cBioPortal

Bioinformatics

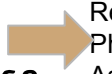
(http://www.cbioportal.org/)

It enables the exploration of

- ➔ Genetics alterations across samples
- ➔ Genes
- ➔ Pathways

Integrated genomics data types

- ★ Somatics mutations
- ★ DNA copy-number alterations(CNA)
- ★ mRNA and miRNA expression
- ★ DNA methylation
- ★ Protein abundance
- ★ Phosphoprotein abundance



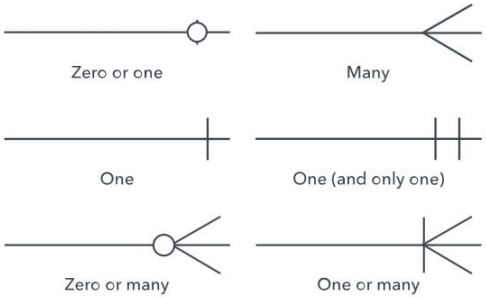
Datasets from

- Published cancer studies
- The Cancer Genome Atlas Project (TCGA)
- International Cancer Genome Consortium (ICGC)
- Cancer Cell Line Encyclopedia (CCLE)

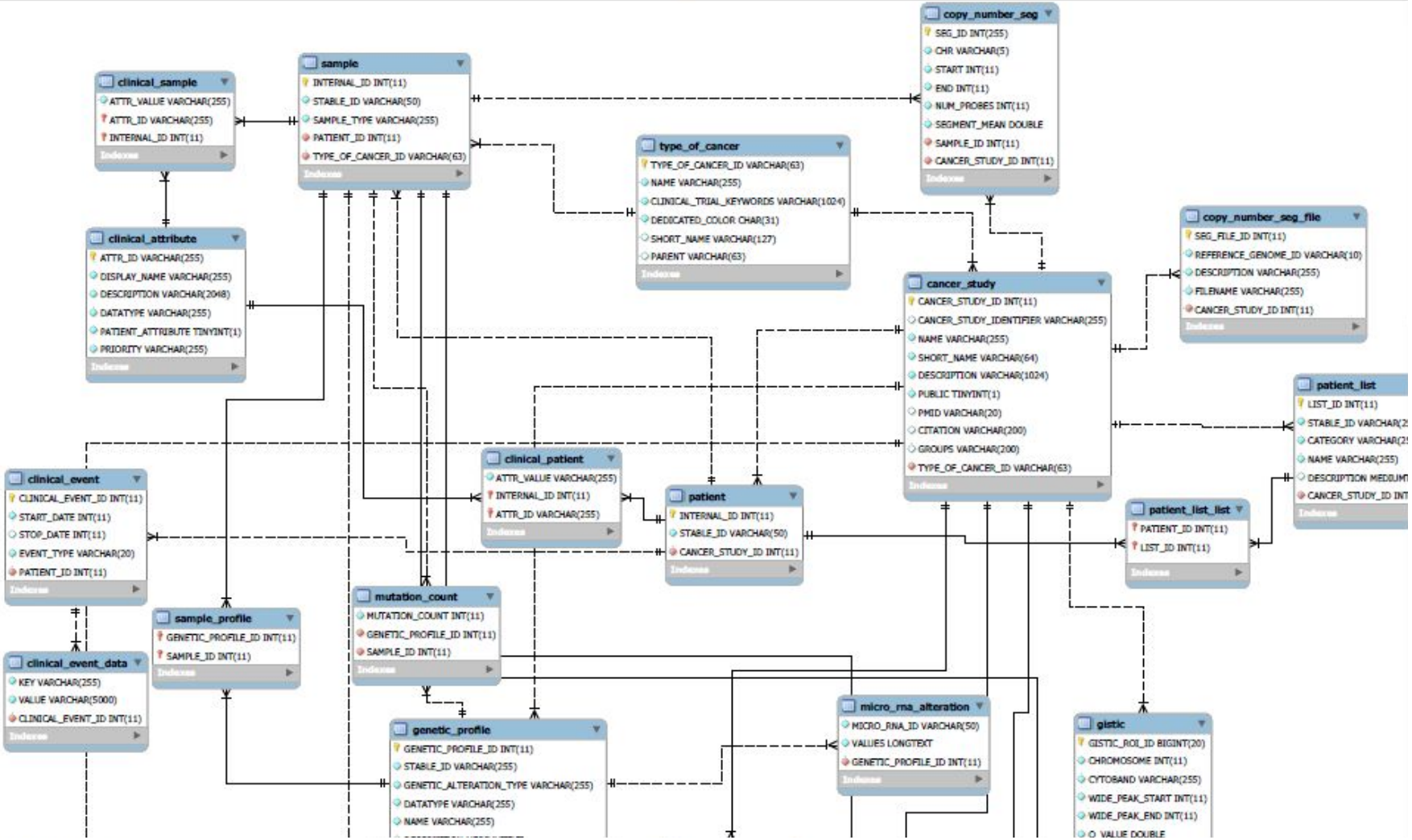
Querying

1. Single Cancer Study
2. Across Cancer Studies
3. Towards Precision Medicine

Entity-Relationship Diagram of cBioPortal



https://www.lucidchart.com/pages/er-diagrams
https://github.com/cBioPortal/cbioportal/blob/master/db-scripts/src/main/resources/cbioportal-er-diagram.pdf


