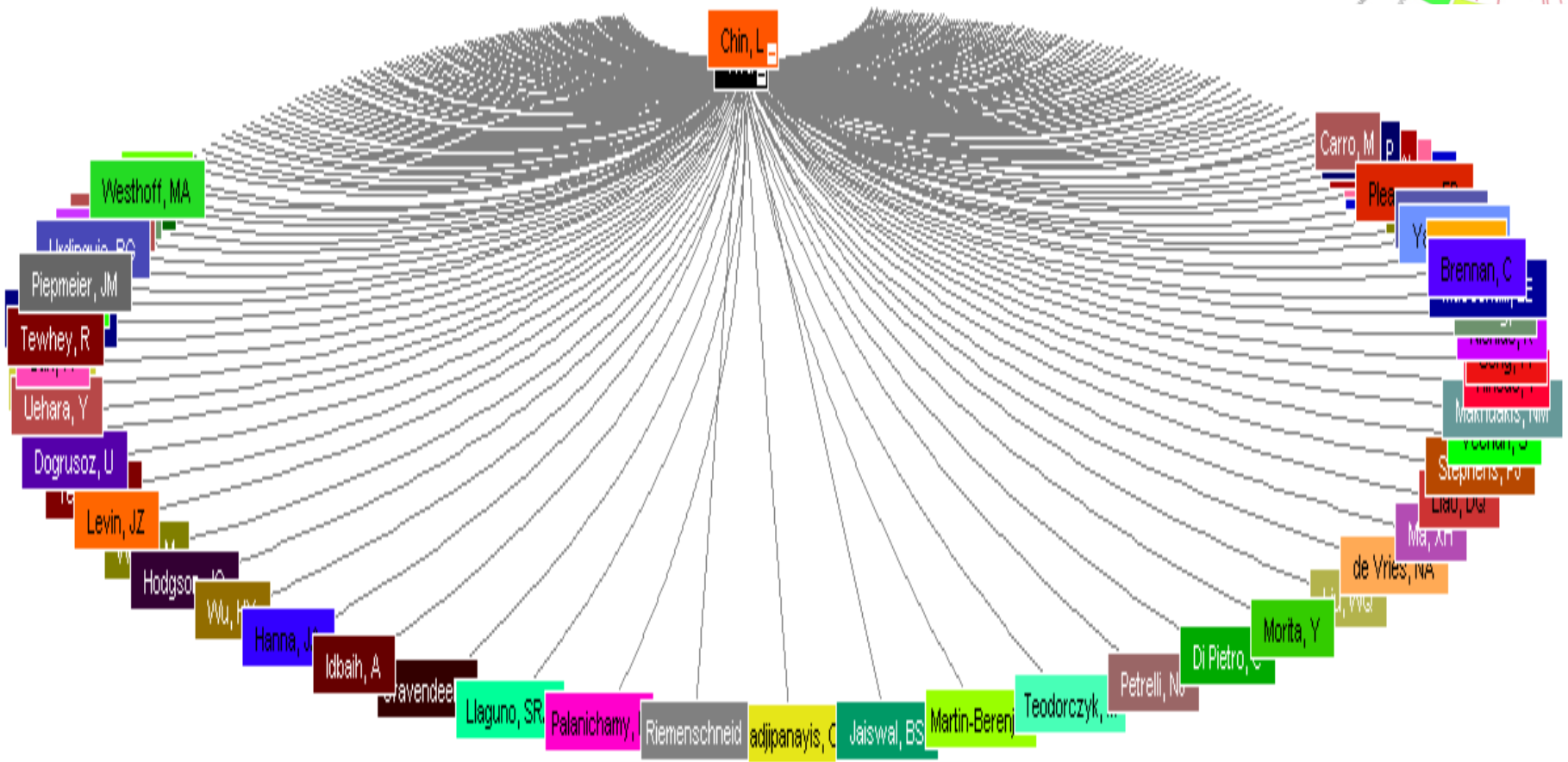



# Sample Criteria Limit 'Askable' Questions




- Primary, *adult* tumors (except for melanoma and triplets)
- Malignant (no *in situ* cases)
- Snap frozen, <60min from clamp to LN2
- ~ 50 mg (biopsies starting to be feasible)
- Pathology review of tissue sent to TCGA
- No more than 20% necrosis ; ≥ 60%* tumor cells
- No prior treatment
- *Matched* source of germline: Blood (buffy coat/white cells)/saliva or skin for liquid tumors
- Clinical annotation; but not pre-analytic variables
- IRB approval for use in TCGA; proactive consenting for genomic studies
- MTA w/out retention of IP

