MAYO CLINIC

Cancer Informatics Lecture

Mayo-UIUC Computational Genomics Course

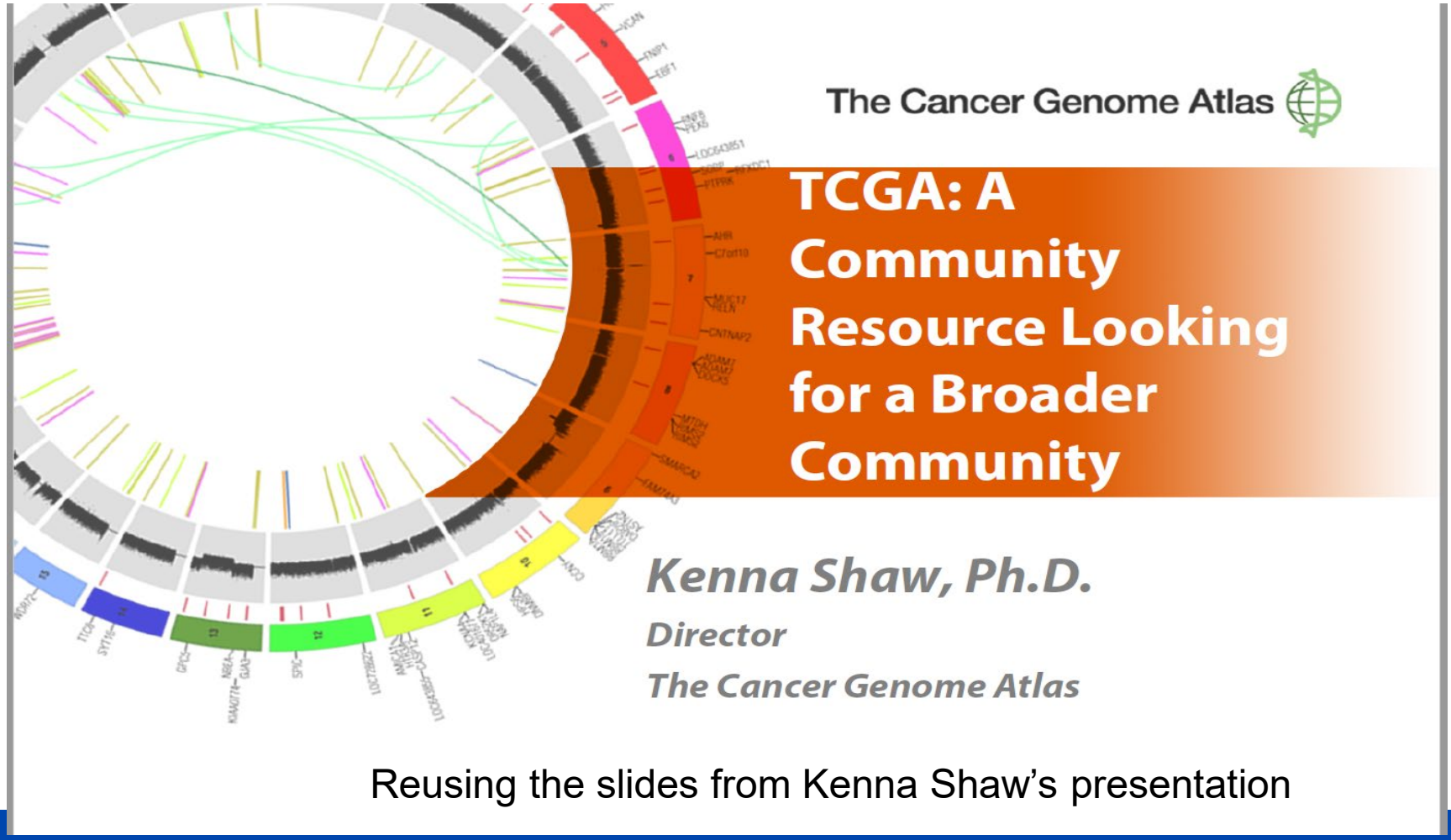
June 12, 2020

Krishna Rani Kalari Ph.D.
Associate Professor

Outline

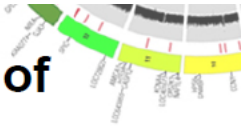
- The Cancer Genome Atlas (TCGA)
- Genomic Data Commons (GDC)
- COSMIC database (mutations database)
- cBioPortal for cancer genomics
- GTEX
- Precision Medicine in Cancer
- Single-cell RNA-Seq
- Proteomics

The Cancer Genome Atlas (TCGA)



Reusing the slides from Kenna Shaw's presentation

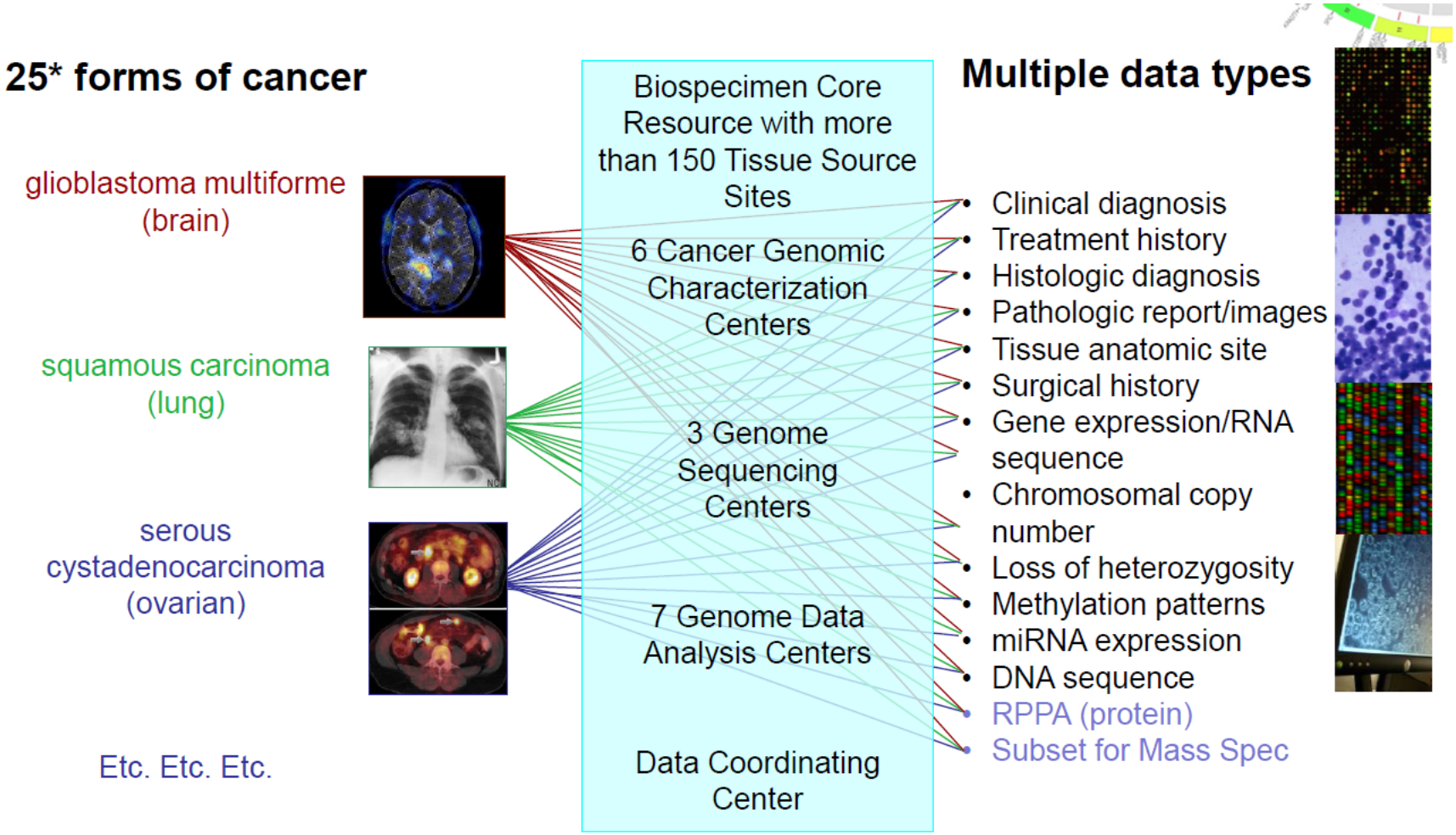
TCGA core objectives



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

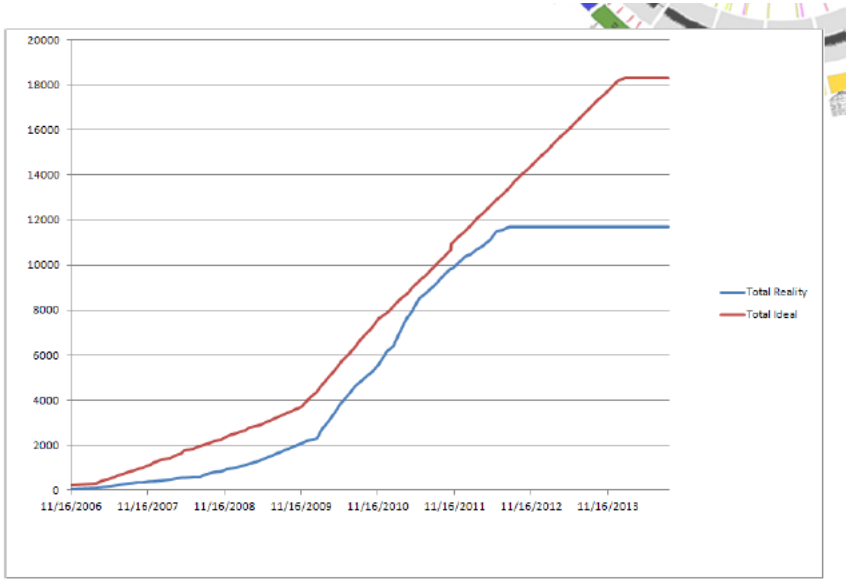
- Establish infrastructure for effective team science
- Develop a scalable "pipeline" beginning with highest quality samples
- Determine the feasibility of a large-scale, high throughput approach to identifying the molecular 'parts-list'
- Evaluate using statistically-robust sample sets
- Make the data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

TCGA multiple data types



Rare tumors projects initiated in 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??

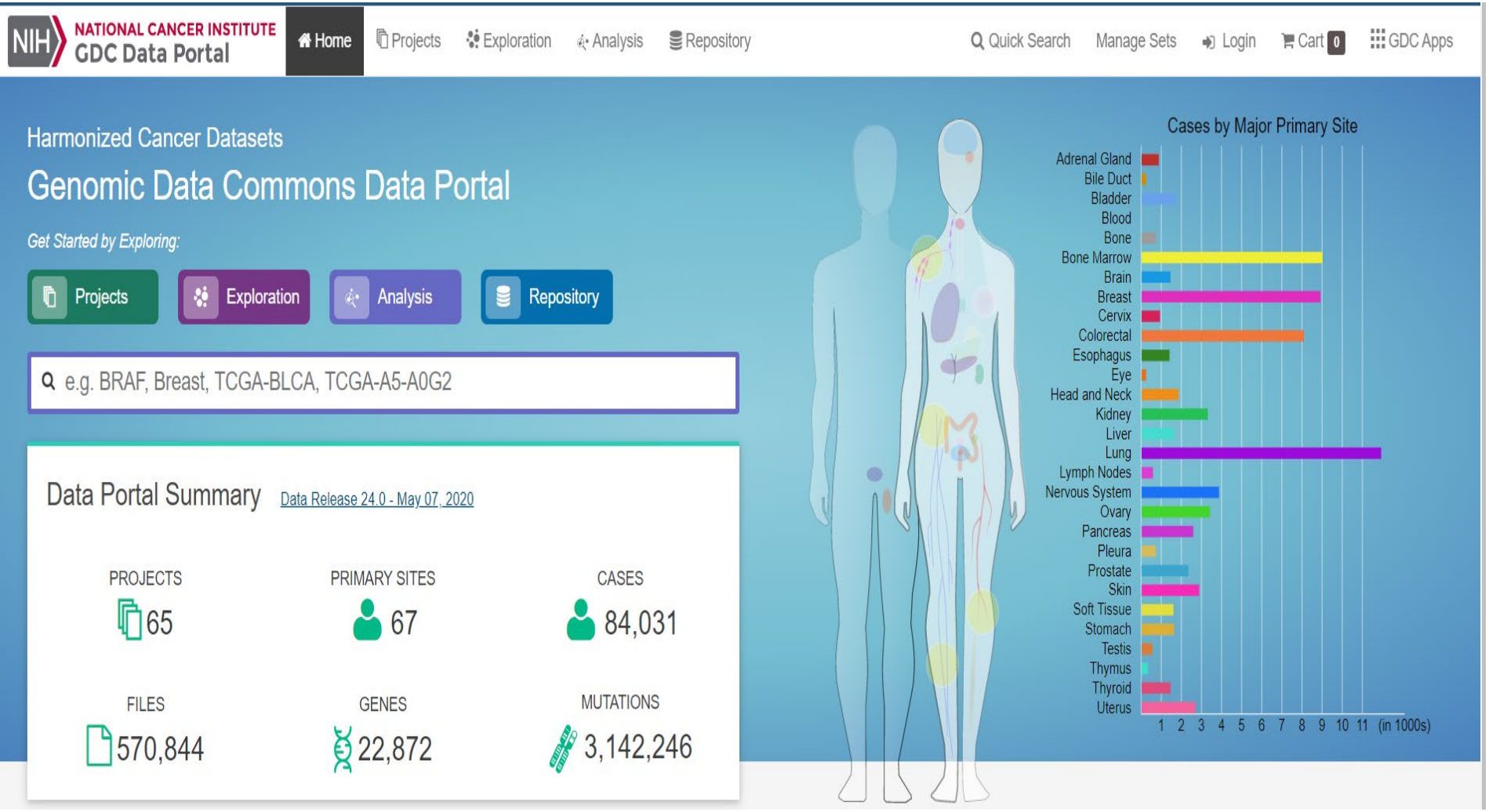


Genomic Data Commons - GDC

Genomic Data Commons

- A NCI repository for The Cancer Genome Atlas and Genomics data.
- It consists of data from 68 primary sites
- >33,605 cases
- >Three million mutations
- 374,699 files

Genomic data commons



Projects

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home | Projects | Exploration | Analysis | Repository | Quick Search | Manage Sets | Login | Cart 0 | GDC Apps

← Start searching by selecting a facet

Project

e.g. TCGA-GBM, Brain

Primary Site

- hematopoietic and reticuloendothelial syst… 18
- kidney 17
- bronchus and lung 15
- colon 15
- heart, mediastinum, and pleura 14

62 More...

Program

- TCGA 33
- TARGET 9
- GENIE 8
- BEATAML1.0 2
- CGCI 2

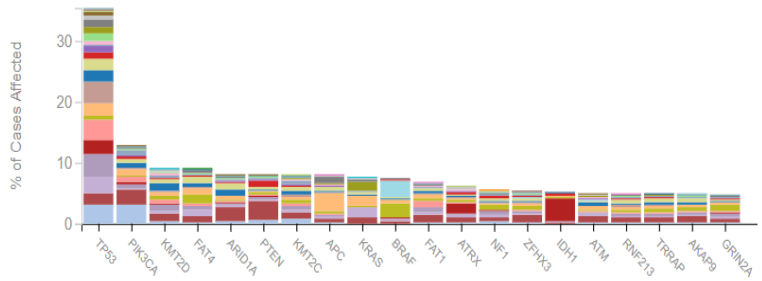
10 More...

Disease Type

- adenomas and adenocarcinomas 33
- cystic, mucinous and serous neoplasms 21

Top Mutated Cancer Genes in Selected Projects

% of Cases Affected | # of Cases Affected



Case Distribution per Project

84,031 Cases across 65 Projects

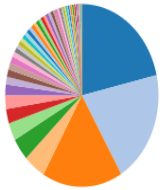


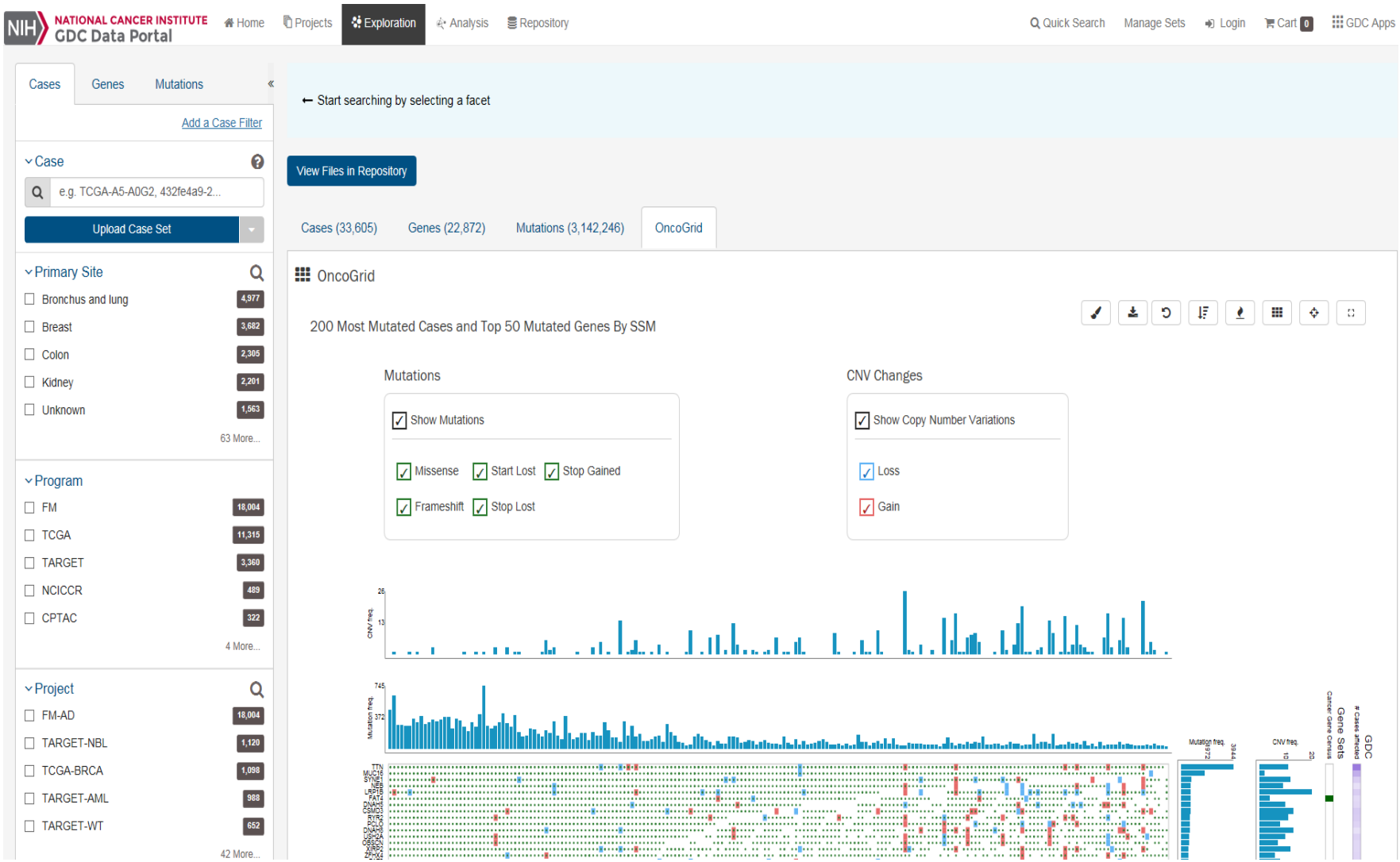
Table | Graph

65 Projects

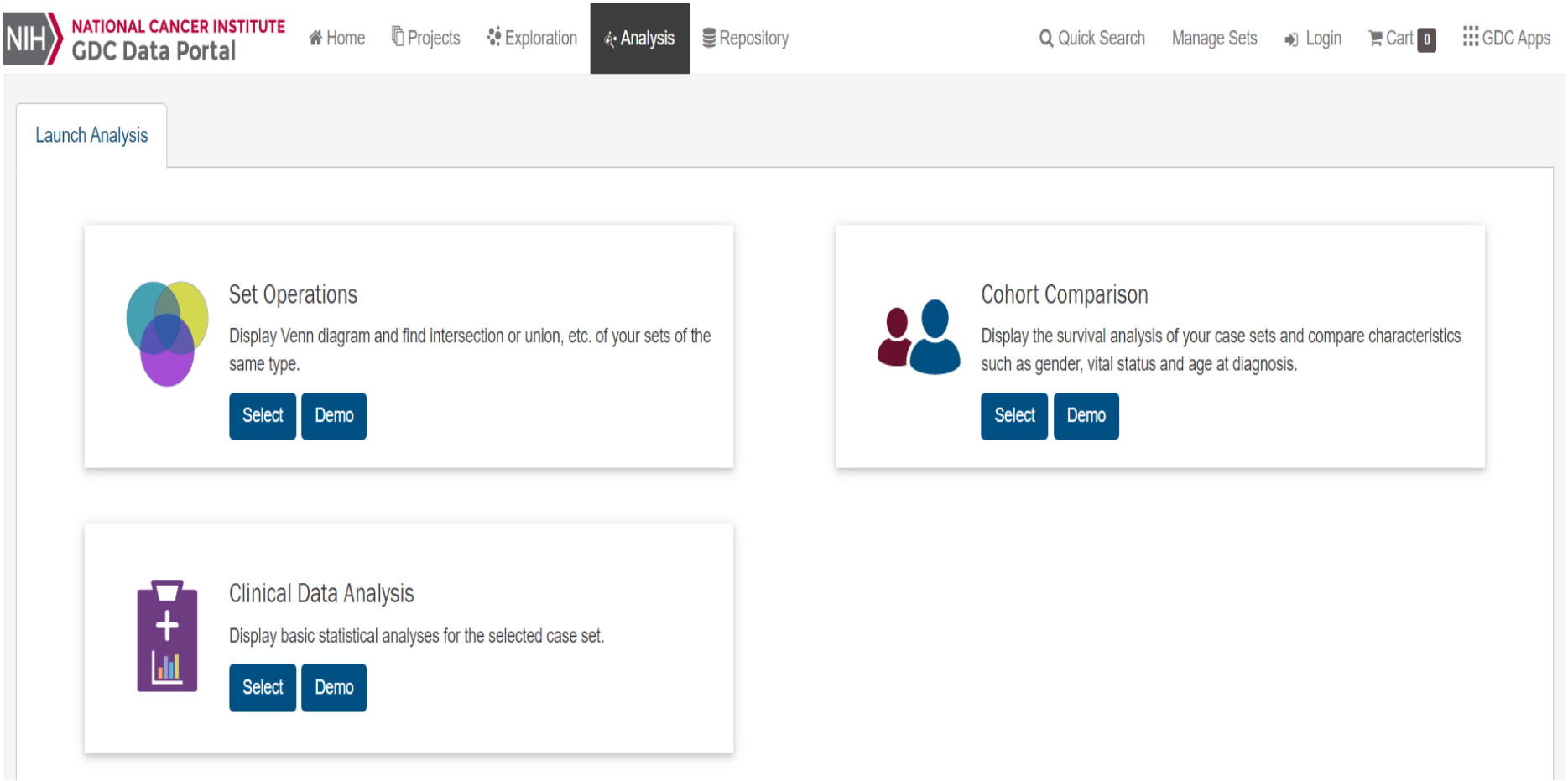
JSON | TSV

Project	Disease Type	Primary Site	Program	Cases	Available Cases per Data Category: Seq	Exp	SNV	CNV	Meth	Clinical	Clinical Supplement	Bio	Bio Supplement	Files
FM-AD	23 Disease Types	42 Primary Sites	FM	18,004	0	0	18,004	0	0	18,004	18,004	18,004	18,004	36,134
GENIE-MSK	49 Disease	49 Primary	GENIE	16,824	0	0	16,823	16,823	0	16,824	0	16,824	0	36,470

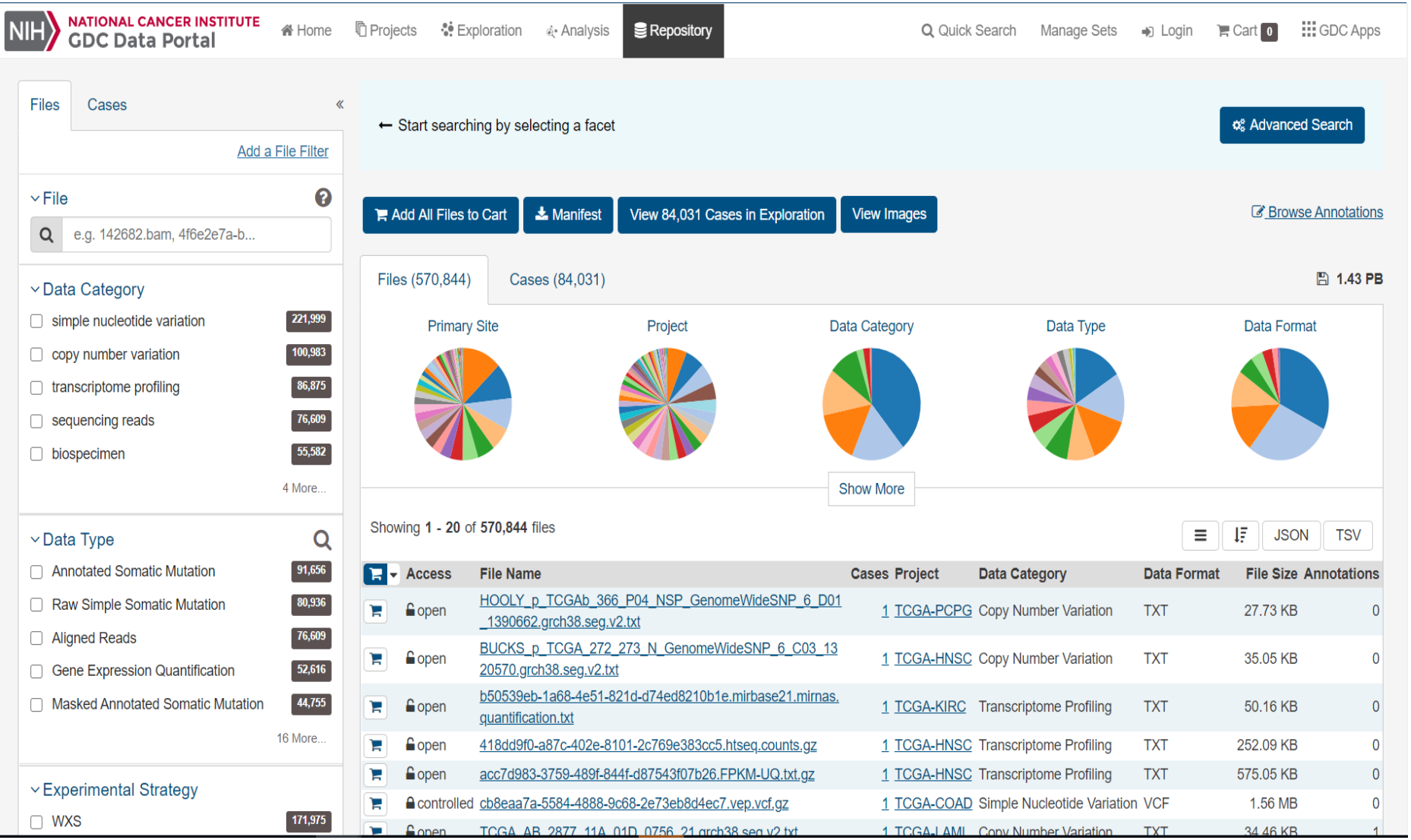
Exploration



Analysis



Repository

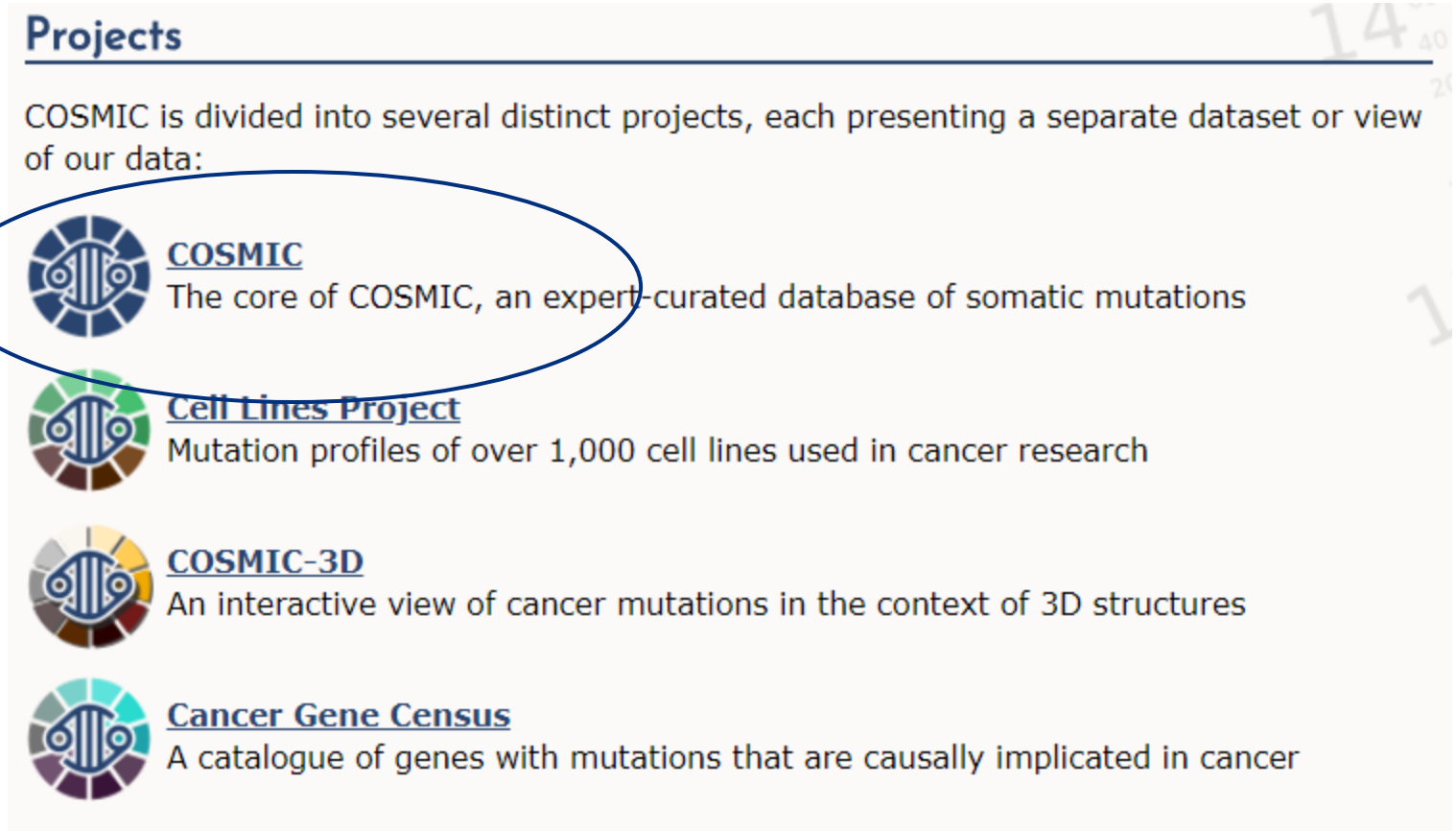


MAYO CLINIC

COSMIC

“COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.”

Expert curated database



COSMIC
Catalogue Of Somatic Mutations In Cancer

Projects | Data | Tools | News | Help | About | Genome Version | Search COSMIC... SEARCH

Gene

KRAS

- Gene view
- Overview
- External links
- Drug resistance
- Tissue distribution
- Genome browser
- Mutation distribution
- Variants
- References

Reset page

Search

Search COSMIC...

Filters

Show advanced filters

Range Show input fields

1 — 190

1 48 96 143 190

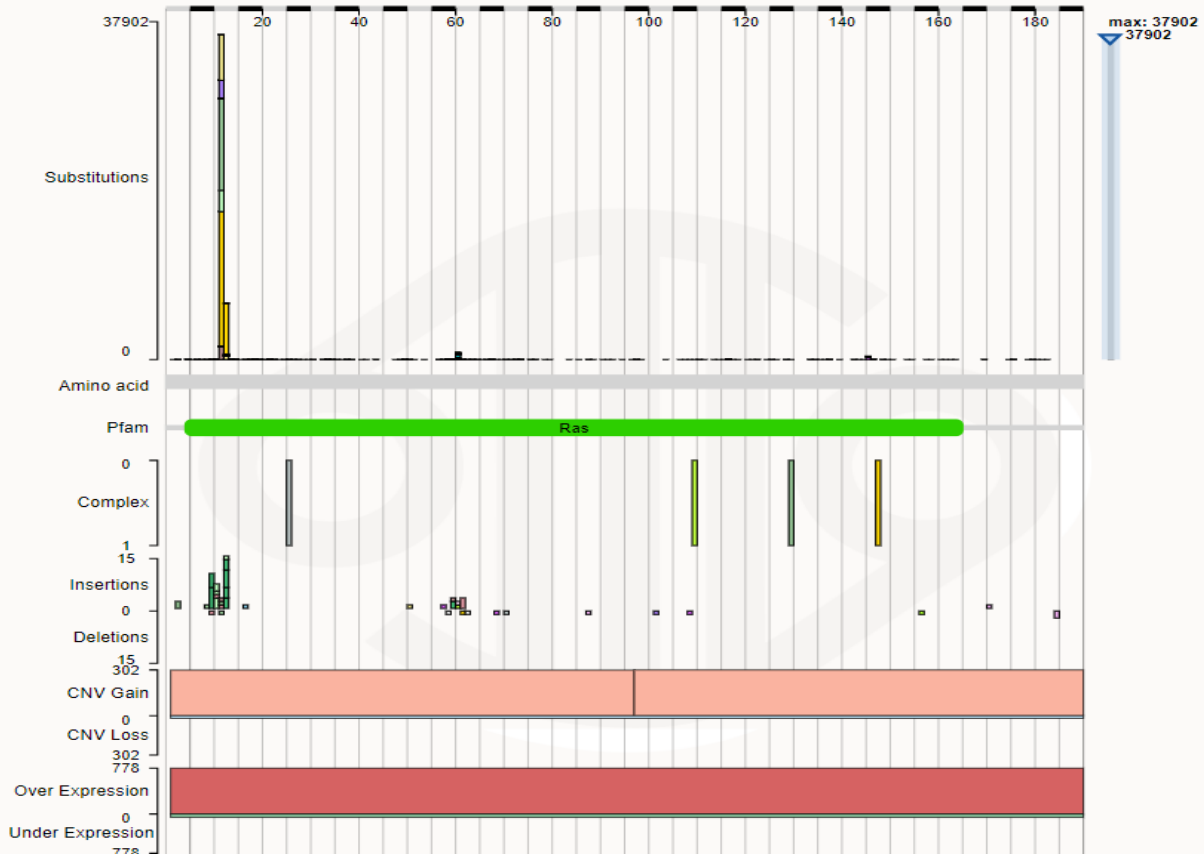
Coordinate system

- (•) Amino-acid
- ( ) cDNA

Apply filters | Reset filters

Gene view

The gene view histogram is a graphical view of mutations across KRAS. These mutations are displayed at the amino
of the gene by dragging across the histogram to highlight the region of interest, or by using the sliders in the filters



Gene
KRAS

- Gene view
- Overview
- External links
- Drug resistance
- Tissue distribution
- Genome browser
- Mutation distribution
- Variants
- References

Reset page

Search

Search COSMIC...

