CANCER GENOMICS CONSORTIUM

Educating for Best Practices in Clinical Cancer Genomics

cBioPortal FOR CANCER GENOMICS

cBioPortal

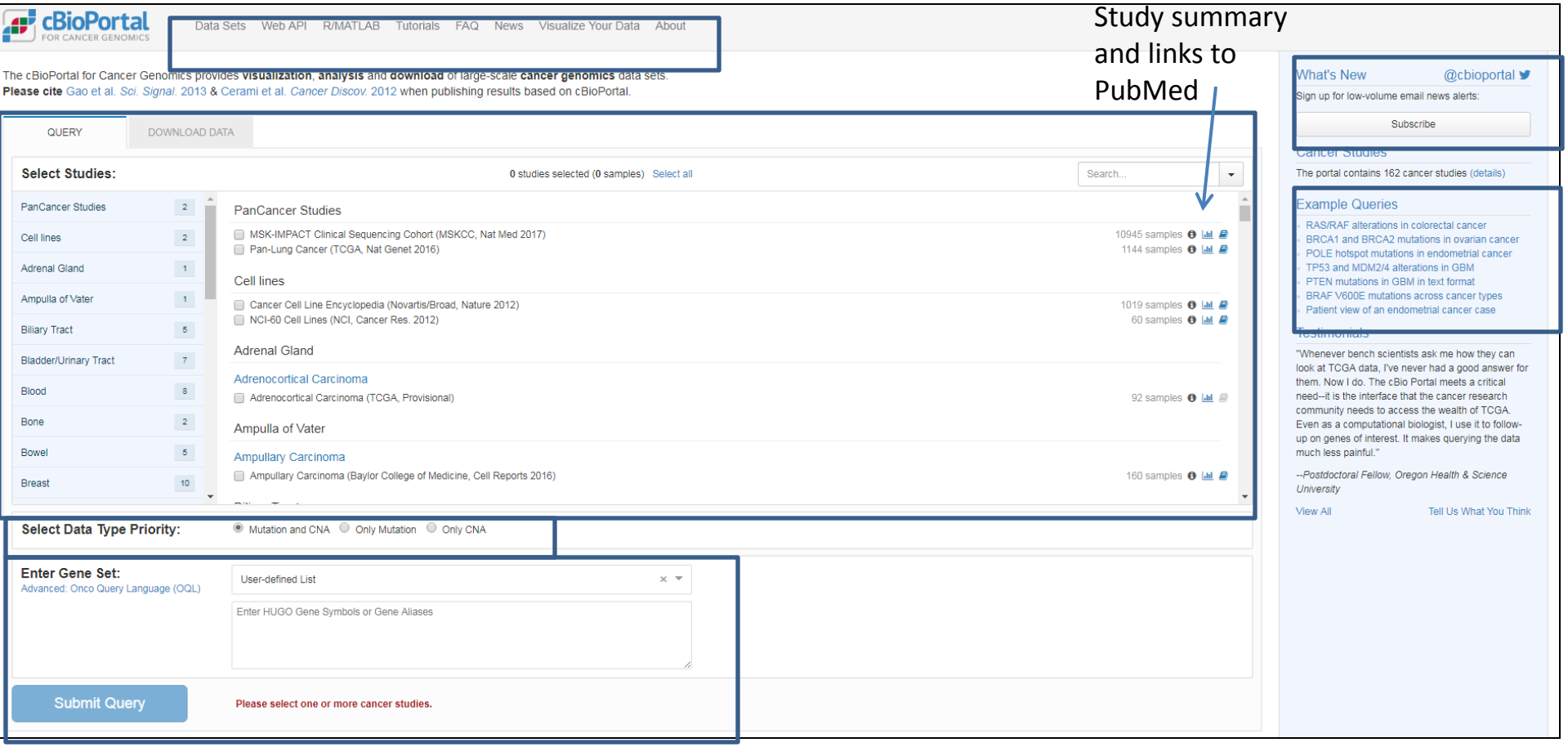
http://www.cbioportal.org/

https://www.ncbi.nlm.nih.gov/pubmed/23550210
http://cancerdiscovery.aacrjournals.org/content/2/5/401

Tutorials

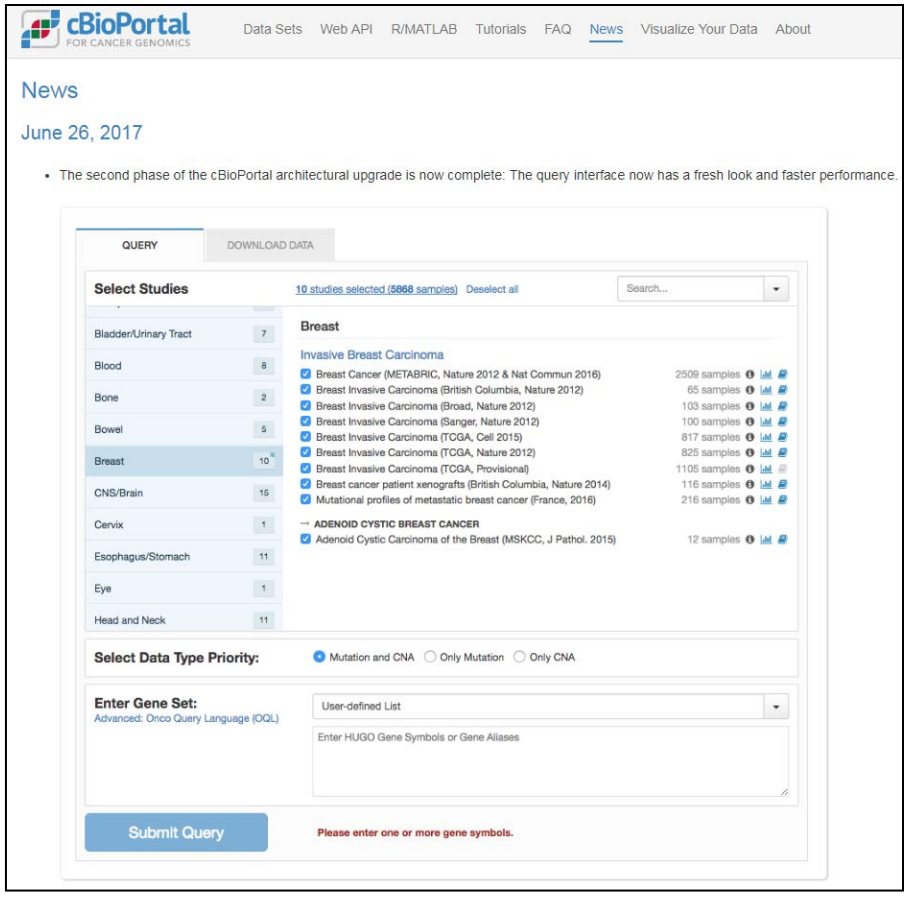
- http://www.cbioportal.org/tutorial.jsp
- http://www.cbioportal.org/faq.jsp
- https://www.ncbi.nlm.nih.gov/pubmed/23550210

Home Page


cBioPortal FOR CANCER GENOMICS
Data Sets
Web API
R/MATLAB
Tutorials
FAQ
News
Visualize Your Data
About
Study summary and links to PubMed
The cBioPortal for Cancer Genomics provides visualization, analysis and download of large-scale cancer genomics data sets.
Please cite Gao et al. Sci. Signal. 2013 & Cerami et al. Cancer Discov. 2012 when publishing results based on cBioPortal.
QUERY
DOWNLOAD DATA
Select Studies:
0 studies selected (0 samples)
Select all
Search...
PanCancer Studies
Cell lines
Adrenal Gland
Ampulla of Vater
Biliary Tract
Bladder/Urinary Tract
Blood
Bone
Bowel
Breast
MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017)
10945 samples
Pan-Lung Cancer (TCGA, Nat Genet 2016)
1144 samples
Cancer Cell Line Encyclopedia (Novartis/Broad, Nature 2012)
1019 samples
NCI-60 Cell Lines (NCI, Cancer Res. 2012)
60 samples
Adrenocortical Carcinoma
Adrenocortical Carcinoma (TCGA, Provisional)
92 samples
Ampullary Carcinoma
Ampullary Carcinoma (Baylor College of Medicine, Cell Reports 2016)
160 samples
Select Data Type Priority:
Mutation and CNA
Only Mutation
Only CNA
Enter Gene Set:
Advanced: Onco Query Language (OQL)
User-defined List
Enter HUGO Gene Symbols or Gene Aliases
Submit Query
Please select one or more cancer studies.
What's New
@cbioportal
Sign up for low-volume email news alerts:
Subscribe
Cancer Studies
The portal contains 162 cancer studies (details)
Example Queries
RAS/RAF alterations in colorectal cancer
BRCA1 and BRCA2 mutations in ovarian cancer
POLE hotspot mutations in endometrial cancer
TP53 and MDM2/4 alterations in GBM
PTEN mutations in GBM in text format
BRAF V600E mutations across cancer types
Patient view of an endometrial cancer case
Testimonials
"Whenever bench scientists ask me how they can look at TCGA data, I've never had a good answer for them. Now I do. The cBio Portal meets a critical need--it is the interface that the cancer research community needs to access the wealth of TCGA. Even as a computational biologist, I use it to follow-up on genes of interest. It makes querying the data much less painful."
--Postdoctoral Fellow, Oregon Health & Science University
View All
Tell Us What You Think

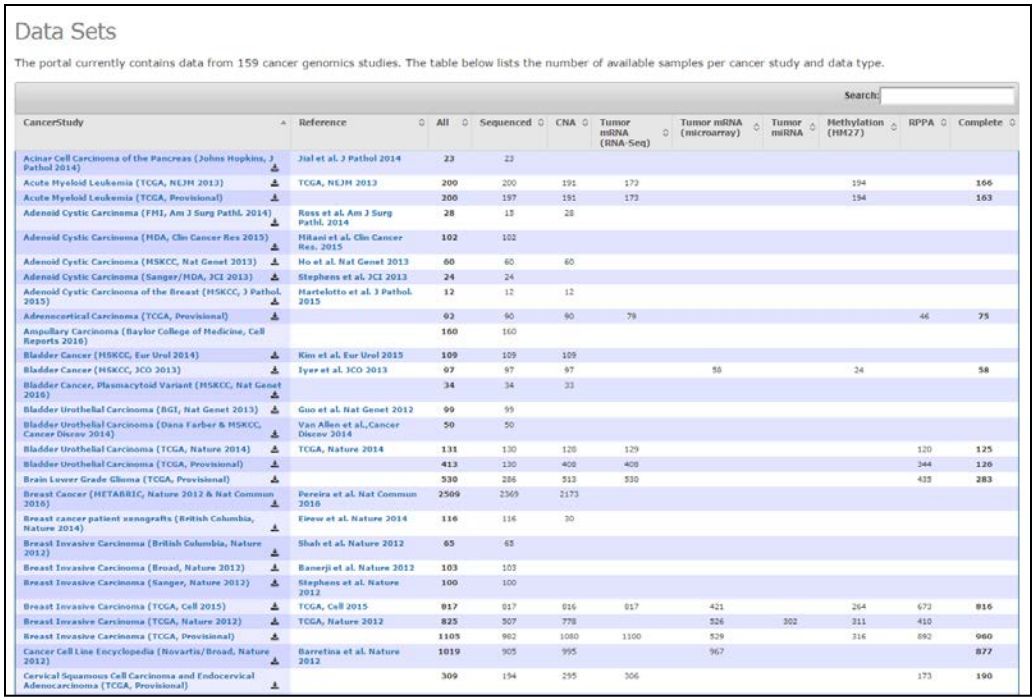
News

- Cilck on “News” in top banner on home page.
- Upgrades to cBioPortal and additions to data are listed.