copy_number_seg
SEG_ID INT(255)
CHR VARCHAR(5)
START INT(11)
END INT(11)
NUM_PROBES INT(11)
SEGMENT_MEAN DOUBLE
SAMPLE_ID INT(11)
CANCER_STUDY_ID INT(11)
clinical_sample
ATTR_VALUE VARCHAR(255)
ATTR_ID VARCHAR(255)
INTERNAL_ID INT(11)
sample
INTERNAL_ID INT(11)
STABLE_ID VARCHAR(50)
SAMPLE_TYPE VARCHAR(255)
PATIENT_ID INT(11)
TYPE_OF_CANCER_ID VARCHAR(63)
type_of_cancer
TYPE_OF_CANCER_ID VARCHAR(63)
NAME VARCHAR(255)
CLINICAL_TRIAL_KEYWORDS VARCHAR(1024)
DEDICATED_COLOR CHAR(31)
SHORT_NAME VARCHAR(127)
PARENT VARCHAR(63)
copy_number_seg_file
SEG_FILE_ID INT(11)
REFERENCE_GENOME_ID VARCHAR(10)
DESCRIPTION VARCHAR(255)
FILENAME VARCHAR(255)
CANCER_STUDY_ID INT(11)
clinical_attribute
ATTR_ID VARCHAR(255)
DISPLAY_NAME VARCHAR(255)
DESCRIPTION VARCHAR(2048)
DATATYPE VARCHAR(255)
PATIENT_ATTRIBUTE TINYINT(1)
PRIORITY VARCHAR(255)
cancer_study
CANCER_STUDY_ID INT(11)
CANCER_STUDY_IDENTIFIER VARCHAR(255)
NAME VARCHAR(255)
SHORT_NAME VARCHAR(64)
DESCRIPTION VARCHAR(1024)
PUBLIC TINYINT(1)
PMID VARCHAR(20)
CITATION VARCHAR(200)
GROUPS VARCHAR(200)
TYPE_OF_CANCER_ID VARCHAR(63)
patient_list
LIST_ID INT(11)
CATEGORY VARCHAR(
NAME VARCHAR(255)
CANCER_STUDY_ID INT
clinical_patient
ATTR_VALUE VARCHAR(255)
INTERNAL_ID INT(11)
ATTR_ID VARCHAR(255)
patient
INTERNAL_ID INT(11)
STABLE_ID VARCHAR(50)
CANCER_STUDY_ID INT(11)
patient_list_list
PATIENT_ID INT(11)
LIST_ID INT(11)
clinical_event
CLINICAL_EVENT_ID INT(11)
START_DATE INT(11)
STOP_DATE INT(11)
EVENT_TYPE VARCHAR(20)
PATIENT_ID INT(11)
mutation_count
MUTATION_COUNT INT(11)
GENETIC_PROFILE_ID INT(11)
SAMPLE_ID INT(11)
sample_profile
GENETIC_PROFILE_ID INT(11)
SAMPLE_ID INT(11)
clinical_event_data
KEY VARCHAR(255)
VALUE VARCHAR(5000)
CLINICAL_EVENT_ID INT(11)
genetic_profile
GENETIC_PROFILE_ID INT(11)
STABLE_ID VARCHAR(255)
GENETIC_ALTERATION_TYPE VARCHAR(255)
DATATYPE VARCHAR(255)
NAME VARCHAR(255)
micro_rna_alteration
MICRO_RNA_ID VARCHAR(50)
VALUES LONGTEXT
GENETIC_PROFILE_ID INT(11)
gistic
GISTIC_ROI_ID BIGINT(20)
CHROMOSOME INT(11)
CYTOBAND VARCHAR(255)
WIDE_PEAK_START INT(11)
WIDE_PEAK_END INT(11)

Atlassian

Bitbucket

Source Code

Further Development

Academic Purposes

```java
package org.mskcc.oncotree.web;

import java.nio.charset.StandardCharsets;
import java.nio.file.Files;
import java.nio.file.Paths;
import java.io.*;
import java.util.List;
import java.util.Properties;
import javax.servlet.*;
import javax.servlet.http.*;

/**
 * Created by zhangh2 on 7/14/15.
 */
public class TumorType extends HttpServlet {
    public void doGet(HttpServletRequest request, HttpServletResponse response)
        throws IOException{

                String propFileName = "properties/config.properties";
                Properties properties = new Properties();
                InputStream inputStream = TumorType.class.getClassLoader().getResourceAsStream(propFileName);

                if (inputStream != null) {
                        properties.load(inputStream);
                } else {
                        throw new FileNotFoundException("property file '" + propFileName + "' not found in the classpath");
                }
                inputStream.close();

                String tumorTypeFilePath = properties.getProperty("tumor_type_file_path");



                response.setContentType("text/plain");

```

https://github.com/cBioPortal

cBioPortal

cBioPortal

MSKCC http://cbioportal.org cbioportal@cbio.mskcc.o...

Repositories People 10

Pinned repositories

cbioportal

cBioPortal for Cancer Genomics

JavaScript 111 141

datahub

A centralized location for storing curated data ready for inclusion in cBioPortal.

6 5

```xml
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.
  <!-- meta data -->
  <parent>
    <artifactId>master</artifactId>
    <groupId>org.mskcc.cbio</groupId>
    <version>1.5.1</version>
  </parent>
  <modelVersion>4.0.0</modelVersion>
  <artifactId>core</artifactId>
  <name>Portal Core</name>
  <description>Core libraries shared among other modules</description>

  <dependencies>
    <dependency>
      <groupId>org.mskcc.cbio</groupId>
      <artifactId>business</artifactId>
      <version>${project.version}</version>
    </dependency>
    <dependency>
      <groupId>org.mskcc.cbio</groupId>
      <artifactId>security-spring</artifactId>
```

Behind the Scene

Default

★ mRNA,miRNA (expression) ,protein(abundance) and phosphoprotein data, z-score= 2 SD from the mean