Filters

Show advanced filters

Range Show input fields

1 — 190

1 48 96 143 190

Coordinate system

- Amino-acid
- cDNA

Apply filters | Reset filters

Overview

This section gives an overview of KRAS, along with links to any related data and resources.

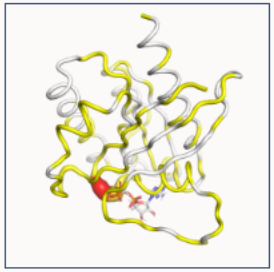
Census gene | Curated gene | Mouse gene | Hallmark gene

COSMIC gene KRAS (COSG97004)

Genomic coordinates 12:25209431..25250803 (negative strand)

Synonyms KRAS1, KRAS2, CCDS8703.1, P01116, ENSG00000133703.11, NM_033360.3, NP_203524

COSMIC-3D There are **103** structures for **KRAS**. View them in COSMIC-3D.



Number of samples
258508 unique samples
44885 unique samples with mutations

Alternative transcripts KRAS_ENST00000557334, KRAS_ENST00000311936, KRAS_ENST00000556131

Sequences You can see various sequences for this gene:
cDNA (ENST00000256078.8)
Protein (KRAS)
Transcript and protein aligned (ENST00000256078.8+KRAS)

Gene fusions KRAS is involved in 1 fusion, with the following gene:
UBE2L3_ENST00000342192 (1 mutation in 1 sample)

Drug sensitivity data Mutations in KRAS are associated with altered sensitivity to the following 19 drugs:
PD0325901, MK-2206, Trametinib
Show all
See all drug sensitivity data for KRAS.

Drug Resistance and tissue distribution

Gene
KRAS

- Gene view
- Overview
- External links
- Drug resistance
- Tissue distribution
- Genome browser
- Mutation distribution
- Variants
- References

Reset page

Search
Search COSMIC...

Filters
Show advanced filters

Range Show input fields

1 — 190

1 48 96 143 190

Coordinate system

- Amino-acid
- cDNA

Apply filters | Reset filters

Drug resistance

This section shows the drugs associated with **KRAS** resistance mutations. In the tabs below you can see any other genes that have resistance mutations to the same drug(s), and the distribution of mutations that occur in those genes.

Alternative transcripts are also displayed here for genes where reported resistant mutations are not located on the canonical transcript but are on the alternative, and also where reported resistant mutations are located at the same genomic position on both the canonical and alternative transcripts or on overlapping genes and/or fusions and share a COSM id.

No targeted therapeutic data has been curated for your selection.

Tissue distribution

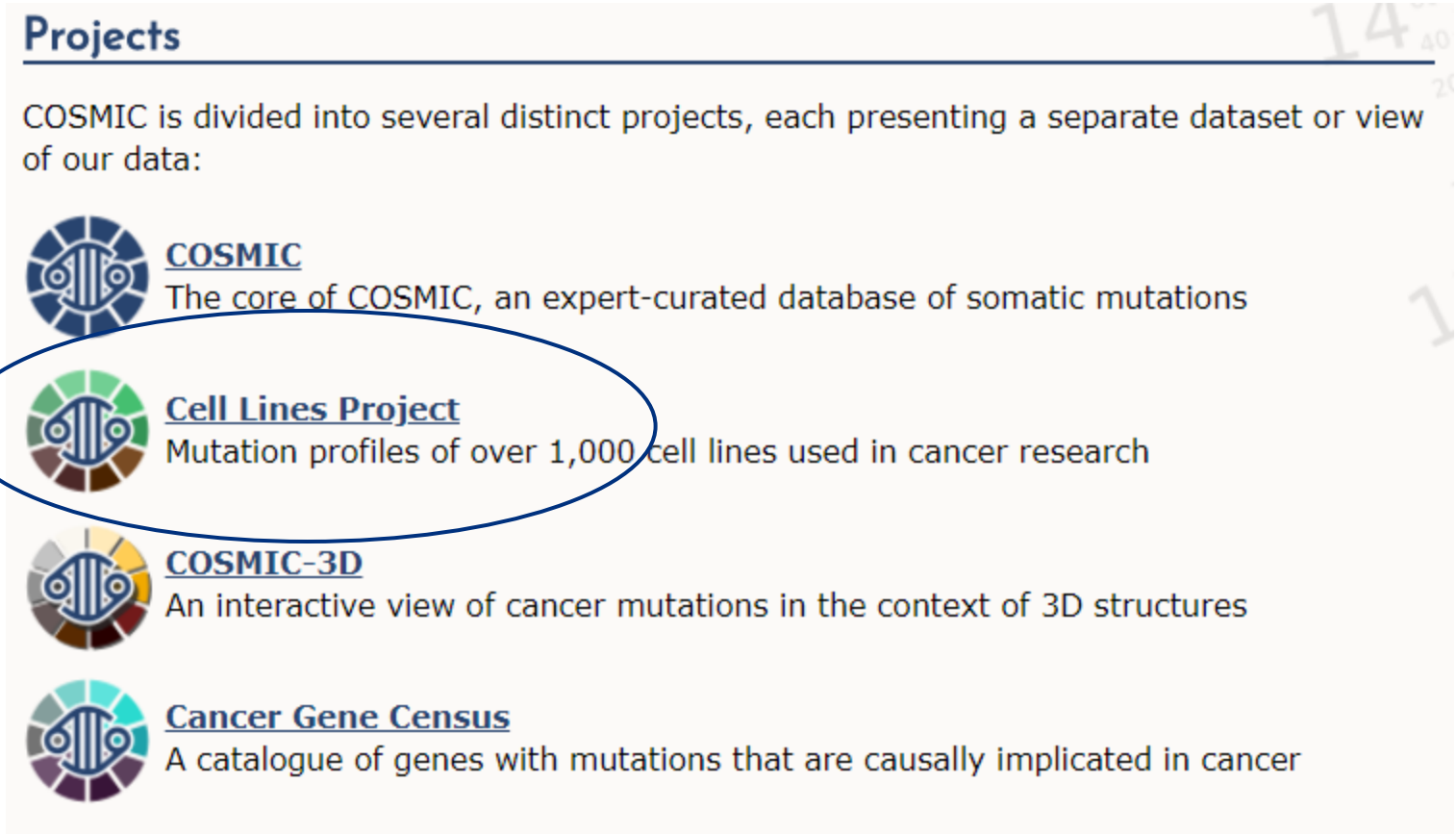
The table shows the distribution of mutations across the primary tissue types that are curated by COSMIC. Histograms show the percentage of mutated samples for point mutations, CNV data and ger expression data. Moving your mouse over the histograms will show additional data. The number of samples tested on this page include samples from the targeted and whole genomes/exome resequencing where all the protein coding genes have been screened for mutations.

You can see additional information about the data presented here in the help pages.

Show All entries Search:

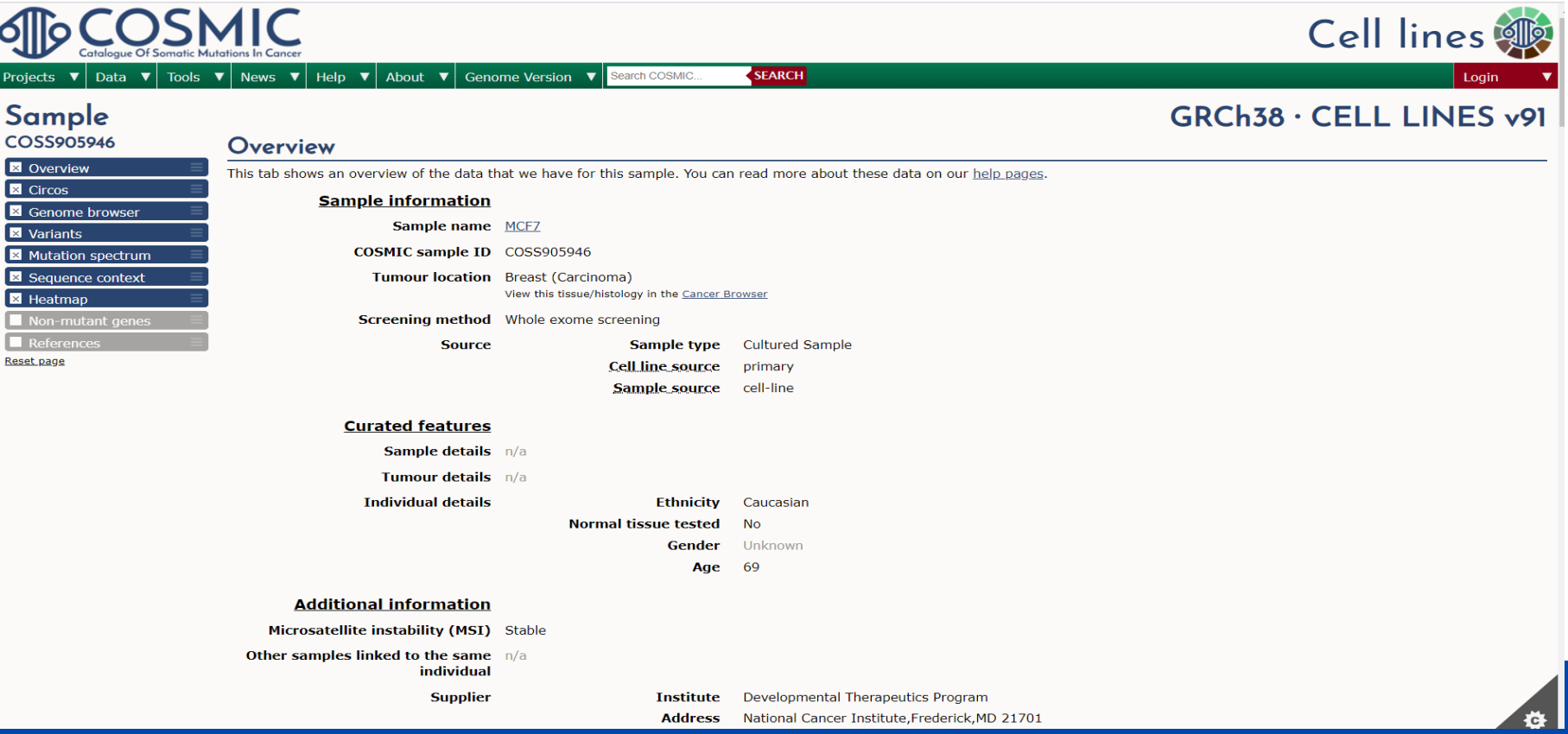
Tissue	Point Mutations		Copy Number Variation		Gene Expression		Methylation	
	% Mutated	Tested	Variant %	Tested	% Regulated	Tested	% Diff. Methylated	Tested
Adrenal gland		1335		267		79		-
Autonomic ganglia		1570		-		-		-
Biliary tract		5355		-		-		-
Bone		1182		-		-		-
Breast		10155		1492		1104		-
Central nervous system		4826		1035		697		-
Cervix		2236		299		307		-
Endometrium		4573		586		602		-
Eye		474		-		-		-
Fallopian tube		9		-		-		-
Female genital tract (site indeterminate)		4		-		-		-
Gastrointestinal tract (site indeterminate)		1083		-		-		-

Cell Lines Project

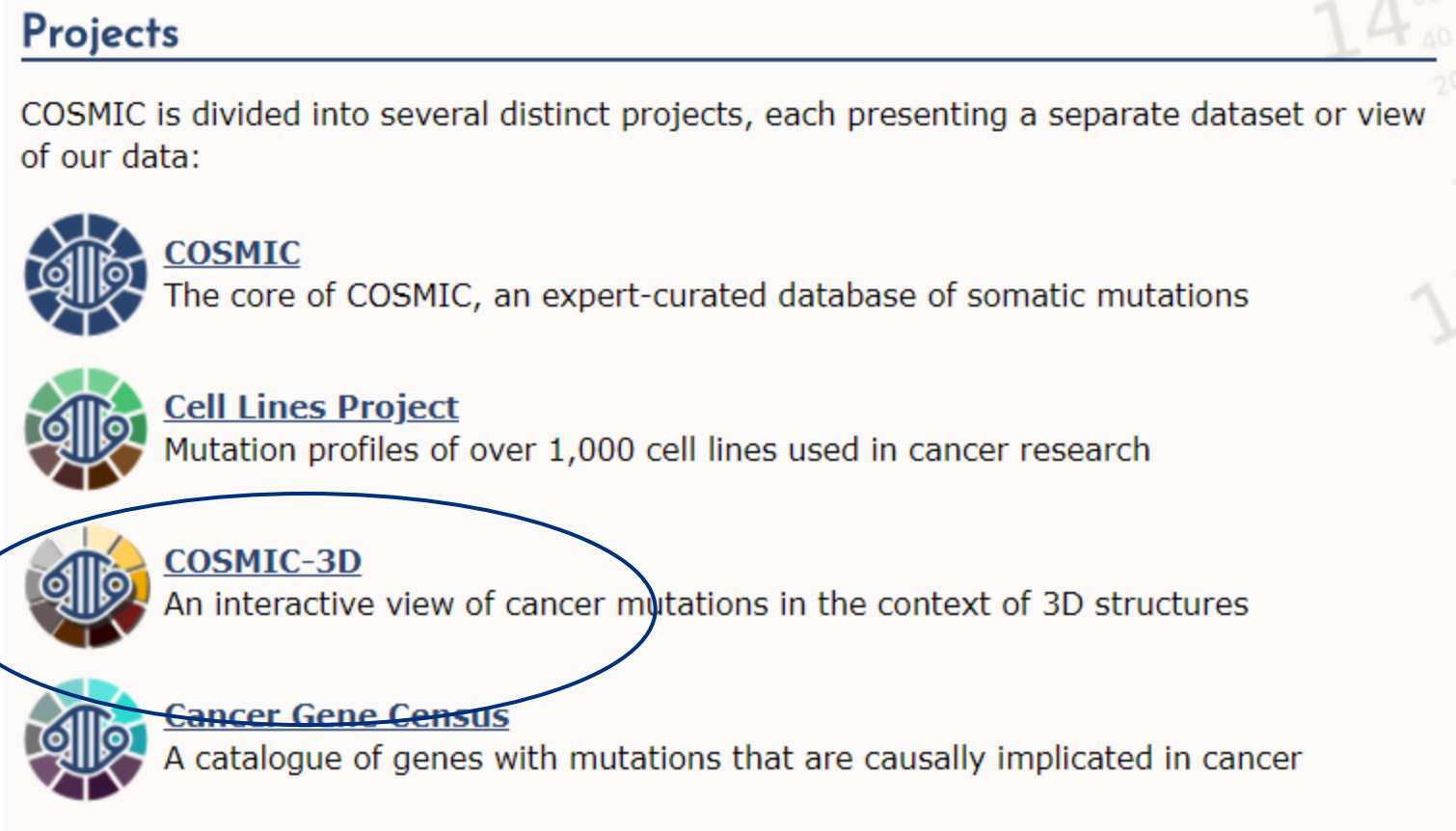


Cell lines project

- Mutation profiles of over 1,000 cell lines used in cancer research
- MCF7

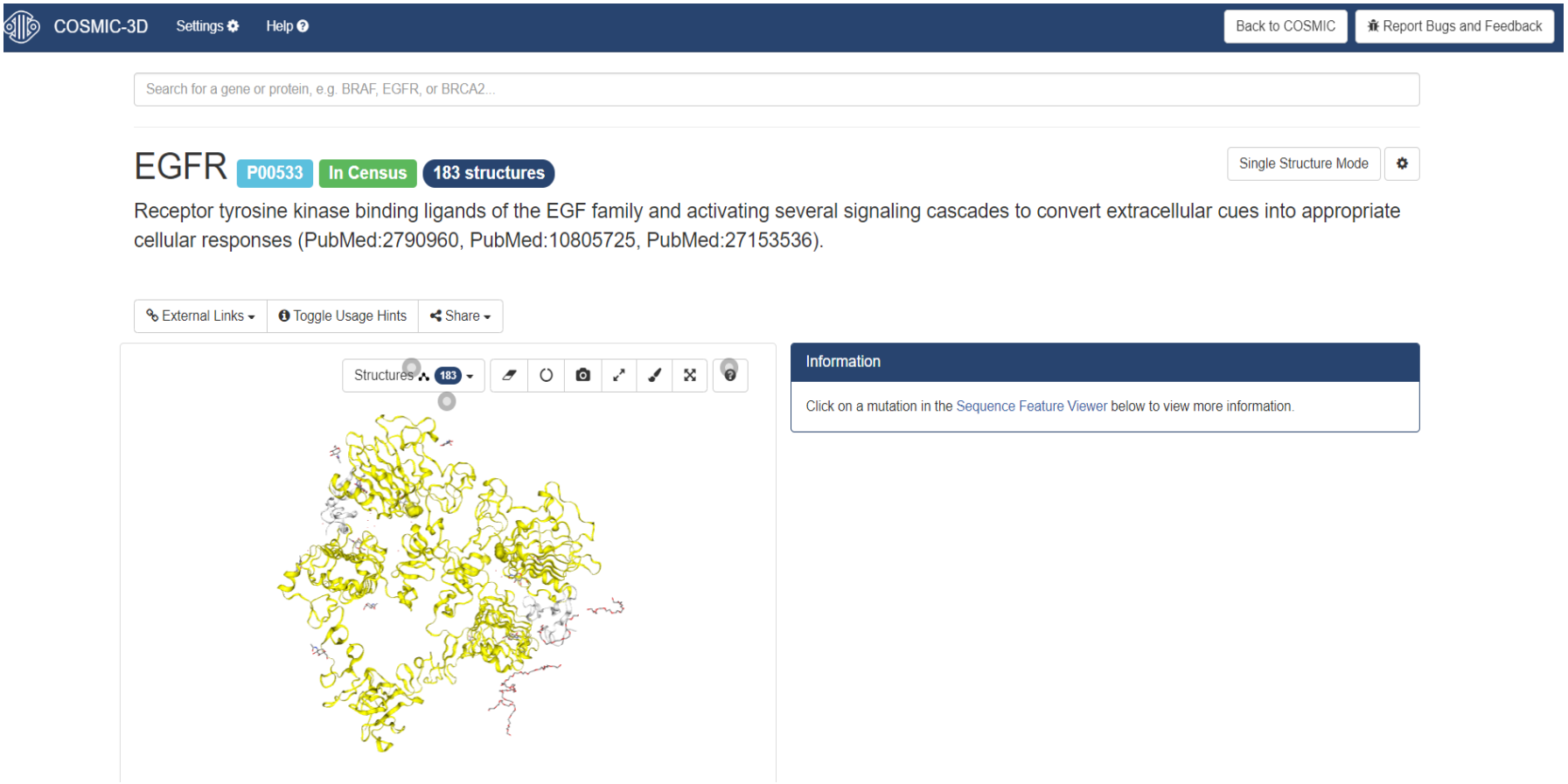


COSMIC-3D

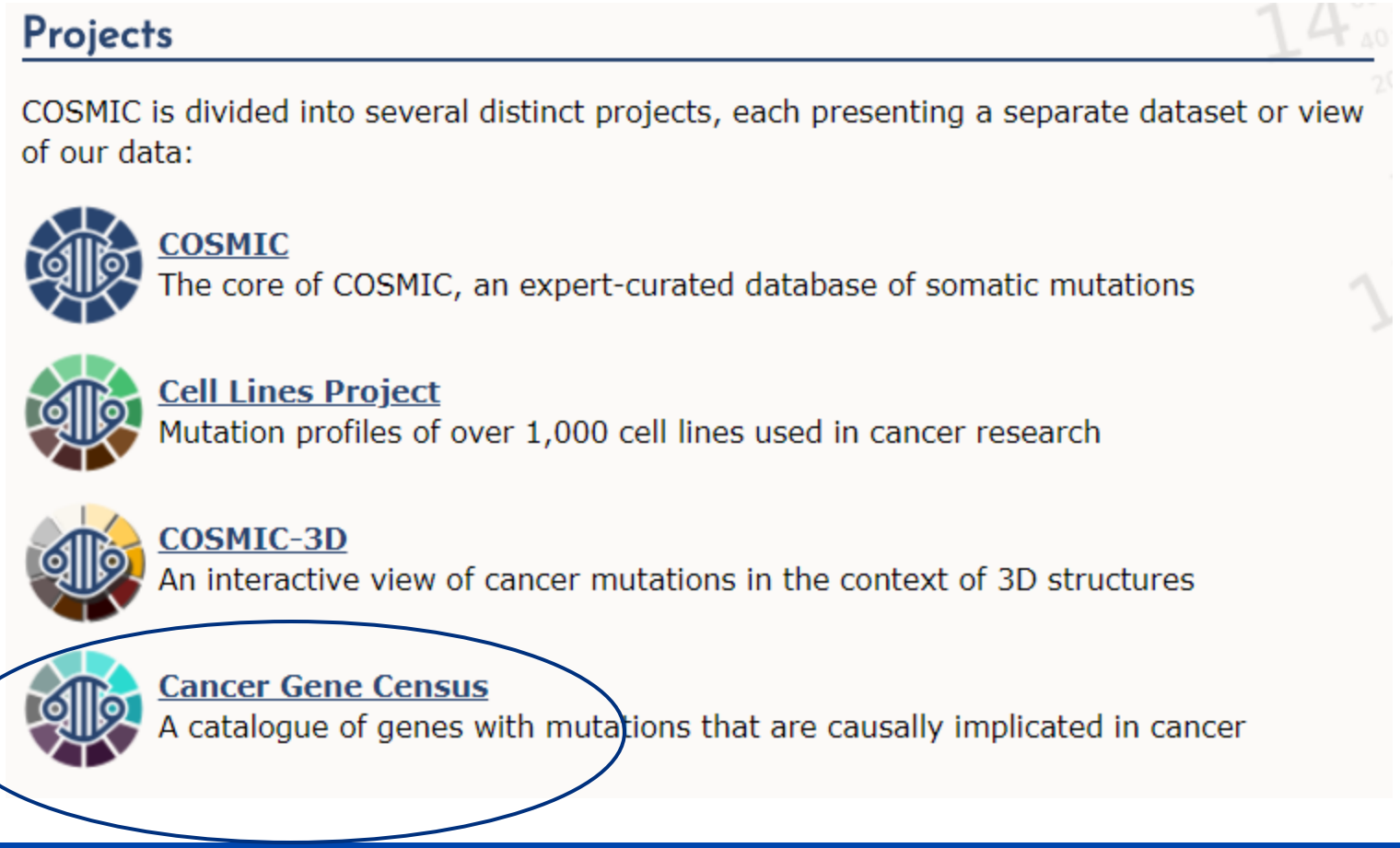


COSMIC-3D

- A platform for understanding cancer mutations in the context of 3D protein structure.



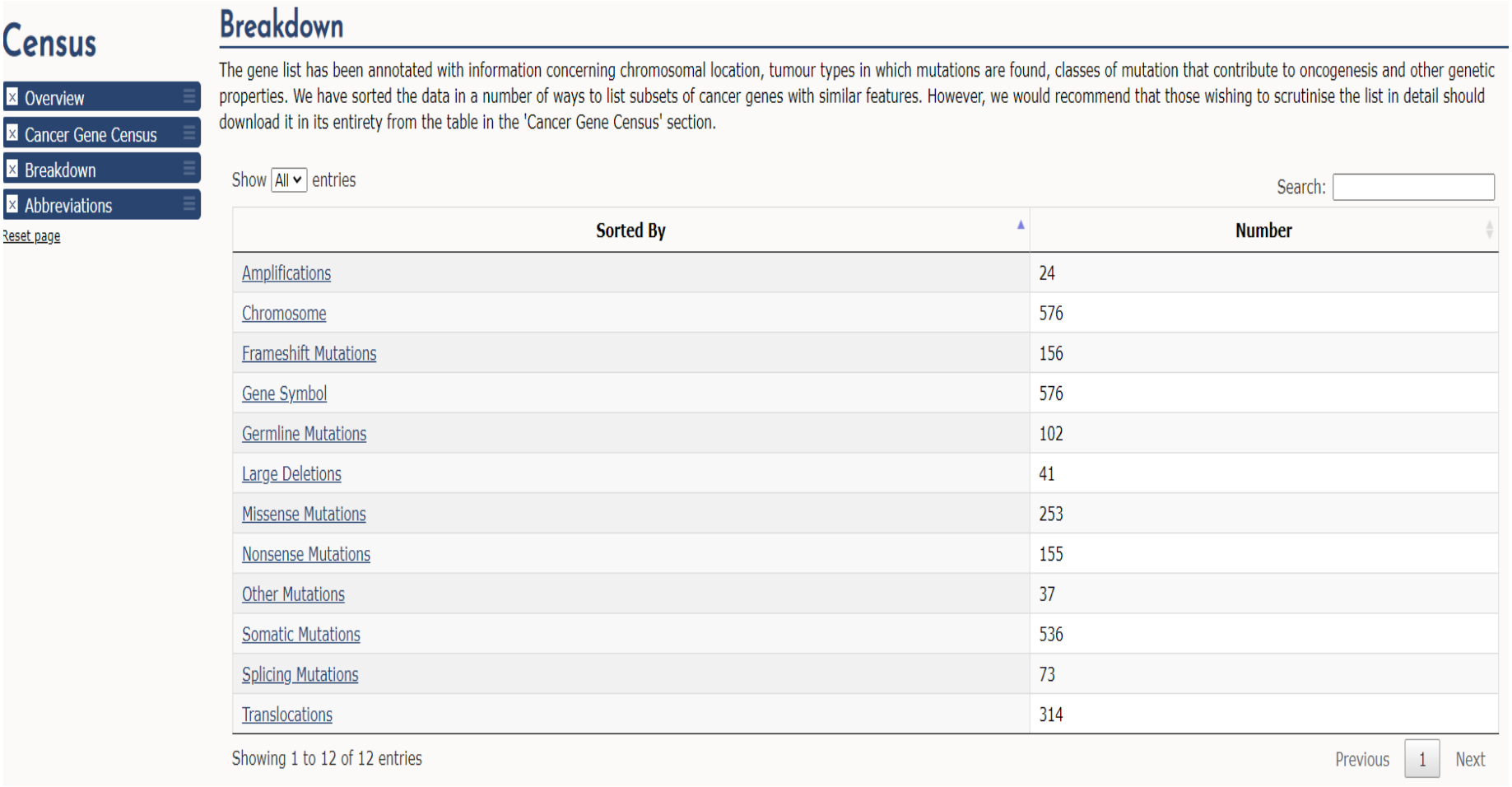
COSMIC



Gene Tiers in Cancer Gene Census

- **Census tiers – 723 genes**
- **Tier 1 –** A gene must possess a documented activity relevant to cancer, along with evidence of mutations in cancer which change the activity of the gene product in a way that promotes oncogenic transformation.
- **Tier 2 -** Consists of genes with strong indications of a role in cancer but with less extensive available evidence.

Breakdown of Genes/mutations



Census

- Overview
- Cancer Gene Census
- Breakdown
- Abbreviations

Reset page

Breakdown

The gene list has been annotated with information concerning chromosomal location, tumour types in which mutations are found, classes of mutation that contribute to oncogenesis and other genetic properties. We have sorted the data in a number of ways to list subsets of cancer genes with similar features. However, we would recommend that those wishing to scrutinise the list in detail should download it in its entirety from the table in the 'Cancer Gene Census' section.

Show All entries

Search:

Sorted By	Number
Amplifications	24
Chromosome	576
Frameshift Mutations	156
Gene Symbol	576
Germline Mutations	102
Large Deletions	41
Missense Mutations	253
Nonsense Mutations	155
Other Mutations	37
Somatic Mutations	536
Splicing Mutations	73
Translocations	314

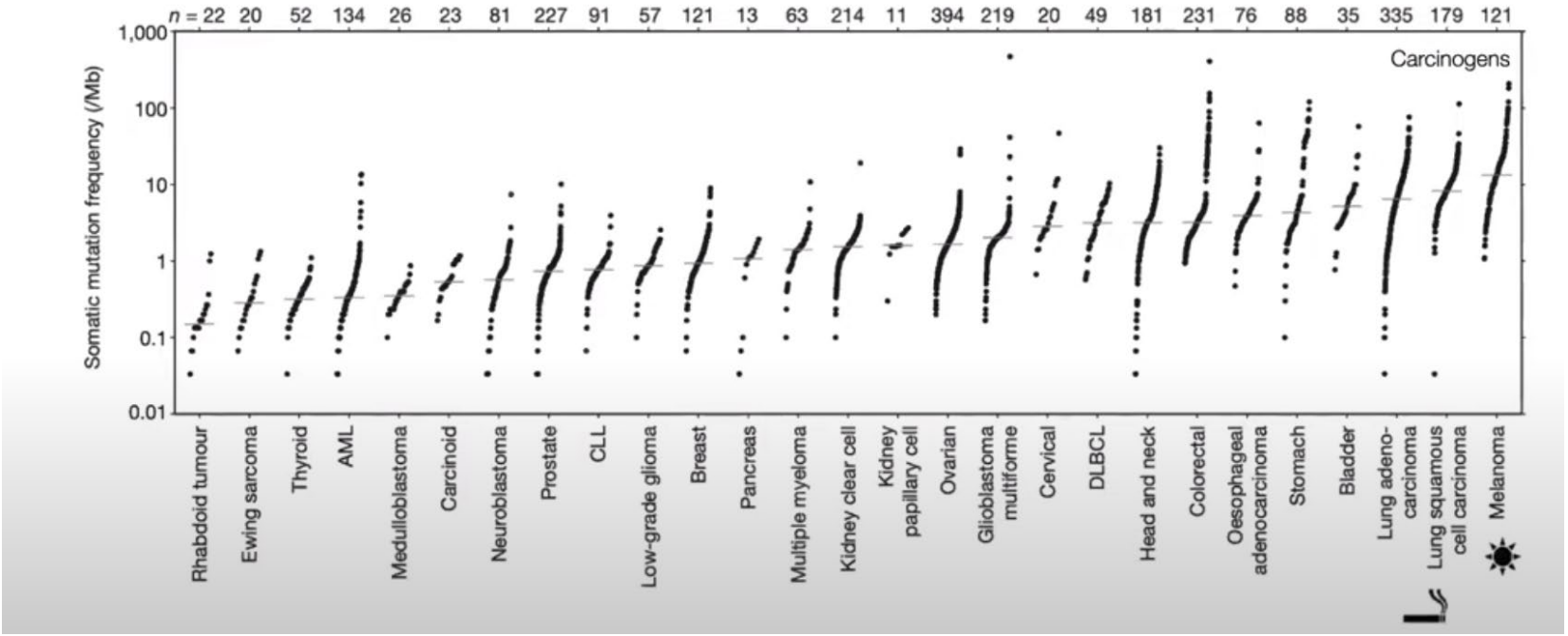
Showing 1 to 12 of 12 entries

Previous 1 Next

MAYO CLINIC

cBioPortal

Somatic mutations in different cancer types



Lawrence MS et al., Nature 2013

Public cancer genomics data for mining

- Cbioportal
- Visualization of multi-omics data
- Conduct simple analysis
- Summary of mutations and other data types
- Walkthrough simple queries
  - Glioma example
  - Breast cancer example
  - Pan-can analysis
  - Other examples

Barrowed few slides from cbioportal

http://www.cbioportal.org/index.do

Overview

- Show how to run a single-study query from the main page
- Walk through each of the data/analysis tabs in a single-study query
  - OncoPrint
  - Cancer Types Summary
  - Mutual Exclusivity
  - Plots
  - Mutations
  - Co-expression
  - Comparison (includes Survival, formerly a separate tab)
  - CN Segments
  - Pathways (replaces the Network tab)
  - Download
- Show how to modify and re-run a query

Overview of Tabs in a Single Study Query

Note that depending on the query run and the data available for a particular study, not all of these will be present (e.g. a study without mRNA expression data will not have a Co-expression tab)

- **OncoPrint:** Overview of genetic alterations per sample in each query gene
- **Cancer Types Summary:** Frequency of alteration in each query gene in the detailed cancer types included in this study
- **Mutual Exclusivity:** Statistical analysis to determine if query genes are mutually exclusively altered
- **Plots:** explore the relationships among genetic alterations, gene expression, protein levels, DNA methylation and available clinical features
- **Mutations:** Details about mutations called in each query gene
- **Co-expression:** Explore which genes have mRNA/protein levels correlated with query genes
- **Comparison:** Explore overlaps, outcomes, clinical attributes and genomic data comparisons among groups of samples as defined by the query
- **CN Segments:** Explore copy number changes with the Integrated Genomics Viewer (IGV)
- **Pathways:** Explore queried genes in TCGA-defined pathways
- **Download:** Download data or copy sample lists

Glioma example

Query overview

Browse available datasets and initiate queries

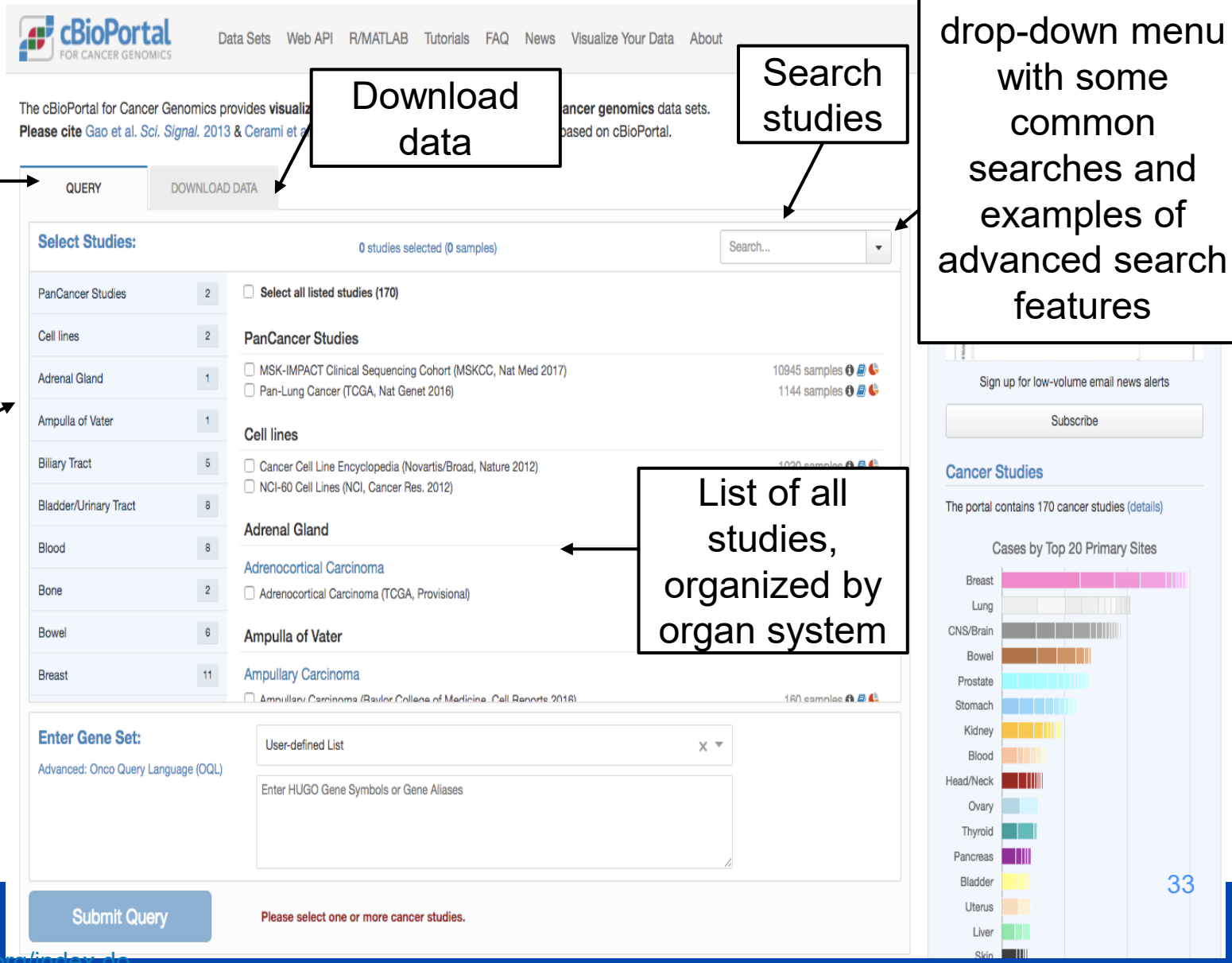
Download data

Search studies

Click here for a drop-down menu with some common searches and examples of advanced search features