Data Sets

- Click on “Data Sets” in top banner on home page.
- All data that are housed in CBioPortal are arranged by Cancer study.
  - Data are available for download for most studies
  - Links to References (PubMed)
  - Summary of data involved in study
    - Total # cases
    - Type of data in study

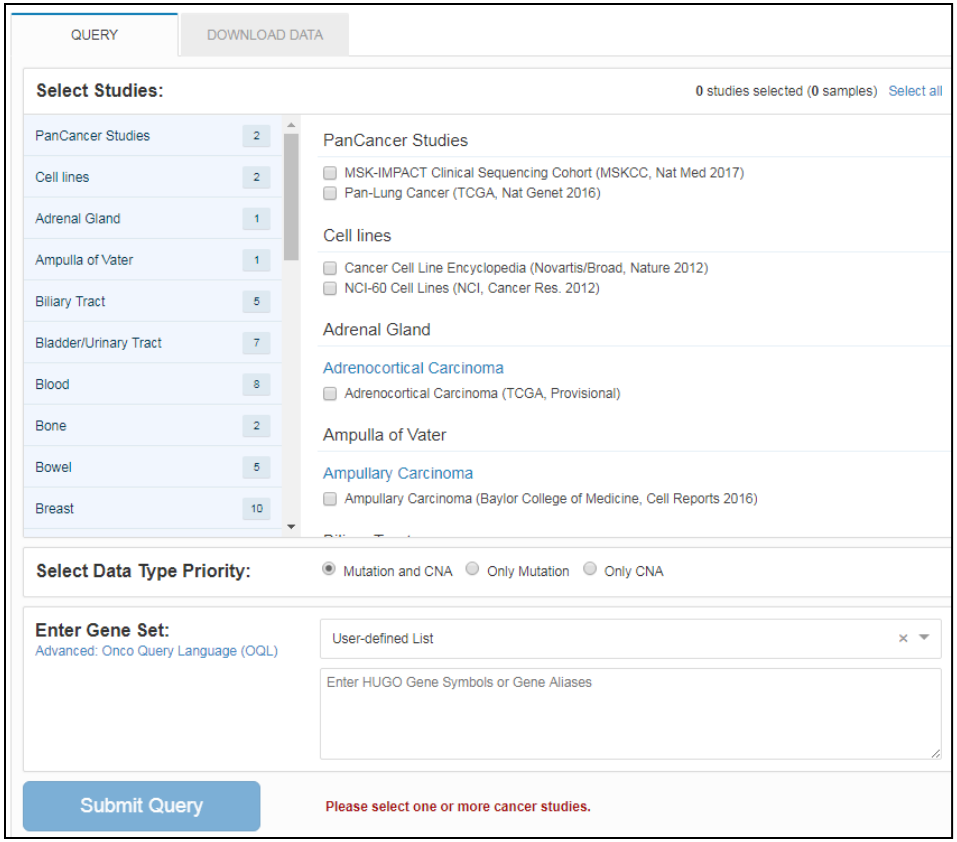


Web API and R/MATLAB Packages

- Click on “Web API” in top banner on home page.
- Web interface API code is provided to connect with the Cancer Genomic Data Server (CGDS)
- Can also access CGDS through R or MATLAB statistical software
  - Installation instructions for packages available under “R/MATLAB” in top banner on home page.
- Create a local instance of cBioPortal for your laboratory data through GitHub (https://github.com/cBioPortal/cbioportal) or Wiki pages (https://github.com/cBioPortal/cbioportal/wiki)

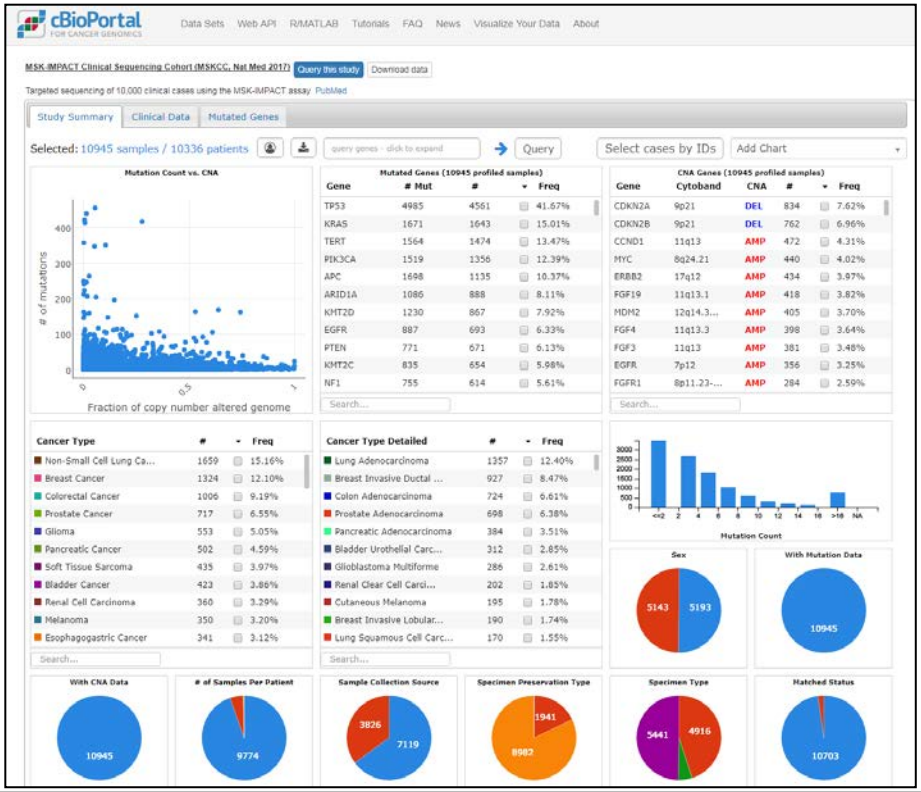
Setting up a Query

- Select Cancer Study / Data Set(s) to query
- Select Data Type(s) of interest
- Select/Enter Gene Set to interrogate
  - User Defined List
    - Can be copied/pasted into window.
    - Can follow Onco Query Language if you are looking for specific alterations in genes.
  - Pre-defined list
    - Will populate in window if selected from drop-down



Study Summary

- Prior to setting up a query based on specific studies, you might want to look at a summary of the data from those studies.
  - Click on  button on home page next to desired study.



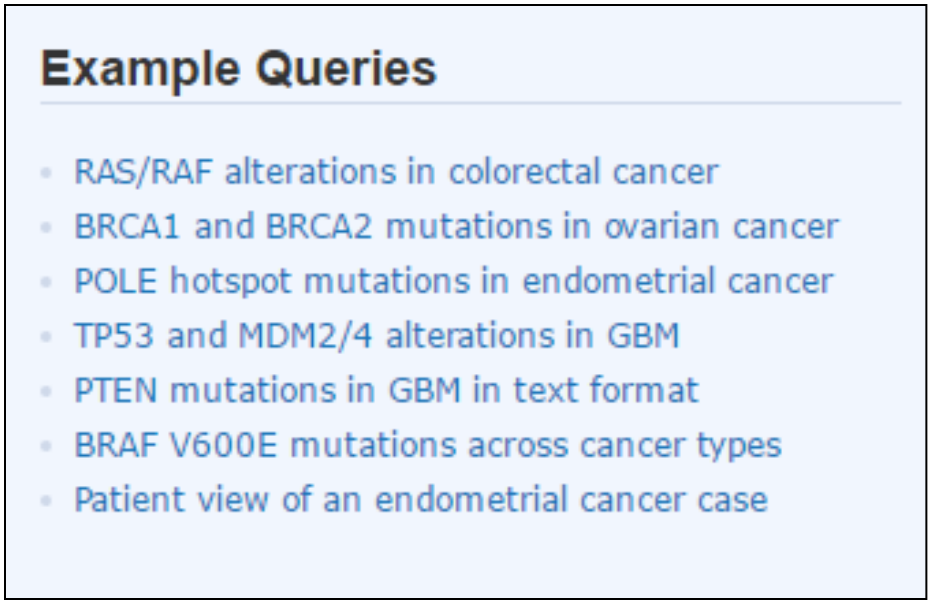
Onco Query Language

- Guide is visible in separate window when link is selected from home page near Gene Set window

Data Type	Keyword	Categories and Levels	Default
Copy Number Alterations	CNA	AMP Amplified HOMDEL Deep Deletion GAIN Gained HETLOSS Shallow Deletion	AMP and HOMDEL
Mutations	MUT	MUT Show mutated cases MUT = X Specific mutations or mutation types.	All somatic, non-synonymous mutations
Fusions	FUSION	FUSION	Show cases with fusions
mRNA Expression	EXP	EXP < -x Under-expression is less than x SDs below the mean. EXP > x Over-expression is greater than x SDs above the mean. The comparison operators <= and >= also work.	At least 2 standard deviations (SD) from the mean.
Protein/phosphoprotein level (RPPA)	PROT	PROT < -x Protein-level under-expression is less than x SDs below the mean. PROT > x Protein-level over-expression is greater than x SDs above the mean. The comparison operators <= and >= also work.	At least 2 standard deviations (SD) from the mean.

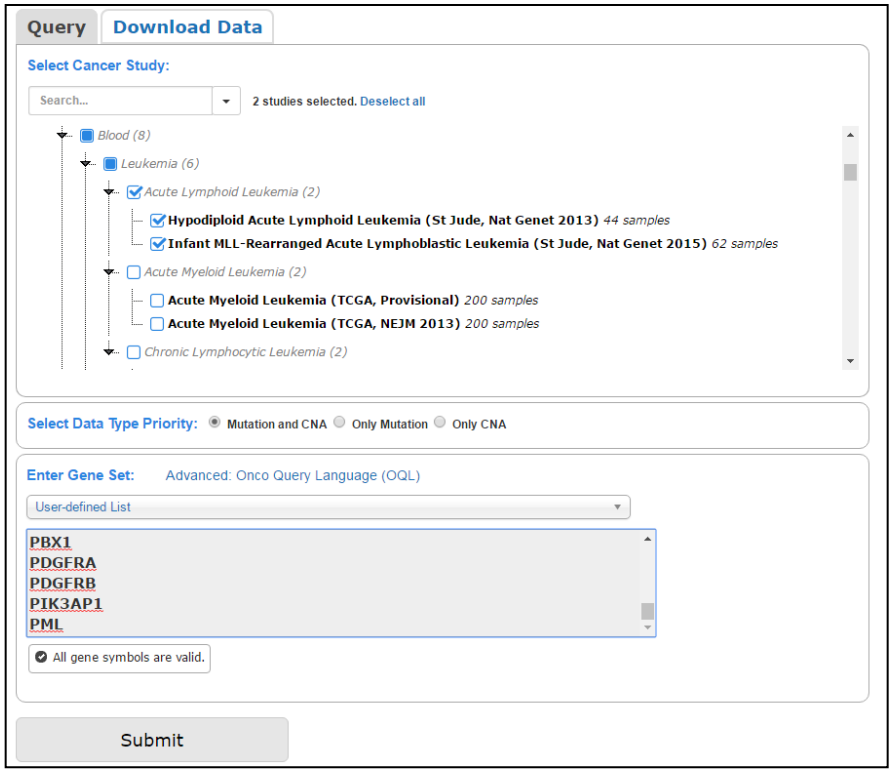
Example Queries

- Example Queries are available on bottom right panel of home page.
- Clicking on these links will bring you straight to the data and will not show you the query screen unless you click “Modify Query” on resulting screen.
- Examples are useful if you are running a similar query.