Correlation Plots

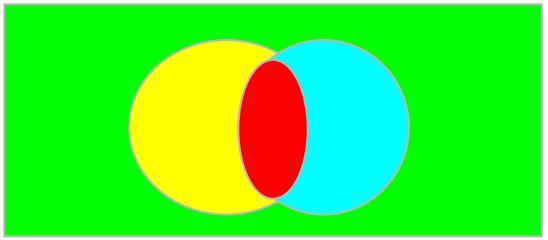
➔ Pearson, Spearman Correlation
➔ eg. methylation beta value >estimate for methylation level of CpG locus based on intensity ratio between methylated and unmethylated alleles

Copy-number status

❖ Homozygously deleted, heterozygously deleted, diploid, gained(amplification event with few copies) , amplified(high copy amplification)

Behind the Scene

Mutual Exclusivity vs. Co-occurence (An Event)



- Based on log odds ratio (OR) score- likelihood of an event

  Let's say, we have two query genes(G1 and G2)

  A: # cases altered in both genes

  D: # cases altered in none of them

  B: # cases altered in G1 but not G2

  C: # cases altered in G2 but not G1

  $$OR= (A*D) / (B*C)$$

- Significance for relationship each gene pair is determined by Fisher's Exact Test (means not an assumption towards infinity, based on p-value)

Behind the Scene

Mutations (just for nonsynonymous)

- Graphical summary gives position and frequency of all mutations as Pfam(Protein family) protein domain(functional regions-e.g binding region)
- Mutations affecting canonical pathways are chosen from Oncotator(Cancer Variant Annotation Tool) (annotate human point mutations, insertion and deletions with relevant cancer)
- # mutation at this position in COSMIC(Catalogue ofSomatic Mutations in Cancer)
- functional impact prediction of missense mutations from Mutation Assessor

Behind the Scene

Protein Changes

- ➢ Proteomics data available from RPPA (Reverse Phase Protein Assay) platform for 12 TCGA patients
- ➢ Protein versus mRNA expression via scatter plot at Plots tab
- ➢ Differential analysis for available RPPA data correlation w/genomic alterations
- ➢ two sided, two-sample Student's t test to identify differences

Survival

- overall vs. disease-free survival in Kaplan-Meier plot w/ p-values from logrank test.
- Kaplan-Meier plot is an estimation plot

Behind the Scene

Network

- Pathways from Human Reference Protein Database(HRPD)
- Reactome
- National Cancer Institue (NCI)-Nature
- Memorial Sloan-Kettering Cancer Center(MSKCC)
- Cancer Cell Map

all derived from Pathway Commons Project

- ❖ Detailed visualization via Cytoscape
- ❖ gene-centric drug target info from DrugBank, KEGG Drug,NCI Cancer Drugs

Behind the Scene

Download

You can also download all the genomic data and per-sample alteration events by using Download option for further analysis

You can visualize CN details via IGV

Bookmark

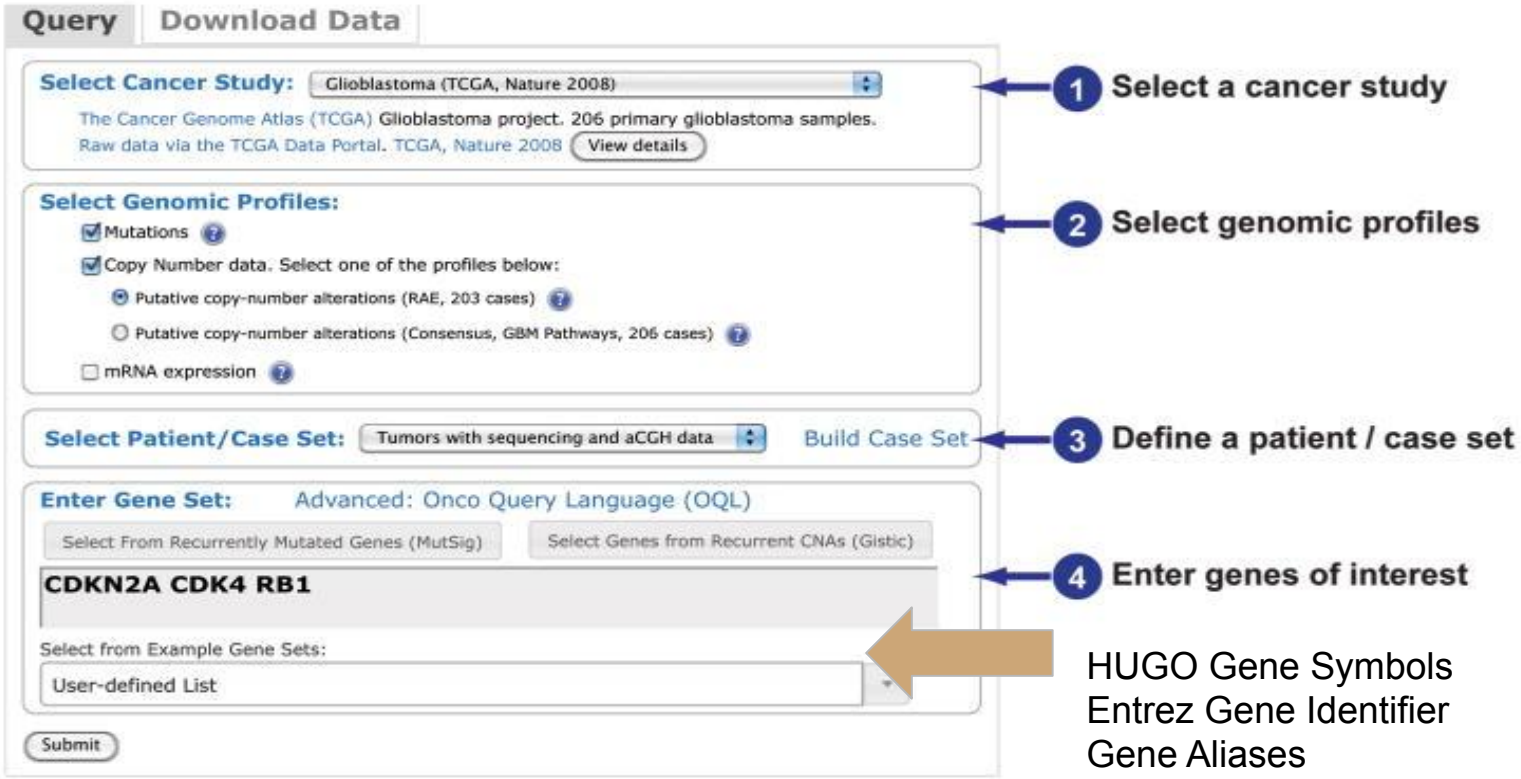
You can store the URL of specific query on your browser as a shortened link, or you can share it w/your collaborator