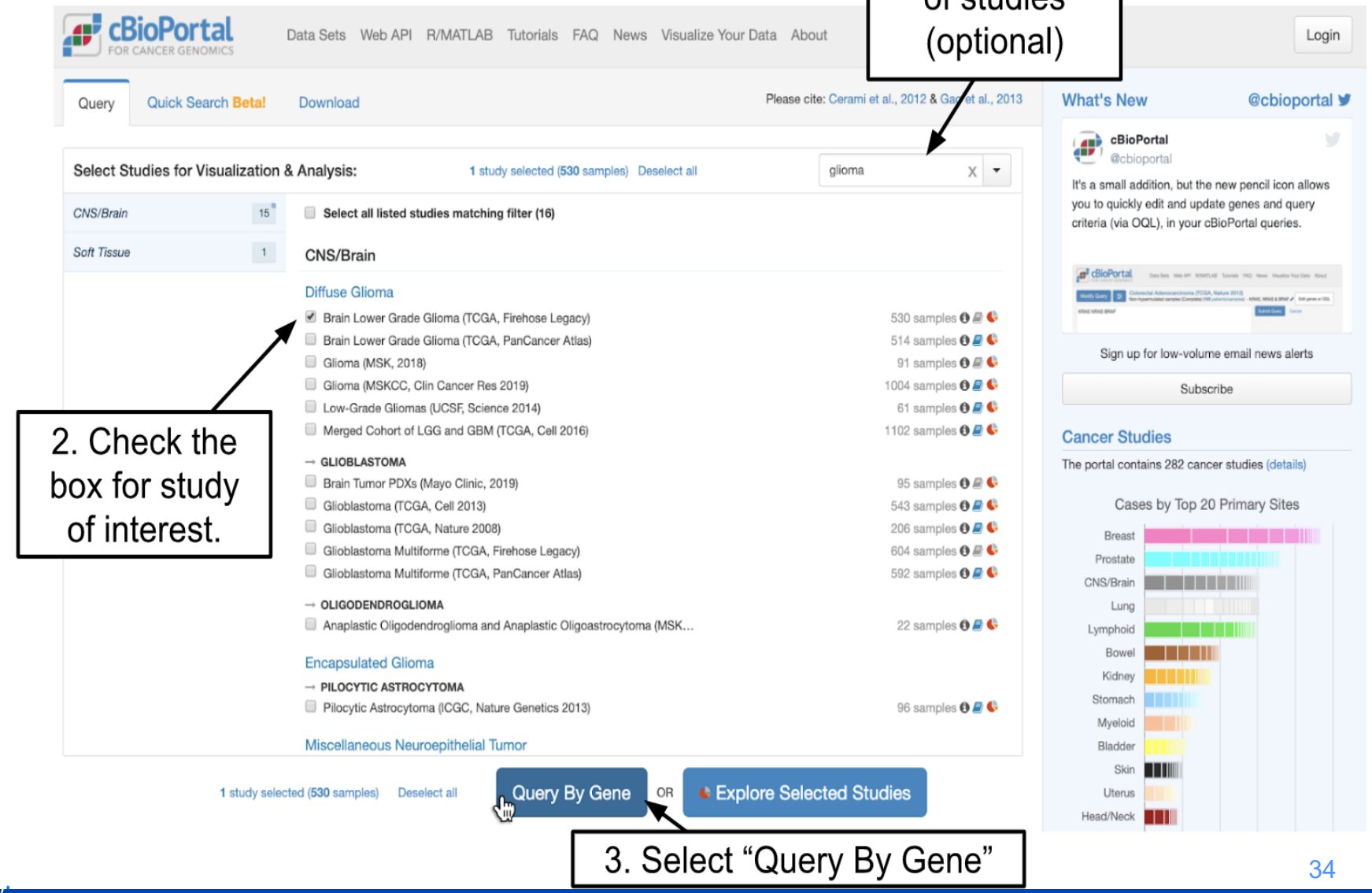
Number of studies for each organ system (click to filter)

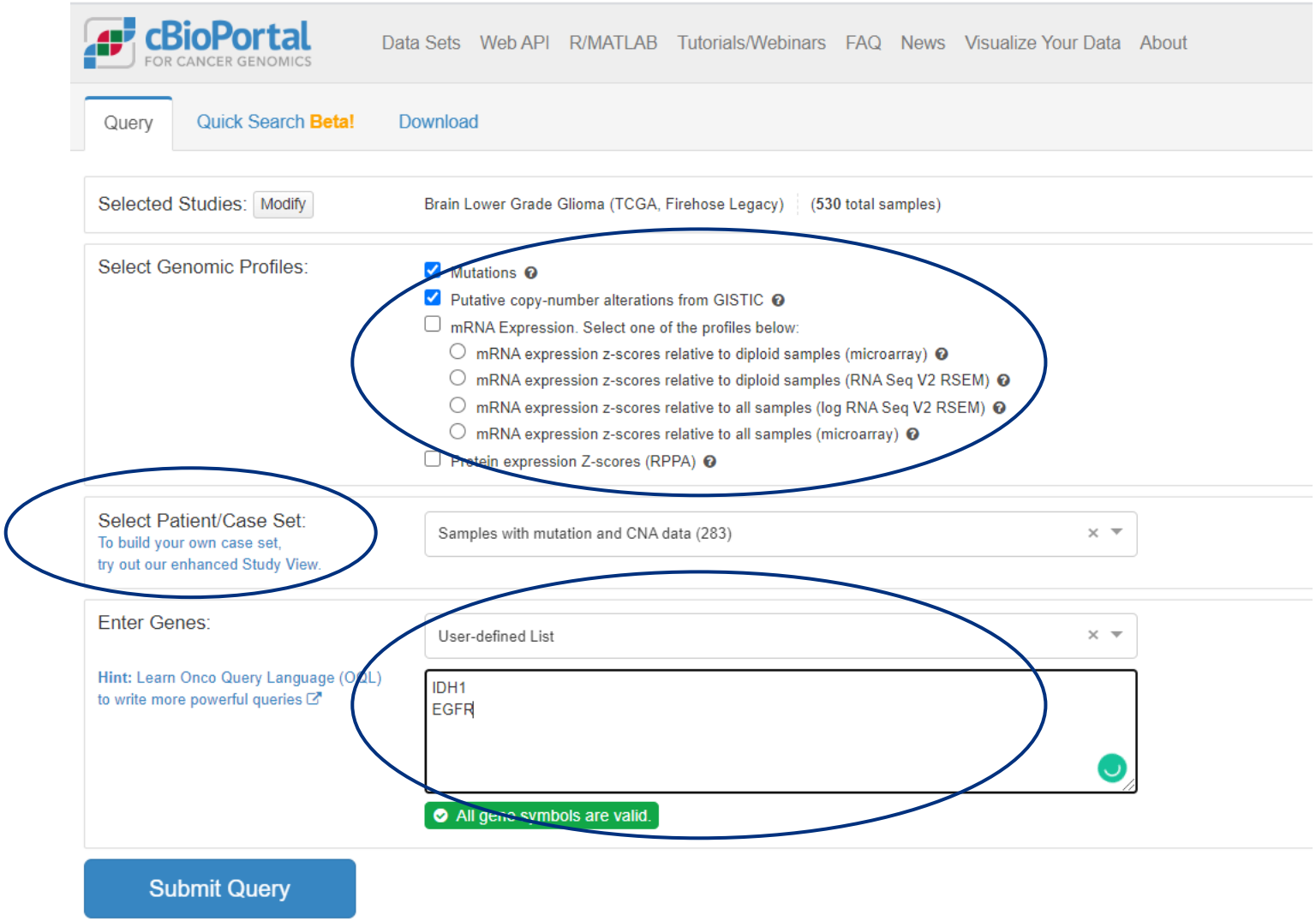
List of all studies, organized by organ system



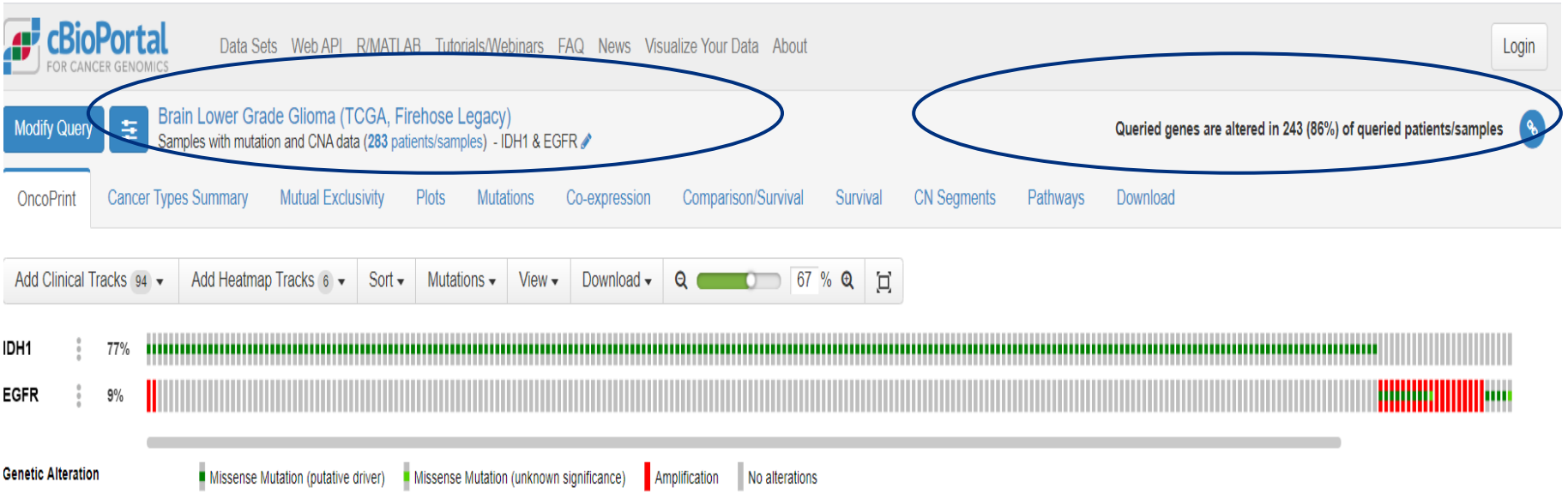
http://www.cbioportal.org/index.do

Single study query

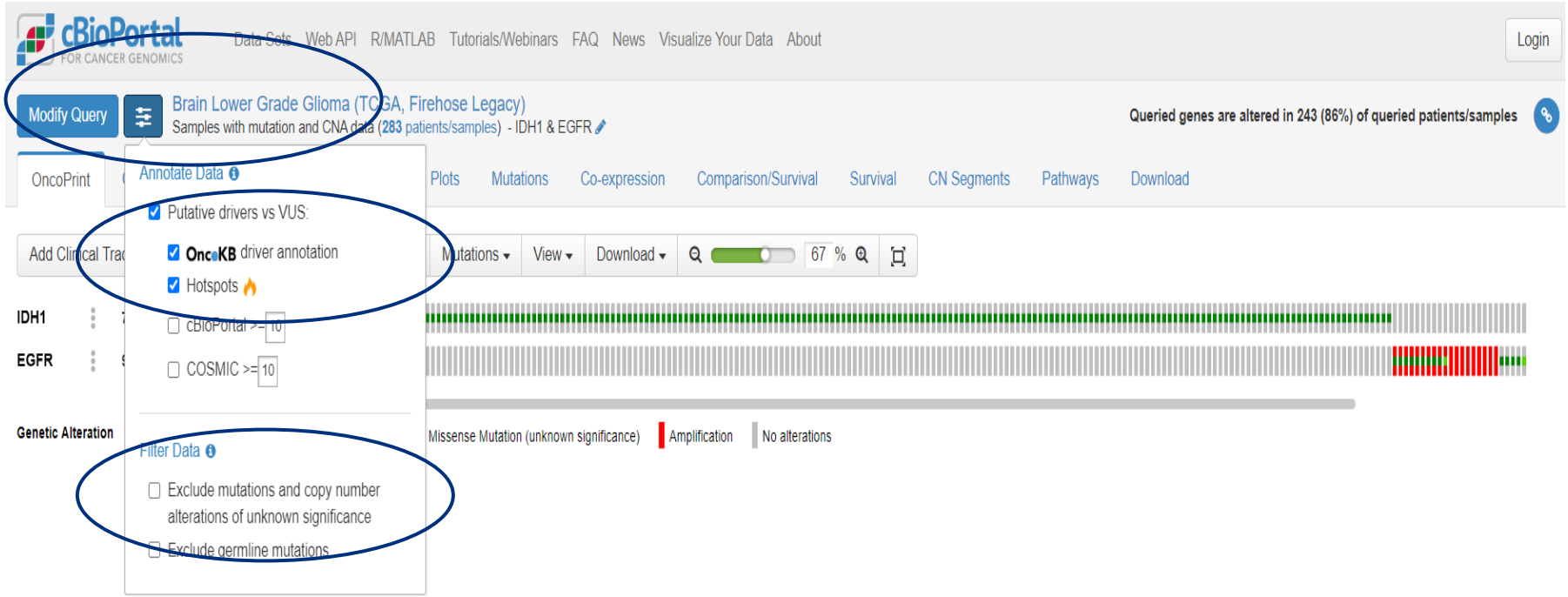




Glioma Query

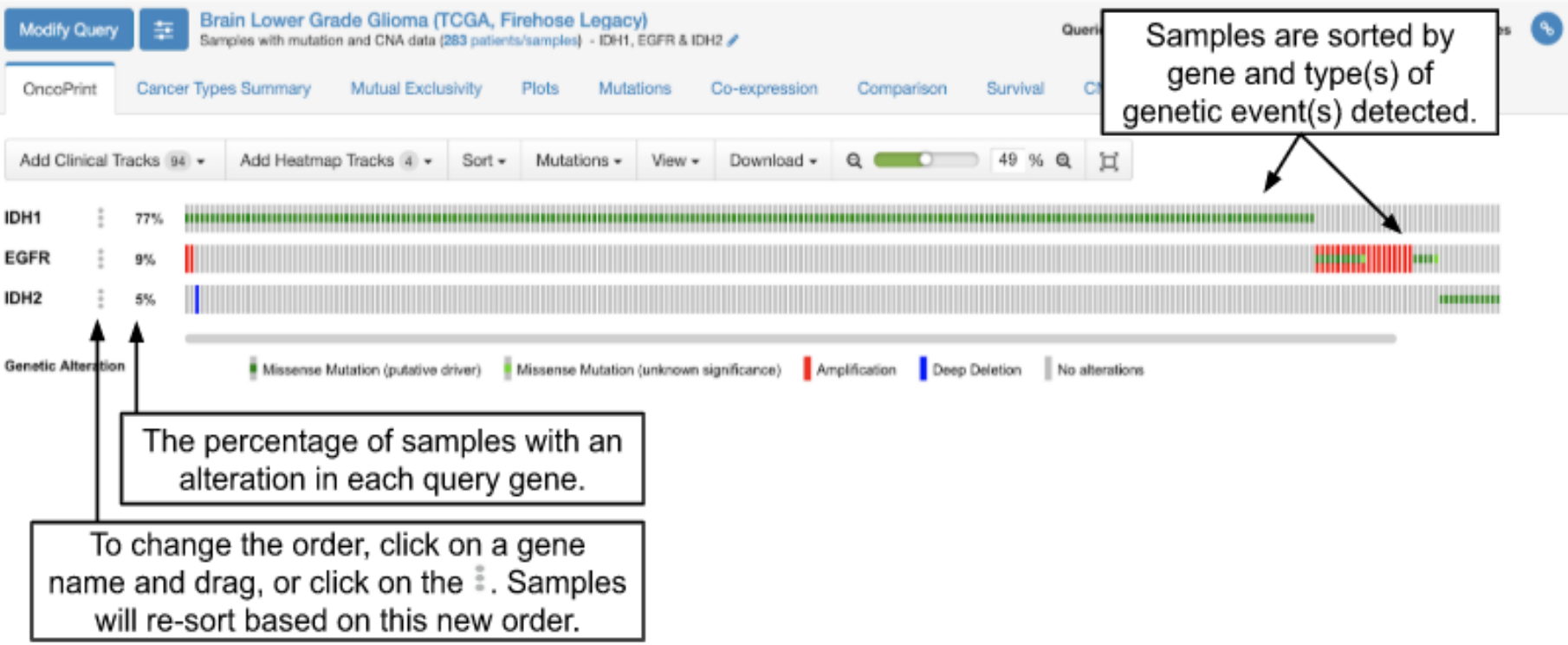


Annotations and Filtering

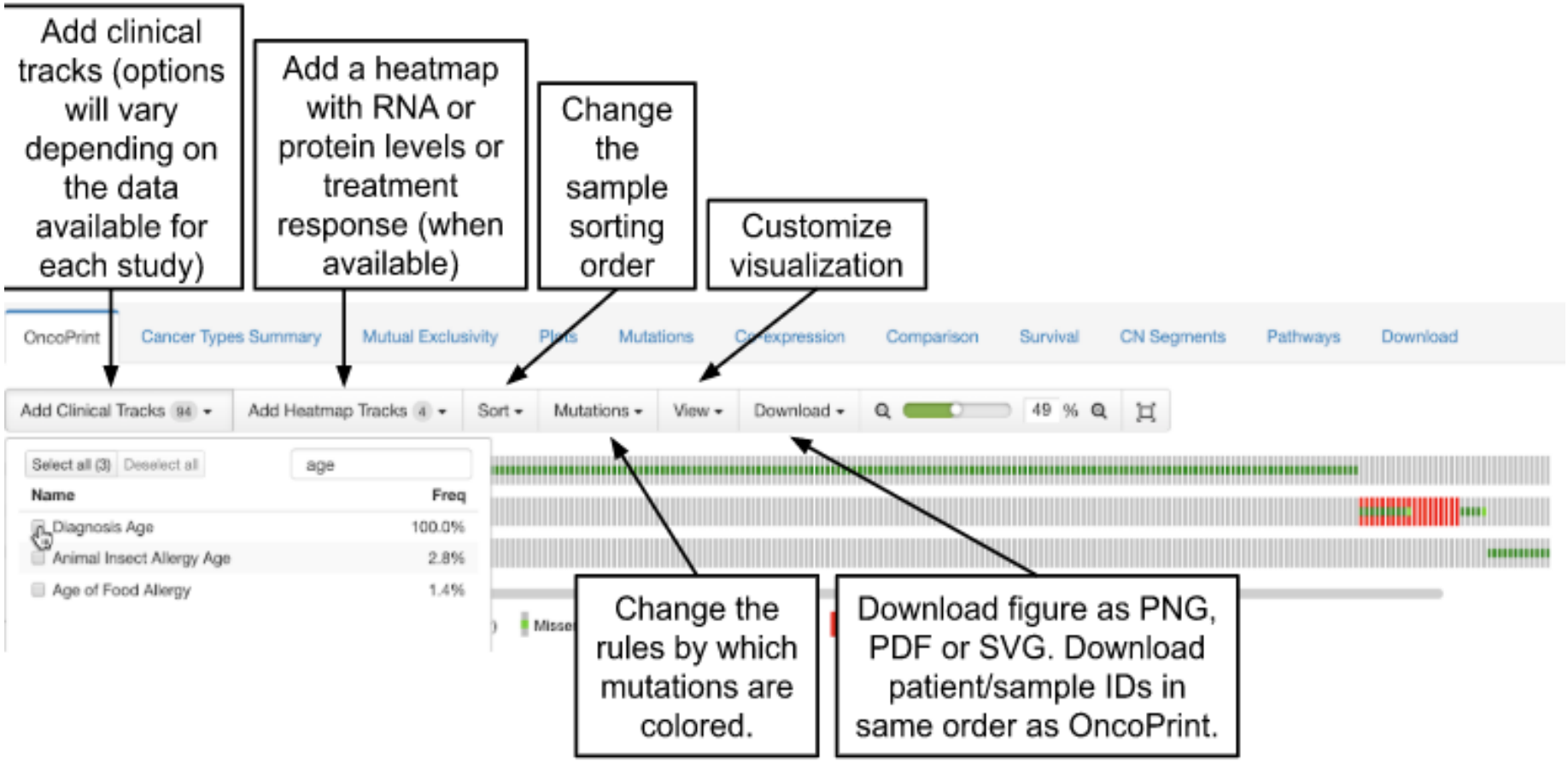


OncoPrint

Summary of alterations per sample. Each sample is a column. Each gene is a row. Different kinds of genetic alterations are highlighted with different colors.



OncoPrint: Features



OncoPrint: What can we learn?

Mutually exclusive – alterations in one gene tend to not have alterations in other genes

Patients with alterations in EGFR tend to be older than patients with IDH1/2 alterations.

Mutual Exclusivity with Glioblastoma example

All pairwise combinations of query genes analyzed for mutual exclusivity or co-occurrence in the queried samples.

On the OncoPrint tab we could see visually that alterations in these three query genes tended to be mutually exclusive. Here we can address that same question with a statistical analysis.

Click on any column header to sort. Hover over the column names for more details about how values are calculated.

A positive value here suggests that alterations in these genes co-occur in the same samples, while a negative value suggests that alterations in these genes are mutually exclusive and occur in different samples.

p-Value comes from Fisher Exact Test. Note that this is an unadjusted p-value and may need to be corrected for multiple hypothesis testing.

$$\log_2\left(\frac{\text{odds of alteration in B given alteration in A}}{\text{odds of alteration in B given lack of alteration in A}}\right)$$

Plots

Depending on available data types for a given study, this tab allows for plots comparing copy number, gene expression, protein levels and DNA methylation of query genes, along with any available clinical attributes.

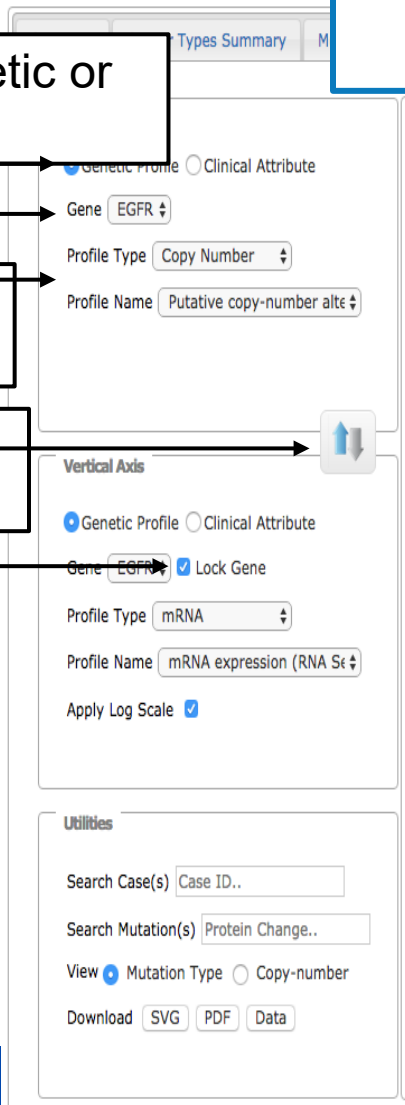
Choose genetic or clinical

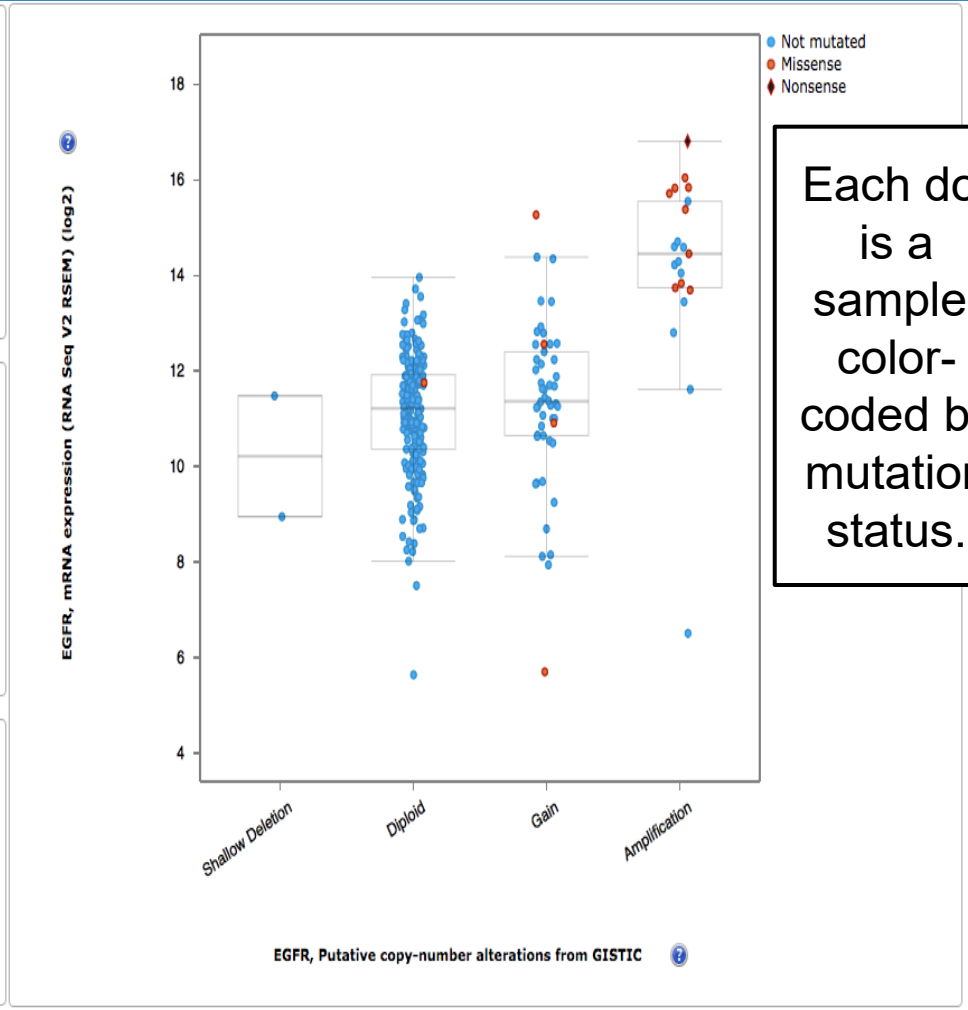
Select a query gene

Select data type and processing

Swap horizontal & vertical axis

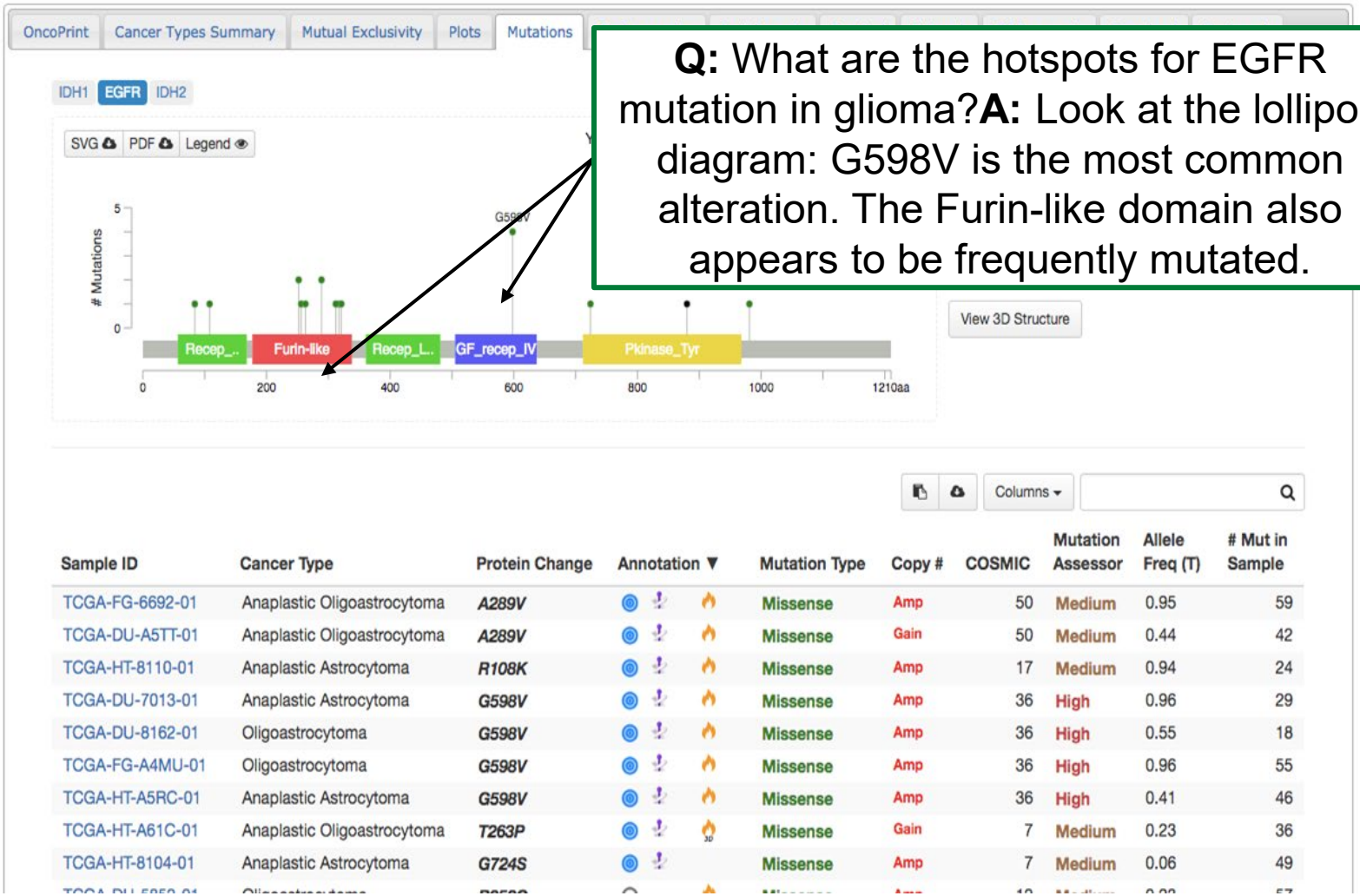
If checked, vertical axis will automatically show the same gene as horizontal axis.





Each dot is a sample, color-coded by mutation status.

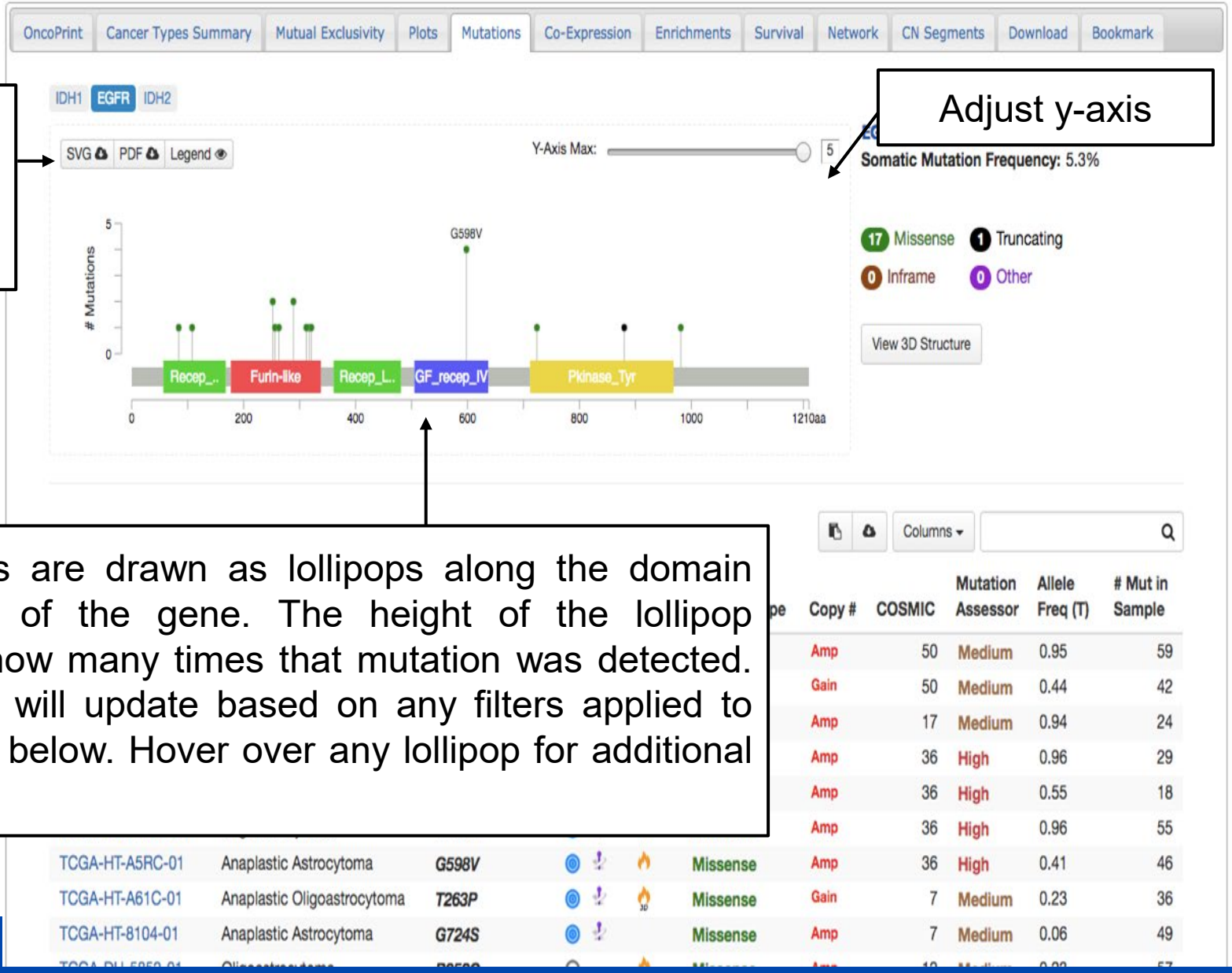
Mutations



Q: What are the hotspots for EGFR mutation in glioma?**A:** Look at the lollipop diagram: G598V is the most common alteration. The Furin-like domain also appears to be frequently mutated.