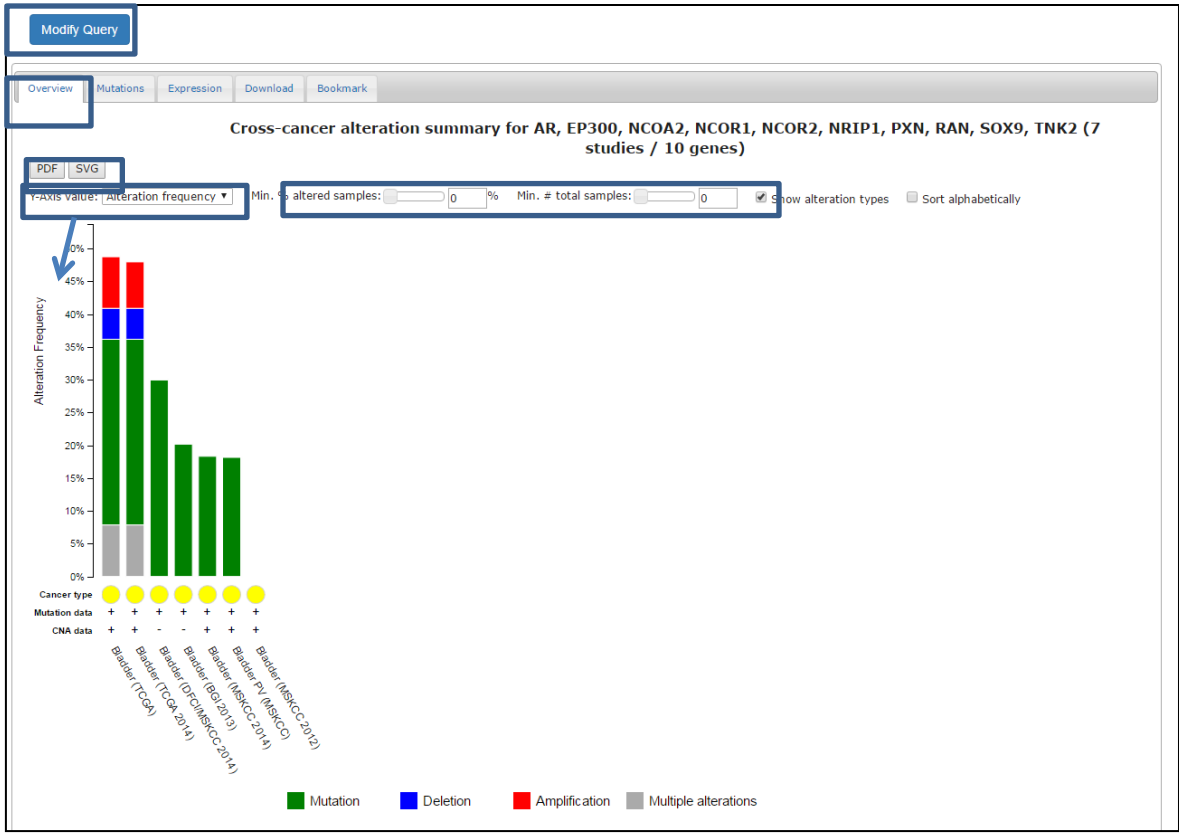
Query

- Simple Query for mutations and CNAs in several B-ALL genes across all ALL studies in cBioPortal.



Query Results: Overview

- Overview of alterations in your desired data set are viewed as bar graph and can be downloaded as PDF or SVG.
- Y-axis can be changed to illustrate absolute counts rather than alteration frequency.
- If interrogating many data sets, the minimum % altered samples and minimum total samples can be adjusted to show only the most pertinent data.
- Alteration types are color-coded
- Query can be modified to produce more/less results.

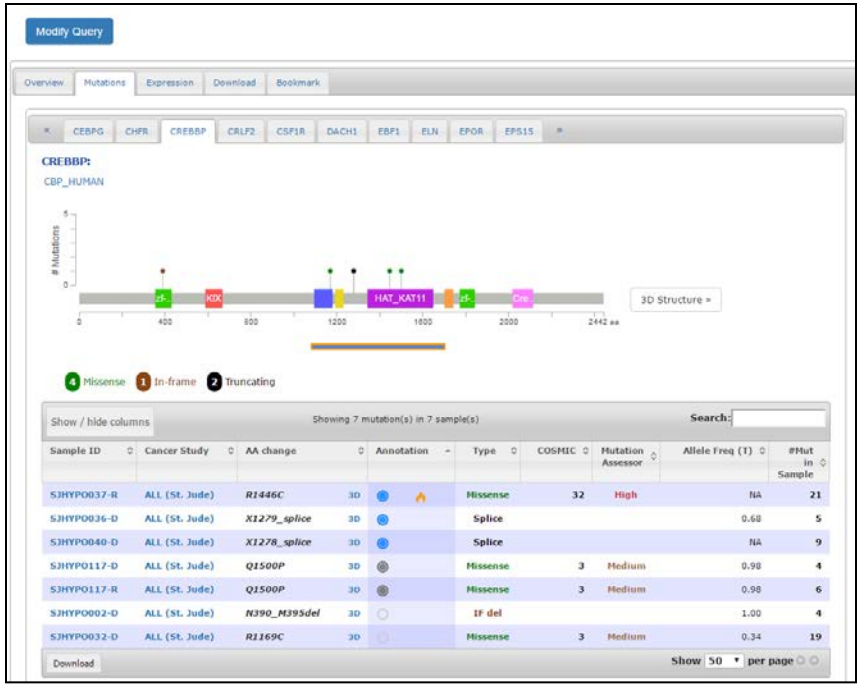


Query Results: Download and Bookmark

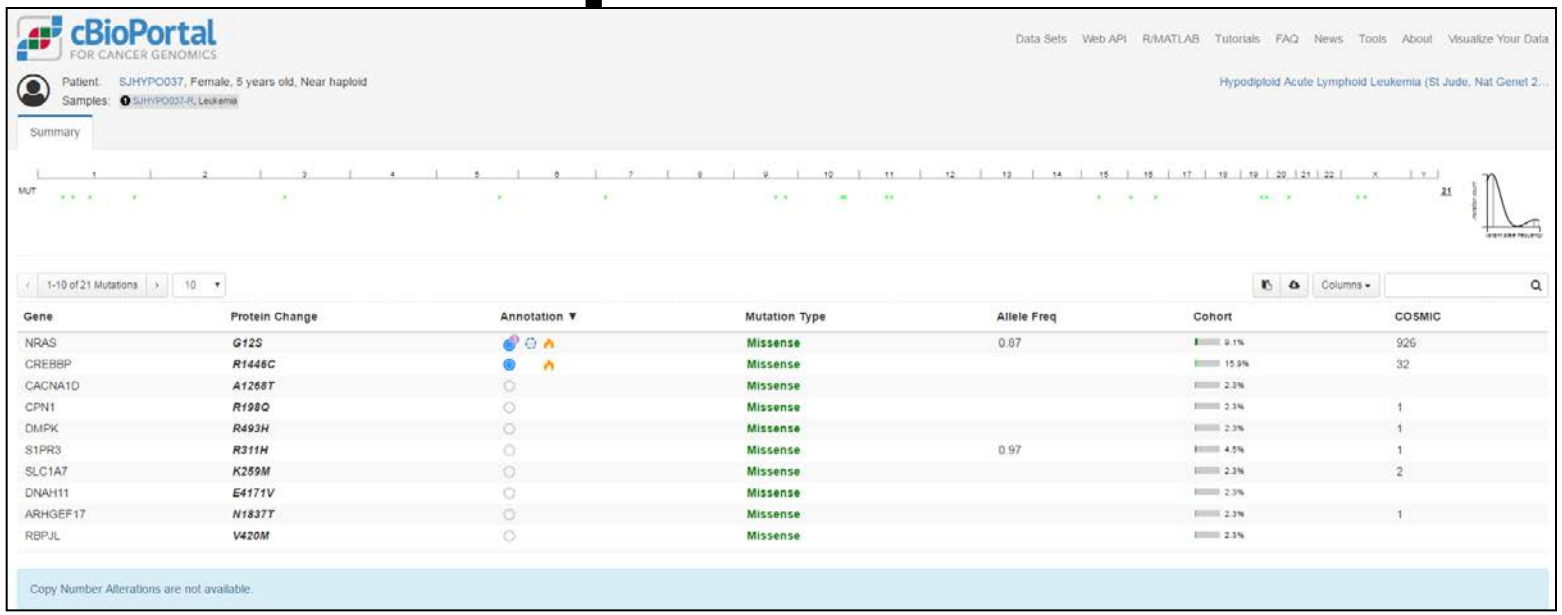
- Click on “Download” tab to copy and paste summary results into Excel.
- Click on “Bookmark” for links you can bookmark to save your query results

Query Results: Mutations

- Each gene in your query will have its own tab.
  - Within the tab will be
    - Gene schematic with visualization of domains and observed mutations
      - 3D protein structure visualization available
    - Table of all observed mutations with sample and mutation information.
      - Available for download
  - If no mutations found, tab will be empty



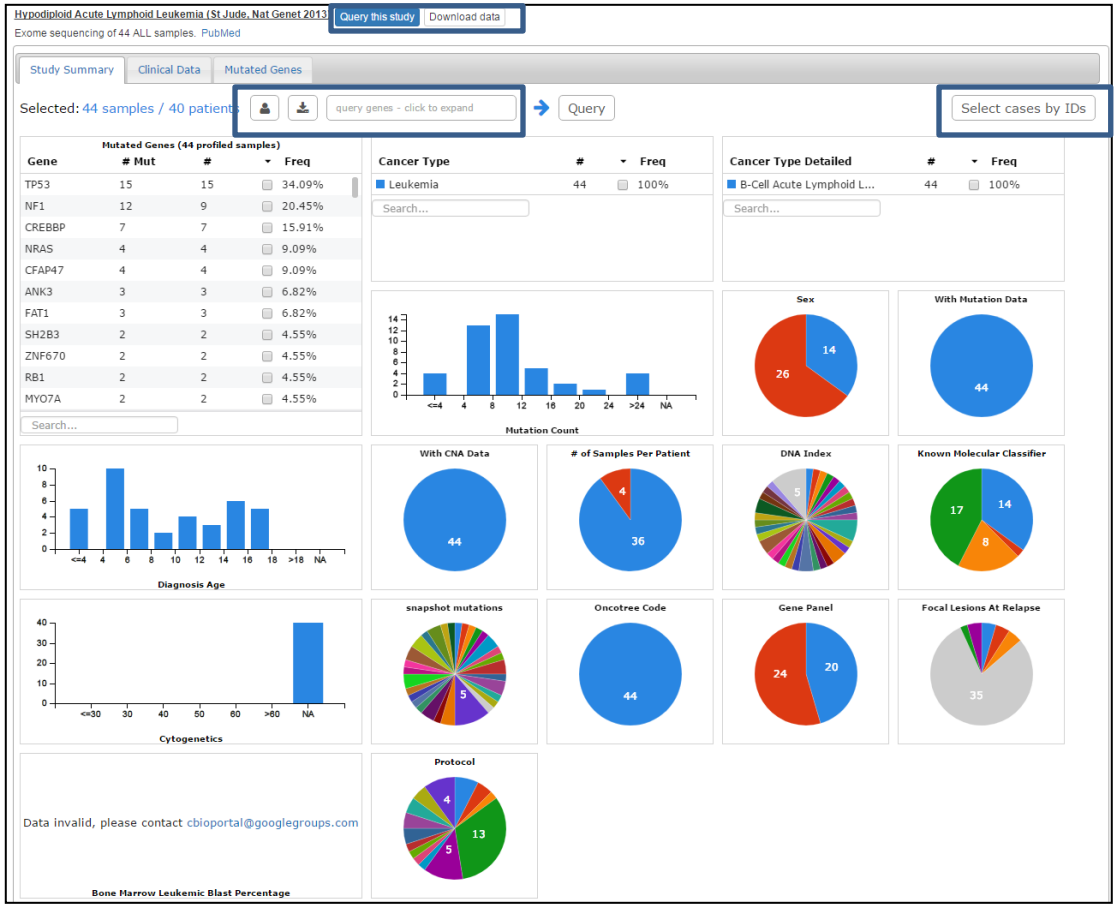
Mutation Information: Sample ID



- Click on Sample ID to see
  - Patient demographics
  - All mutations within a given sample by chromosomes and in a table (download available)
  - Mutation annotation – OncoKB, My Cancer Genome, Cancer HotSpots
  - % samples in study with same/similar mutations
  - COSMIC hits