# End Goal: Making an Exhaustible Resource Inexhaustible

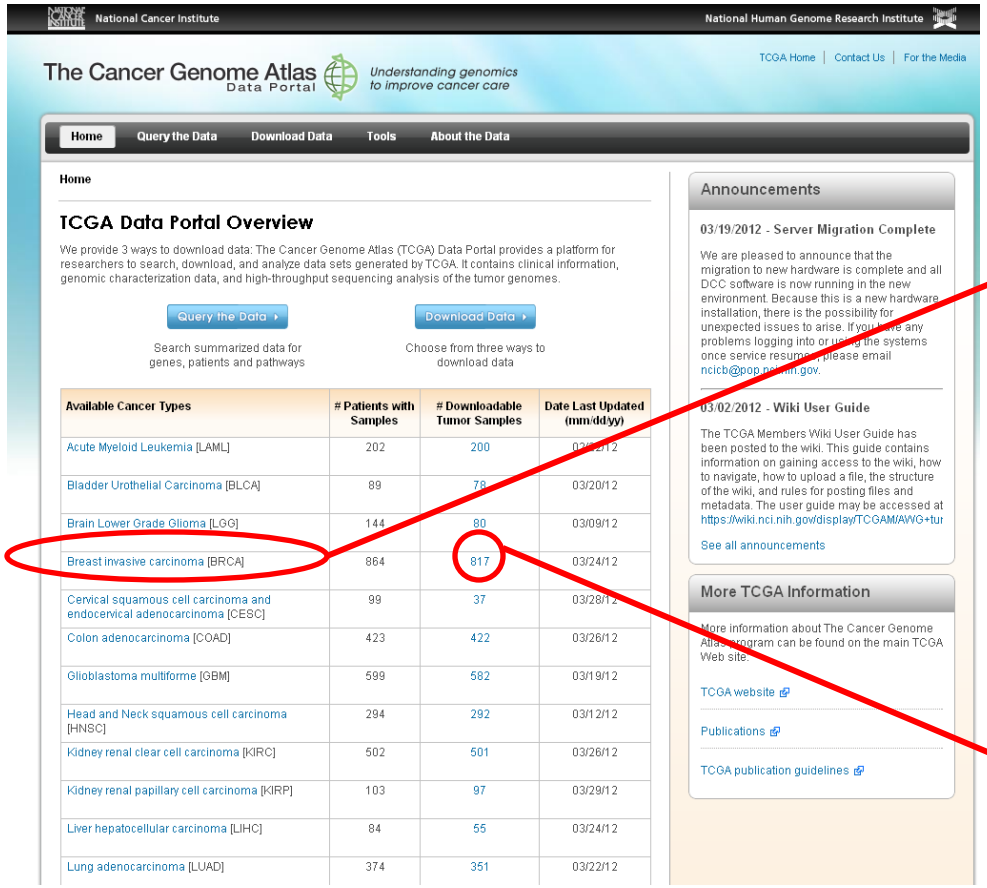


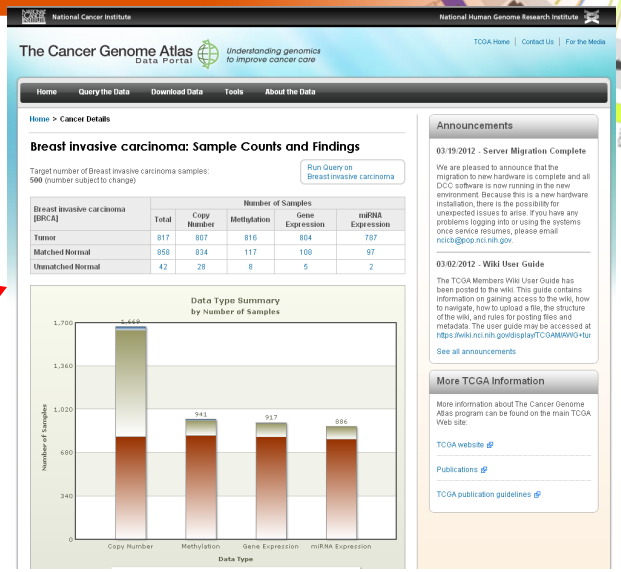

Source: ISI Web of Knowledge™, , www.thomsonscientific.com

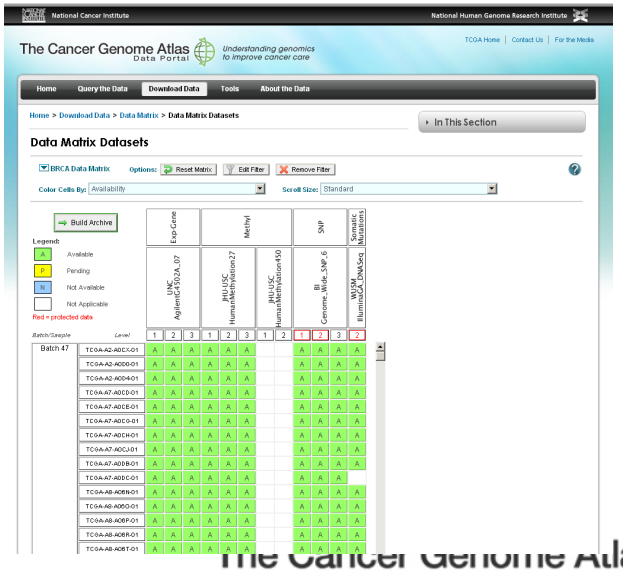


The Cancer Genome Atlas

# TCGA Data Portal

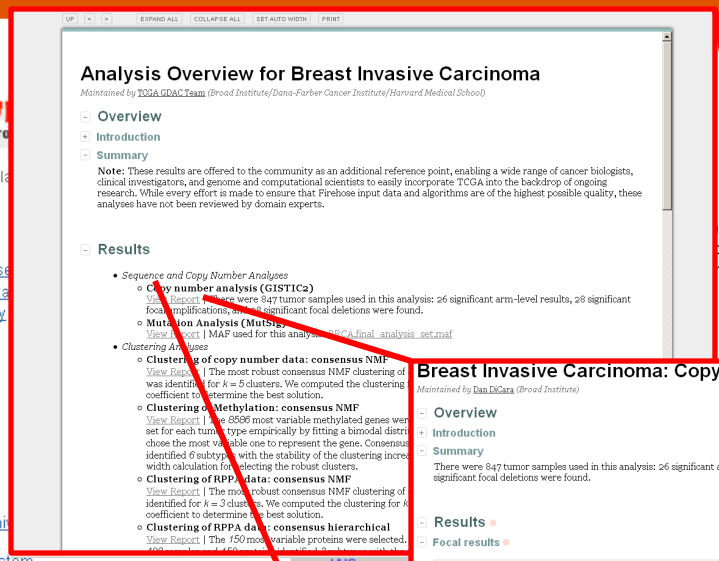



https://tcga-data.nci.nih.gov/tcga/

# Ongoing Pipeline Analysis Available





Tools

8 Added by Aaron Ball, la

- AWG Reps
- Contact Us
- Dashboard-Analyse
- Dashboard-Stddata
- Data Usage Policy
- DCC Interactions
- Download
- FAQ
- Nozzle
- Pipeline Docs
- Presentations
- ProcessFlow
- QualityControl