Sample ID	Cancer Type	Protein Change	Annotation ▼	Mutation Type	Copy #	COSMIC	Mutation Assessor	Allele Freq (T)	# Mut in Sample
TCGA-FG-6692-01	Anaplastic Oligoastrocytoma	*A289V*		Missense	Amp	50	Medium	0.95	59
TCGA-DU-A5TT-01	Anaplastic Oligoastrocytoma	*A289V*		Missense	Gain	50	Medium	0.44	42
TCGA-HT-8110-01	Anaplastic Astrocytoma	*R108K*		Missense	Amp	17	Medium	0.94	24
TCGA-DU-7013-01	Anaplastic Astrocytoma	*G598V*		Missense	Amp	36	High	0.96	29
TCGA-DU-8162-01	Oligoastrocytoma	*G598V*		Missense	Amp	36	High	0.55	18
TCGA-FG-A4MU-01	Oligoastrocytoma	*G598V*		Missense	Amp	36	High	0.96	55
TCGA-HT-A5RC-01	Anaplastic Astrocytoma	*G598V*		Missense	Amp	36	High	0.41	46
TCGA-HT-A61C-01	Anaplastic Oligoastrocytoma	*T263P*		Missense	Gain	7	Medium	0.23	36
TCGA-HT-8104-01	Anaplastic Astrocytoma	*G724S*		Missense	Amp	7	Medium	0.06	49

Mutations



Download figure as SVF or PDF

Adjust y-axis

Mutations are drawn as lollipops along the domain structure of the gene. The height of the lollipop reflects how many times that mutation was detected. This plot will update based on any filters applied to the table below. Hover over any lollipop for additional details.

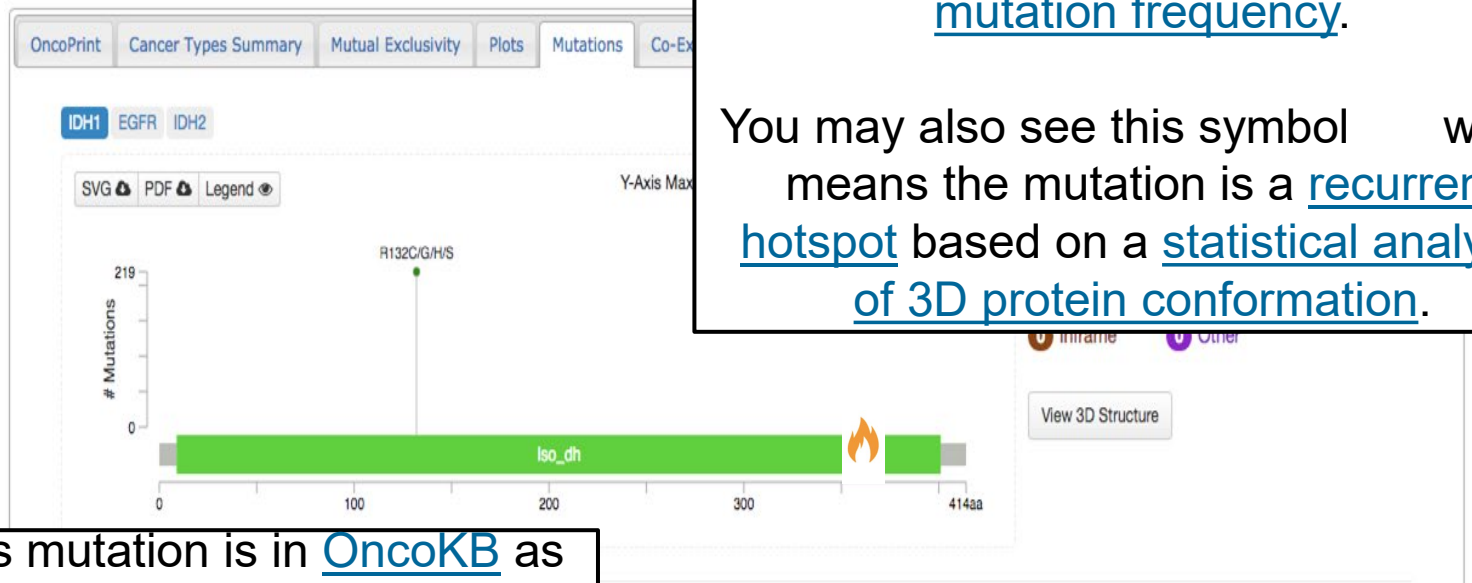
http://www.cbioportal.org/index.do

Mutations

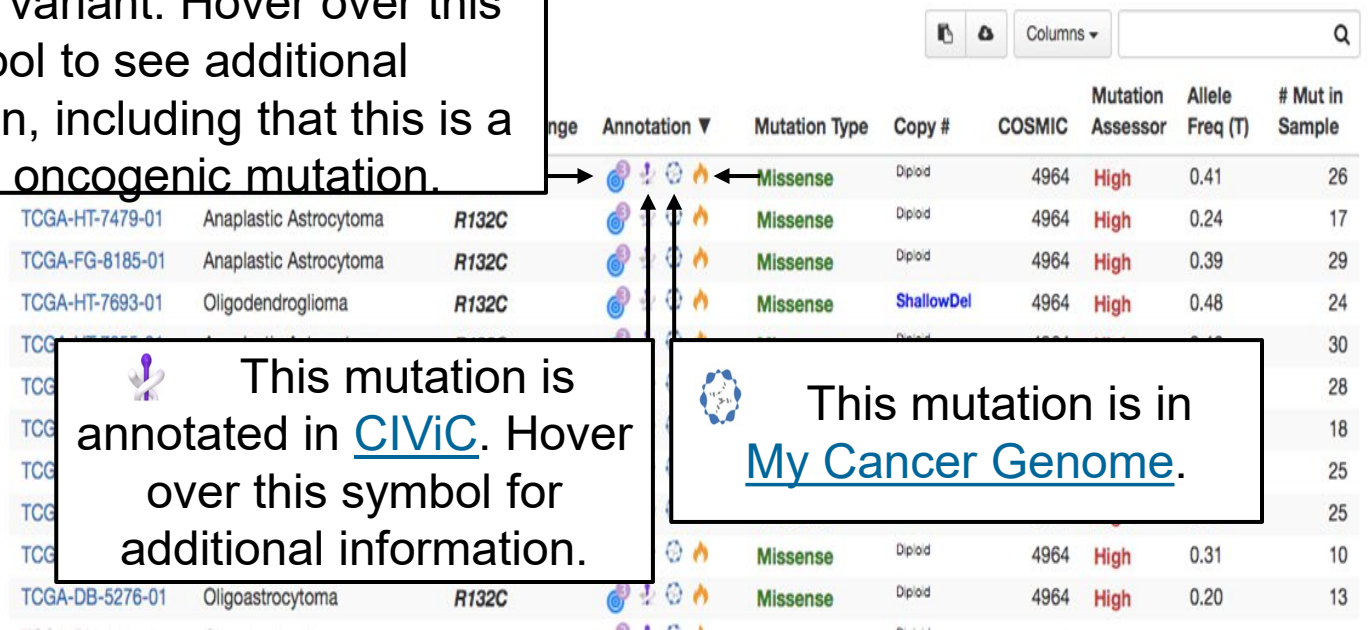
http://www.cbioportal.org/index.do

This mutation is a recurrent hotspot based on a statistical analysis of mutation frequency.

You may also see this symbol which means the mutation is a recurrent hotspot based on a statistical analysis of 3D protein conformation.



This mutation is in OncoKB as a Level 3 variant. Hover over this symbol to see additional information, including that this is a known oncogenic mutation.



This mutation is annotated in CIViC. Hover over this symbol for additional information.

This mutation is in My Cancer Genome.

Breast cancer example

Modify Query

Breast Invasive Carcinoma (TCGA, PanCancer Atlas)

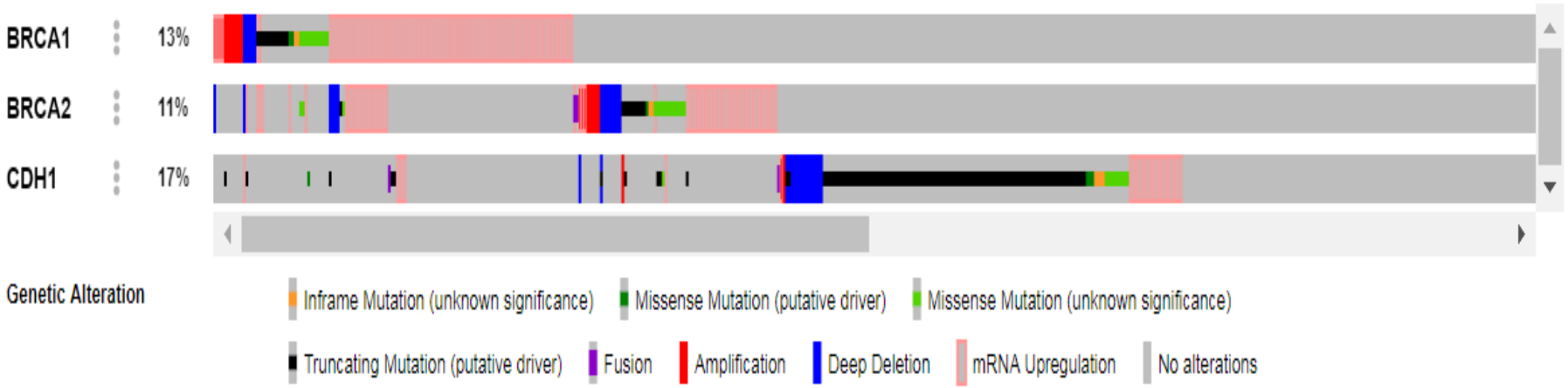
All Complete Tumors (**993** samples) / **3** Genes

Gene Set / Pathway is altered in 361 (36.4%) of queried samples

OncoPrint | Cancer Types Summary | Mutual Exclusivity | Plots | Mutations | Co-Expression | Enrichments | Network | CN Segments | Download | Bookmark

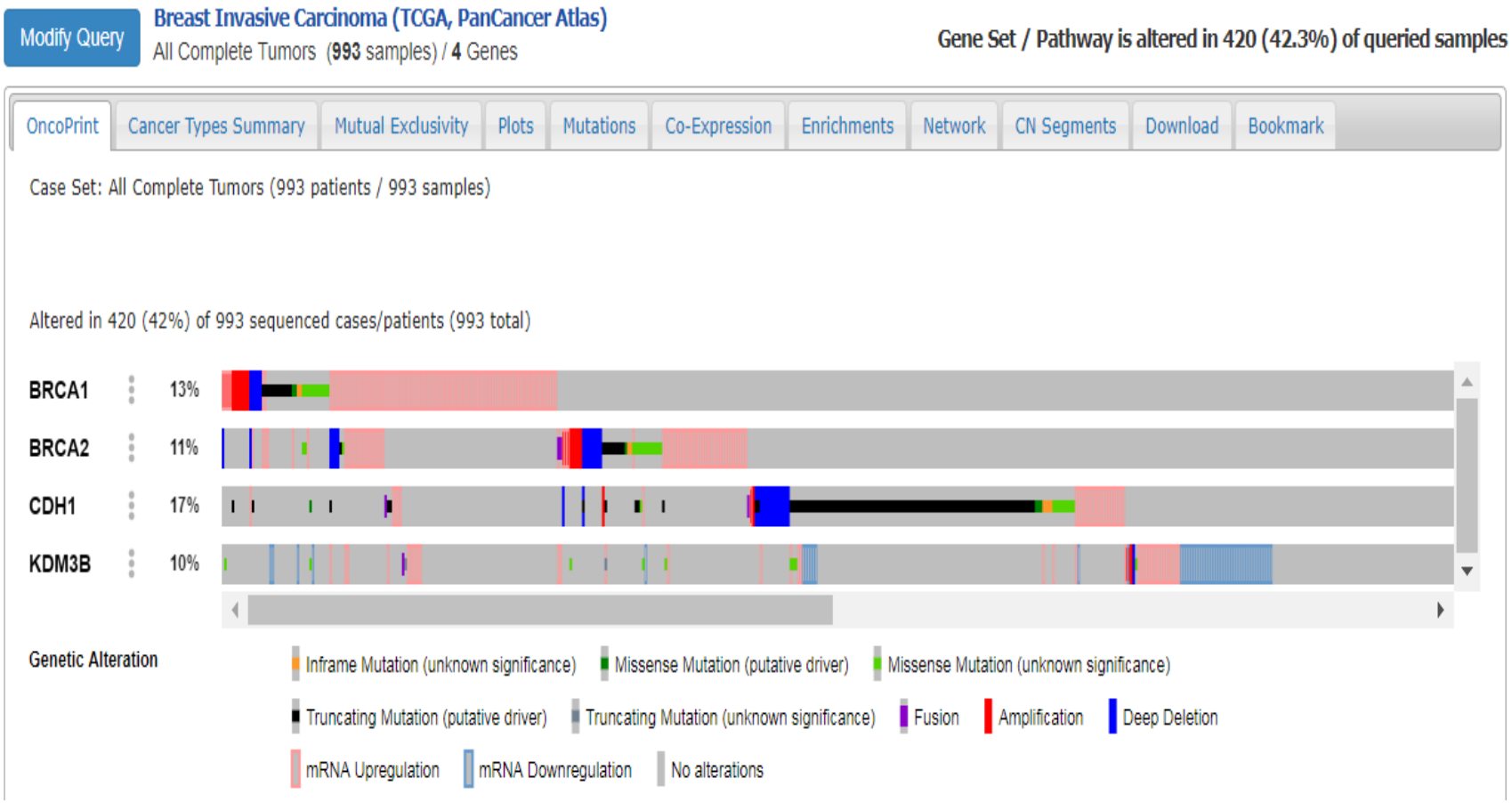
Case Set: All Complete Tumors (993 patients / 993 samples)

Altered in 361 (36%) of 993 sequenced cases/patients (993 total)



http://www.cbioportal.org/index.do

CNV



mRNA overexpressed

Co-Expression

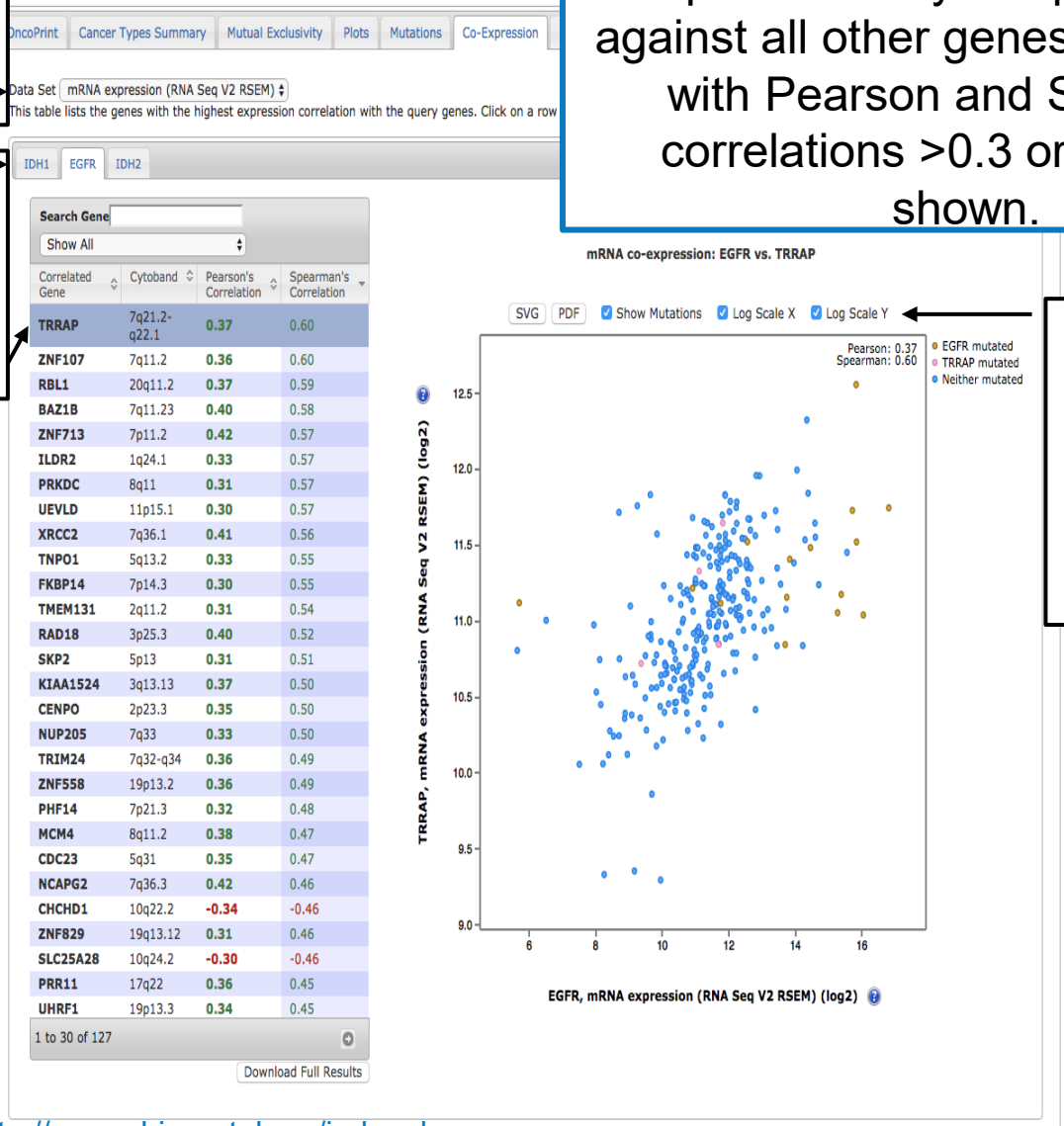
Select from available data types

Each gene appears on a separate tab

Click on a gene name to see correlation plot

Compares mRNA/protein level expression of your query genes against all other genes. Only genes with Pearson and Spearman correlations >0.3 or <-0.3 are shown.

Check boxes to color-code sample dots by mutation status or change x- or y-axis to log scale



http://www.cbioportal.org/index.do

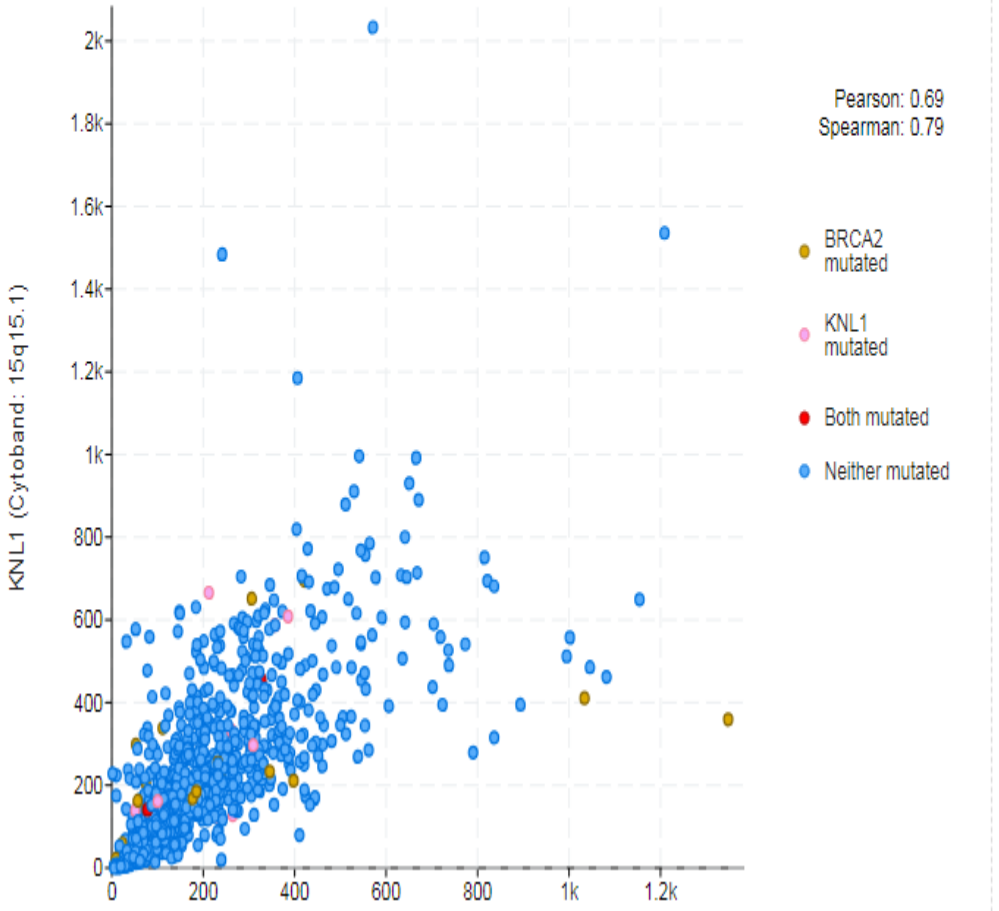
BRCA1 BRCA2

Show Any Correlation | Enter gene or cytoband.

Correlated Gene	Cytoband	Pearson's Correlation	Spearman's Correlation ▼
KNL1	15q15.1	0.69	0.79
GAS2L3	12q23.1	0.63	0.75
CLSPN	1p34.3	0.60	0.73
SMC2	9q31.1	0.66	0.72
PRR11	17q22	0.19	0.72
ARHGAP11B	15q13.2	0.63	0.71
RBL1	20q11.23	0.50	0.71
CKAP2L	2q14.1	0.66	0.71
KIF20B	10q23.31	0.61	0.71
CENPI	Xq22.1	0.61	0.70
CKAP2	13q14.3	0.64	0.69
ECT2	3q26.31	0.64	0.69
XPO1	2p15	0.65	0.69
ATAD5	17q11.2	0.60	0.68
ASPM	1q31.3	0.62	0.68
NUP155	5p13.2	0.57	0.67
WDHD1	14q22.2-q22.3	0.59	0.67
MKI67	10q26.2	0.61	0.67
CEP152	15q21.1	0.54	0.66
SGO2	2q33.1	0.61	0.66
TOPBP1	3q22.1	0.60	0.66

Show Mutations | Log Scale | SVG | PDF

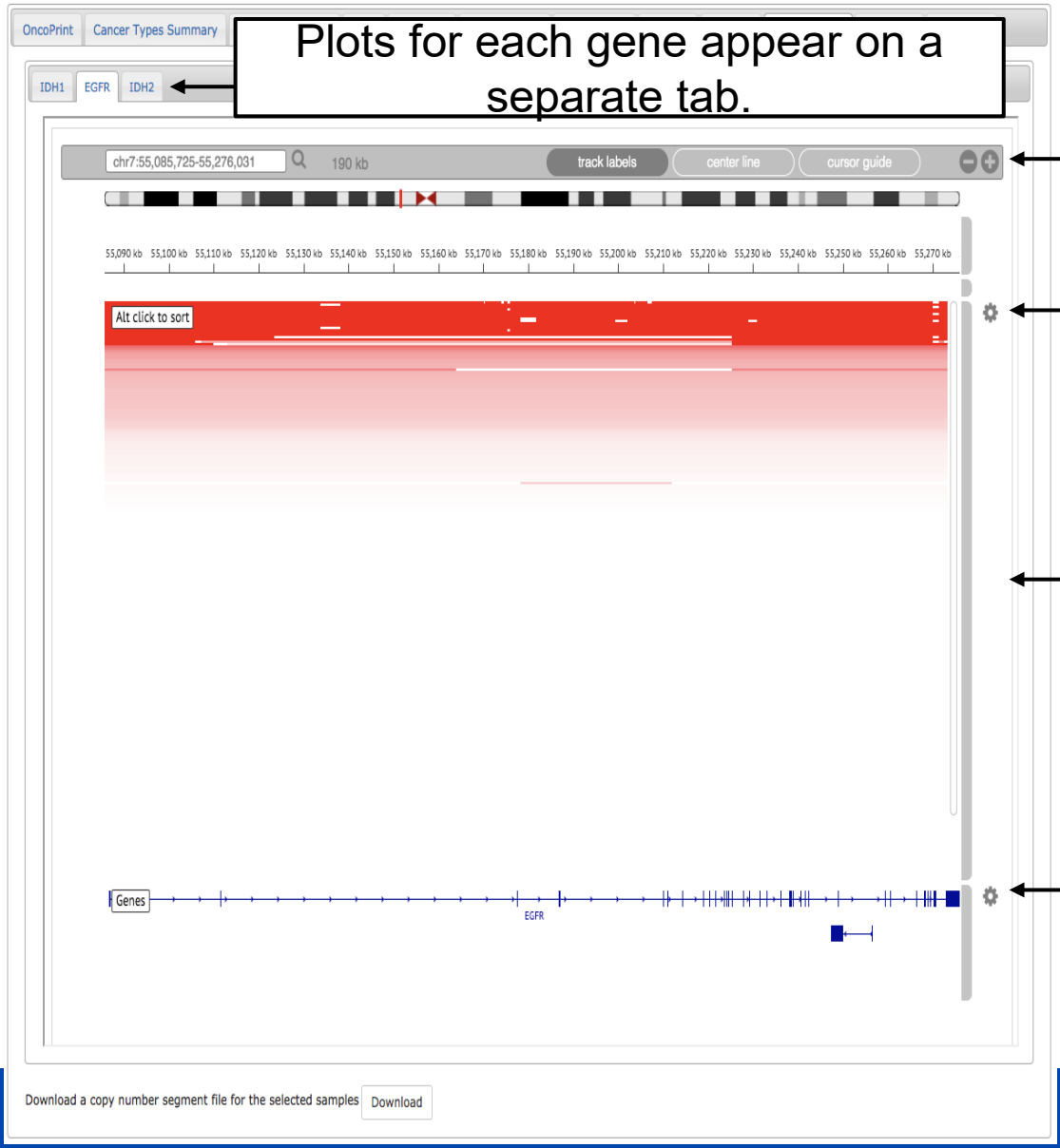
mRNA Expression Batch Normalized/Merged from Illumina HiSeq_RNASeqV2 syn4976369: BRCA2 vs



http://www.cbioportal.org/index.do

CN Segments

View copy number for each sample at each query gene via the Integrated Genomics Viewer (IGV).



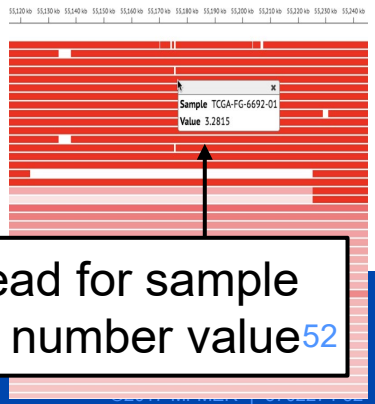
Plots for each gene appear on a separate tab.

Toggle track labels, a vertical line marking the center of the viewing screen, and a vertical line that moves with your cursor. Use to zoom in or out.

Click for track settings, including expanding the height of each sample (see below)

Each row is a single sample

Gene structures

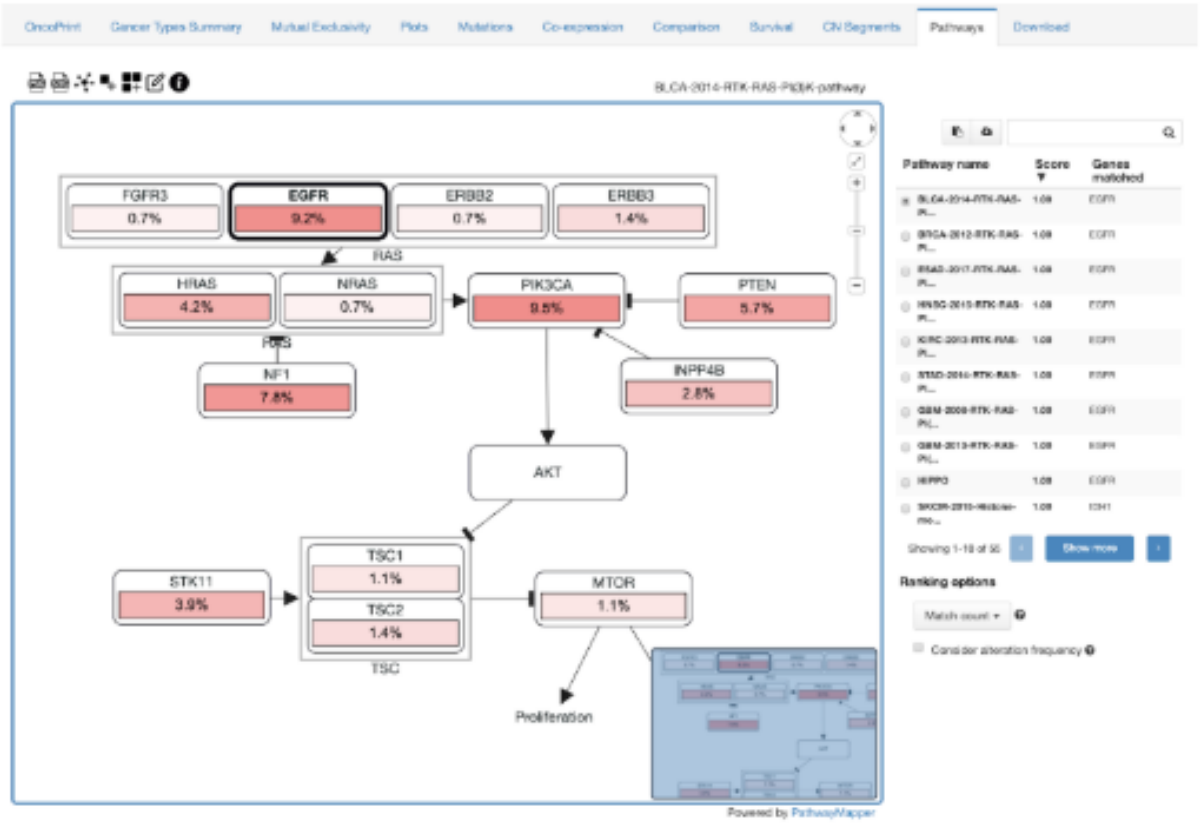


Click on a read for sample ID and copy number value

Link to this page

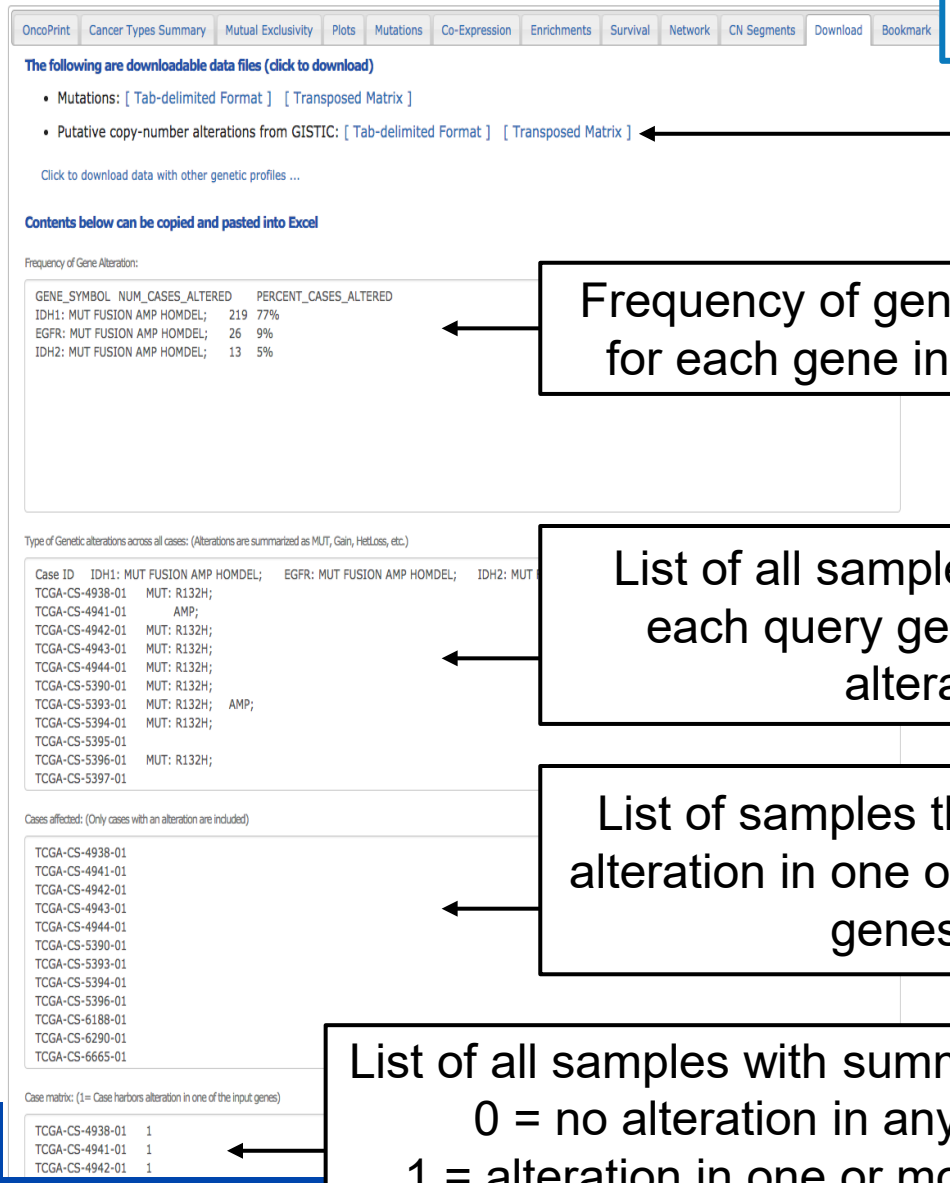
Pathways

The Pathways tab replaces the now retired “Network” tab. This tab in an integration with PathwayMapper. The tab enables exploration of the queried genes in the context of Pathways defined by TCGA. For more detail on this tab, refer to the Pathways Tutorial.



Link to this page

Download



Download data or copy lists of samples.