Mutation Information: Cancer Study: Study Summary

- Can Filter by Any of the criteria on Study Summary Page
  - Mutated Genes: number and % frequency
  - Diagnosis Age
  - Cytogenetics – chromosome count
  - Cancer type – Broad, Detailed, or Oncotree
  - Sex
  - % and # Cases with Mutation data and CNA data
  - Molecular Classification
  - Gene Panel (if applicable)
  - Focal Lesions at Relapse
  - Snapshot Mutations
  - Treatment Protocol
  - DNA Index
  - Blast % (if applicable)
- Link to PubMed reference
- Can Download information
- Can Select Patients to view
- Can query specific gene(s) within study

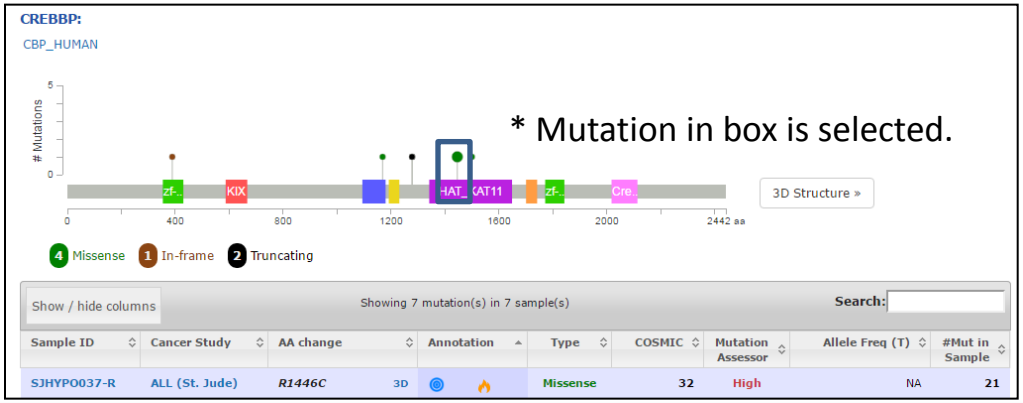


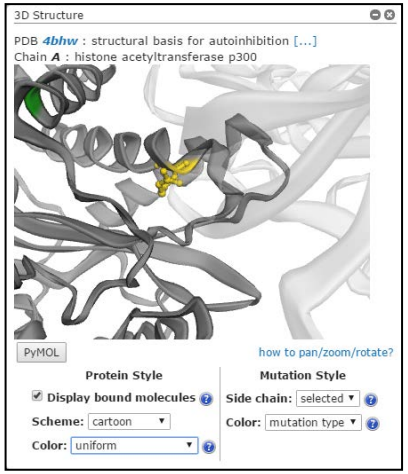
Mutation Information: Cancer Study: Clinical Data and Mutated Genes

- Clinical data table is available for each study in cBioPortal. Can filter by:
  - Patient or Sample ID
  - Blast %
  - Name of cancer study or cancer type
- Mutated genes table contains information about gene
  - Cytoband
  - Gene size (nts)
  - # mutations observed in study
  - # mutations per nucleotide

Mutation Information: AA Change

- Click on this row to highlight mutation marker on gene schematic.
- Click on 3D link to view mutation in context of protein structure.
  - Several visualization options for protein structure.



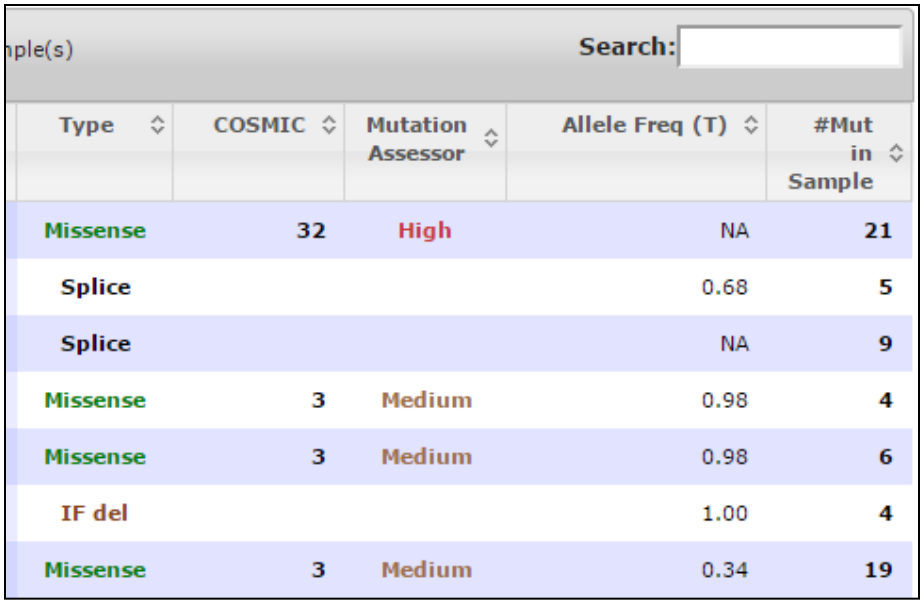


Mutation Information: Annotation

- Click on Blue/Gray Target for OncoKB annotations 
  - Numbers attached to target indicates therapeutic information available – See levels of evidence in OncoKB presentation
  - Citations included in pop-up window
- Click on My Cancer Genome symbol for MCG links 
- Flame icon indicates recurrent cancer hotspot (from http://cancerhotspots.org/) 

Other Mutation Information

- Mutation tab will contain
  - Type of mutation
  - # COSMIC entries
  - Mutation Assessor
    - Predicts functional impact
    - Link to http://mutationassessor.org
    - Allele Frequency (mutated reads/total reads)
  - # Mutations in sample



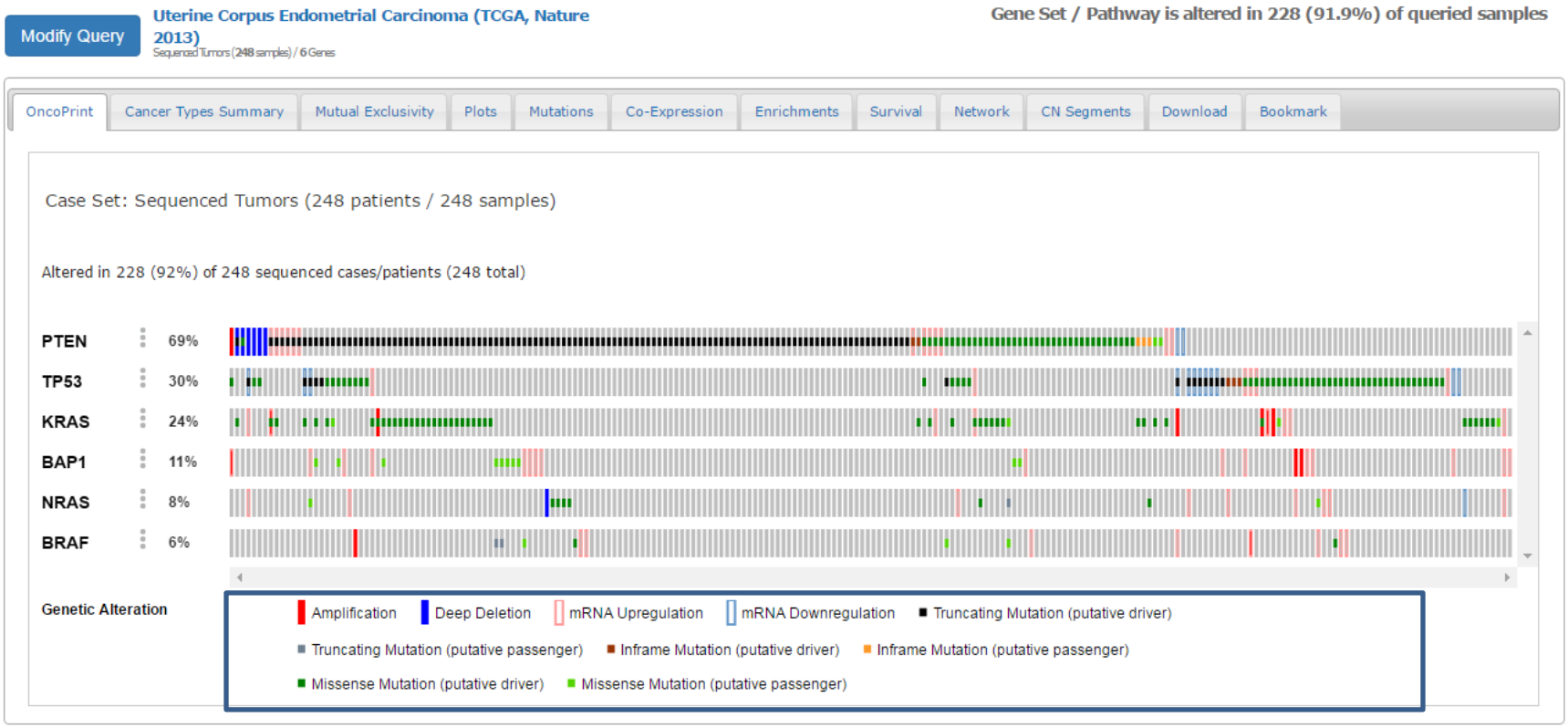
ple(s) Search:

Type	COSMIC	Mutation Assessor	Allele Freq (T)	#Mut in Sample
Missense	32	High	NA	21
Splice			0.68	5
Splice			NA	9
Missense	3	Medium	0.98	4
Missense	3	Medium	0.98	6
IF del			1.00	4
Missense	3	Medium	0.34	19

Expression Data

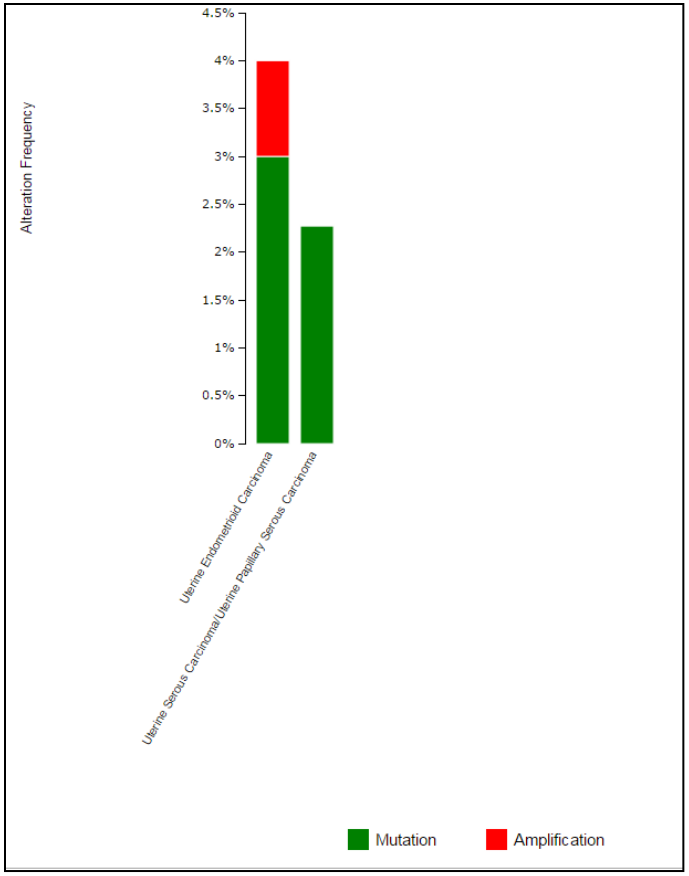
- Can be used to provide evidence for mutations affecting expression levels.
- Can view one gene at a time
- Data is downloadable

OncoPrint as Query Result



Cancer Types Summary

- If study contains multiple cancer types, data will stratify mutation by cancer type per gene in this tab resulting from query.