Email Archi

Tracking System

ter (GDAC). On behalf of The Cancer Genome Atlas (TCGA), we've designed and
c datasets through scores of quantitative algorithms, in the hope of accelerating the
nthly runs, or this presentation for more background information. Note that

2012_08_25 analyses Run

| | Successful | Download |
|---|---|---|
| LIHC | 100% | Open Protected |
| LUAD | 100% | Open Protected |
| LUSC | 100% | Open Protected |
| OV | 100% | Open Protected |
| PAAD | 100% | Open Protected |
| PRAD | 100% | Open Protected |
| SARC | 100% | Open Protected |
| SKCM | 100% | Open Protected |
| STAD | 100% | Open Protected |
| TGCA | 100% | Open Protected |
| UCEC | 100% | Open Protected |
| PANCANCER | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 100% | Open Protected |
| | 96% | Open Protected |
| | 94% | Open Protected |
| | 88% | Open Protected |
| | 42% | Open Protected |

get.

Labels: None

13 Child Pages

- Contact Us
- DCC Interactions
- Nozzle
- Pipeline Docs
- Presentations

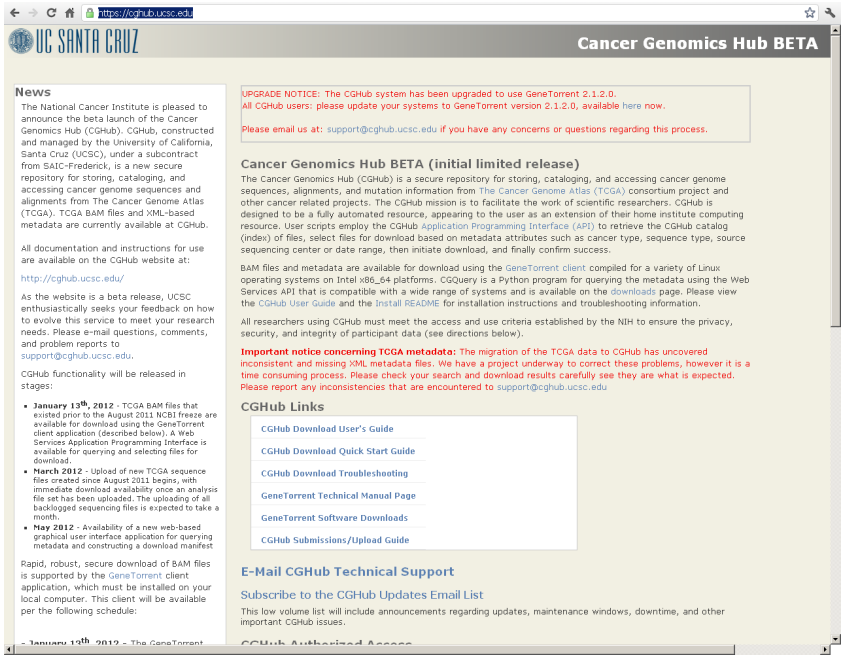
**http://gdac.broadinstitute.org**



The Cancer Genome Atlas

# Where to find TCGA Sequence Data

- Moved from Short Read Archive (SRA at NCBI) to UCSC
- Open for downloads as of January 2012