Download mutations and copy number

Frequency of gene alteration for each gene in the query

List of all samples with status of each query gene (blank = no alteration)

List of samples that have an alteration in one or more query genes

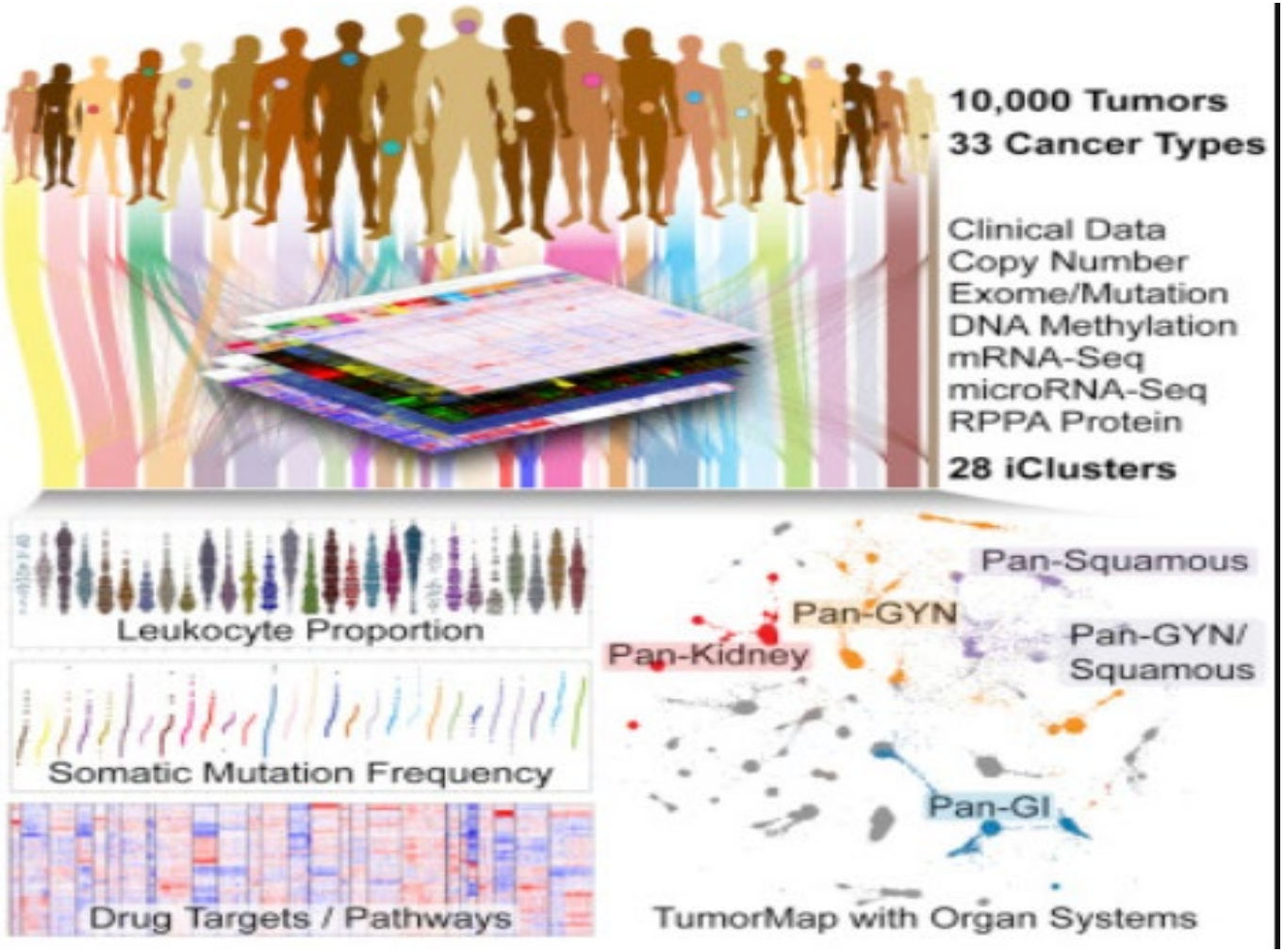
Advanced feature: use this list as a custom sample list to run a new query in only the subset of samples with a particular genetic alteration.

List of all samples with summary classification:
0 = no alteration in any query gene
1 = alteration in one or more query genes

Link to this page

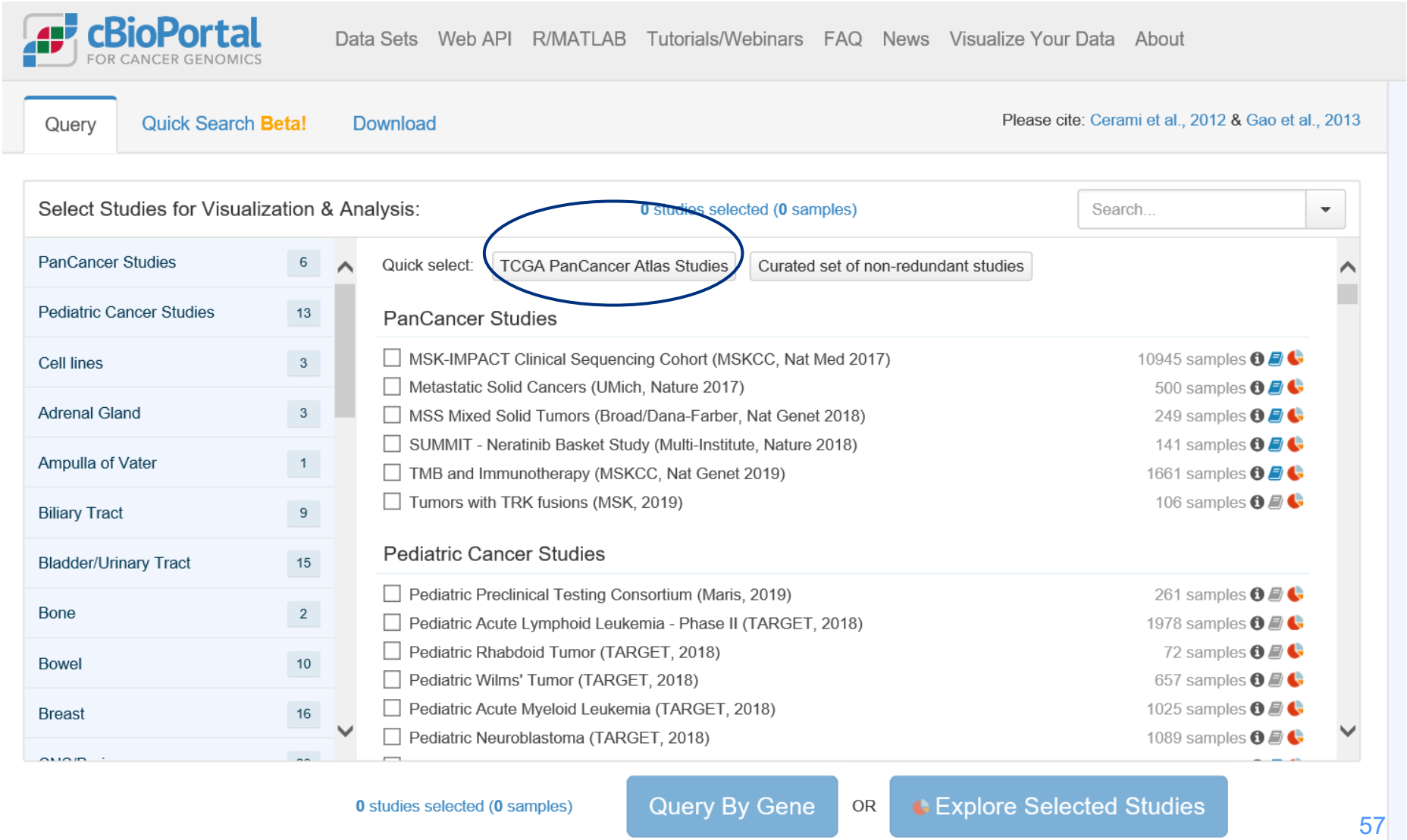
PAN-can example

PAN-Can datasets and analyses

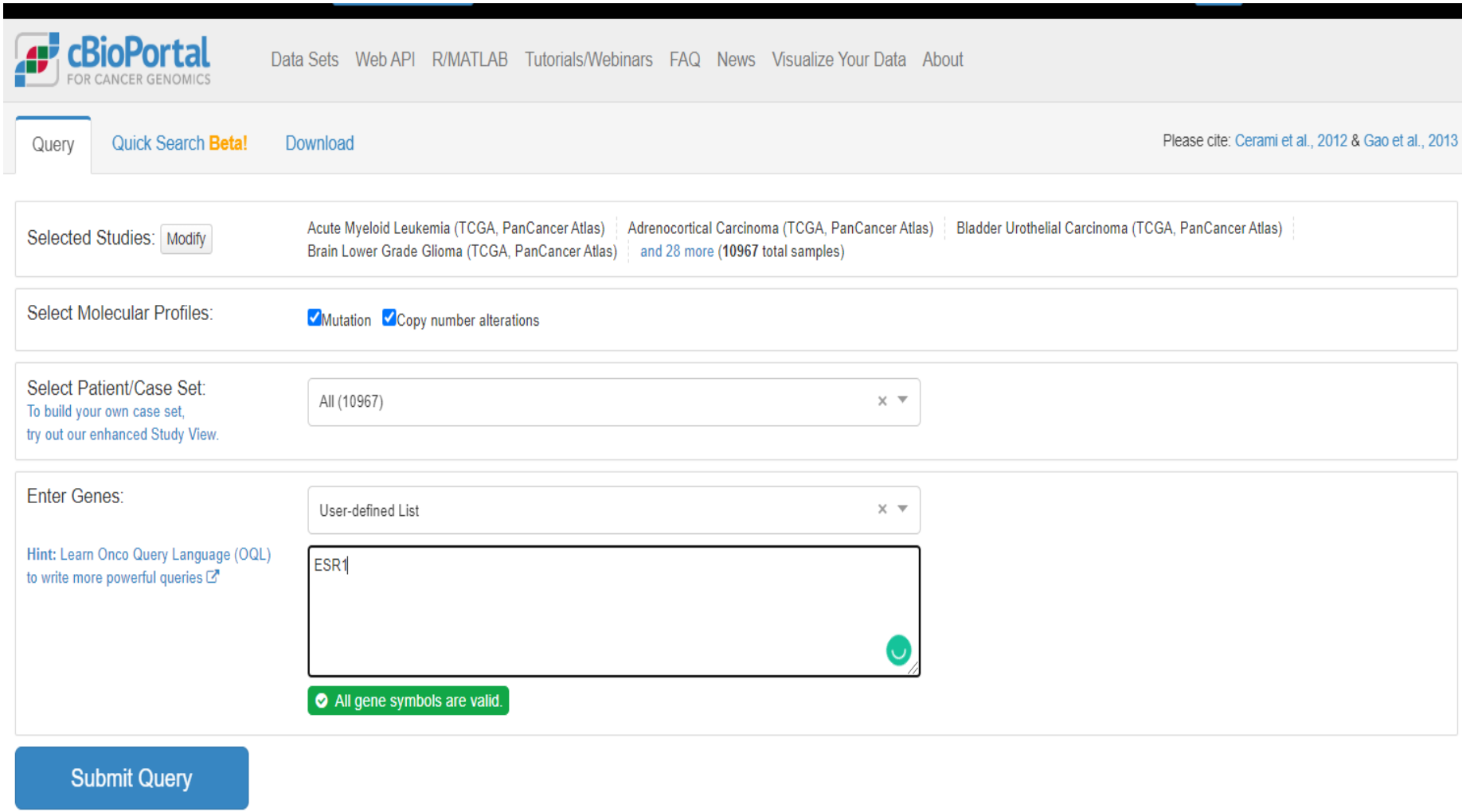


https://www.sciencedirect.com/science/article/pii/S0092867418303027

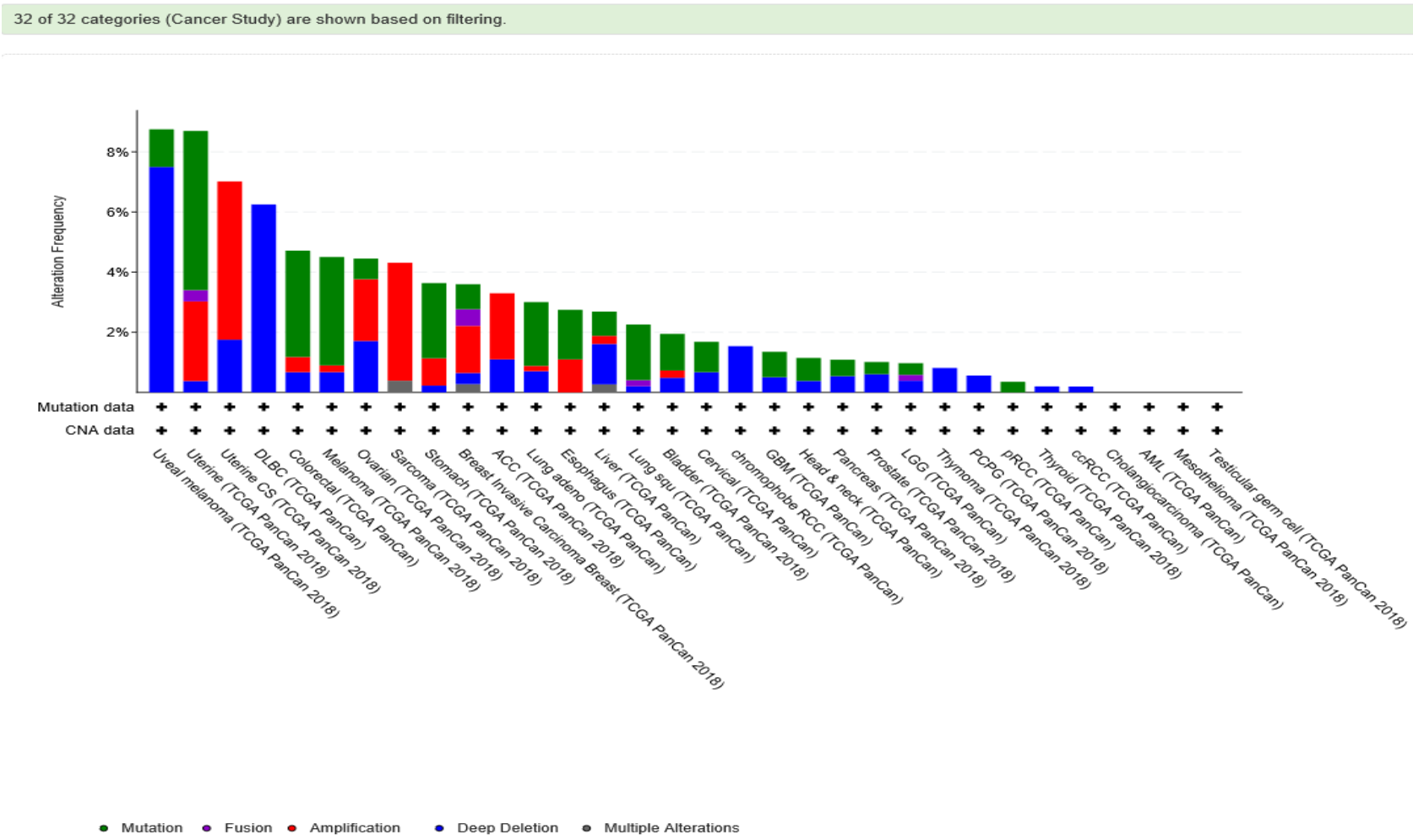
PAN-Can datasets and analyses



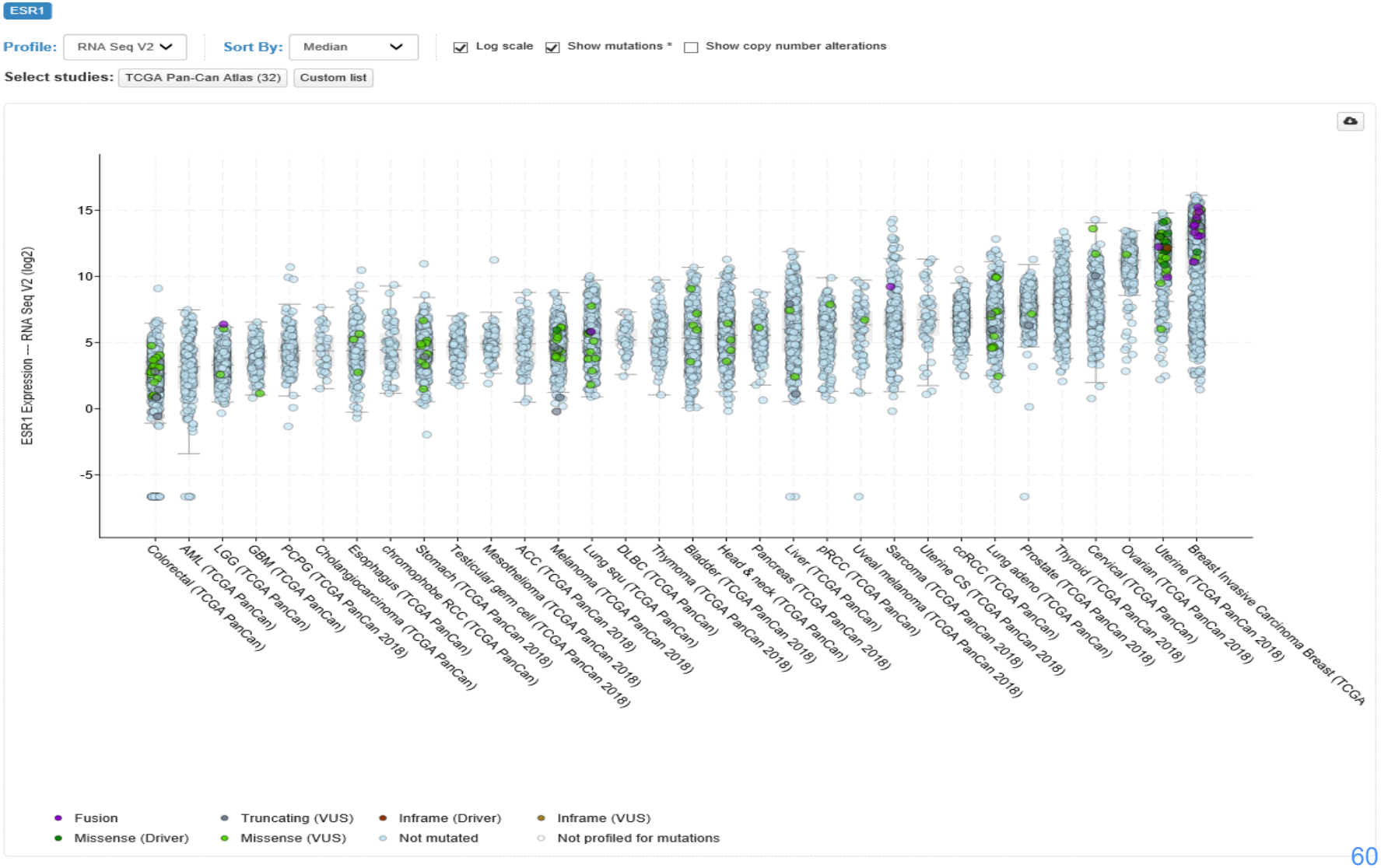
PAN-Can datasets and analyses



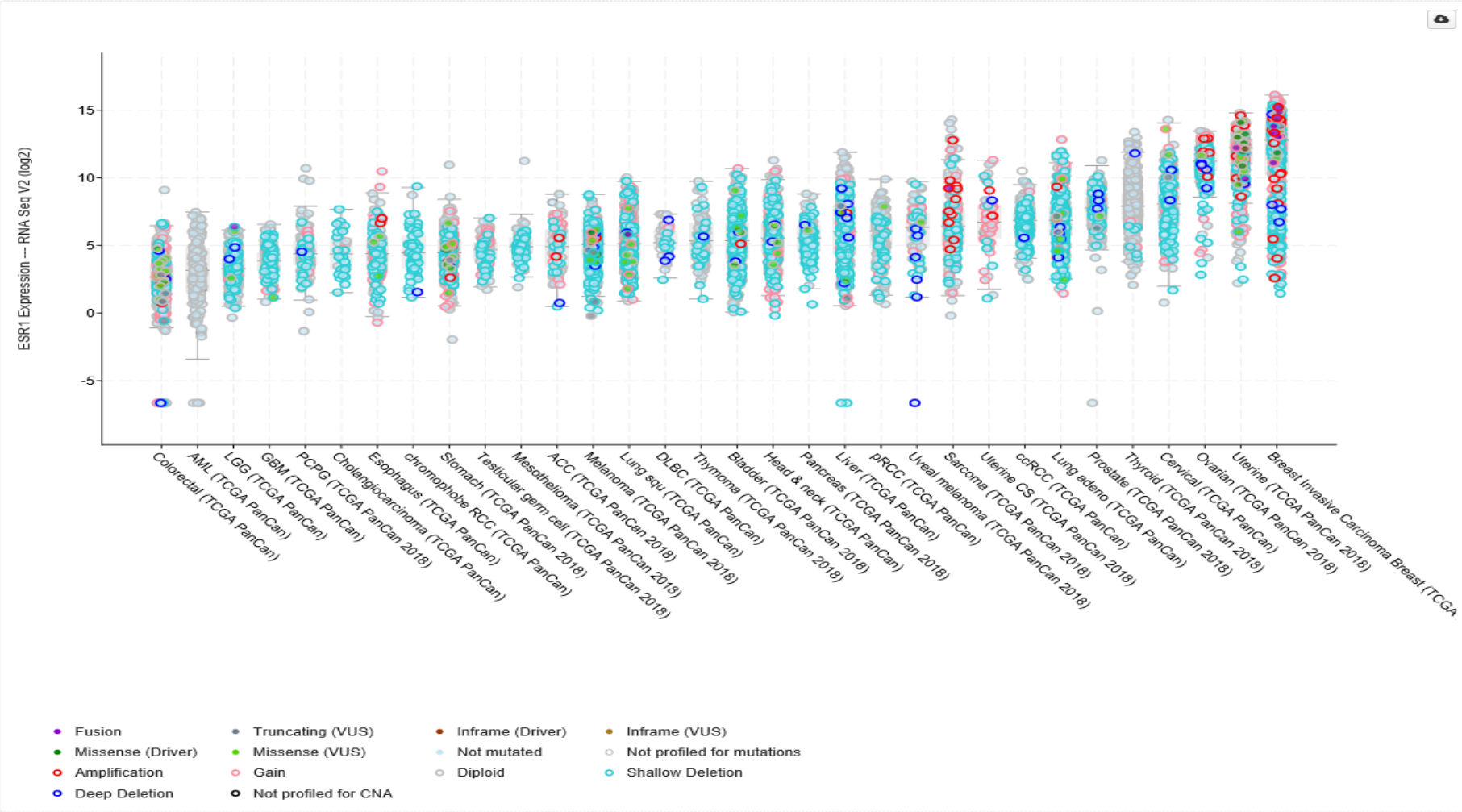
PAN-Can datasets and analyses



PAN-Can datasets and analyses

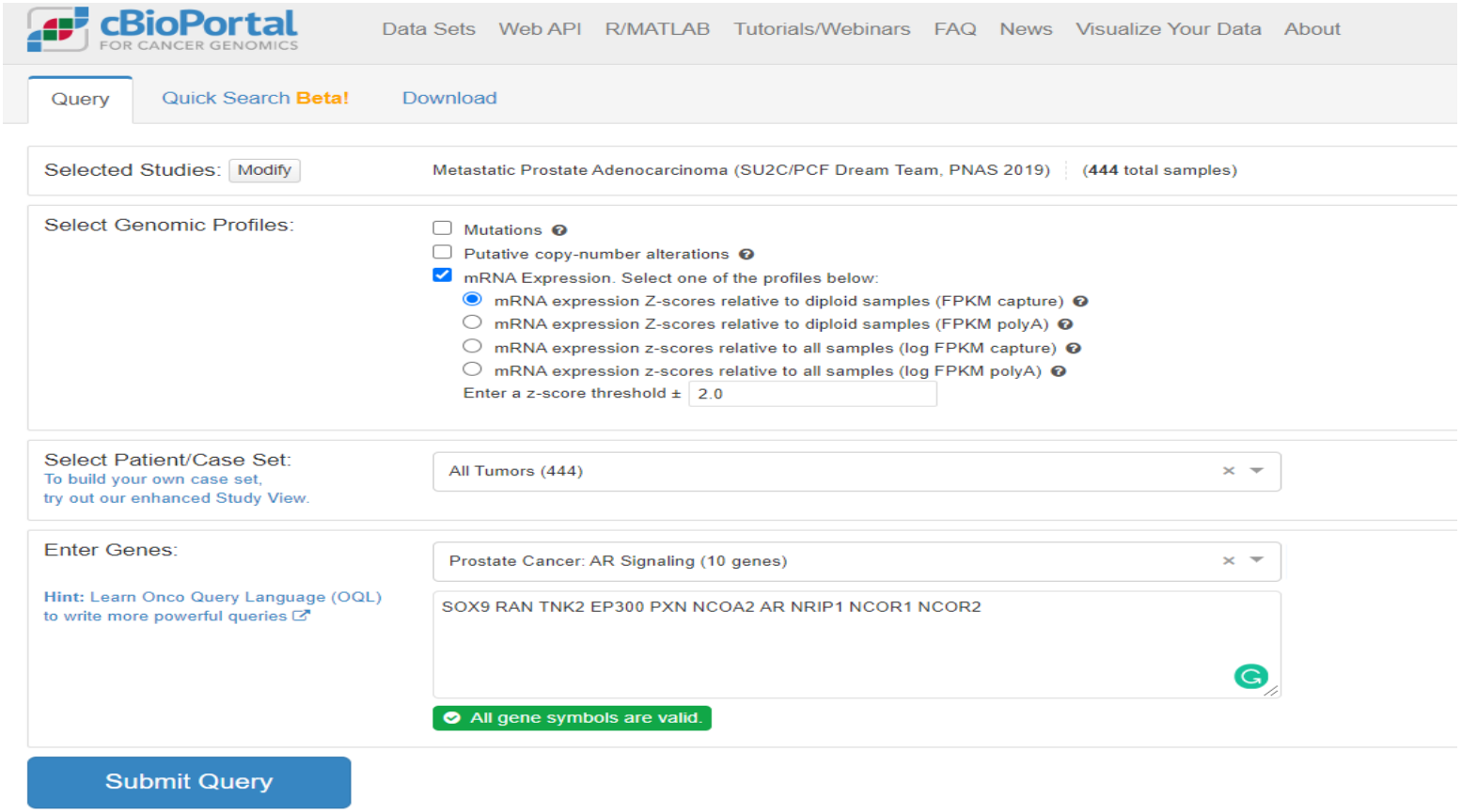


PAN-Can datasets and analyses

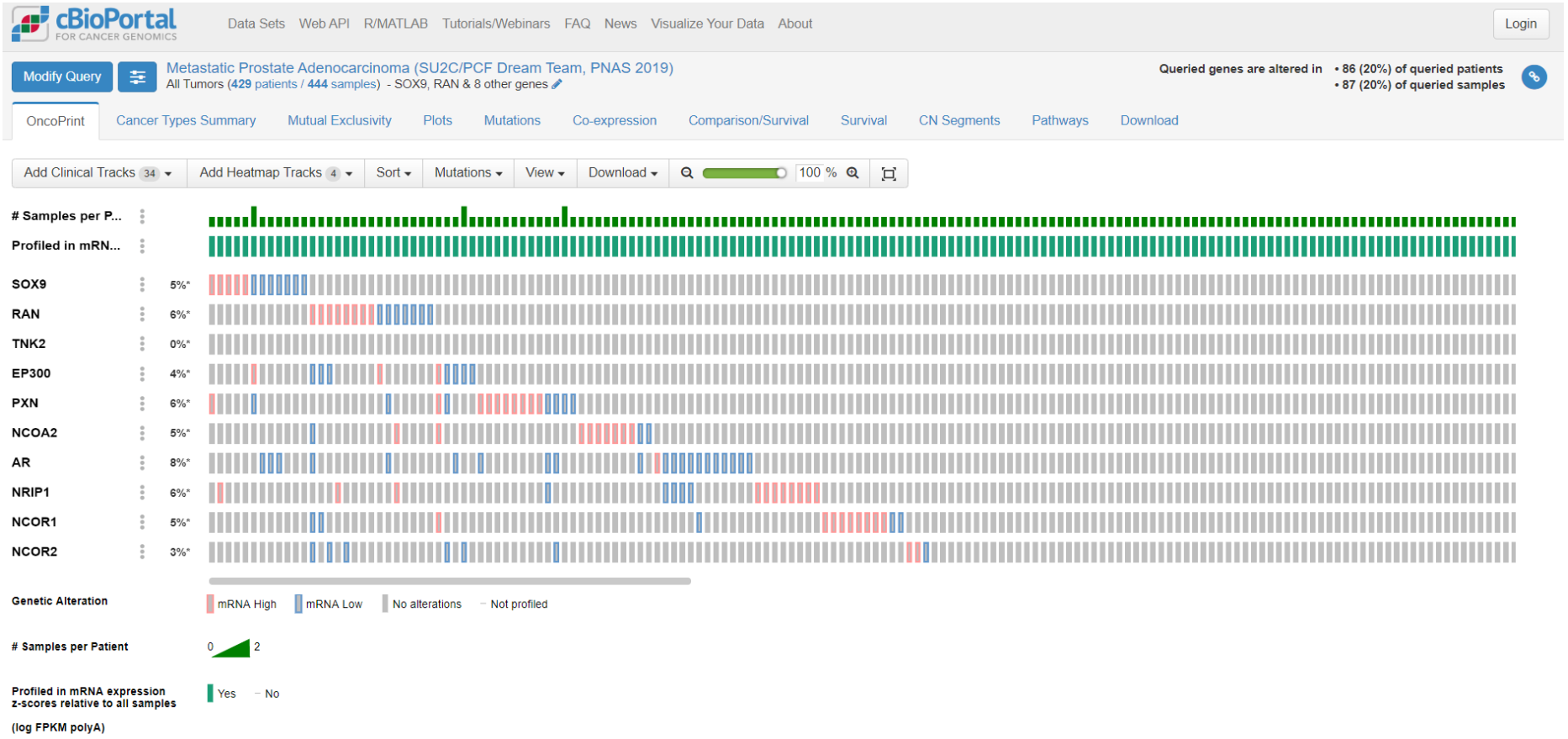


* Driver annotation settings are located in the Mutation Color menu of the Oncoprint.