Mutual Exclusivity

- Tab in query result will show which gene pairs have mutually exclusive mutations and which gene pairs have co-occurring mutations
- P-value of less than 0.05 = significant
- Results are downloadable

The query contains **6** gene pairs with mutually exclusive alterations (**1** significant), and **9** gene pairs with co-occurrent alterations (**2** significant).

☑ Mutual exclusivity ☑ Co-occurrence ☐ Significant pairs **Search Gene**

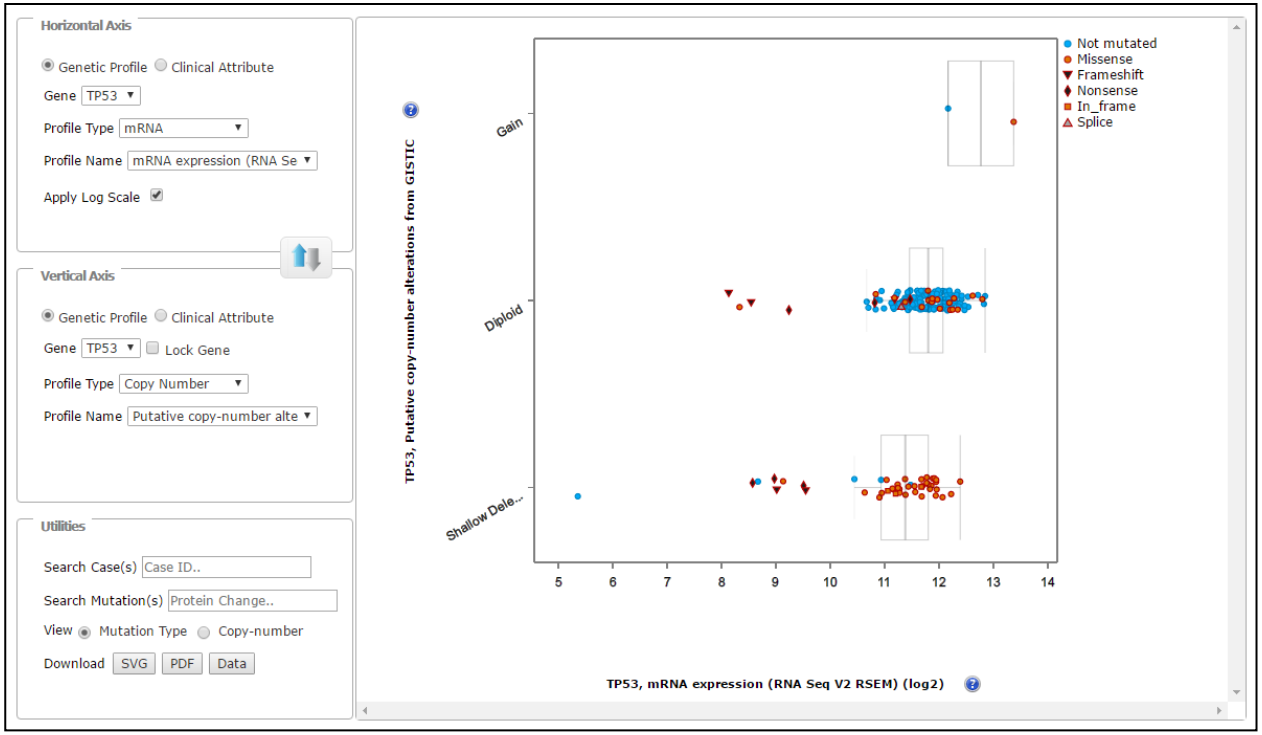
Gene A	Gene B	p-Value	Log Odds Ratio	Association
PTEN	TP53	**<0.001**	-2.282	Tendency towards mutual exclusivity **Significant**
TP53	BAP1	**0.038**	0.815	Tendency towards co-occurrence **Significant**
NRAS	TP53	**0.040**	0.941	Tendency towards co-occurrence **Significant**
BRAF	TP53	0.066	0.922	Tendency towards co-occurrence
KRAS	PTEN	0.080	0.528	Tendency towards co-occurrence
KRAS	BAP1	0.142	-0.717	Tendency towards mutual exclusivity
BRAF	KRAS	0.209	-0.847	Tendency towards mutual exclusivity
KRAS	TP53	0.219	-0.316	Tendency towards mutual exclusivity
BRAF	BAP1	0.265	0.648	Tendency towards co-occurrence
PTEN	NRAS	0.267	-0.409	Tendency towards mutual exclusivity
BRAF	NRAS	0.377	0.530	Tendency towards co-occurrence
NRAS	BAP1	0.398	0.360	Tendency towards co-occurrence
PTEN	BAP1	0.544	-0.036	Tendency towards mutual exclusivity
KRAS	NRAS	0.557	0.047	Tendency towards co-occurrence
BRAF	PTEN	0.613	0.010	Tendency towards co-occurrence

Showing 1 to 15 of 15 entries

Download Full Result

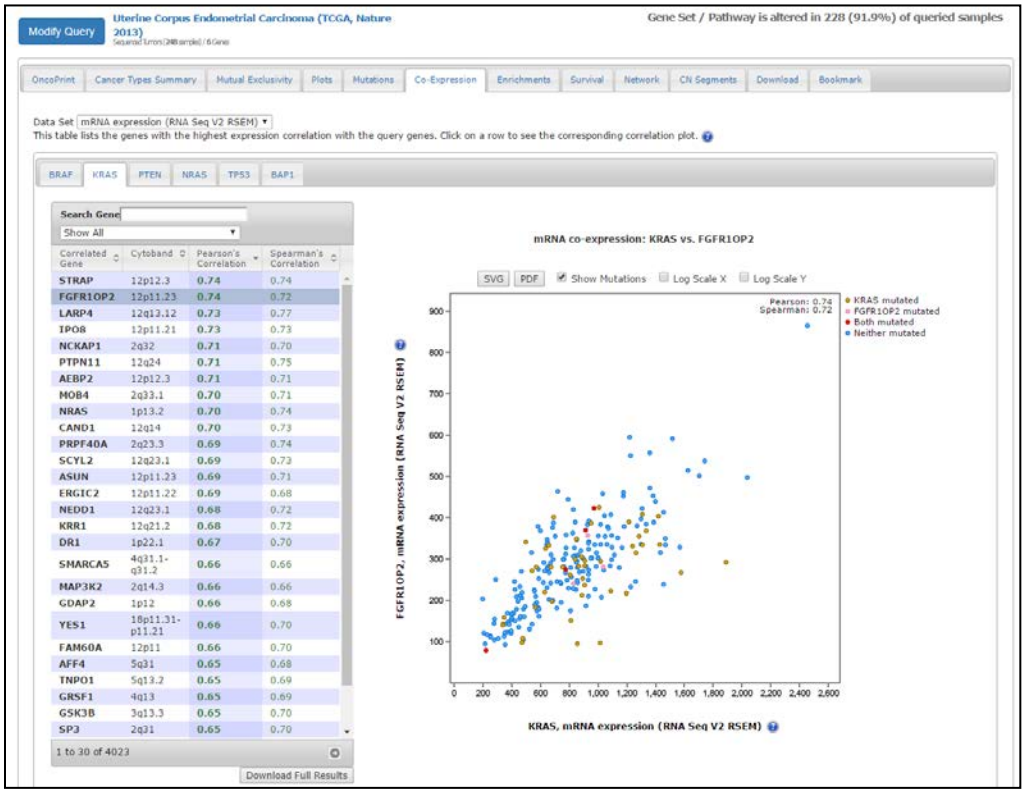
Plots

- Can choose data type and gene for both horizontal and vertical axes.
- Can also search case IDs if you are curious where data will lie in context of cohort
- Plots are available for download



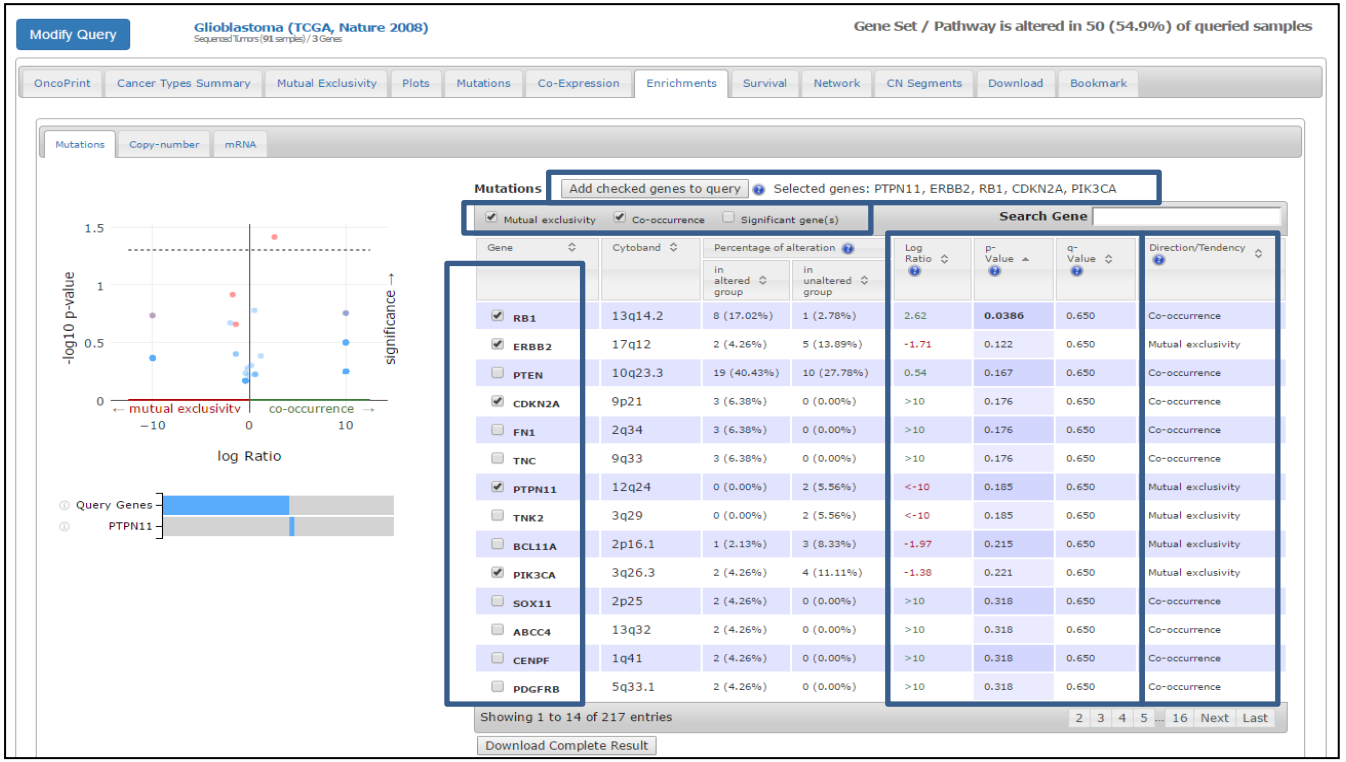
Co-Expression

- Separate tab for each gene in query.
- Can see co-expression data for many genes.
  - Use ‘Search Gene’ function or sort columns to find desired target.
  - Can download full results per gene