R-MATLAB Packages

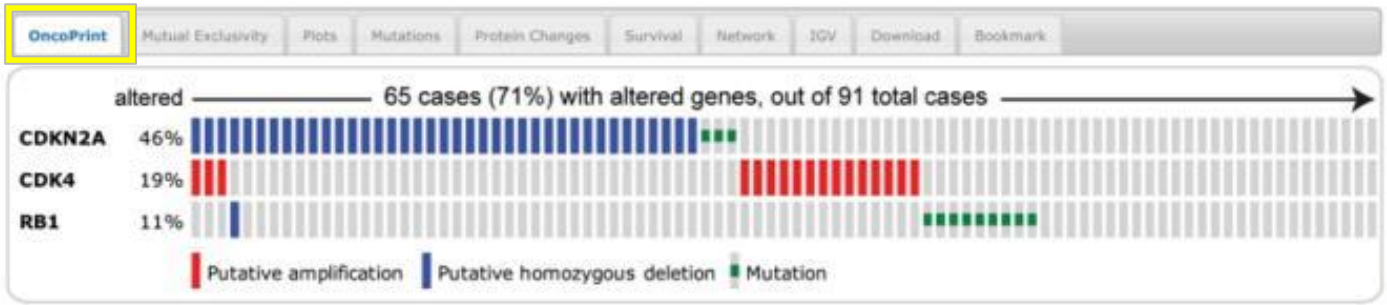
data retrieval for further analysis
CRAN> CGDS-R package | MATLAB CGDS Cancer Genomics Toolbox

DEMO

1-Querying Individual Cancer Studies



1-Querying Individual Cancer Studies

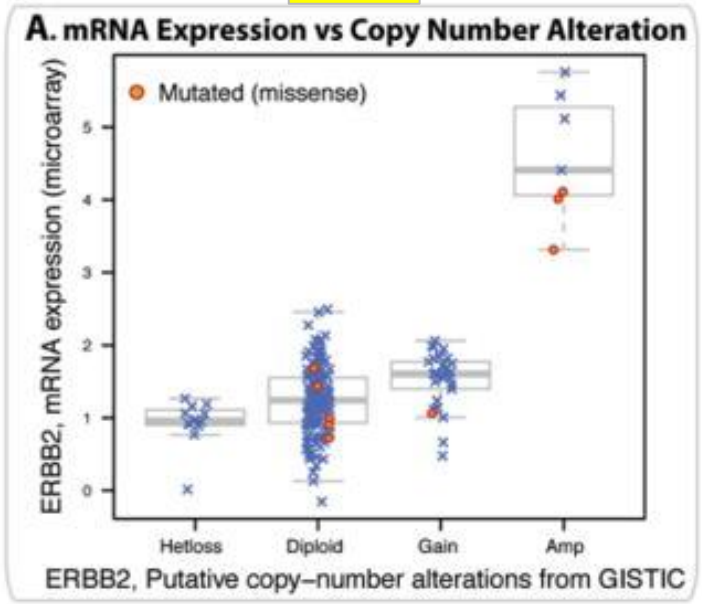


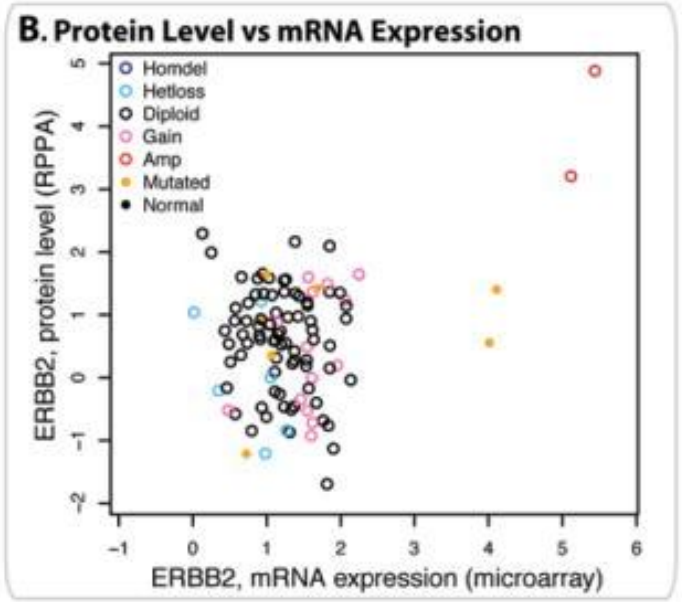
Query:
1. Glioblastoma (TCGA, Nature 2008)
2. Mutations and Copy Number Data
3. Tumors with sequencing and aCGH data
4. CDKN2A CDK4 RB1

Query:
1. Glioblastoma (TCGA, Nature 2008)
2. Mutations and Copy Number Data
3. Tumors with sequencing and aCGH data
4. CDKN2A CDK4 RB1

1-Querying Individual Cancer Studies

OncoPrint | Mutual Exclusivity | **Plots** | Mutations | Protein Changes | Survival | Network | IGV | Download | Bookmark