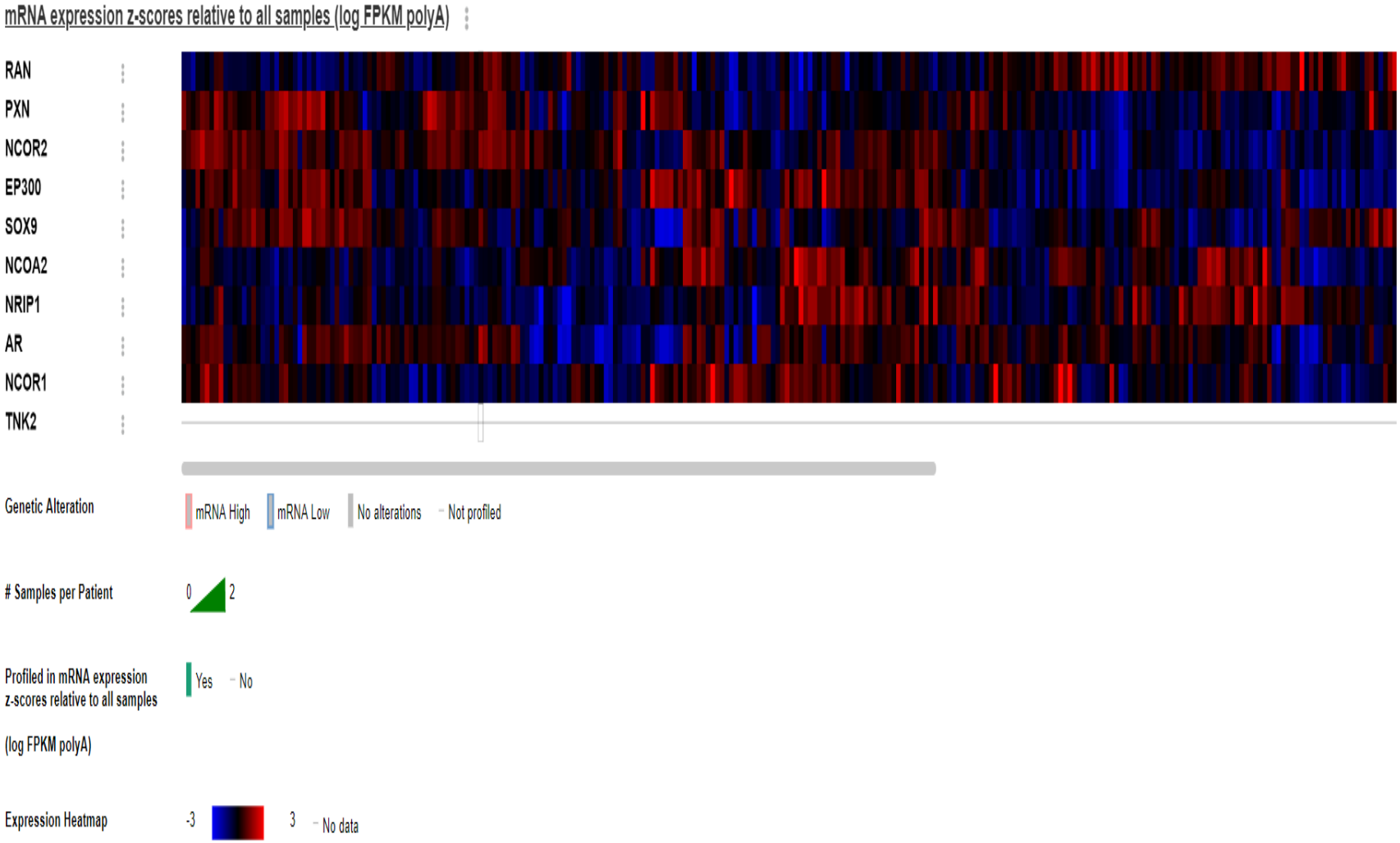
Clustering example



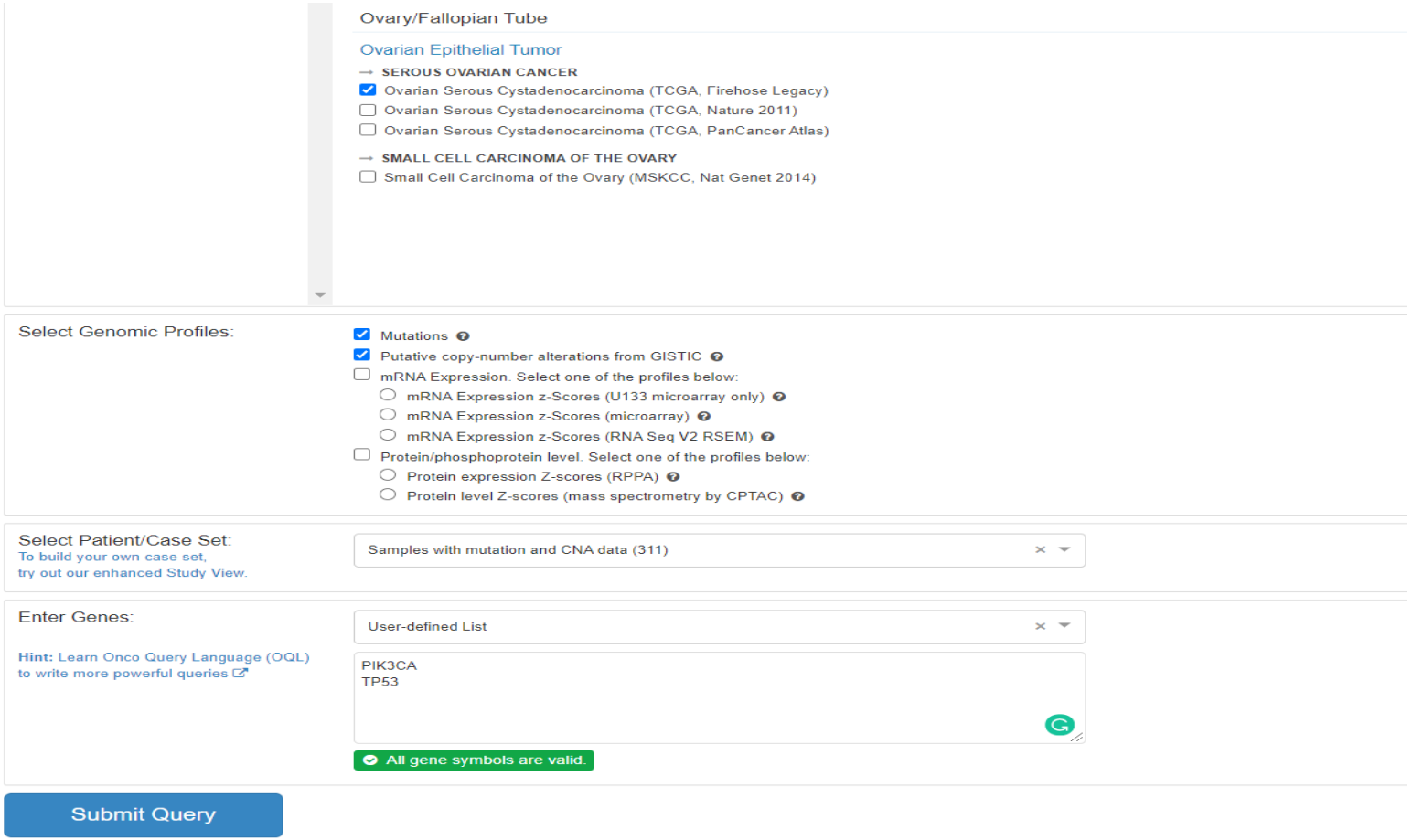
Clustering example



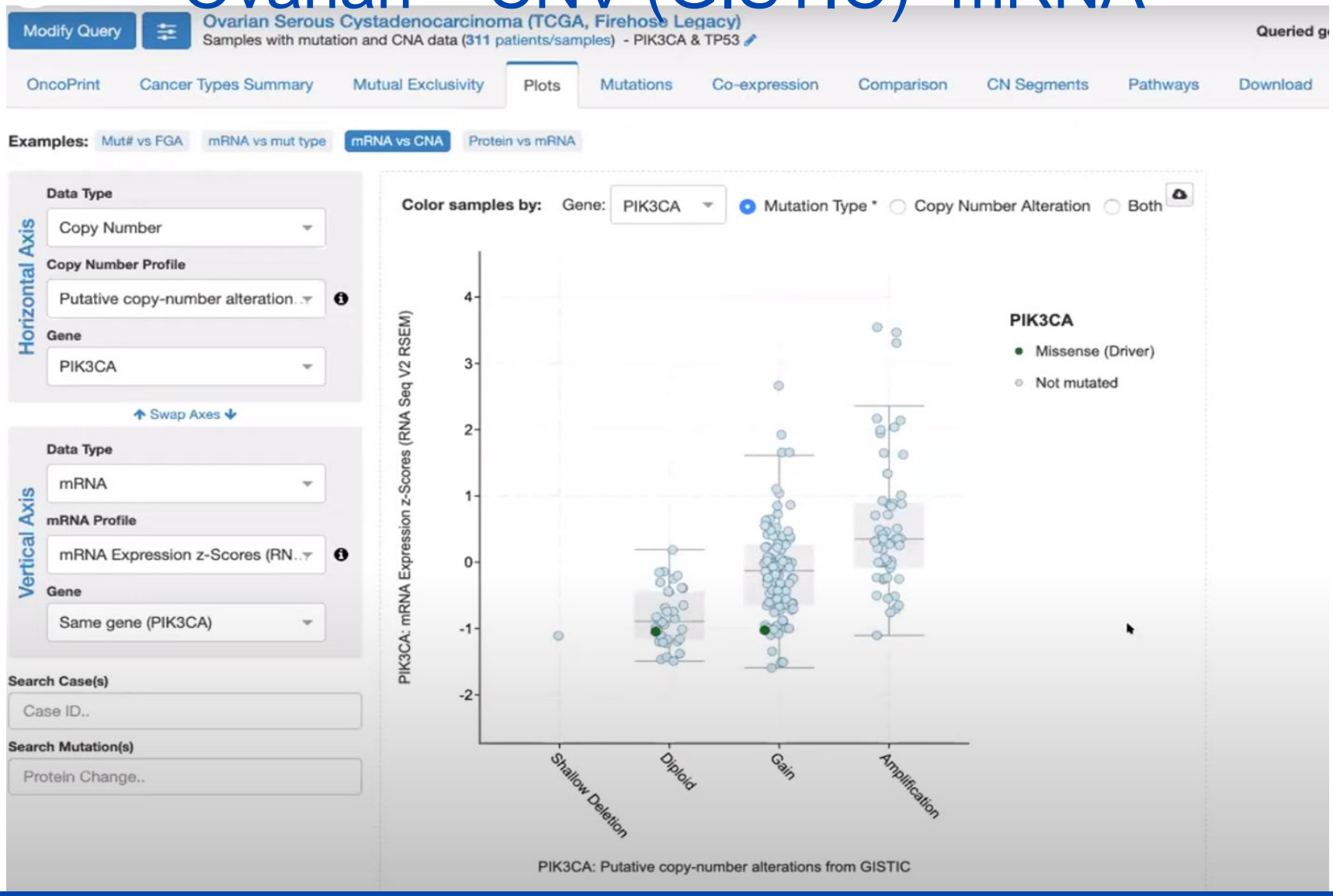
Clustering example



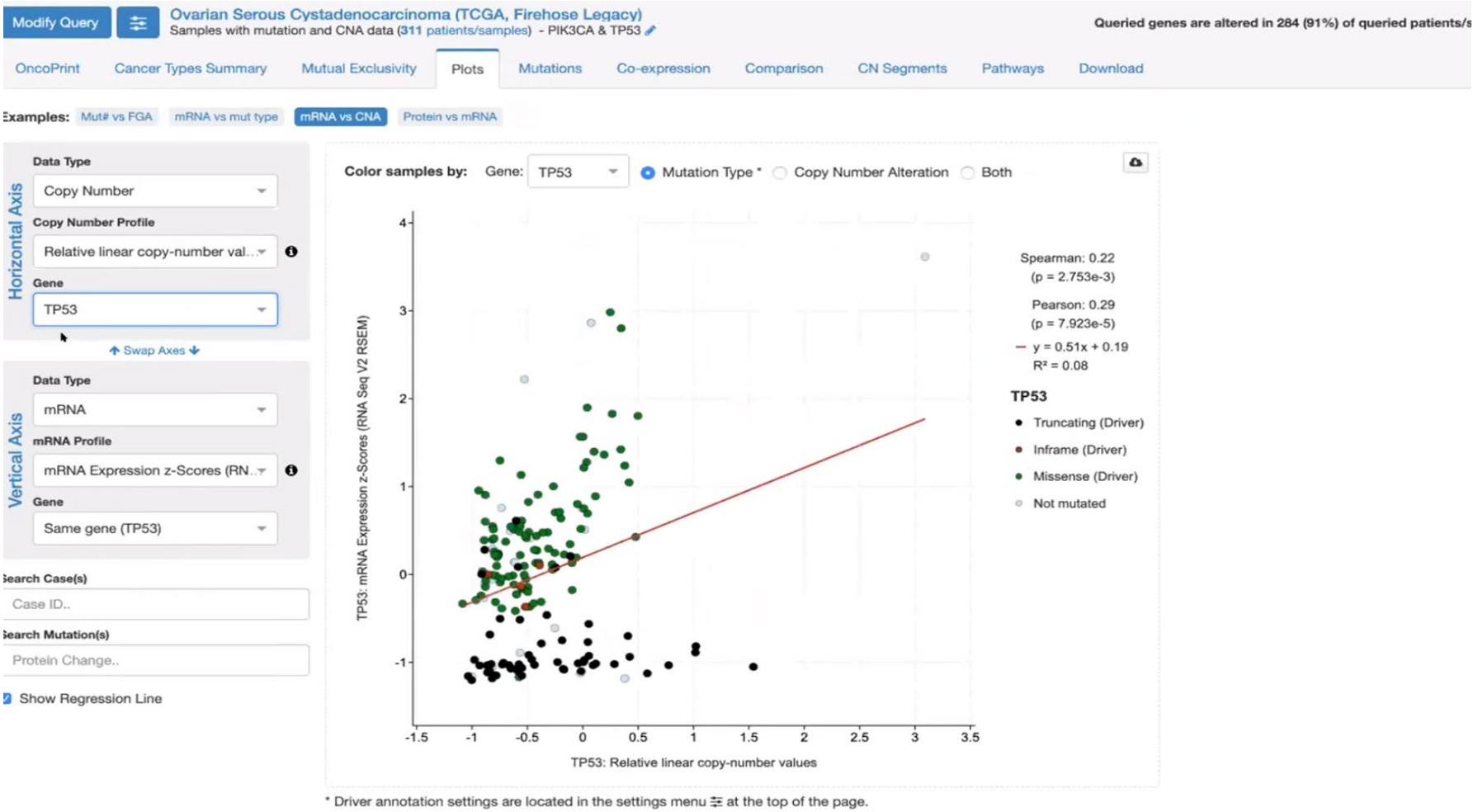
Ovarian – multi-omics example



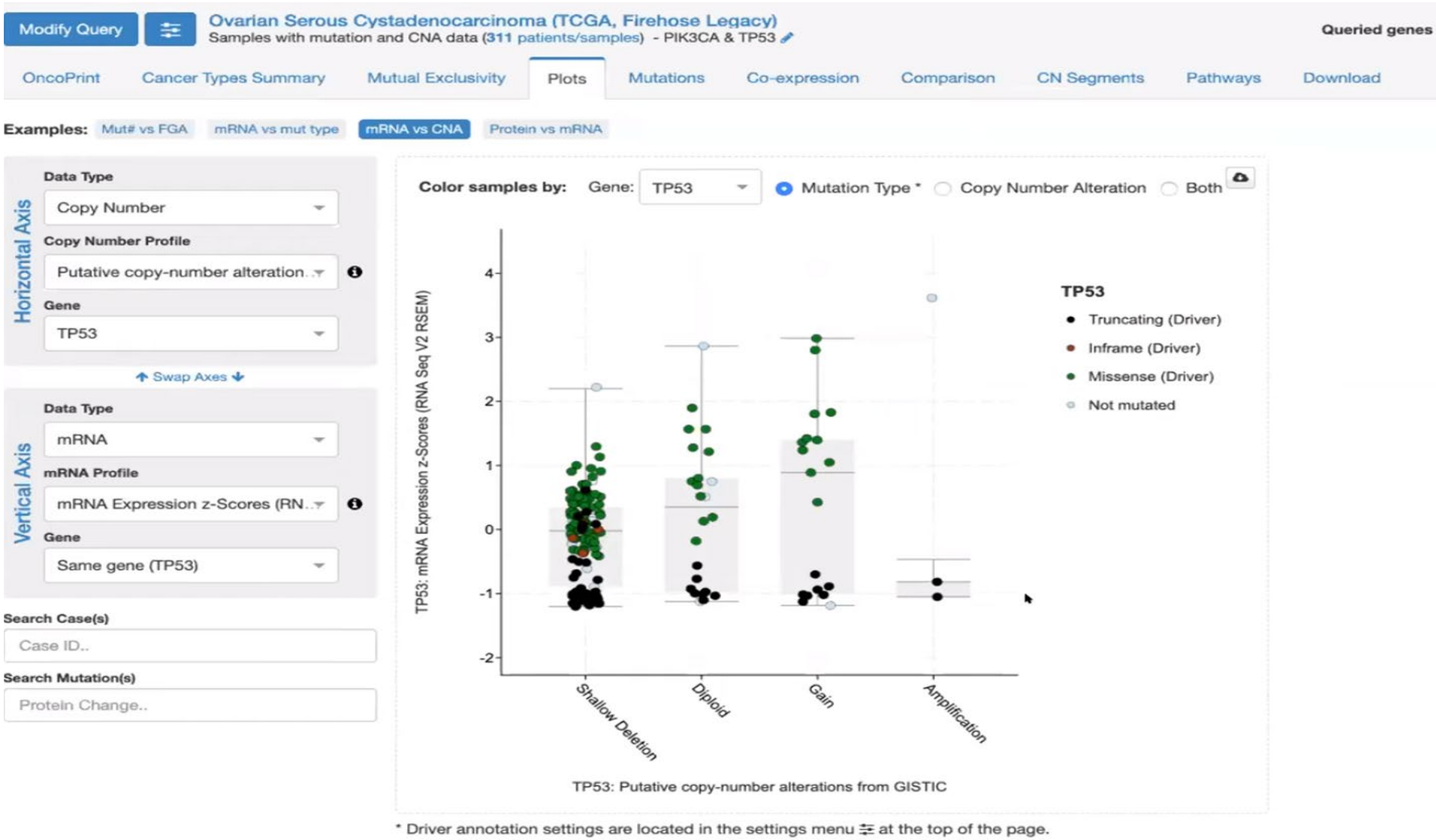
Ovarian – CNV (GISTIC) -mRNA



Ovarian – mRNA-CNV (linear data) (scatter plot)

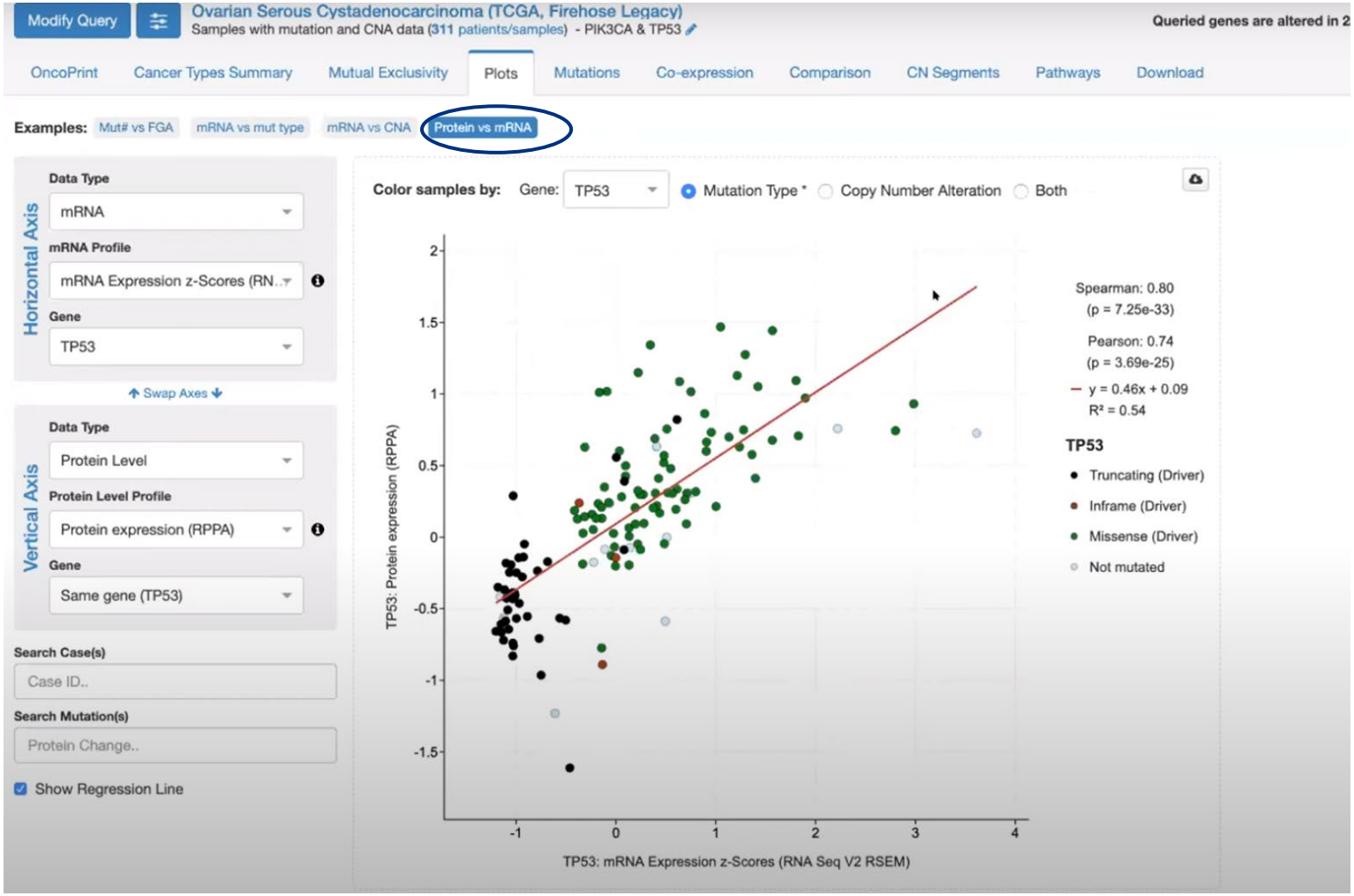


Ovarian – TP53 mRNA,CNV and mutations

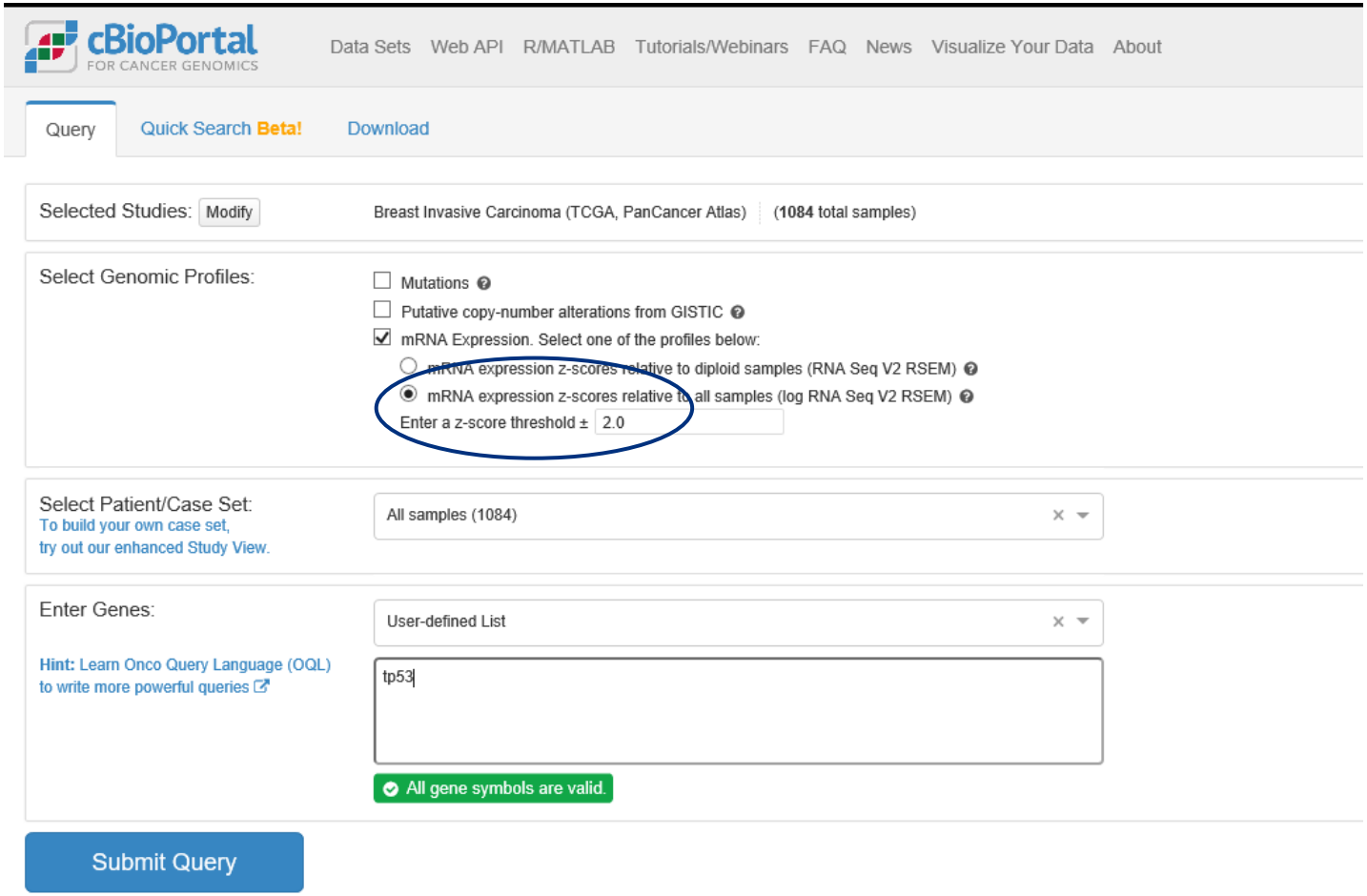


Ovarian – mRNA-mutation

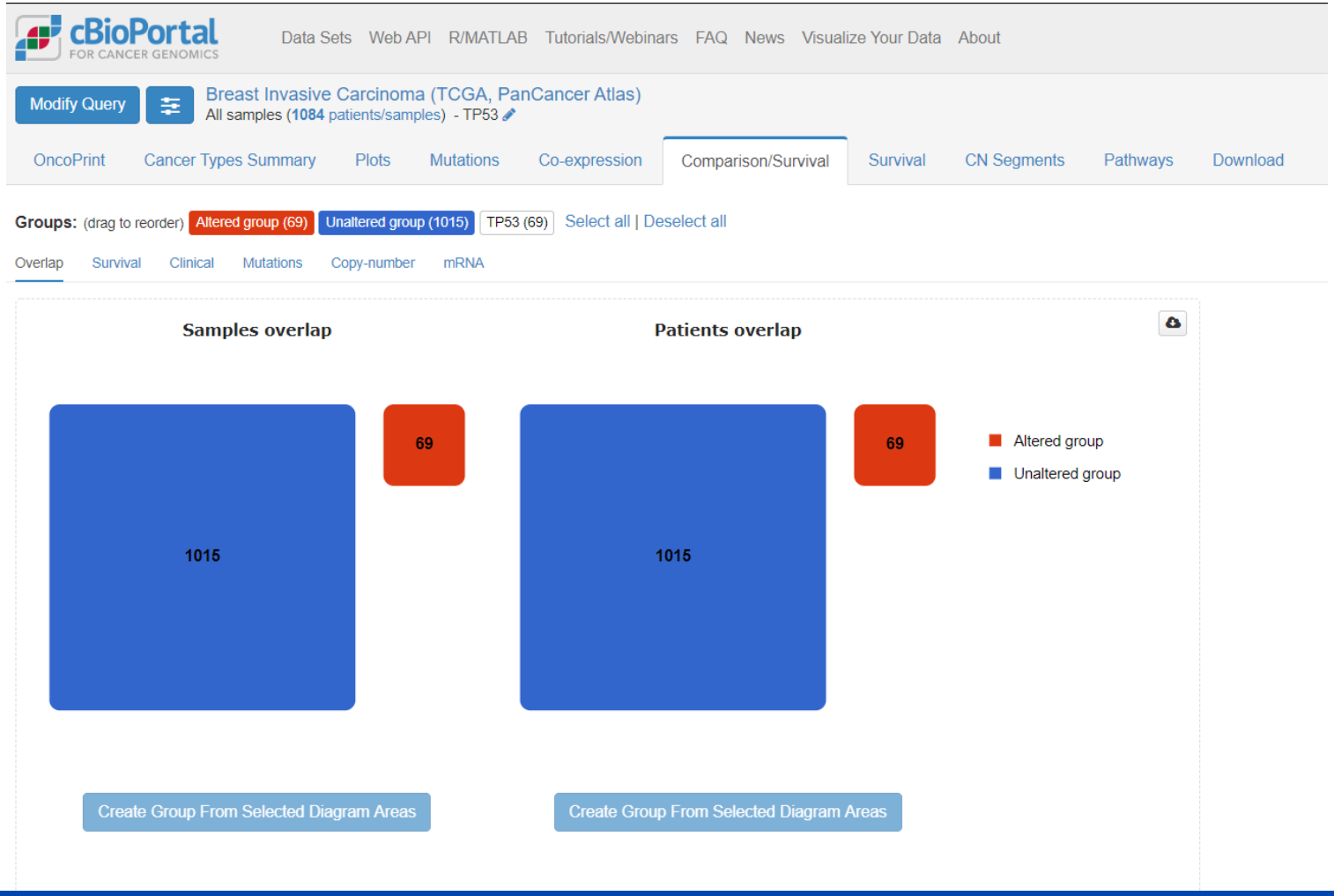
Ovarian – mRNA-protein



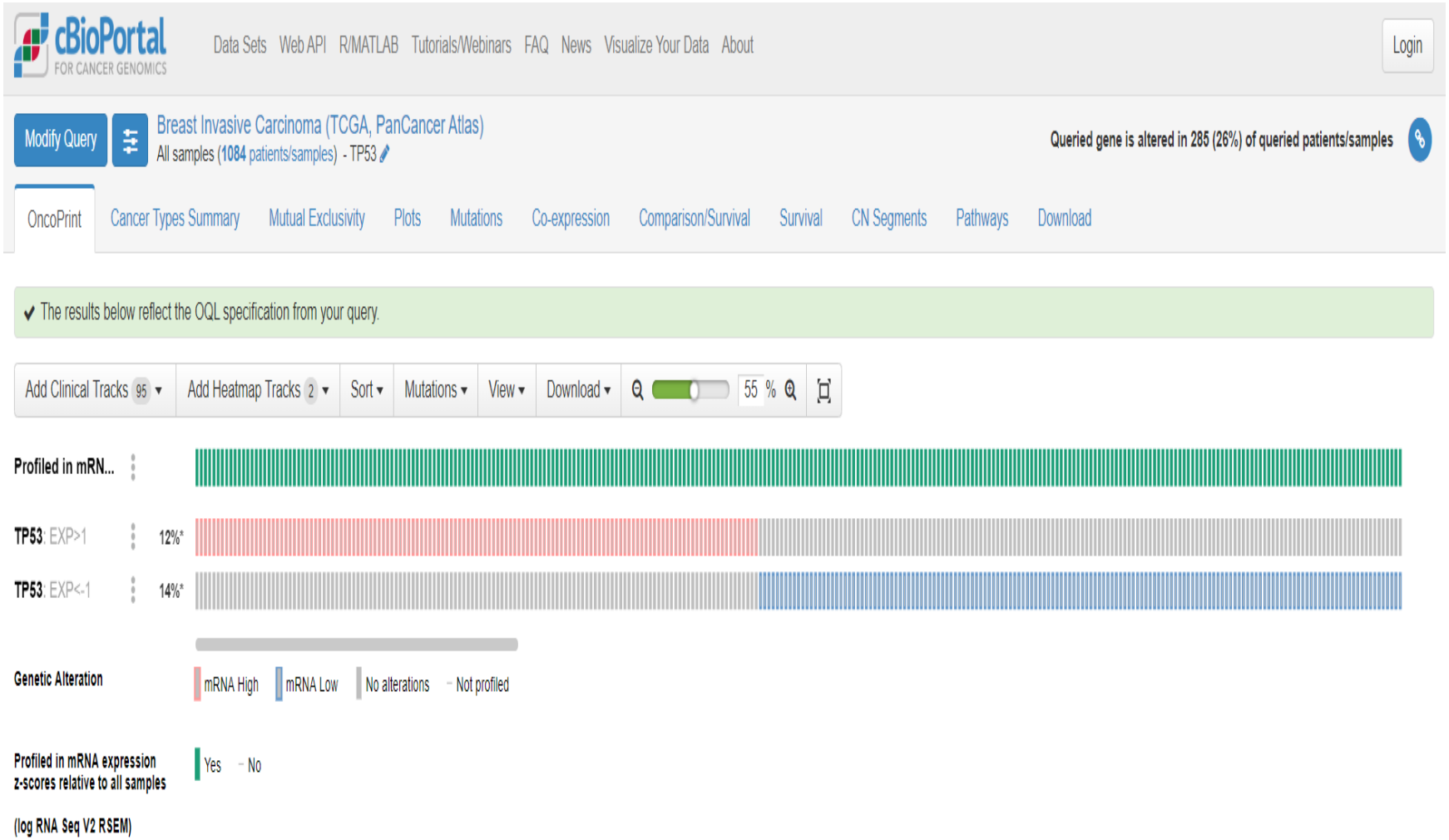
Breast Cancer (TP53 + and TP53 - analysis)



Breast Cancer (TP53+/TP53-) analysis

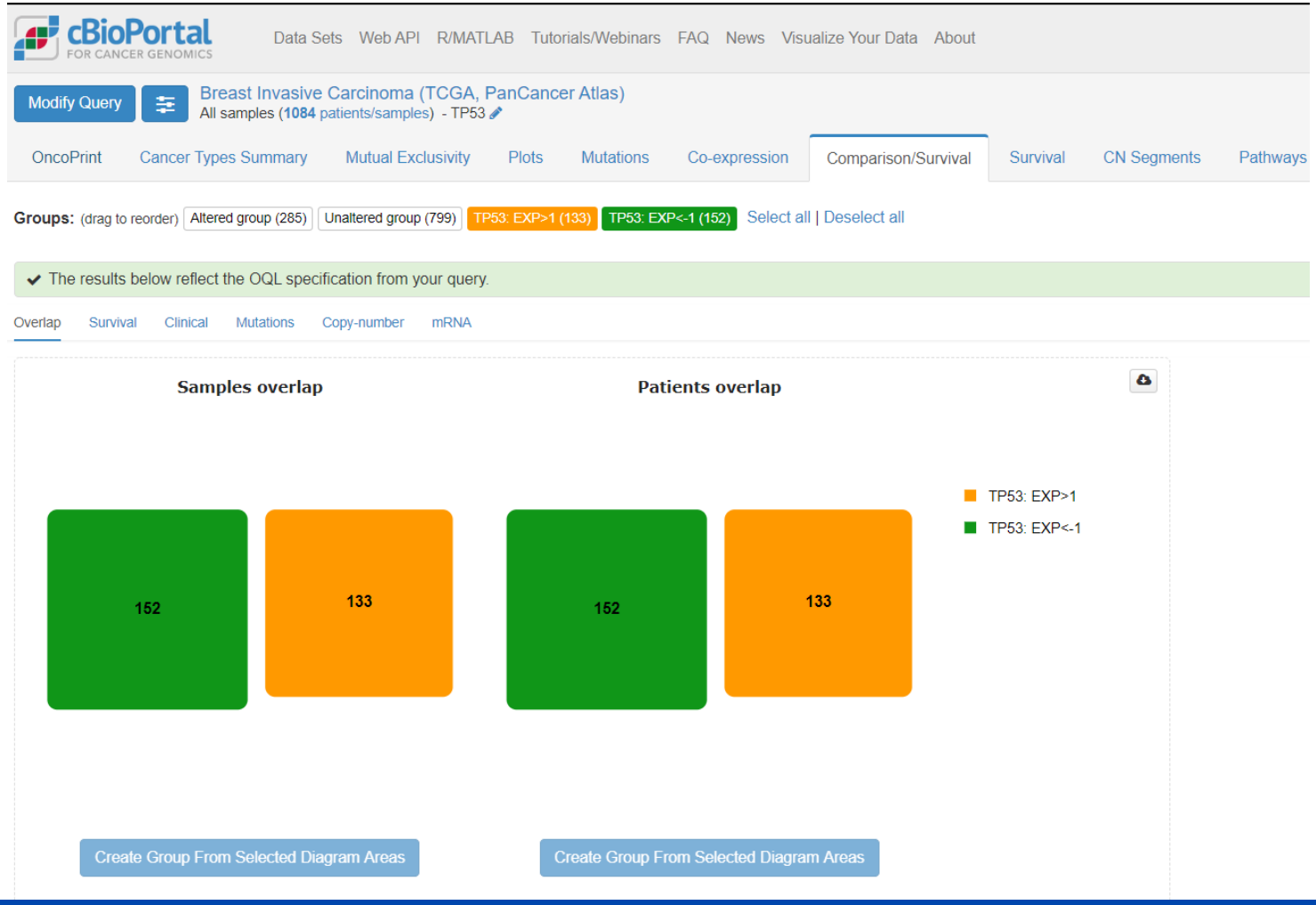


Breast Cancer (TP53 high and TP53 low)

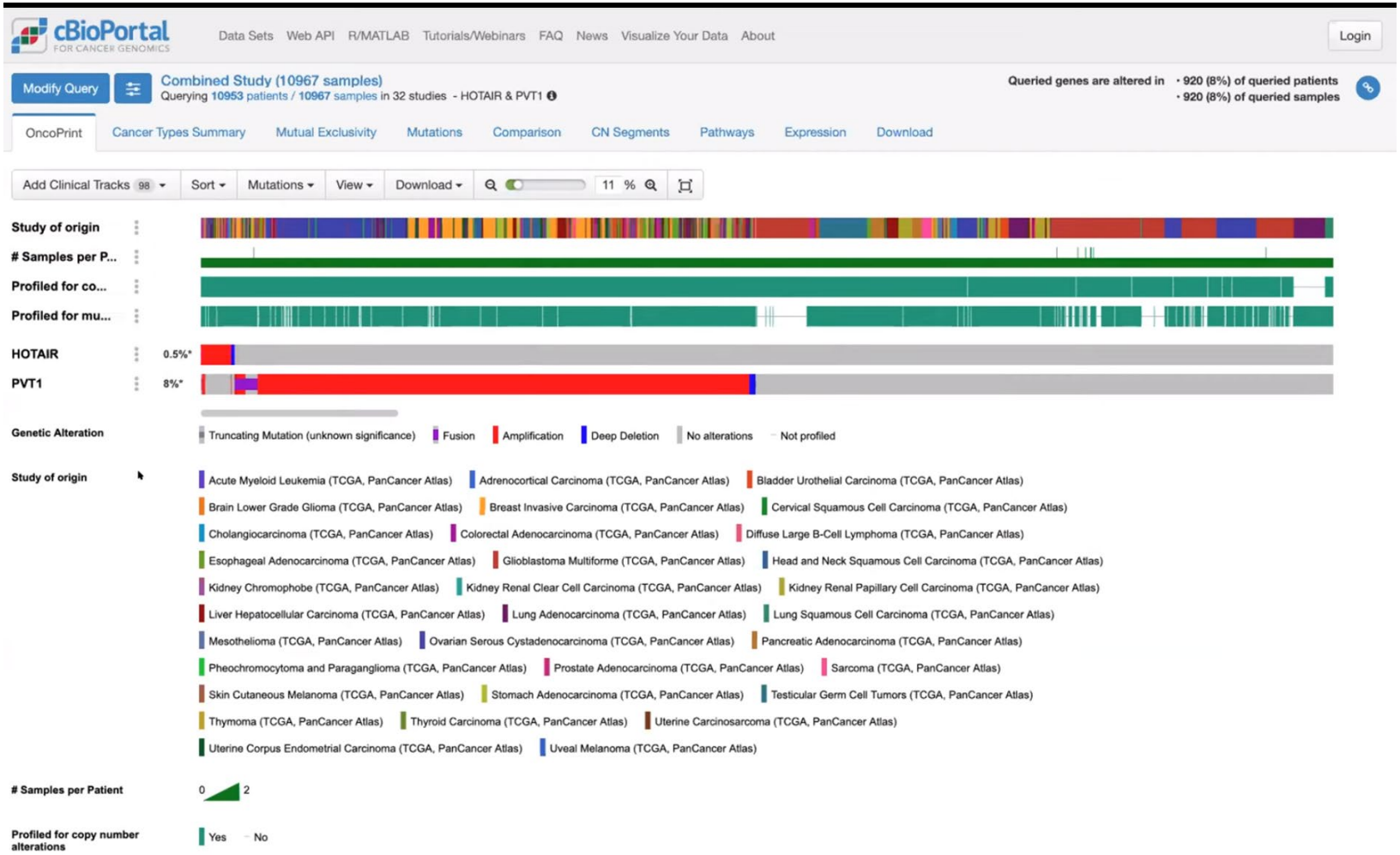


Breast Cancer (TP53 high and TP53 low)