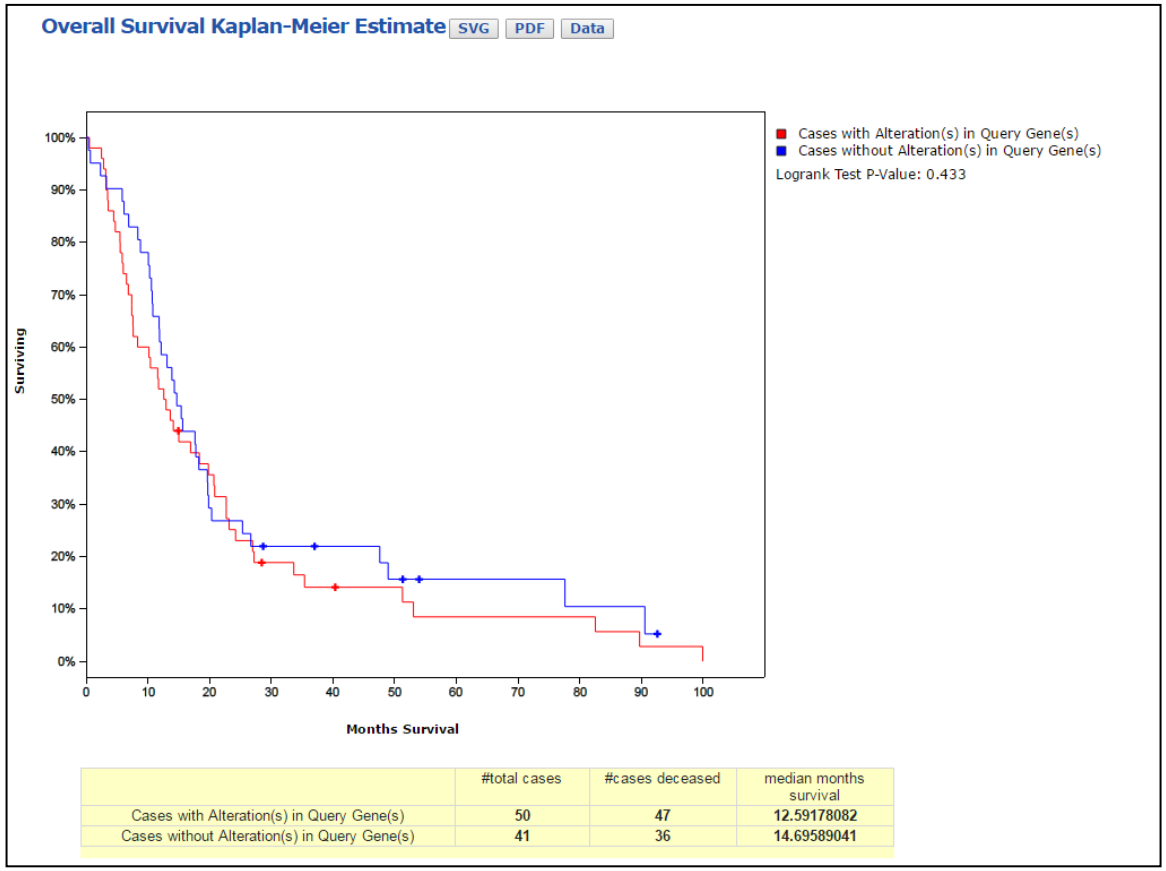
Enrichments

- Separate tabs for identifying mutations, copy number, and changes in mRNA/microRNA expression that either co-occur or are mutually exclusive to your query.



Survival

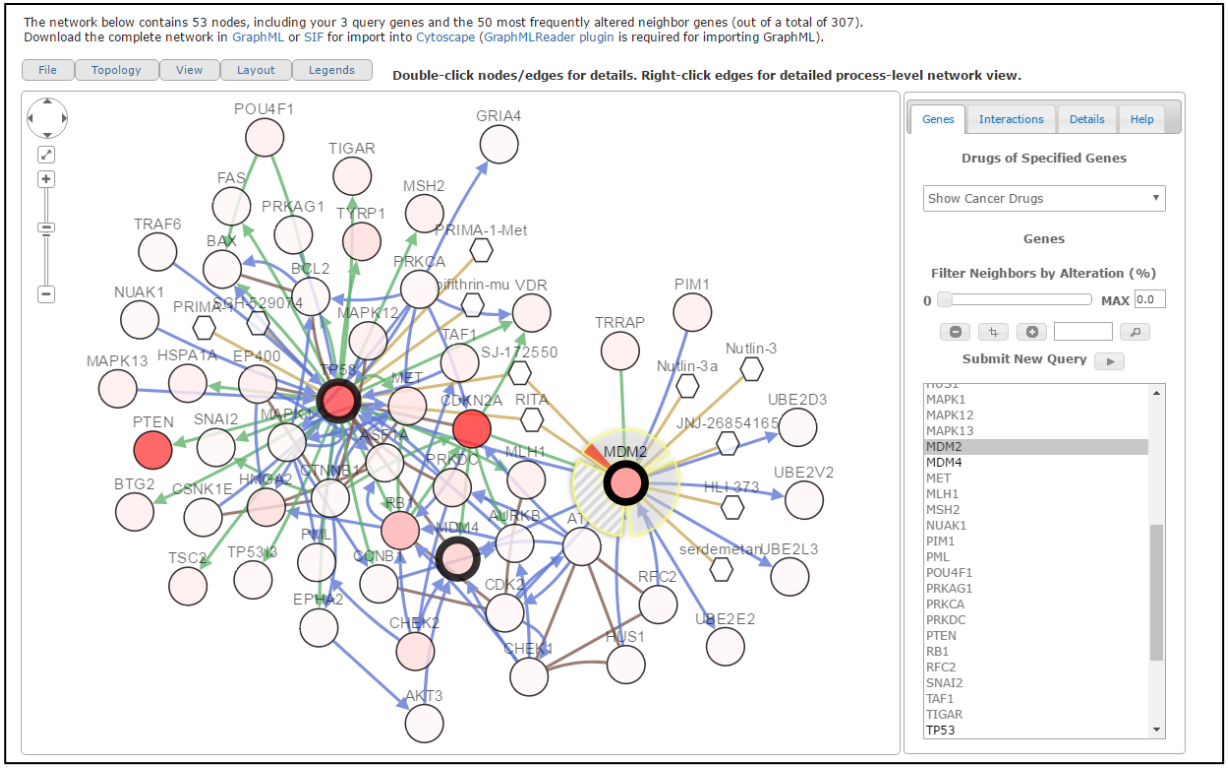
- If the query selected contains studies with survival information, Kaplan-Meier Estimates for disease free survival and overall survival are displayed.



	#total cases	#cases deceased	median months survival
Cases with Alteration(s) in Query Gene(s)	50	47	12.59178082
Cases without Alteration(s) in Query Gene(s)	41	36	14.69589041

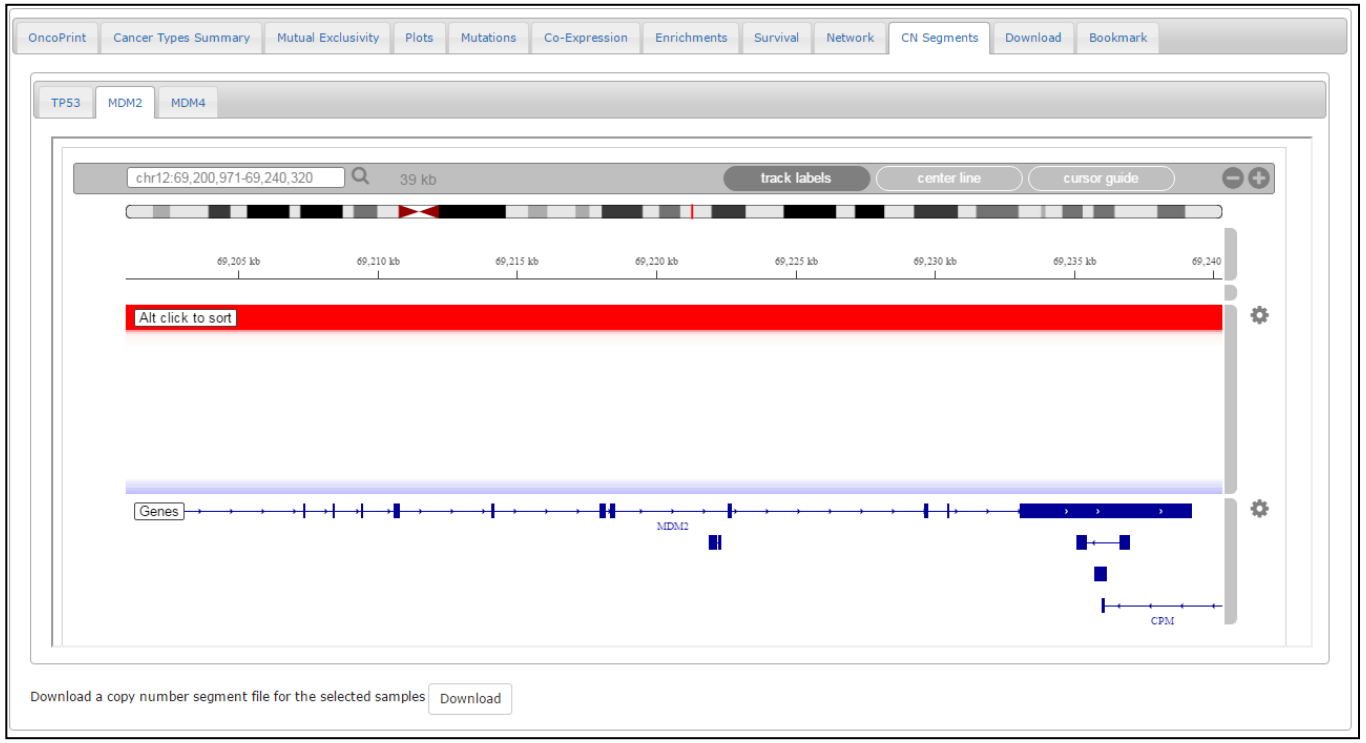
Network

- Will display your queried genes and their 50 most frequently altered neighbors.
- Can alter image by filtering neighbors by alteration % on Gene subtab on right panel and buttons along top of network window



CN Segments

- Will show Copy number data on all genes in your query within your selected studies.
  - Each gene is on a separate tab.
    - Red = CN Gain
    - Blue = CN Loss



OncoPrinter Visualization Tool

- This tool can be used to visualize your own data.
- http://www.cbioportal.org/oncoprinter.jsp
- Click on link above and view link for "Data Format" (or see next page) and Format a tab-delimited file to the specifications listed.
- Copy and paste data into "Input Mutation Data" window or "Choose File" to select tab-delimited file from saved location.
- If desired, order the genes and/or samples as you would like them to be shown on the graph.