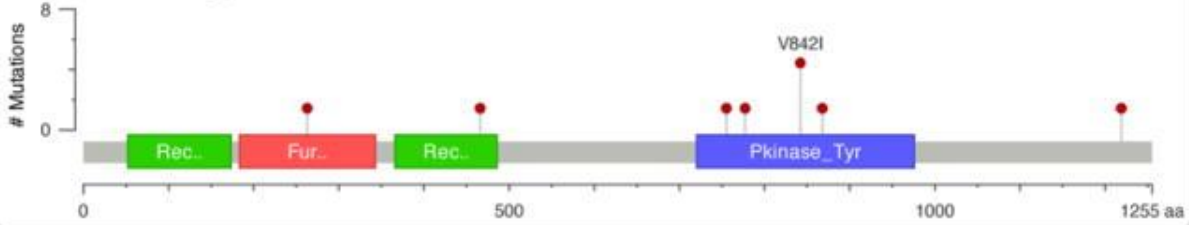


Query:
1. Colon and Rectum Adenocarcinoma (TCGA, Provisional)
2. Mutations and Copy Number Data
3. All Complete Tumors
4. ERBB2

1-Querying Individual Cancer Studies

OncoPrint | Mutual Exclusivity | Plots | **Mutations** | Protein Changes | Survival | Network | IGV | Download | Bookmark

A. Mutation diagram of ERBB2 in colorectal cancer (TCGA)



B. Mutations of ERBB2 in colorectal cancer (TCGA)

Case ID	AA Change	Type	COSMIC	FIS	3D	MS	VS	Center
TCGA-AA-3864	*I263T*	Missense		N	3D	S	U	Broad
TCGA-AA-3977	*A466T*	Missense		L	3D	S	V	Broad
TCGA-AA-3869	*L755S*	Missense	6	M	3D	S	V	Broad
TCGA-AA-3994	*V777L*	Missense	8	N	3D	S	V	Broad
TCGA-AA-3994	*V842I*	Missense	2	N	3D	S	V	Broad
TCGA-AG-3599	*V842I*	Missense	2	N	3D	S	V	Baylor
TCGA-AG-3898	*V842I*	Missense	2	N	3D	S	V	Broad
TCGA-AG-3902	*V842I*	Missense	2	N	3D	S	V	Broad
TCGA-AG-A002	*R868W*	Missense		H	3D	S	U	Broad
TCGA-AA-3947	*N1219S*	Missense		L		S	V	Broad

Query:
1. Colon and Rectum Adenocarcinoma (TCGA, Provisional)
2. Mutations
3. Sequenced Tumors
4. ERBB2

1-Querying Individual Cancer Studies

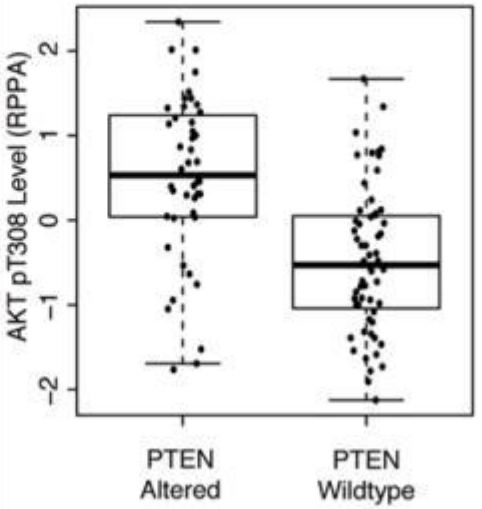
OncoPrint | Mutual Exclusivity | Plots | Mutations | **Protein Changes** | Survival | Network | IGV | Download | Bookmark

A. Phopshoproteins with different abundence levels between PTEN-altered samples and others in GBM

Antibody Type: phosphorylation Search:

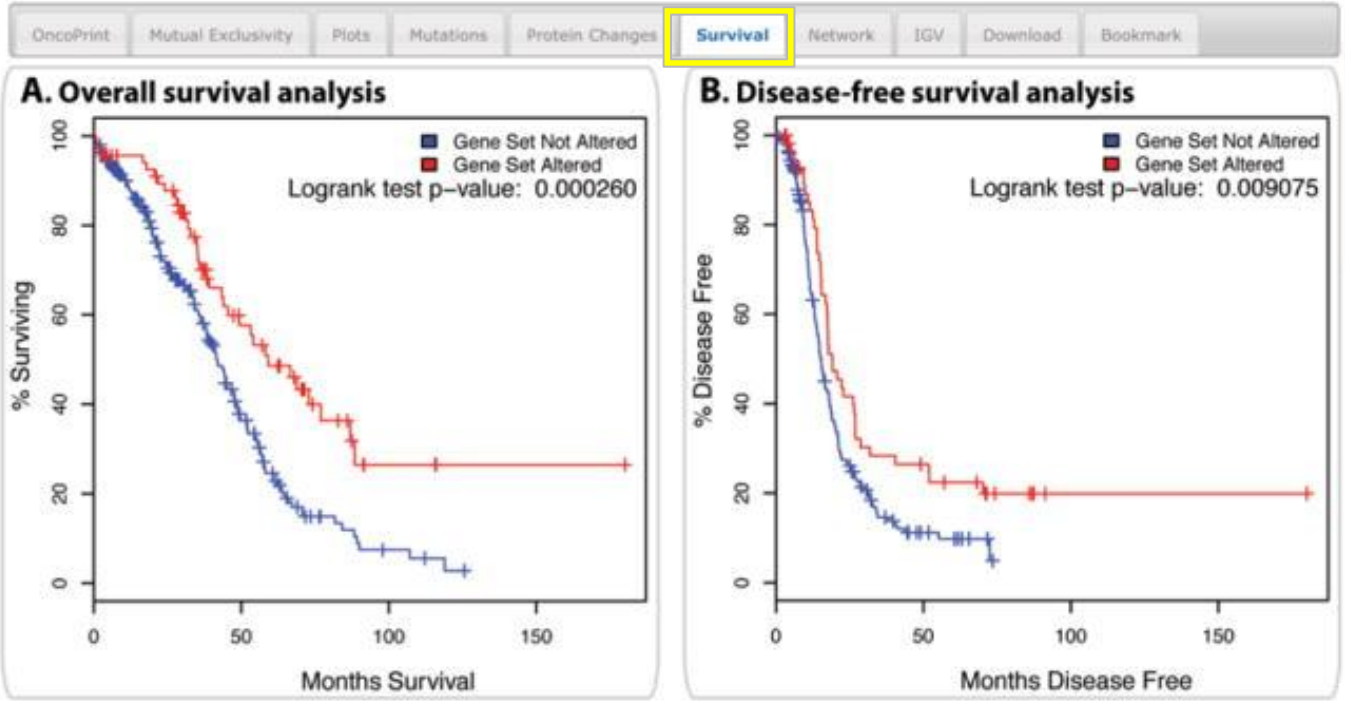
Target		Ave. Abundance		p-value	Plot
Protein	Residue	Unaltered	Altered		
AKT1/AKT2/AKT3	pT308	-0.47	0.51	**9.92e-7**	
AKT1/AKT2/AKT3	pS473	-0.40	0.45	**9.48e-6**	
EGFR	pY992	0.34	-0.11	**0.017**	
SHC1	pY317	0.15	-0.26	**0.027**	
GSK3A/GSK3B	pS9	-0.29	0.15	**0.031**	
EGFR	pY1173	0.26	-0.09	0.067	
EGFR	pY1068	0.26	-0.06	0.095	
RPS6	pS235	-0.08	0.23	0.103	
RPS6	pS240	-0.09	0.22	0.121	
GSK3A/GSK3B	pS21	-0.23	0.08	0.141	
EIF4EBP1	pT37	-0.13	0.15	0.147	
RPS6KA1	pT359	-0.00	-0.28	0.149	

B. AKT pT308 levels of PTEN-altered samples versus others in GBM



Query:
1. Glioblastoma Multiforme (TCGA, Provisional)
2. Mutations and Copy Number Data
3. Tumors with sequencing and aCGH data
4. PTEN

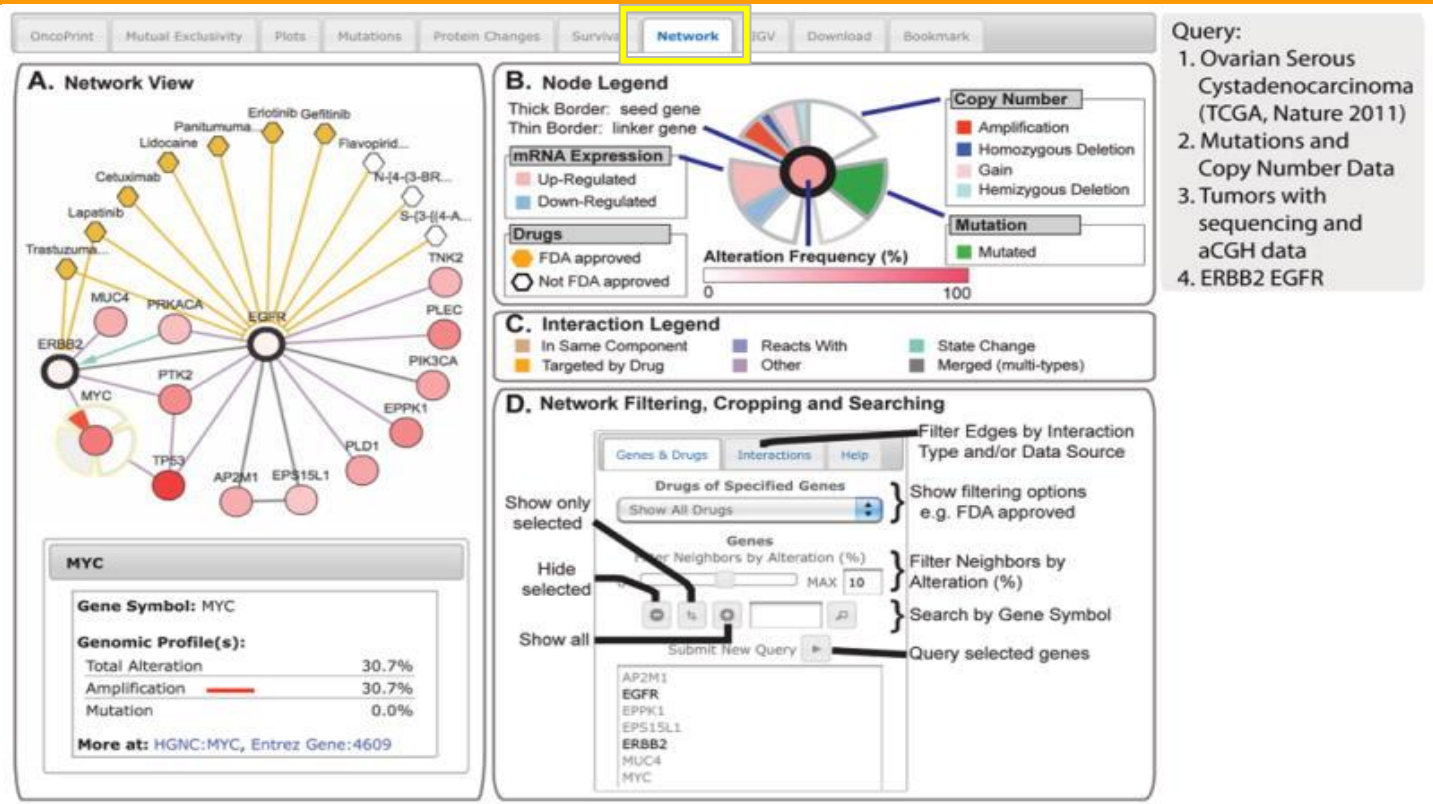
1-Querying Individual Cancer Studies



Query:

1. Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011)
2. Mutations
3. Sequenced Tumors
4. BRCA1 BRCA2

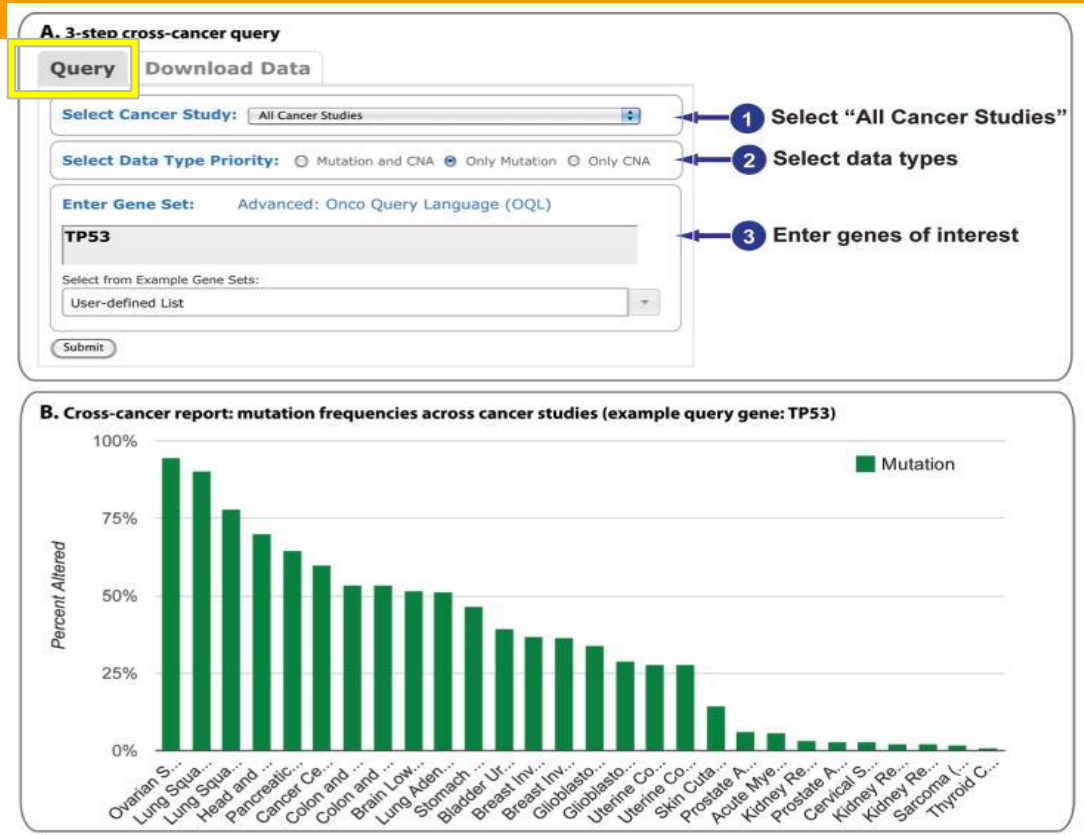
1-Querying Individual Cancer Studies



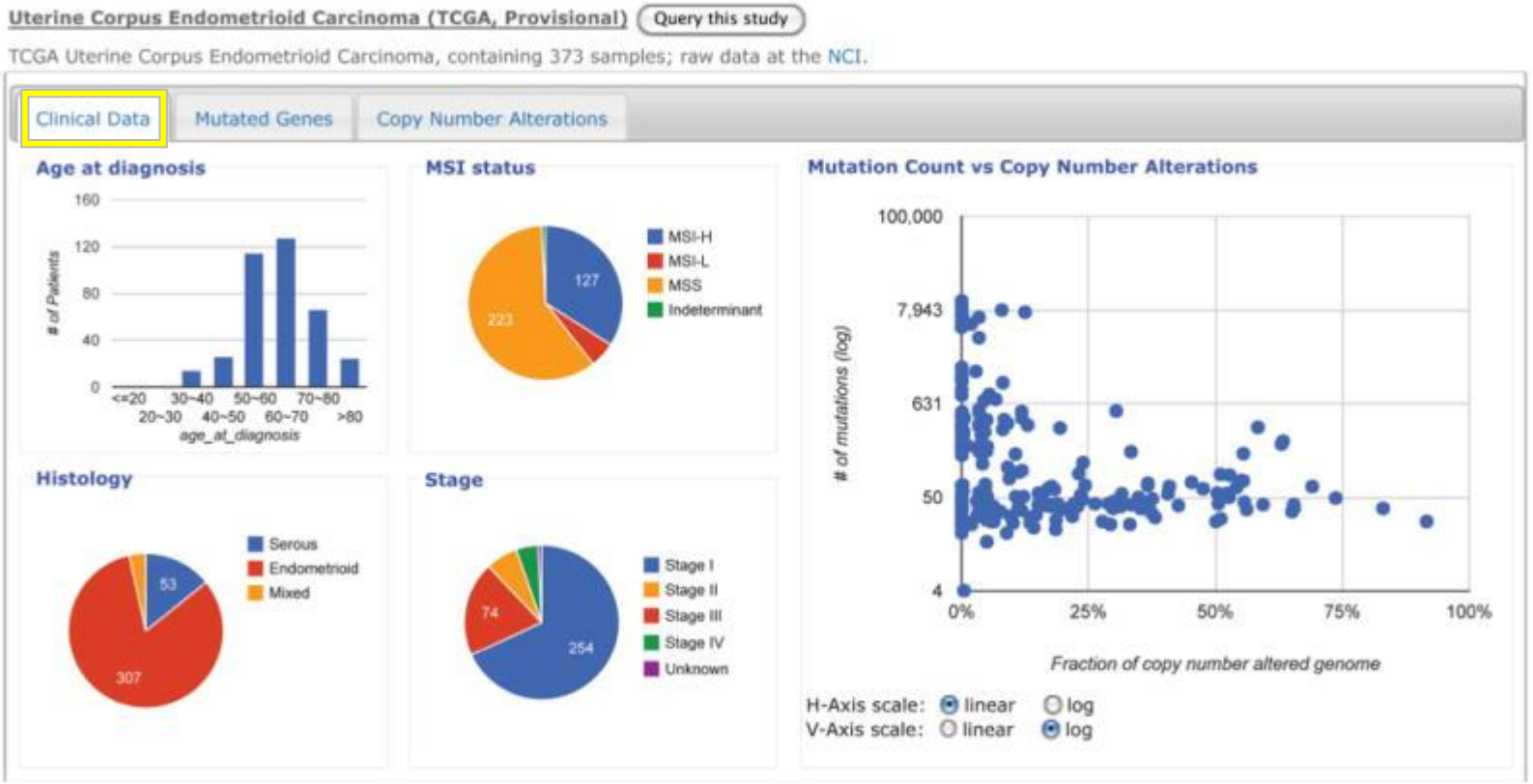
Query:

1. Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011)
2. Mutations and Copy Number Data
3. Tumors with sequencing and aCGH data
4. ERBB2 EGFR

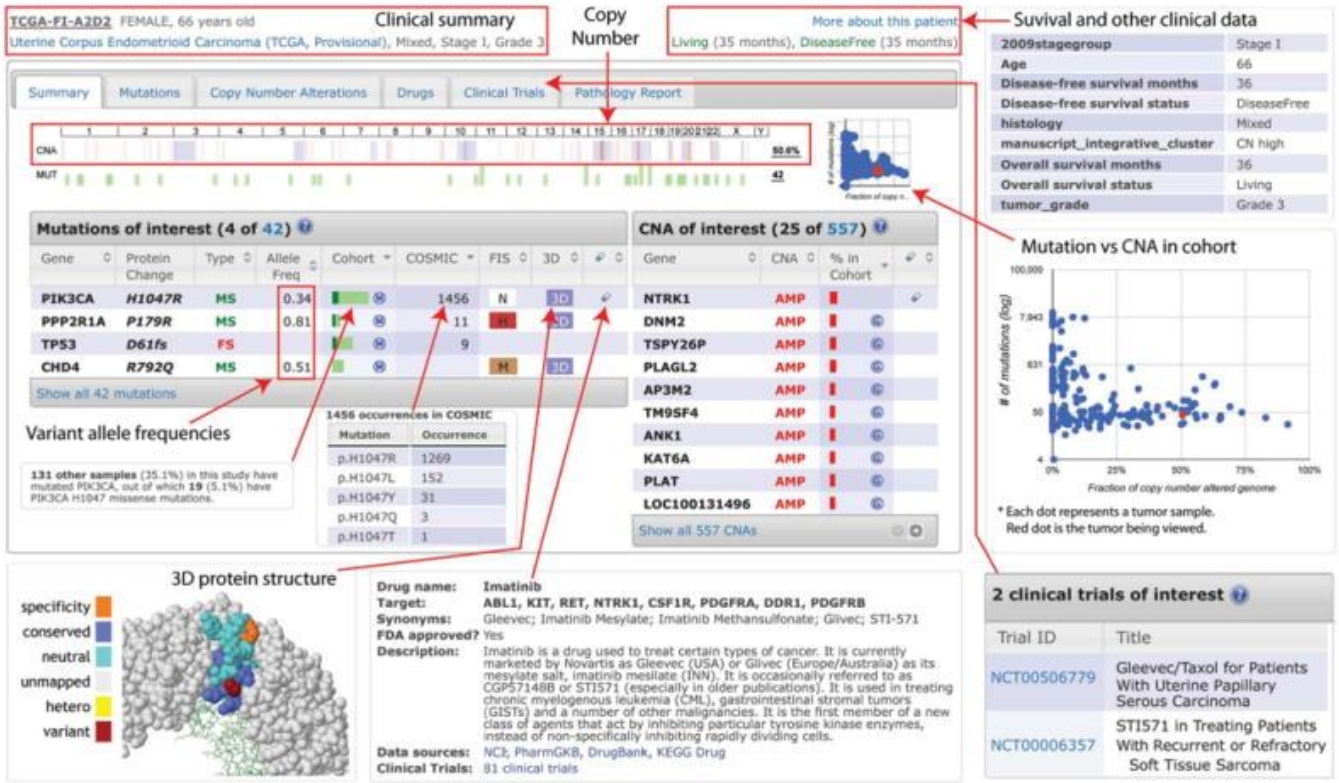
2-Cross Cancer Queries



Cancer Study Summary View



Individual Patient View



References

Cerami et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*. May 2012 2; 401.

Gao et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, pl1 (2013)

Gallagher, R. I., & Espina, V. (2014). Reverse Phase Protein Arrays: Mapping the path towards personalized medicine. *Molecular Diagnosis & Therapy*, *18*(6), 619–630. http://doi.org/10.1007/s40291-014-0122-3

http://www.cbioportal.org/

https://www.lucidchart.com/pages/er-diagrams

https://github.com/cBioPortal/cbioportal/