Breast Cancer (TP53 high and TP53 low)

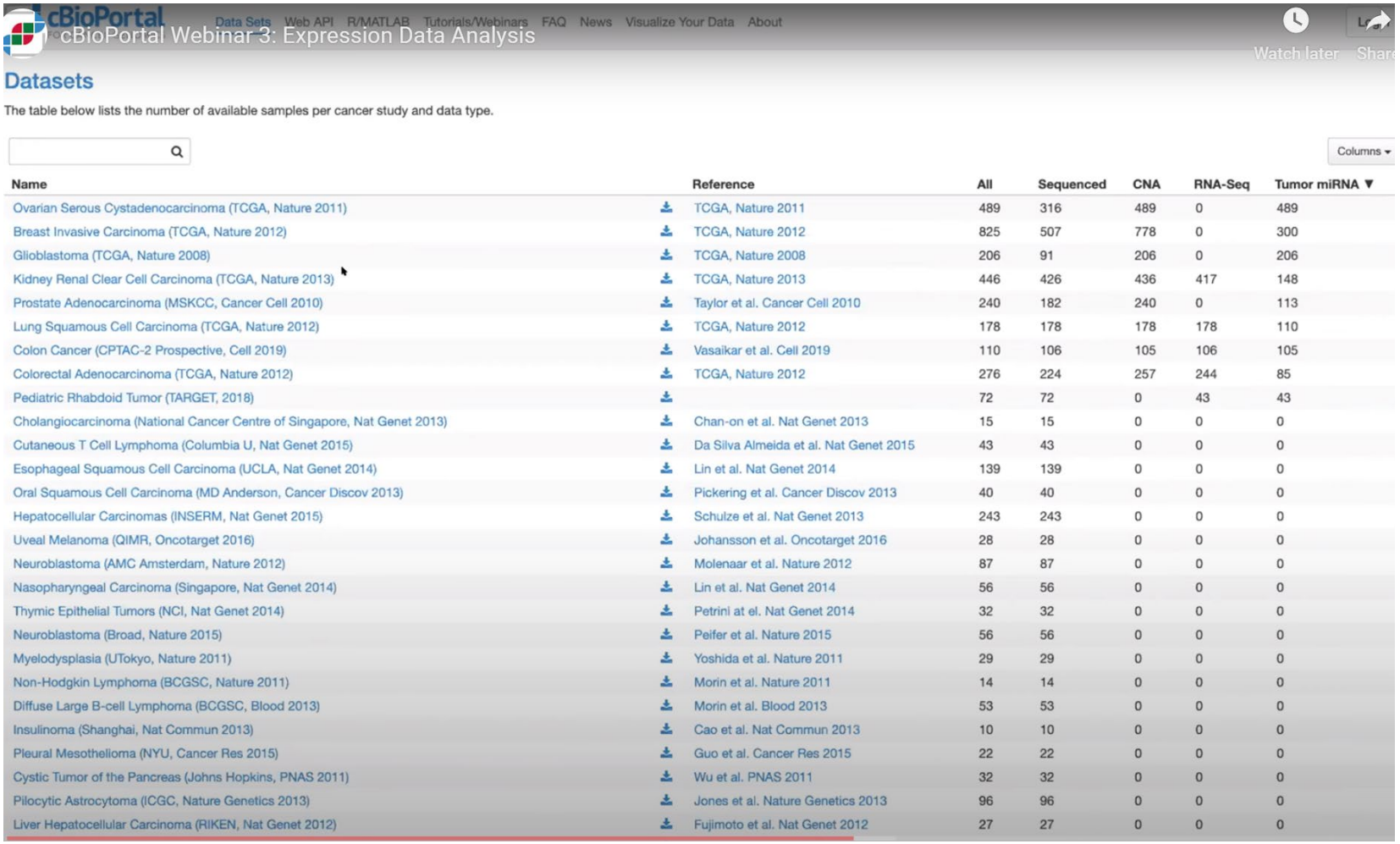


Long-noncoding RNA example



Long-noncoding RNA example

miRNA example



Name	Reference	All	Sequenced	CNA	RNA-Seq	Tumor miRNA ▼
Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011)	TCGA, Nature 2011	489	316	489	0	489
Breast Invasive Carcinoma (TCGA, Nature 2012)	TCGA, Nature 2012	825	507	778	0	300
Glioblastoma (TCGA, Nature 2008)	TCGA, Nature 2008	206	91	206	0	206
Kidney Renal Clear Cell Carcinoma (TCGA, Nature 2013)	TCGA, Nature 2013	446	426	436	417	148
Prostate Adenocarcinoma (MSKCC, Cancer Cell 2010)	Taylor et al. Cancer Cell 2010	240	182	240	0	113
Lung Squamous Cell Carcinoma (TCGA, Nature 2012)	TCGA, Nature 2012	178	178	178	178	110
Colon Cancer (CPTAC-2 Prospective, Cell 2019)	Vasaikar et al. Cell 2019	110	106	105	106	105
Colorectal Adenocarcinoma (TCGA, Nature 2012)	TCGA, Nature 2012	276	224	257	244	85
Pediatric Rhabdoid Tumor (TARGET, 2018)		72	72	0	43	43
Cholangiocarcinoma (National Cancer Centre of Singapore, Nat Genet 2013)	Chan-on et al. Nat Genet 2013	15	15	0	0	0
Cutaneous T Cell Lymphoma (Columbia U, Nat Genet 2015)	Da Silva Almeida et al. Nat Genet 2015	43	43	0	0	0
Esophageal Squamous Cell Carcinoma (UCLA, Nat Genet 2014)	Lin et al. Nat Genet 2014	139	139	0	0	0
Oral Squamous Cell Carcinoma (MD Anderson, Cancer Discov 2013)	Pickering et al. Cancer Discov 2013	40	40	0	0	0
Hepatocellular Carcinomas (INSERM, Nat Genet 2015)	Schulze et al. Nat Genet 2013	243	243	0	0	0
Uveal Melanoma (QIMR, Oncotarget 2016)	Johansson et al. Oncotarget 2016	28	28	0	0	0
Neuroblastoma (AMC Amsterdam, Nature 2012)	Molenaar et al. Nature 2012	87	87	0	0	0
Nasopharyngeal Carcinoma (Singapore, Nat Genet 2014)	Lin et al. Nat Genet 2014	56	56	0	0	0
Thymic Epithelial Tumors (NCI, Nat Genet 2014)	Petrini at el. Nat Genet 2014	32	32	0	0	0
Neuroblastoma (Broad, Nature 2015)	Peifer et al. Nature 2015	56	56	0	0	0
Myelodysplasia (UTokyo, Nature 2011)	Yoshida et al. Nature 2011	29	29	0	0	0
Non-Hodgkin Lymphoma (BCGSC, Nature 2011)	Morin et al. Nature 2011	14	14	0	0	0
Diffuse Large B-cell Lymphoma (BCGSC, Blood 2013)	Morin et al. Blood 2013	53	53	0	0	0
Insulinoma (Shanghai, Nat Commun 2013)	Cao et al. Nat Commun 2013	10	10	0	0	0
Pleural Mesothelioma (NYU, Cancer Res 2015)	Guo et al. Cancer Res 2015	22	22	0	0	0
Cystic Tumor of the Pancreas (Johns Hopkins, PNAS 2011)	Wu et al. PNAS 2011	32	32	0	0	0
Pilocytic Astrocytoma (ICGC, Nature Genetics 2013)	Jones et al. Nature Genetics 2013	96	96	0	0	0
Liver Hepatocellular Carcinoma (RIKEN, Nat Genet 2012)	Fujimoto et al. Nat Genet 2012	27	27	0	0	0

RNA and miRNA example

cBioPortal FOR CANCER GENOMICS — Data Sets | Web API | R/MATLAB | Tutorials/Webinars | FAQ | News | Visualize Your Data | About | Login

Modify Query — Breast Invasive Carcinoma (TCGA, Nature 2012)
All samples (825 patients/samples) - MGMT

Queried gene is altered in 12 (1%) of queried patients/samples

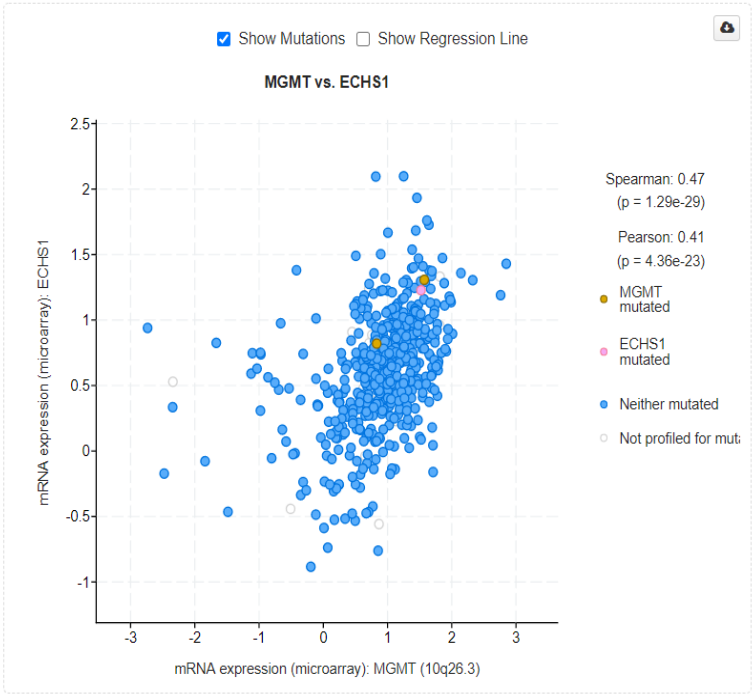
OncoPrint | Cancer Types Summary | Plots | Mutations | Co-expression | Comparison/Survival | Survival | CN Segments | Pathways | Download

MGMT

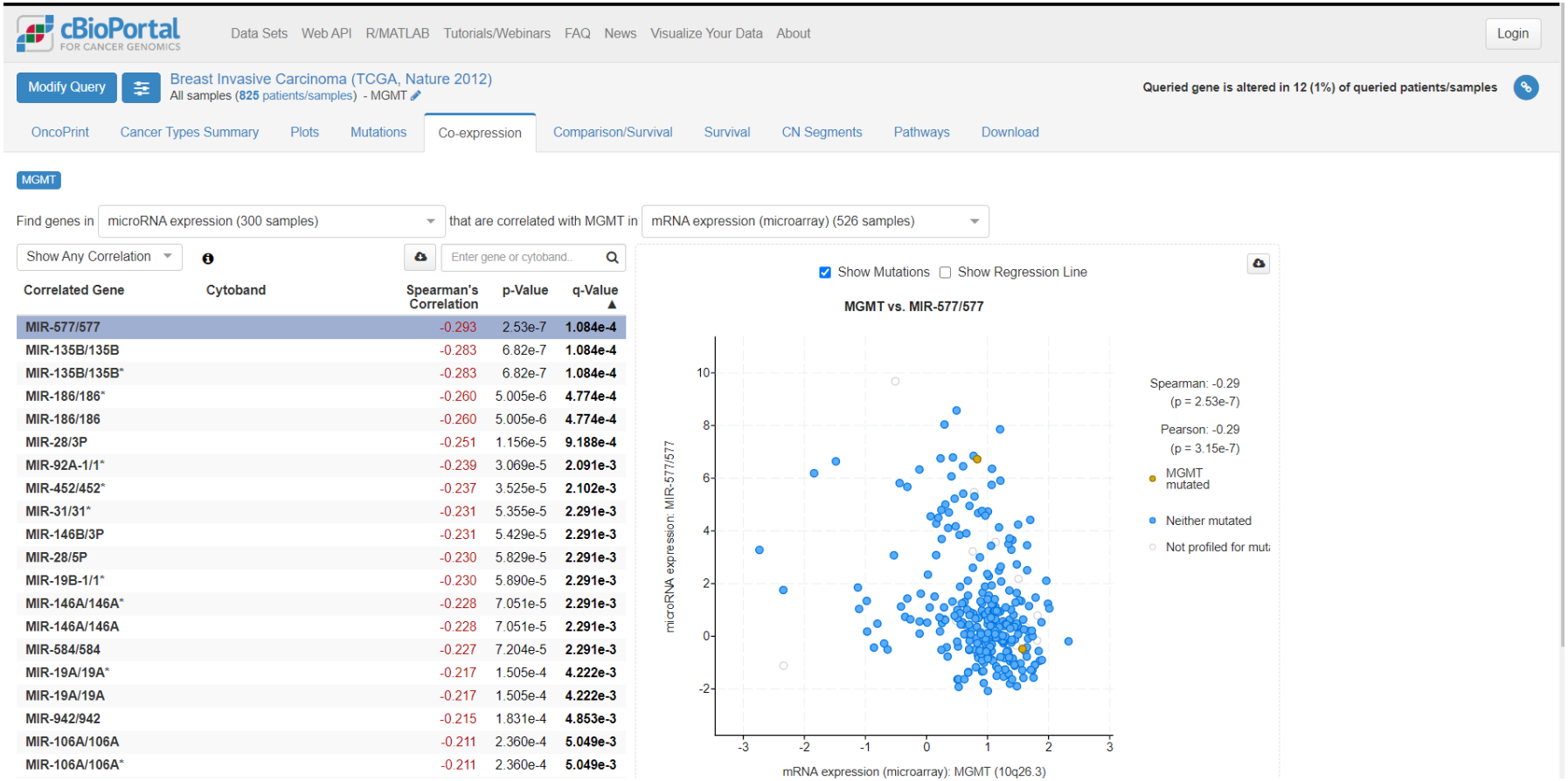
Find genes in mRNA expression (microarray) (526 samples) that are correlated with MGMT in mRNA expression (microarray) (526 samples)

Show Any Correlation — Enter gene or cytoband..

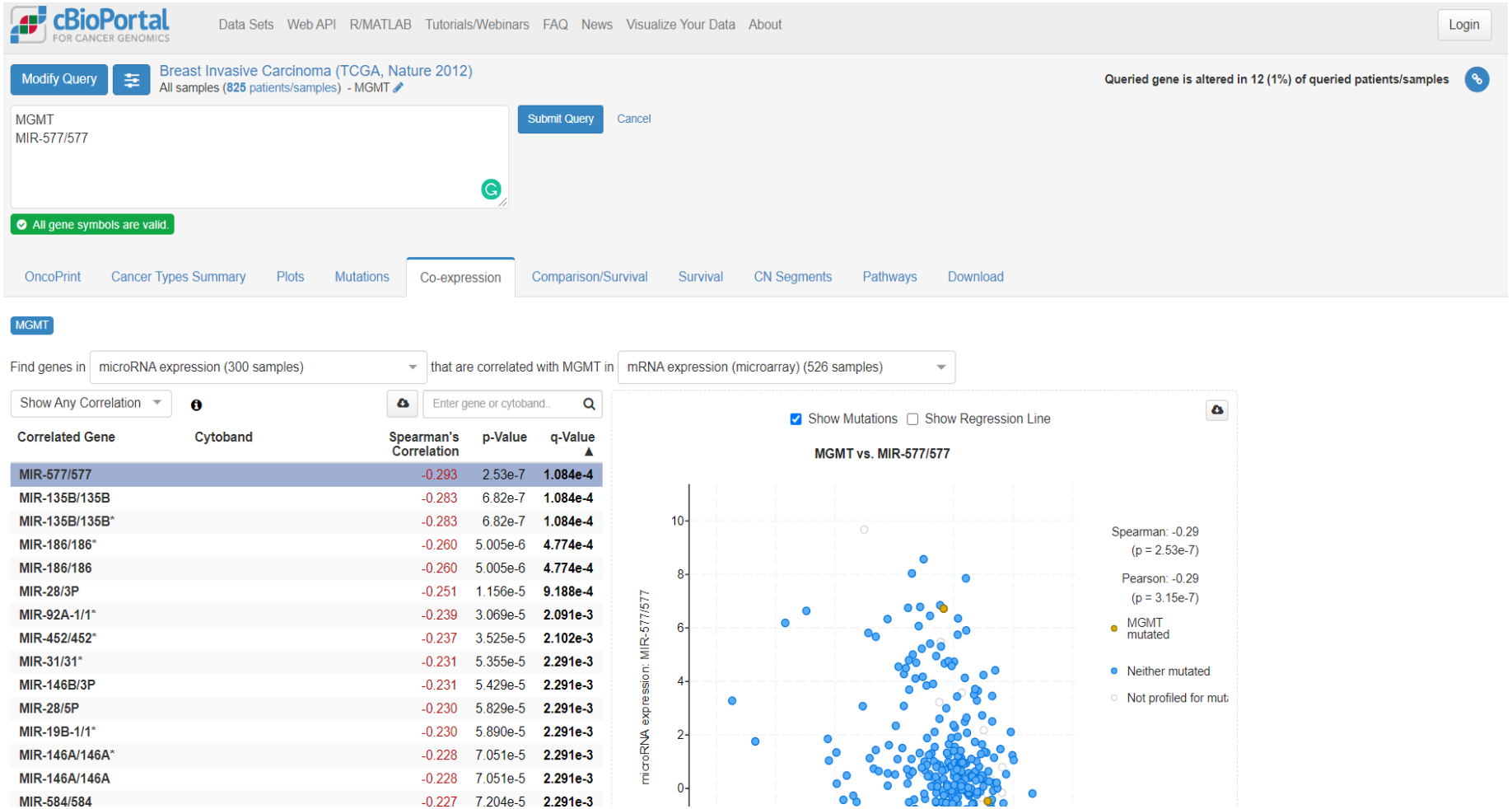
Correlated Gene	Cytoband	Spearman's Correlation	p-Value	q-Value ▲
ECHS1	10q26.3	0.465	1.29e-29	**2.23e-25**
PAOX	10q26.3	0.437	6.04e-26	**5.21e-22**
LRRC27	10q26.3	0.410	9.95e-23	**5.72e-19**
BTRC	10q24.32	0.396	3.26e-21	**1.41e-17**
HIF1AN	10q24.31	0.388	2.30e-20	**7.91e-17**
MRPL43	10q24.31	0.383	8.85e-20	**2.32e-16**
CHCHD5	2q14.1	0.382	9.43e-20	**2.32e-16**
GSTO2	10q25.1	0.376	4.53e-19	**9.77e-16**
ZNF511	10q26.3	0.375	5.47e-19	**1.05e-15**
CCS	11q13.2	0.367	3.02e-18	**5.21e-15**
PPP2R2D	10q26.3	0.366	3.99e-18	**6.00e-15**
C12ORF10	12q13.13	0.366	4.18e-18	**6.00e-15**
PWWP2B	10q26.3	0.362	9.27e-18	**1.23e-14**
FBXL15	10q24.32	0.361	1.16e-17	**1.42e-14**
TUBGCP2	10q26.3	0.359	1.73e-17	**1.98e-14**
PEX11G	19p13.2	0.358	2.46e-17	**2.65e-14**
MAP4K4	2q11.2	-0.353	6.71e-17	**6.80e-14**
FAM204A	10q26.11	0.350	1.29e-16	**1.24e-13**
ARL3	10q24.32	0.349	1.79e-16	**1.62e-13**
FLYWCH2	16p13.3	0.348	1.91e-16	**1.65e-13**
SFXN2	10q24.32	0.346	3.17e-16	**2.60e-13**



RNA and miRNA example



RNA and miRNA example



miRNA correlation with mRNA example

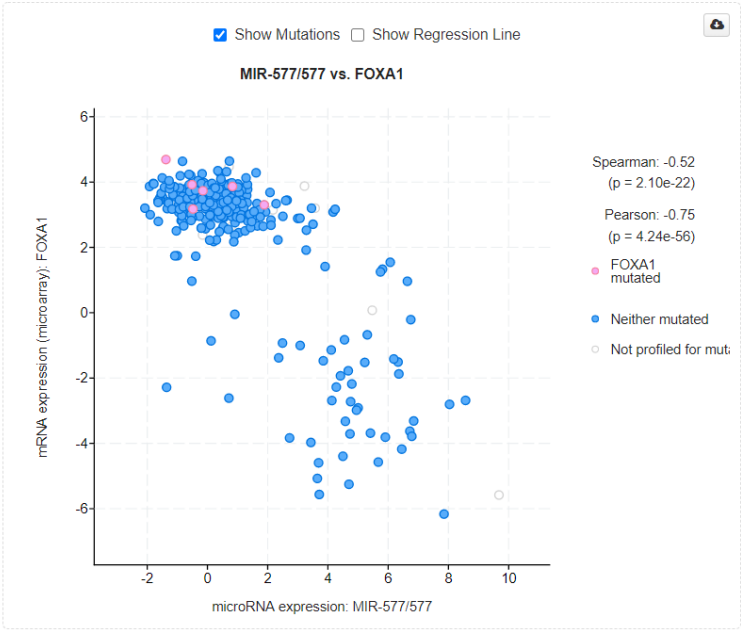
MGMT MIR-577/577

Find genes in mRNA expression (microarray) (526 samples) that are correlated with MIR-577/577 in microRNA expression (300 samples)

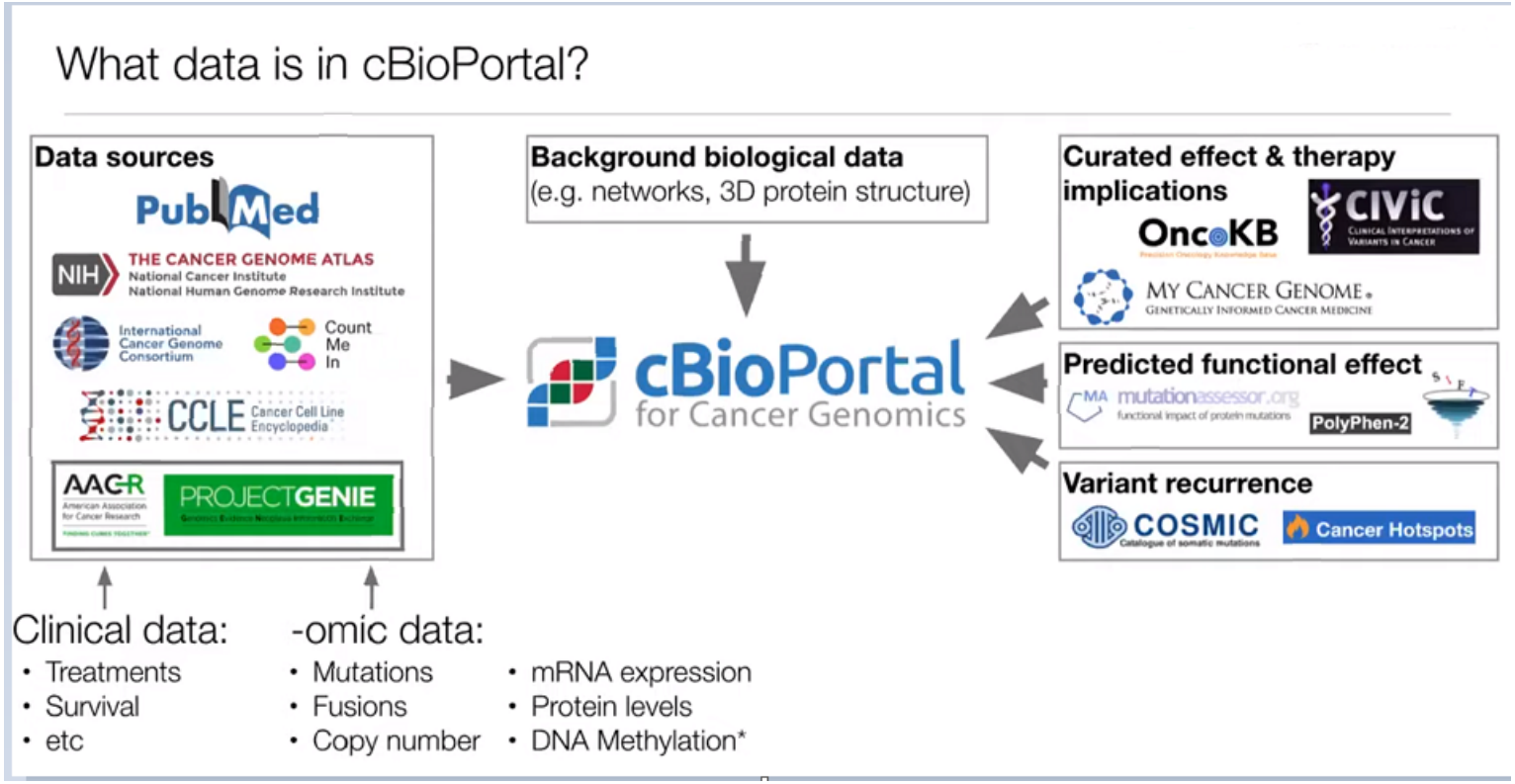
Show Any Correlation | Enter gene or cytoband..

Correlated Gene	Cytoband	Spearman's Correlation	p-Value	q-Value ▲
FOXA1	14q21.1	-0.523	2.10e-22	**1.90e-18**
UGT8	4q26	0.523	2.21e-22	**1.90e-18**
BCL11A	2p16.1	0.515	1.13e-21	**6.47e-18**
GATA3	10p14	-0.511	2.74e-21	**1.18e-17**
CDC20	1p34.2	0.510	3.62e-21	**1.25e-17**
YBX1	1p34.2	0.507	6.58e-21	**1.89e-17**
PSAT1	9q21.2	0.498	3.51e-20	**8.64e-17**
KIF2C	1p34.1	0.496	5.56e-20	**1.20e-16**
NUDT12	5q21.2	-0.494	8.69e-20	**1.66e-16**
SLC7A8	14q11.2	-0.491	1.48e-19	**2.55e-16**
CDCA8	1p34.3	0.489	2.09e-19	**3.28e-16**
HAPLN3	15q26.1	0.487	3.07e-19	**4.41e-16**
AGR3	7p21.1	-0.486	3.99e-19	**5.29e-16**
LPIN1	2p25.1	0.483	6.82e-19	**8.39e-16**
RAB17	2q37.3	-0.482	8.39e-19	**9.64e-16**
CENPA	2p23.3	0.480	1.14e-18	**1.17e-15**
CYB5R1	1q32.1	-0.480	1.16e-18	**1.17e-15**
ATAD3A	1p36.33	0.476	2.37e-18	**2.27e-15**
RUNDC1	17q21.31	-0.475	2.90e-18	**2.63e-15**
TBC1D9	4q31.21	-0.475	3.33e-18	**2.74e-15**
DYNLRB2	16q23.2	-0.475	3.34e-18	**2.74e-15**
NCAPD2	12p13.31	0.474	3.61e-18	**2.83e-15**
SH2D2A	1q23.1	0.472	5.56e-18	**4.04e-15**
MATN4	20q13.12	0.472	5.82e-18	**4.04e-15**
CTPS1	1p34.2	0.472	5.86e-18	**4.04e-15**

Showing 1-25 of 17237 ‹ Show more ›



Cbioportal.org



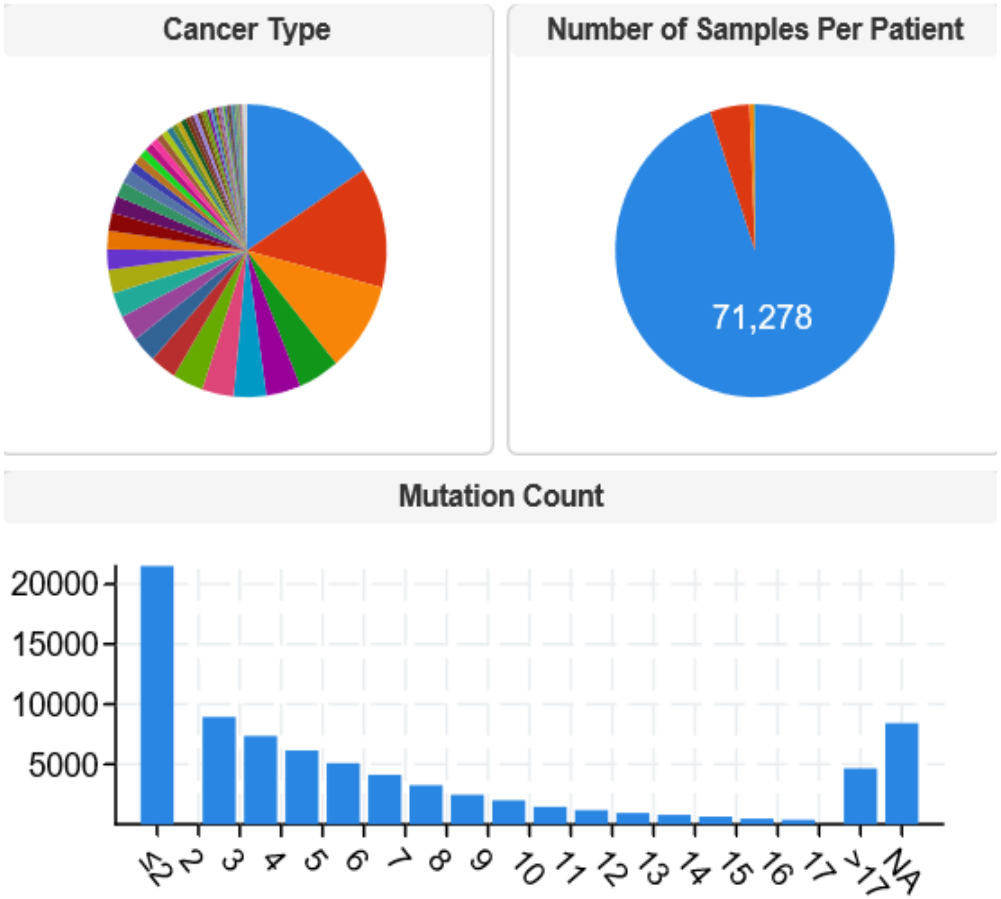
Slide borrowed from cbioportal website

AACR Project GENIE

- Clinical sequencing data from 19 cancer centers worldwide.
- It consists of primary and metastatic tumor unlike TCGA where they only have primary and untreated tumors
- For some samples, GENIE also consists of pre and post treatment.
- Targeted gene panels (# of genes targeted varied across the cancer centers)
- Majority of them have mutations and some have CNV

GENIE cbioportal

- https://genie.cbioportal.org/
- It consists of more 79K samples



Search cancer types

Cancer Type

	#	Freq
Non-Small Cell Lung Cancer	12,525	15.7%
Breast Cancer	10,634	13.3%
Colorectal Cancer	7,911	9.9%
Glioma	3,817	4.8%
Melanoma	3,195	4.0%
Prostate Cancer	2,988	3.7%
Ovarian Cancer	2,933	3.7%
Pancreatic Cancer	2,818	3.5%
Endometrial Cancer	2,426	3.0%
Cancer of Unknown Primary	2,309	2.9%
Soft Tissue Sarcoma	2,298	2.9%

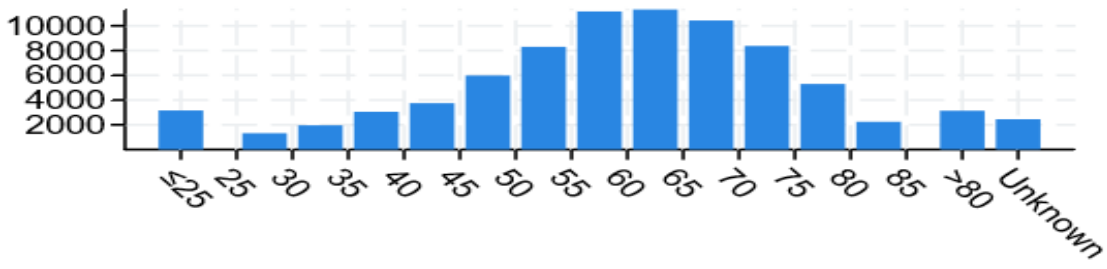
Search...

Sequence Assay ID

	#	Freq
MSK-IMPACT468	19,634	24.6%
MSK-IMPACT410	10,011	12.6%
DFCI-ONCOPANEL-2	8,289	10.4%
DFCI-ONCOPANEL-3	6,625	8.3%
JHU-50GP	5,094	6.4%
DFCI-ONCOPANEL-1	3,125	3.9%
MSK-IMPACT341	2,816	3.5%
CRUK-TS	2,420	3.0%
MSK-IMPACT-HEME-400	2,414	3.0%
MDA-50-V1	1,872	2.3%
VICC-01-T7	1,622	2.0%

Search...