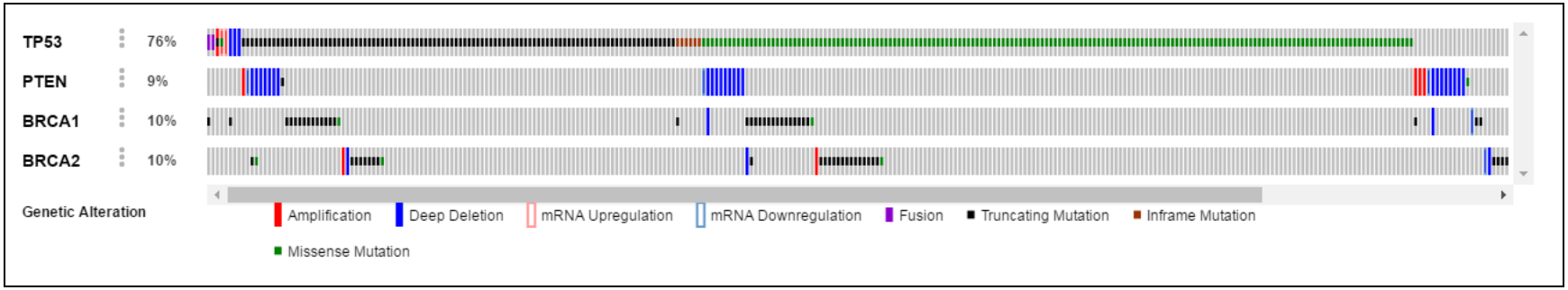
OncoPrinter: Data Format

Data format

The data should contain three tab-delimited columns. The first row is a header row, which contains: `Sample Gene Alteration Type`. Each following row contains a single genomic event in a single sample. You can also list samples without any events at the end of the list so that the percentages can be properly calculated. Note: Any row which has an entry in the Gene (2nd) column must also have an entry in the Alteration (3rd) and Type (4th) columns

1. Sample: Sample ID
2. Gene: Gene symbol (or other gene identifier)
3. Alteration: Definition of the alteration event
    - Mutation event: amino acid change or any other information about the mutation
    - Fusion event: fusion information
    - Copy number alteration (CNA) - please use one of the four events below:
        - AMP: high level amplification
        - GAIN: low level gain
        - HETLOSS: shallow deletion
        - HOMDEL: deep deletion
    - mRNA expression - please use one of the two events below:
        - UP: expression up
        - DOWN: expression down
4. Type: Definition of the alteration type. It has to be one of the following.
    - For a mutation event, please use one of the three mutation types below:
        - MISSENSE: a missense mutation
        - INFRAME: a inframe mutation
        - TRUNC: a truncation mutation
    - FUSION: a fusion event
    - CNA: a copy number alteration event
    - EXP: a expression event

OncoPrinter Results



- Percent mutated samples per gene displayed adjacent to table
- Genetics alterations are color-coded
- Options for visualization appear in a menu near upper right-hand side of graph when you hover over graph
  - Color-code mutations and sort by type OR show all mutations with same color
  - Hide unaltered cases
  - Remove white spaces between columns
  - Zoom in/out – by buttons or slider bar
  - Zoom to fit altered cases in screen
  - Downloads: PDF, PNG, SVG, or .txt file with order of patients in graph

MutationMapper

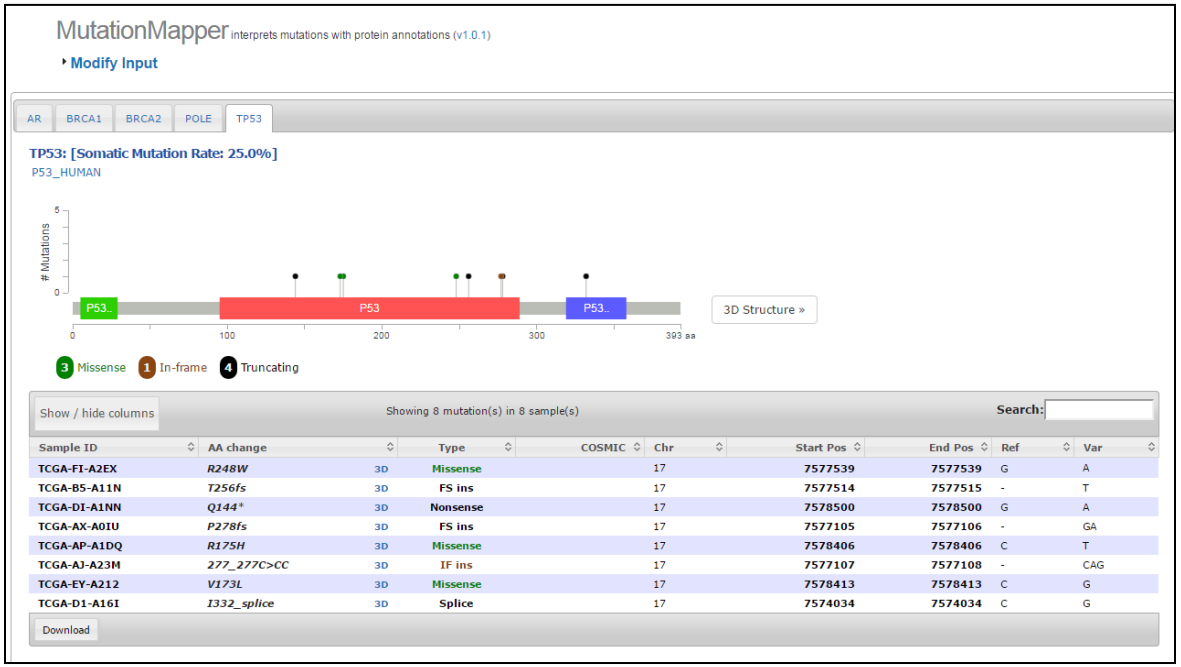
- Interprets mutations with protein annotations
- Input tab-delimited file needs at least two columns
  - HUGO gene name
  - Protein change

List of valid input headers:

Column Header	Description	Example
Hugo_Symbol	HUGO symbol for the gene	TP53
Protein_Change	Amino acid change	V600E
Sample_ID	Tumor sample ID	TCGA-B5-A11E
Mutation_Type	Translational effect of variant allele	Missense_Mutation, Nonsense_Mutation, etc.
Chromosome	Chromosome number	X, Y, M, 1, 2, etc.
Start_Position	Lowest numeric position of the reported variant on the genomic reference sequence	666
End_Position	Highest numeric position of the reported variant on the genomic reference sequence	667
Reference_Allele	The plus strand reference allele at this position	A
Variant_Allele	Tumor sequencing (discovery) allele	C
Validation_Status	Second pass results from orthogonal technology	Valid
Mutation_Status	Mutation status	Somatic, Germline, etc.
Center	Center/Institute reporting the variant	mskcc.org

MutationMapper: Resulting Data

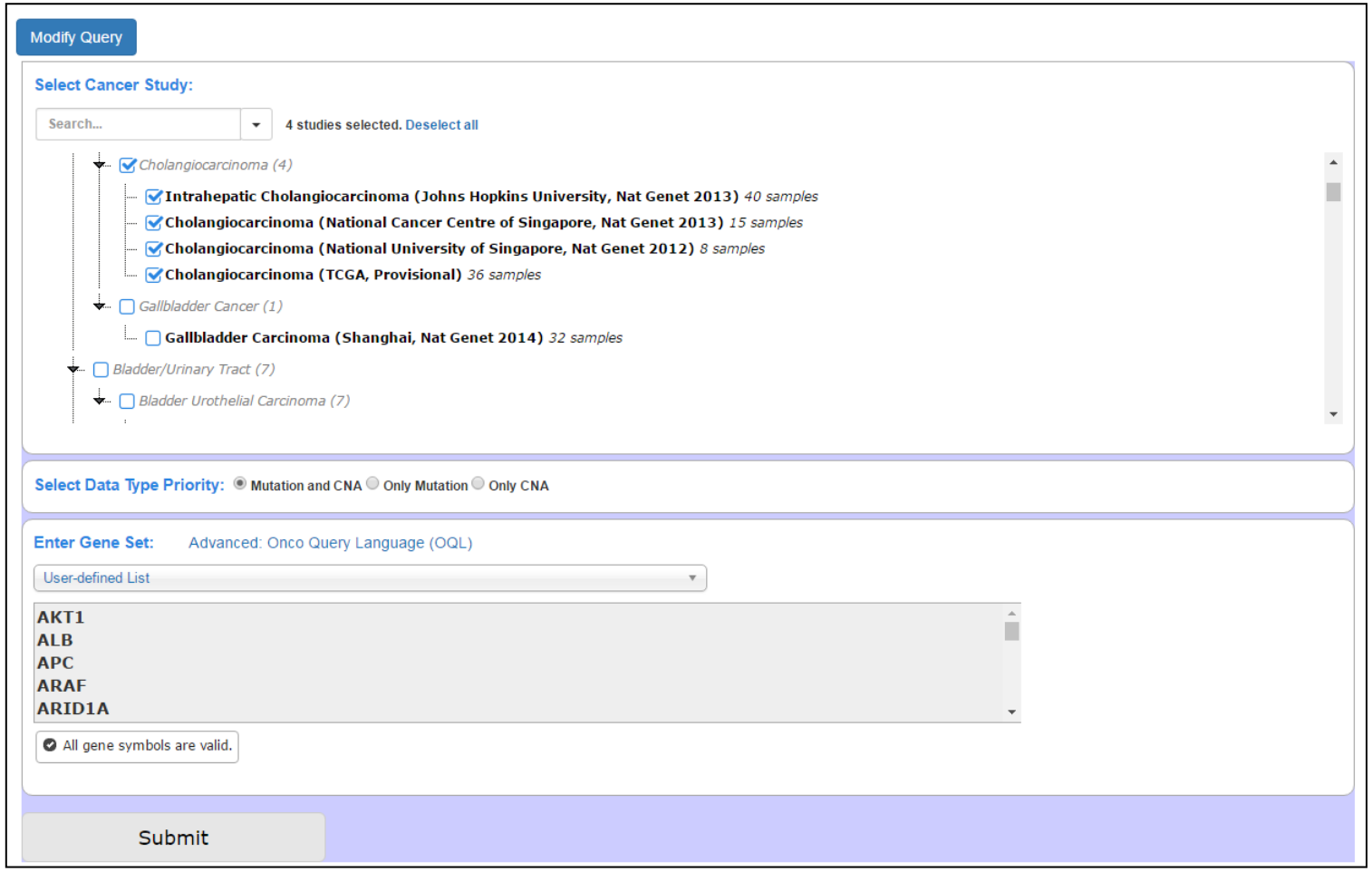
- Similar to Mutation window in Query results
  - Does not have annotations like query results
- Can modify input data
- Table can be downloaded as TSV



Scenario #1

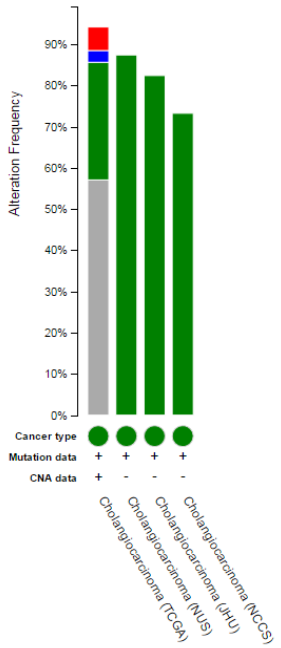
- You would like to validate one of your lab's disease specific gene lists against the data in cBioPortal.
  - Run Query by selecting relevant studies and enter all genes in Gene Set window.

Design Query



View Overview

- Is your gene list sensitive enough to have a high number and percentage of alterations?
  - Qualitatively, yes.