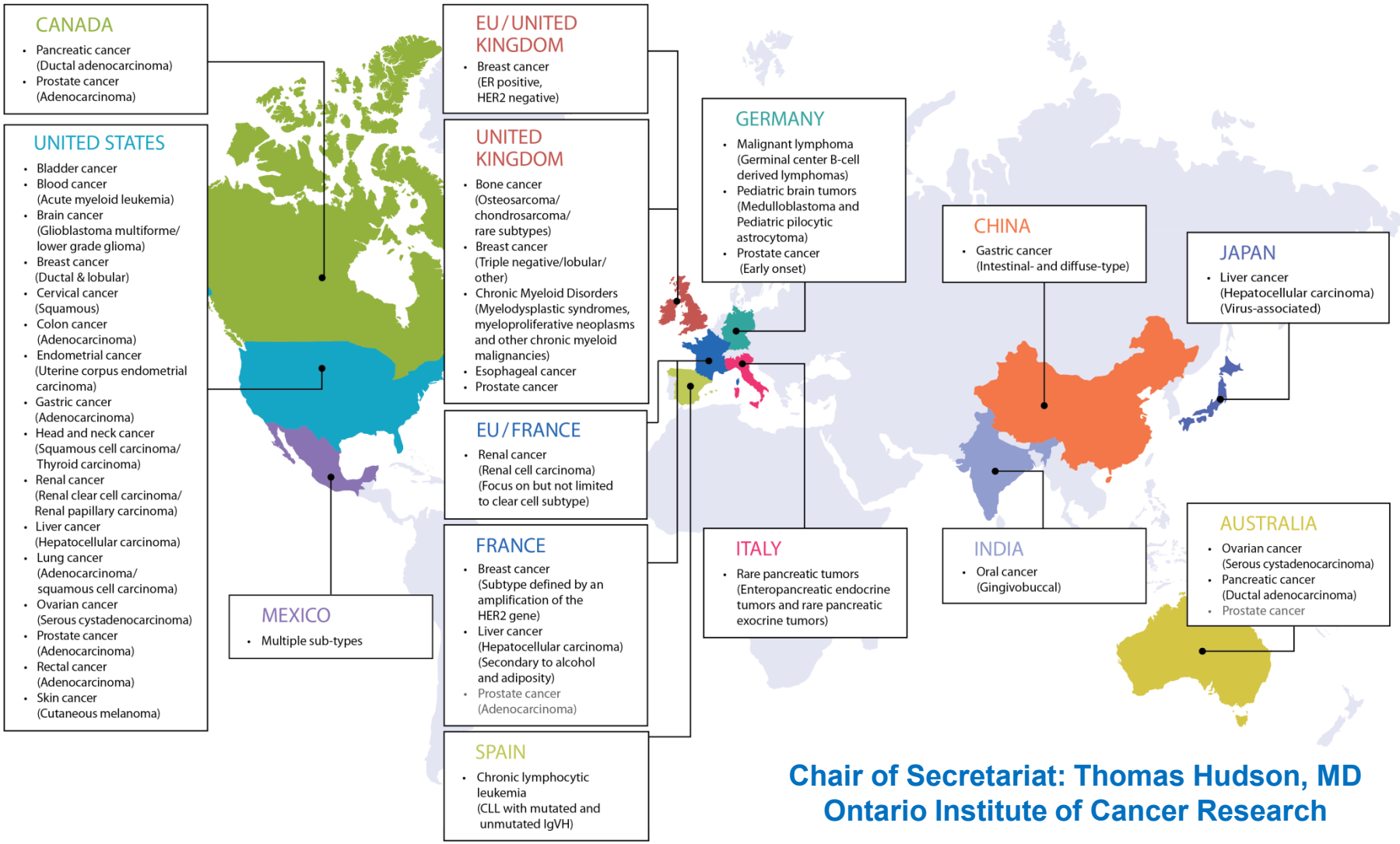



**https://cghub.ucsc.edu**

- 2 petabytes now, 5Pb ($5 \times 10^{15}$) total by 2014
- General Parallel File System, Dual RAID 6 subsystems, Redundant I/O paths
- Currently holds 10,000 files; expected to double in next 3 months
- Co-location opportunities in same data center for groups who want to compute on the data
- User support: support@cghub.ucsc.edu

# ICGC

## March 2011 : International Cancer Genome Consortium Projects




**Chair of Secretariat: Thomas Hudson, MD**
**Ontario Institute of Cancer Research**

# Acknowledgements

**NCI Program Office**

Kenna Shaw, Liming Yang, Roy Tarnuzzer, Zhining Wang, Emma Spaulding, Margi Sheth, John Demchok, Julia Zhang, Martin Ferguson, Greg Eley

**Center for Cancer Genomics**

Stephen Chanock & Lou Staudt

**NHGRI**

Brad Ozenberger, Heidi Sofia, Lindsay Lund, Mark Guyer