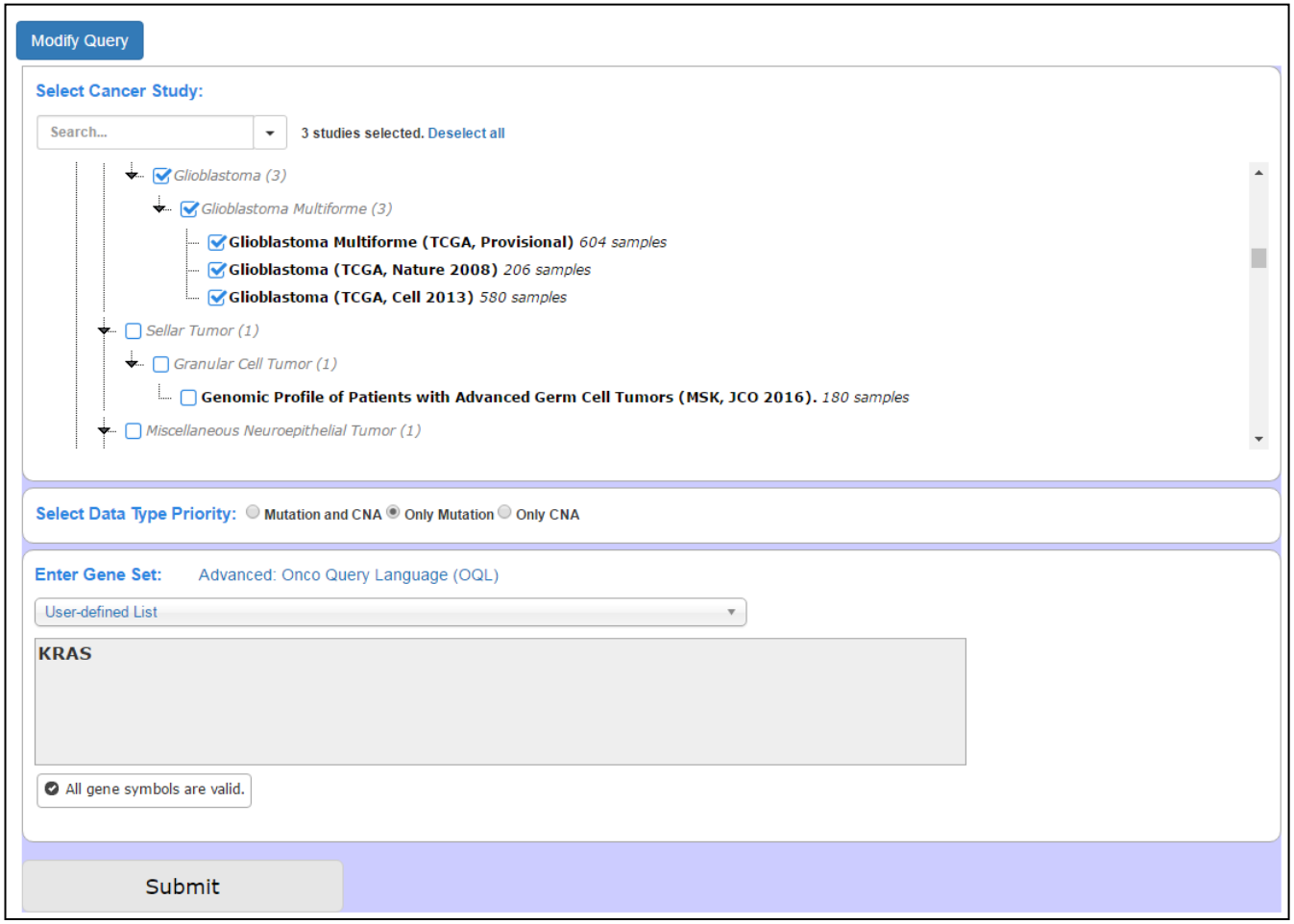
View Mutations

- How many genes have observed mutations out of your gene list?
  - 41/62 genes have hits and are validated through cBioPortal. Look through literature or other knowledgebase resources to validate (or not validate) the other 21.

Scenario #2

- You are studying KRAS mutation frequency and co-mutations in glioblastoma.
  - What genes are commonly co-mutated in KRAS mutant GBM?

Query KRAS mutations in GBM



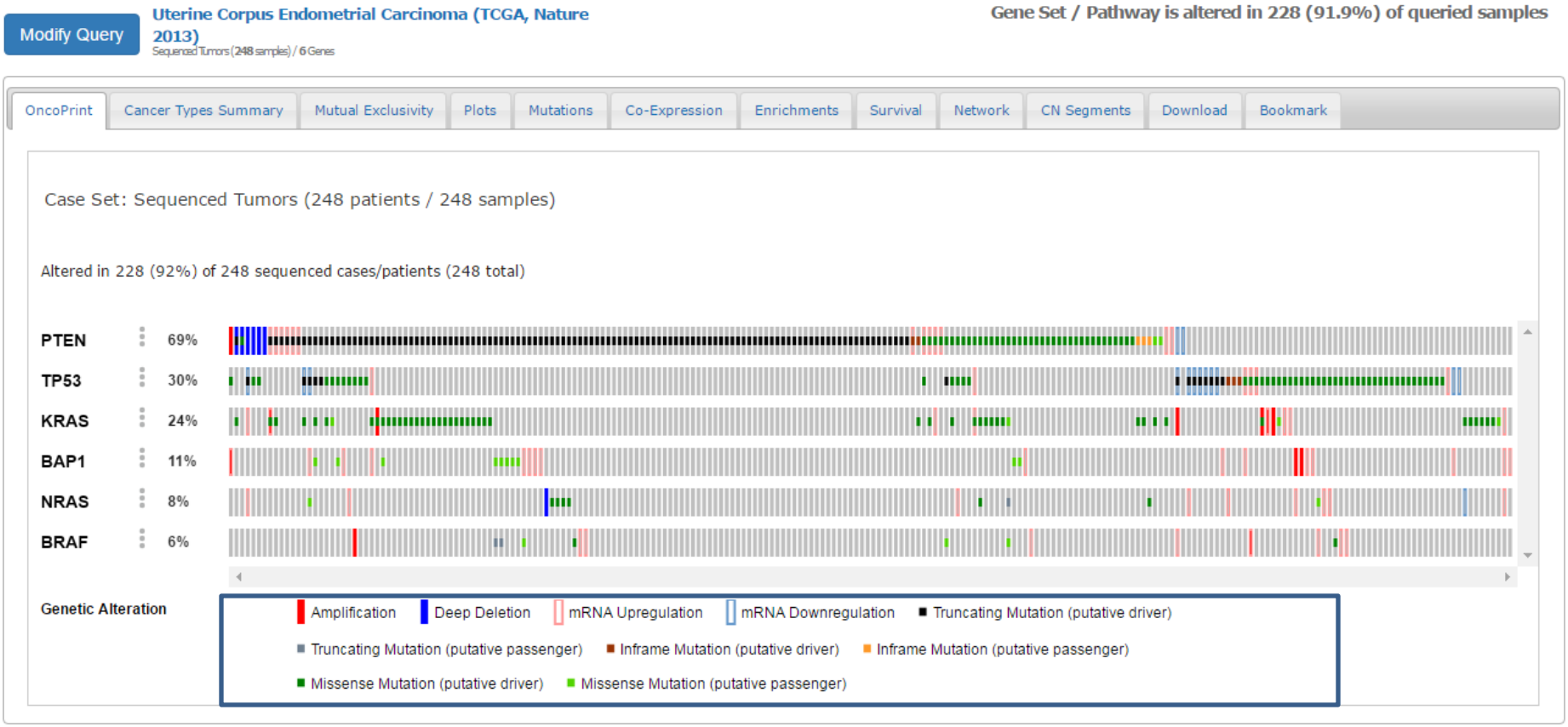
Go to Mutations Page and Click on Sample ID in list

- Repeat for all patients
- If necessary, download CSV files for CNAs and/or mutations
- Identify recurring mutations in cohort besides KRAS
  - 4/5 GBM patients with KRAS mutation also have PTEN mutation
  - 4/5 GBM patients with KRAS mutation also have MDM4 amplification, CDKN2A and CDKN2B deep deletions

Scenario #2B

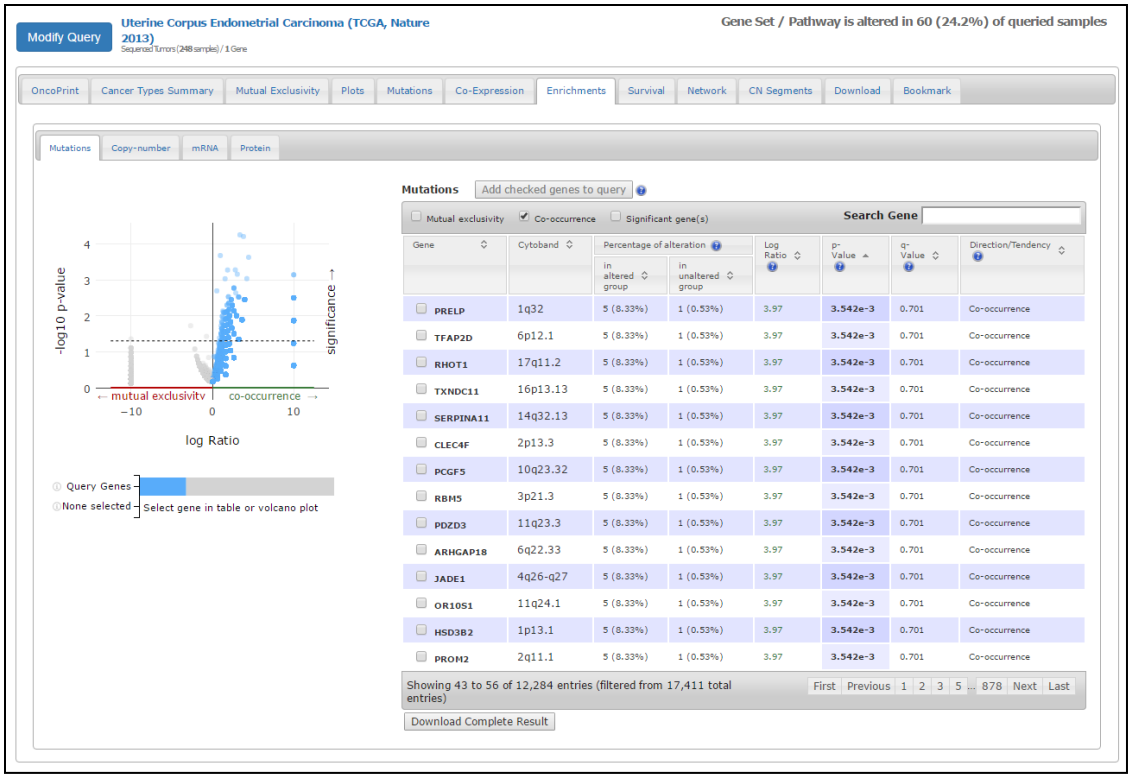
- Same scenario – instead of GBM, look for co-mutations in uterine endometrial carcinoma via query.
  - OncoPrint as Query Result
  - Enrichments using Modified Query

OncoPrint as Query Resu



Enrichments

- Modify query to only include gene of interest (KRAS)
- Click on Enrichments tab in Query Results
- Make sure you are on Mutations sub-tab within Enrichments tab.
- Download complete result by pressing button at bottom of window.
  - Determine p-value to use for data acceptance.
  - Can also toggle through data points on log Ratio graphs on left side of window.



Scenario #3

- You would like to contribute to cBioPortal efforts.
  - cBioPortal accepts coding contributions
    - https://github.com/cBioPortal/cbioportal/blob/master/CONTRIBUTING.md
    - See link above to learn how to contribute through GitHub.

Contacts

- Twitter Handle - @cbioportal
- cbioportal@googlegroups.com