Age at Which Sequencing was Reported



GTEX

Genotype Tissue-Expression Project

- Genome-wide association studies (GWAS)
- Cases vs controls
- ~95% of SNPs located in non-coding regions
- 53 tissue sites

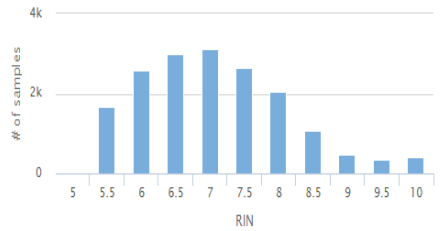
Dataset Summary of Analysis Samples

Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)

V8 Release	# Tissues	# Donors	# Samples
Total	54	948	17382
With Genotype	54	838	15253
Has eQTL Analysis*	49	838	15201

** Number of samples with genotype >= 70*

Sample RIN Histogram

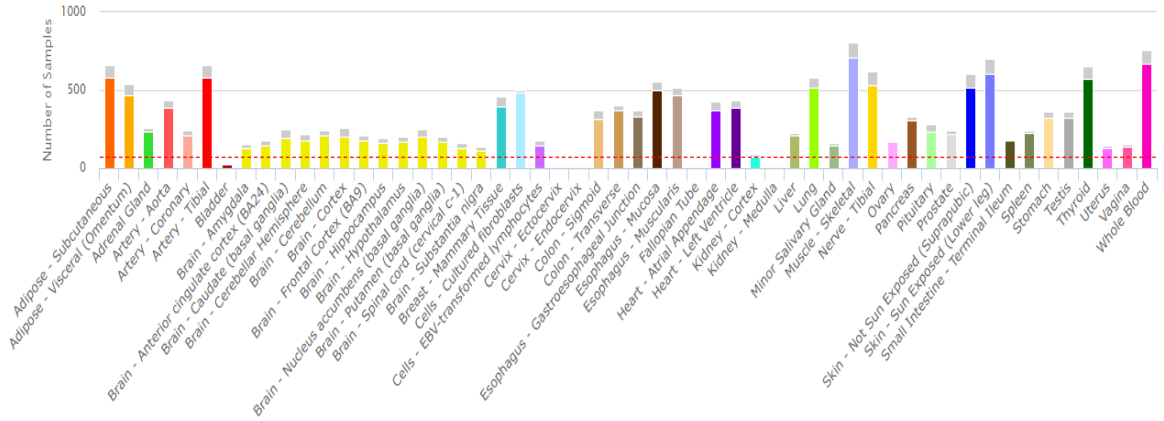


V8 Sample Counts by Tissues

Sort tissues by: Tissue alphabetical

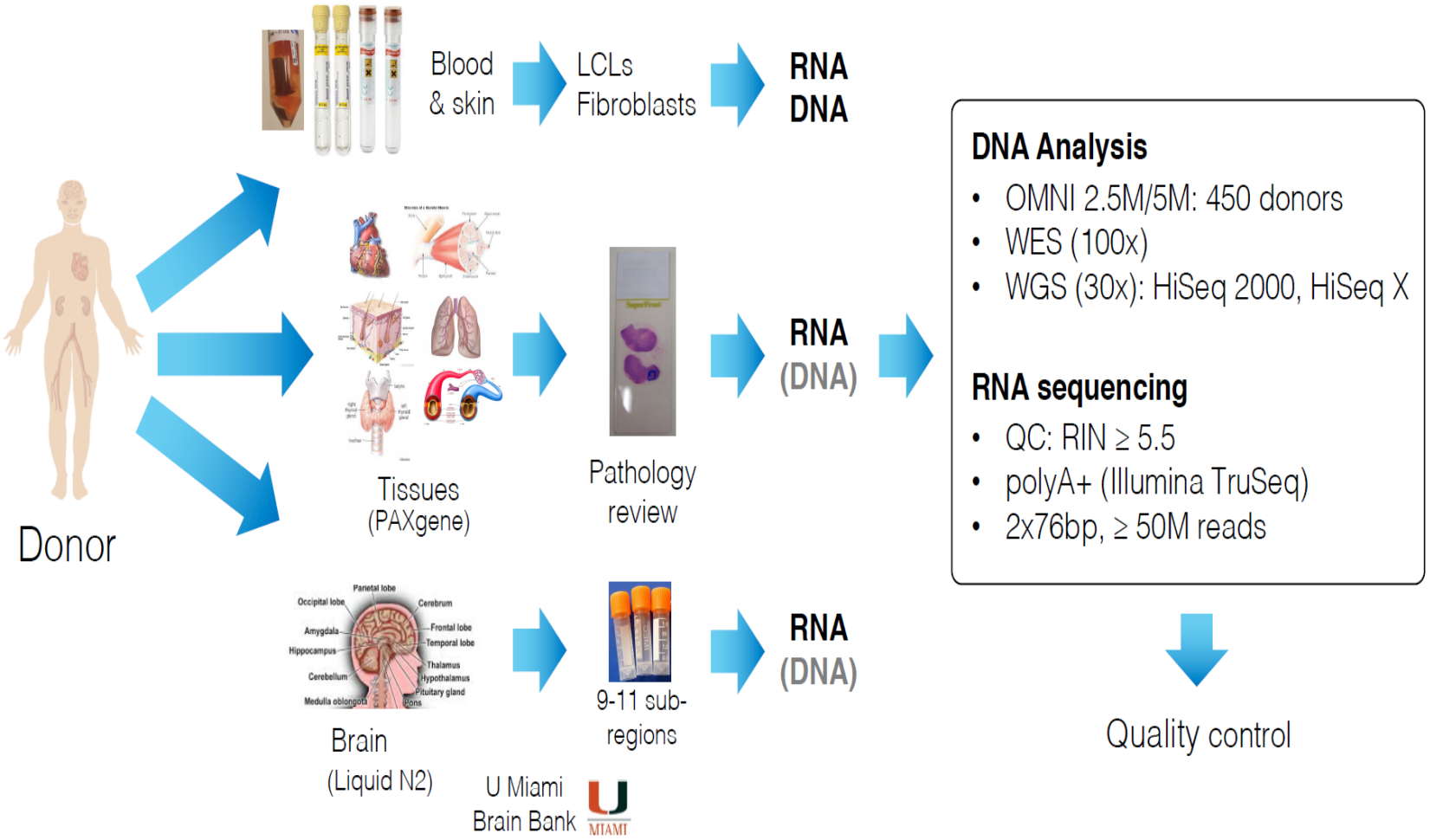
Download

View as: Plot | Table



Note: Cells - Transformed fibroblasts from previous releases has been corrected to Cells - Cultured fibroblasts.

Sample and data processing overview



https://storage.googleapis.com/ashg-workshop-files/GTEx_ASHG_workshop_101817_final.pdf

RNA-seq and eQTL pipeline details

Current public release

Release	V6p	V7	V8	V9
Genome build	GRCh37	**GRCh37**	**GRCh38**	GRCh38
GENCODE annotation	v19	**v19**	**v26**	v26
Aligner	TopHat 1.4.1	STAR 2.4.2a	STAR 2.5.3a	STAR 2.5.3a
Gene expression	RNA-SeQC 1.1.8	RNA-SeQC 1.1.9	RNA-SeQC 1.1.9	RNA-SeQC 1.1.9
Transcript expression	FluxCapacitor 1.6	RSEM 1.2.22	RSEM 1.3.0	RSEM 1.3.0
Quality control metrics	RNA-SeQC 1.1.8	RNA-SeQC 1.1.9	RNA-SeQC 1.1.9	RNA-SeQC 1.1.9
QTL mapper	FastQTL			

- Pipeline components selected and updated based on internal and published benchmarks (e.g., Teng et al., Genome Biology, 2016).

https://storage.googleapis.com/ashg-workshop-files/GTEx_ASHG_workshop_101817_final.pdf

Overview of GTEx resources: open-access data

- Expression
  - Gene-level expression (TPM, counts)
  - Transcript-level expression (TPM, counts, isoform proportions)
  - Exon read counts
- QTLs
  - Single-tissue eQTLs (*cis*- and *trans*-)
  - Multi-tissue eQTLs
  - Future: splicing QTLs
- Histology images
- De-identified public access sample and subject metadata

All open-access data is available at gtexportal.org

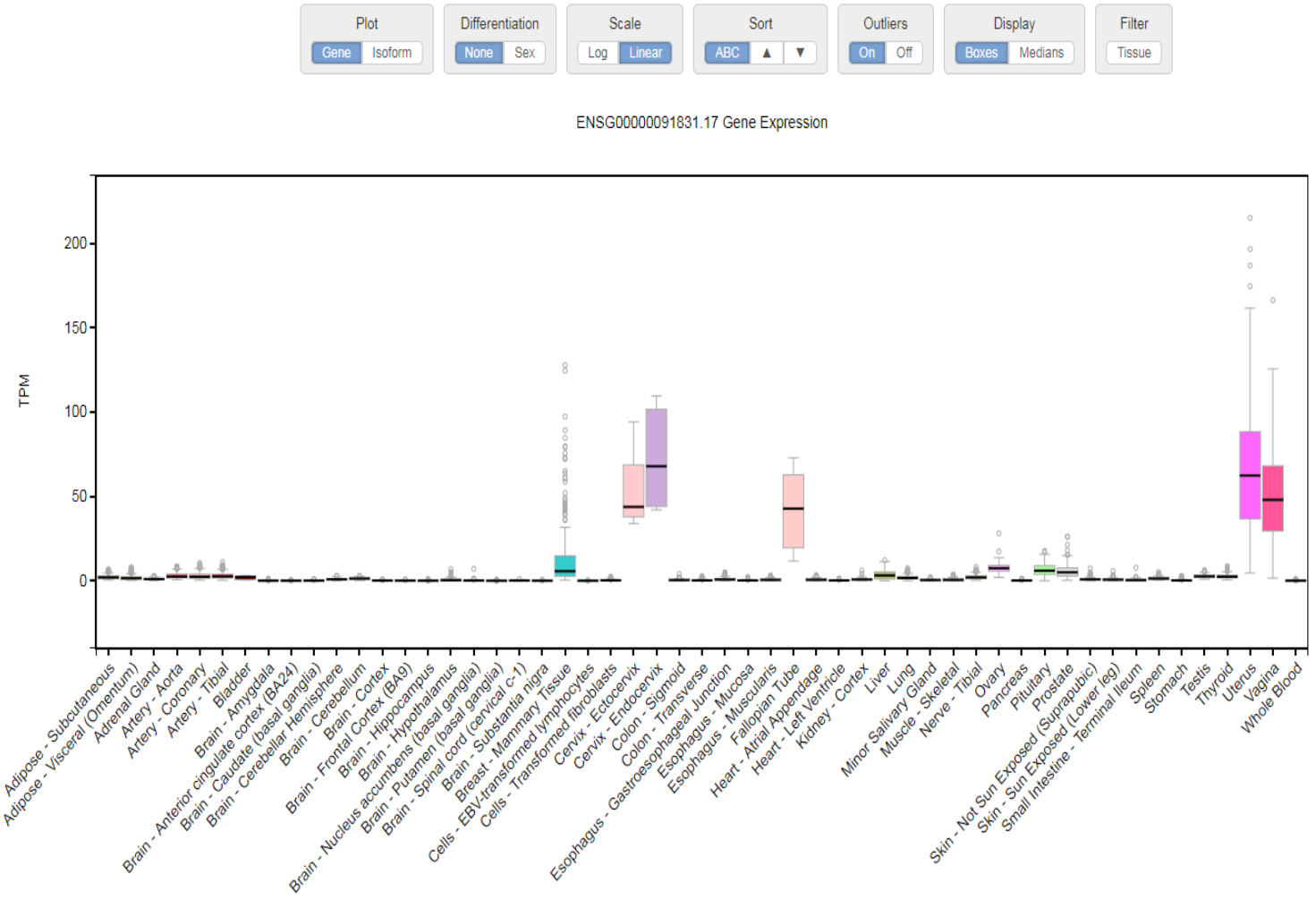
Overview of GTEx resources: protected data

- Sequence data:
  - RNA-seq (2x76 bp, unstranded, >50M reads/sample)
  - WGS (30x coverage) and WES (100x coverage)
  - Illumina Omni2.5/5 microarray genotypes (subset of 450 donors)
- Allele-specific expression (ASE)
- Full sample and subject metadata
- Future: eGTEx sequence data
  - ChIP-seq
  - WGBS-seq

https://storage.googleapis.com/ashg-workshop-files/GTEx_ASHG_workshop_101817_final.pdf

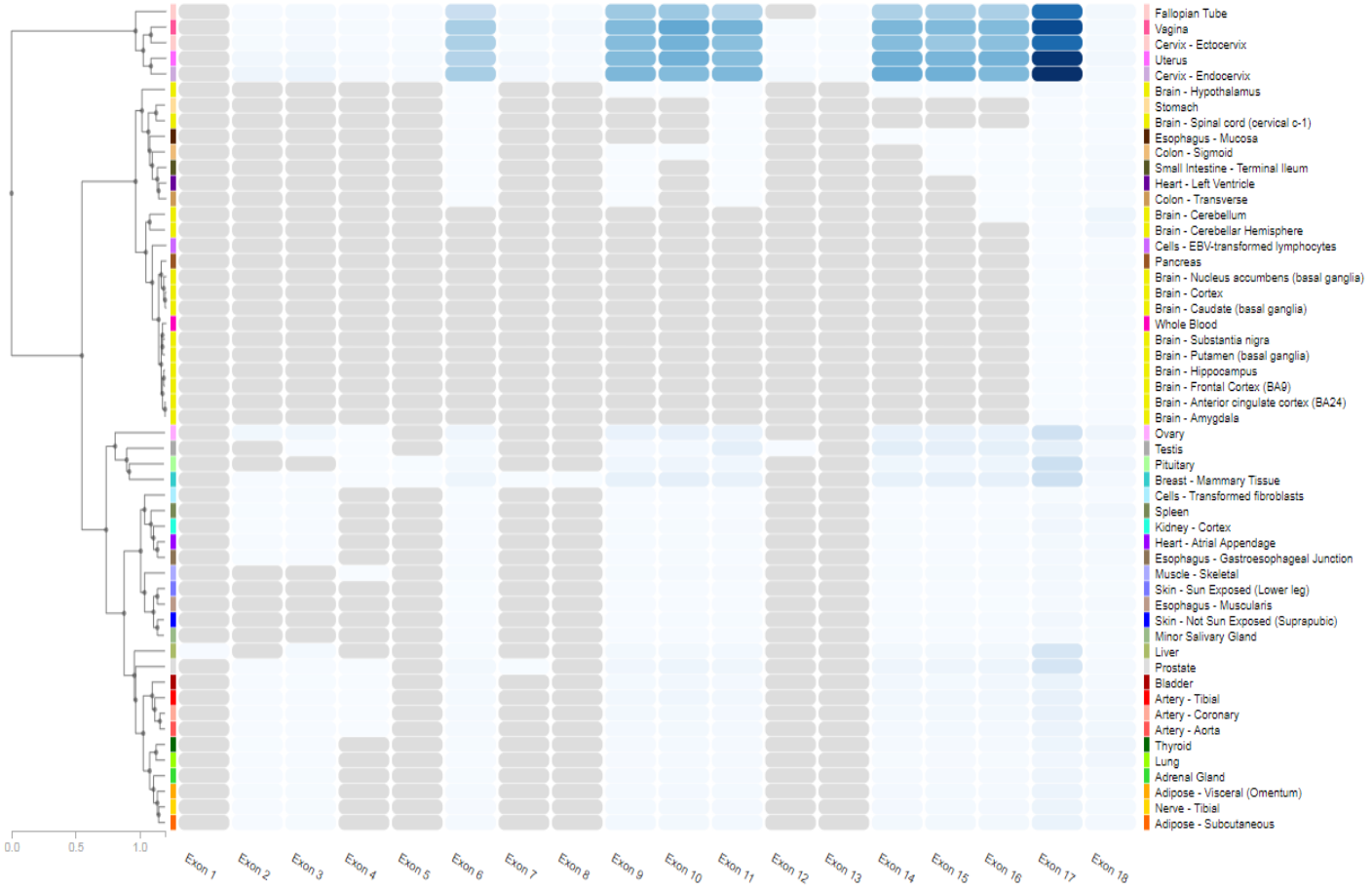
ESR1 query

Top
Gene Expression
Exon Expression
Single-Tissue eQTLs
Splice QTLs
Protein Trunc. Variants



Exon expression

Top
Gene Expression
Exon Expression
Single-Tissue eQTLs
Splice QTLs
Protein Trunc. Variants



ESR1 - eQTLs

Top
Gene Expression
Exon Expression
Single-Tissue eQTLs
Splice QTLs
Protein Trunc. Variants

− Significant Single-Tissue eQTLs for ESR1 (ENSG00000091831.17) in all tissues

Data Source: GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2)

ESR1 Gene eQTL Visualizer

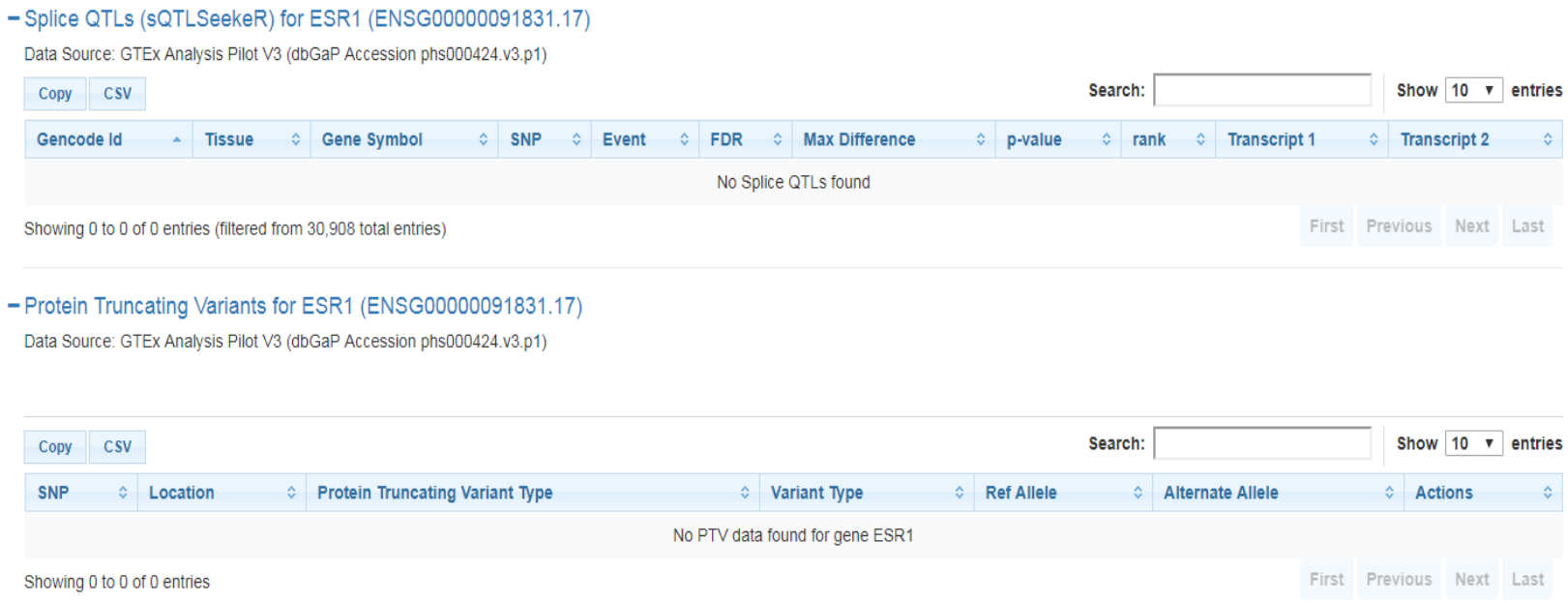
Copy CSV Search: Show 10 entries

Gencode Id	Gene Symbol	Variant Id	SNP		P-Value	NES	Tissue	Actions
ENSG00000091831.17	ESR1	6_151998105_G_A_b37	rs1293942	dbSNP	2.2e-7	-0.21	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_151998085_T_G_b37	rs1293943	dbSNP	2.2e-7	-0.21	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_152346190_TC_T_b37	rs113533024	dbSNP	2.4e-7	0.28	Testis	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_152000028_A_G_b37	rs712220	dbSNP	3.0e-7	-0.20	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_151999603_A_G_b37	rs1293938	dbSNP	3.0e-7	-0.21	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_151999507_C_G_b37	rs1293939	dbSNP	3.1e-7	-0.21	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_151998723_G_A_b37	rs980280	dbSNP	3.1e-7	-0.21	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_151990859_G_A_b37	rs1293956	dbSNP	3.7e-7	-0.21	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_151990954_T_C_b37	rs1293955	dbSNP	3.8e-7	-0.21	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot
ENSG00000091831.17	ESR1	6_151990961_A_G_b37	rs1293954	dbSNP	4.3e-7	-0.21	Thyroid	eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot

Showing 1 to 10 of 204 entries

First Previous 1 2 3 4 5 … 21 Next Last

No splice QTLs and protein truncating variants found for ESR1



– Splice QTLs (sQTLSeekeR) for ESR1 (ENSG00000091831.17)

Data Source: GTEx Analysis Pilot V3 (dbGaP Accession phs000424.v3.p1)

Copy CSV Search: Show 10 entries

Gencode Id	Tissue	Gene Symbol	SNP	Event	FDR	Max Difference	p-value	rank	Transcript 1	Transcript 2
No Splice QTLs found										

Showing 0 to 0 of 0 entries (filtered from 30,908 total entries) First Previous Next Last

– Protein Truncating Variants for ESR1 (ENSG00000091831.17)

Data Source: GTEx Analysis Pilot V3 (dbGaP Accession phs000424.v3.p1)

Copy CSV Search: Show 10 entries

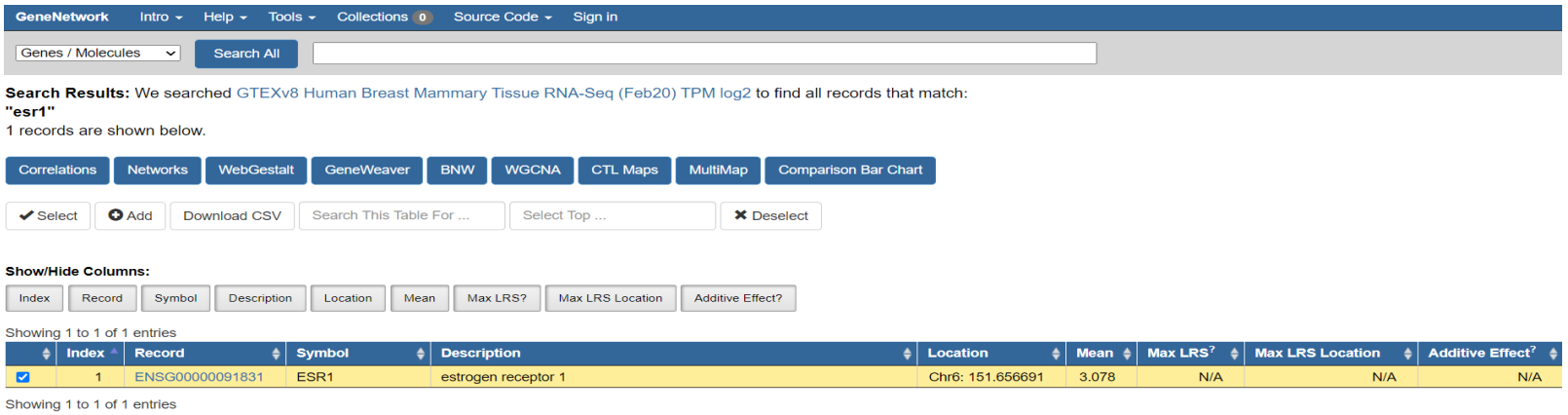
SNP	Location	Protein Truncating Variant Type	Variant Type	Ref Allele	Alternate Allele	Actions
No PTV data found for gene ESR1						

Showing 0 to 0 of 0 entries First Previous Next Last

MAYO CLINIC

WebQTL software

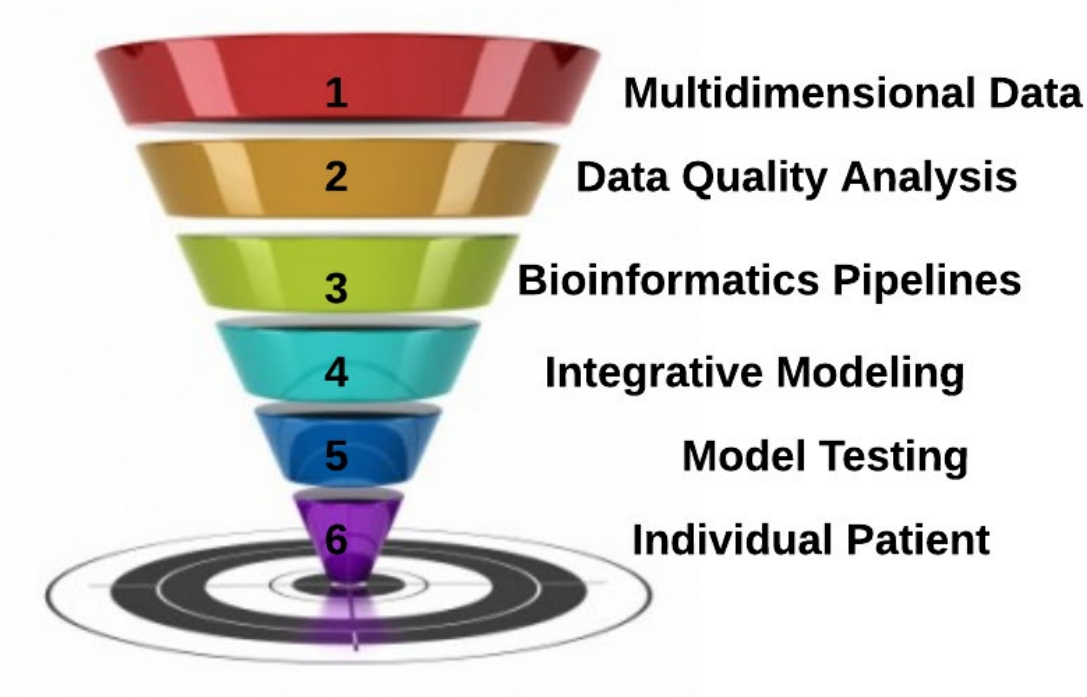
GeneNetwork WebQTL



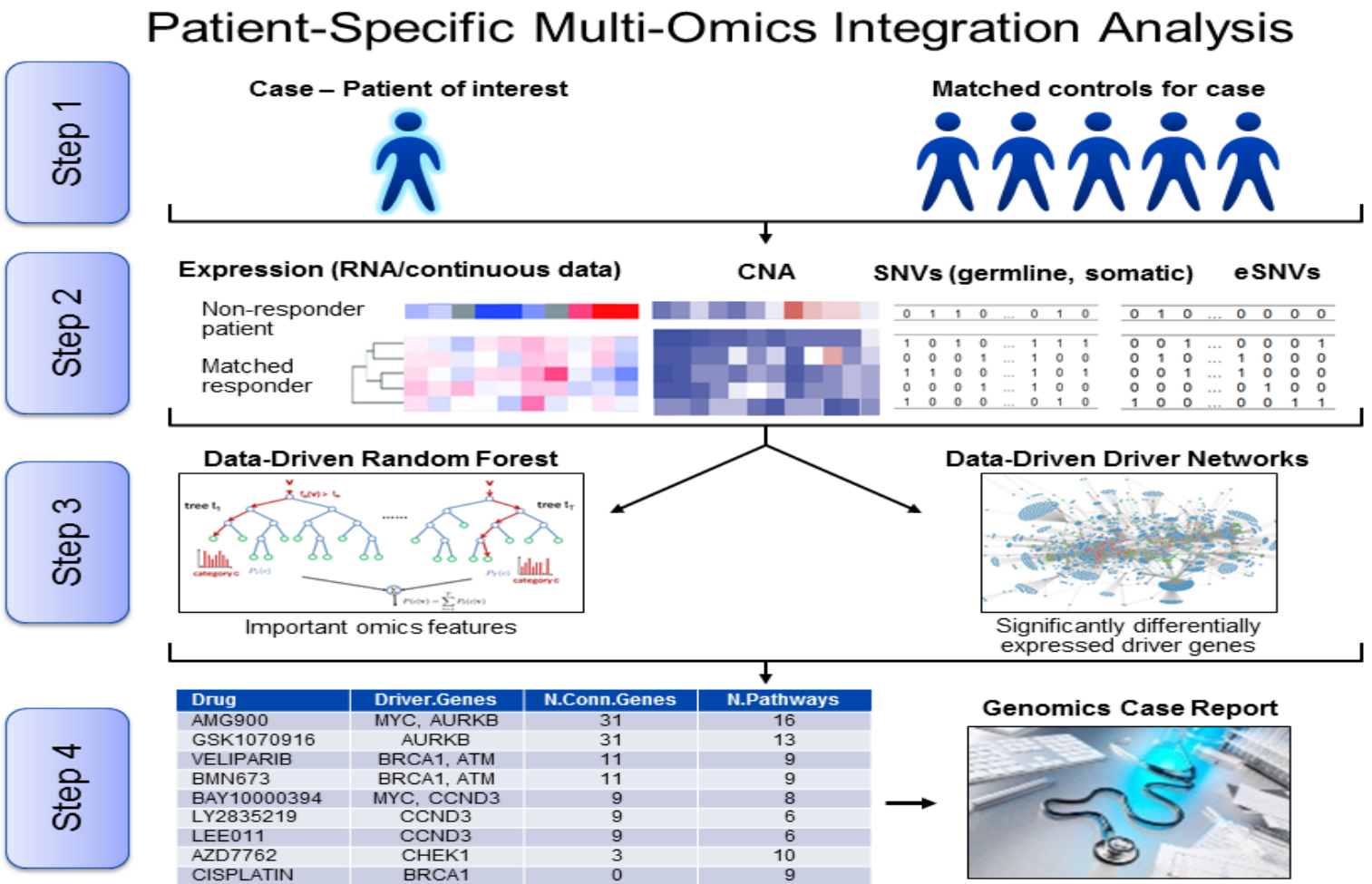
MAYO CLINIC

Precision medicine for cancer patients using clinical and molecular data

Multi-dimensional data to individual patient



PANOPLY – **P**recision c**a**ncer ge**no**mics re**p**ort: single samp**l**e inventor**y**



Drug	Driver.Genes	N.Conn.Genes	N.Pathways
AMG900	MYC, AURKB	31	16
GSK1070916	AURKB	31	13
VELIPARIB	BRCA1, ATM	11	9
BMN673	BRCA1, ATM	11	9
BAY10000394	MYC, CCND3	9	8
LY2835219	CCND3	9	6
LEE011	CCND3	9	6
AZD7762	CHEK1	3	10
CISPLATIN	BRCA1	0	9

Integration of multi-omics data for precision medicine

- PANOPLY- **Precision Cancer Genomic Report: Single Sample Inventory**
- PANOPLY: Omics-Guided Drug Prioritization Method

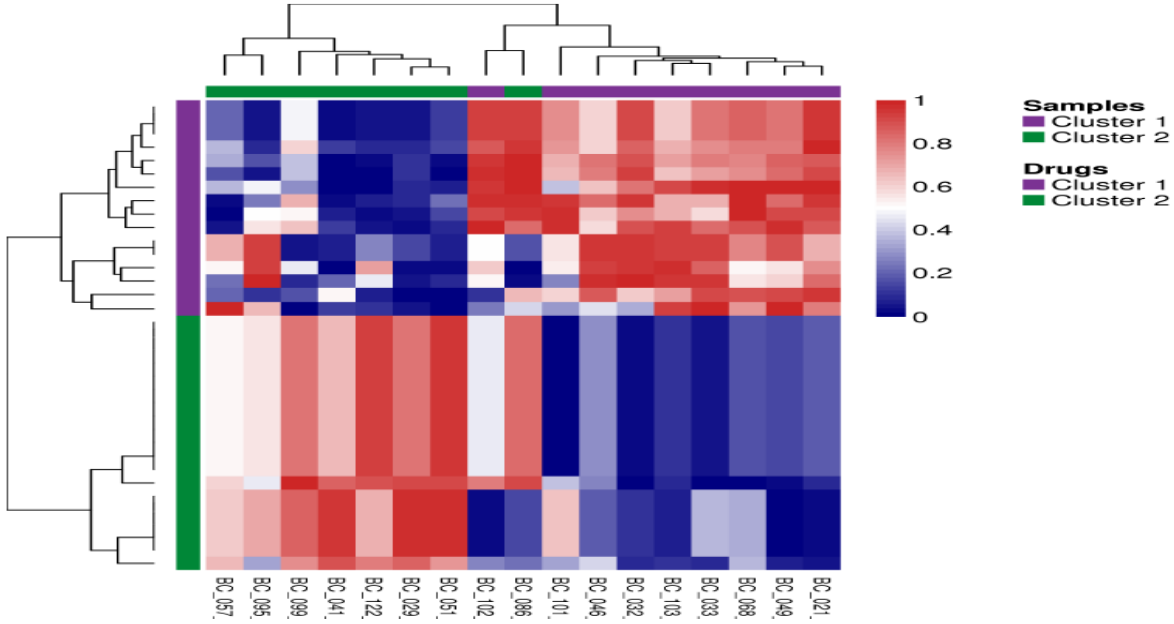
Tailored to an Individual Patient.

DEMO PANOPLY

Kalari KR, Sinnwell JP, Thompson KJ, Tang X, Carlson EE, Yu J, Vedell PT, Ingle JN, Weinshilboum RM, Boughey JC, Wang L, Goetz MP, Suman V. JCO Clin Cancer Inform. 2018 Dec;(2):1-11. doi: 10.1200/CCI.18.00012.

- http://www.kalarikrlab.org/Software/Panoply.html

Integration of 17 non-responder PANOPLY reports

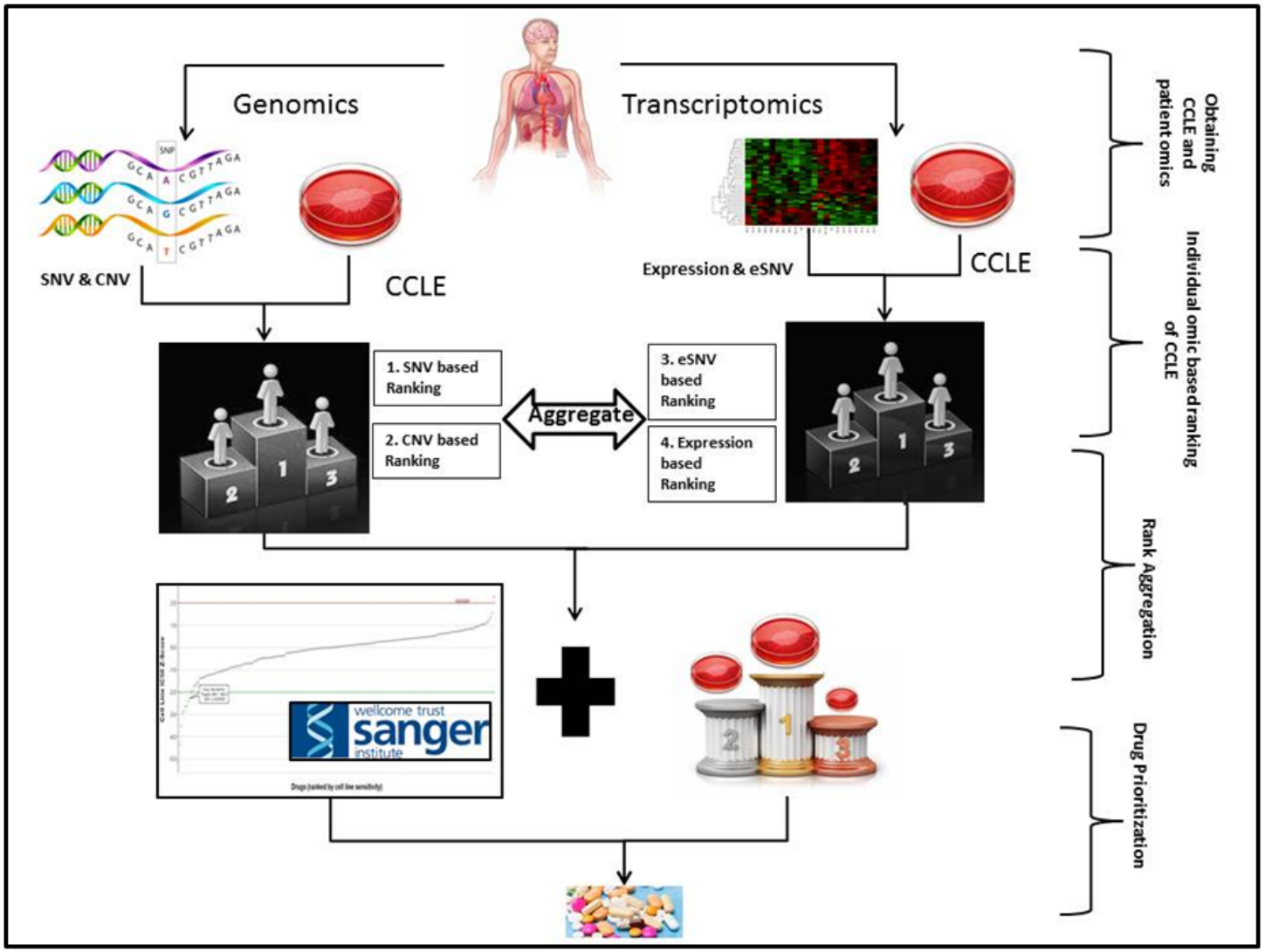


Kalari KR, Sinnwell JP, Thompson KJ, Tang X, Carlson EE, Yu J, Vedell PT, Ingle JN, Weinshilboum RM, Boughey JC, Wang L, Goetz MP, Suman V.JCO Clin Cancer Inform. 2018 Dec;(2):1-11. doi: 10.1200/CCI.18.00012.

MAYO CLINIC

Oncomatch – matching best cancer cell line to a patient

OncoMatch



MAYO CLINIC

Non-coding reads

Unmapped host reads

9. A comprehensive analysis of breast cancer microbiota and host gene expression.
Thompson KJ, Ingle JN, Tang X, Chia N, Jeraldo PR, Walther-Antonio MR, Kandimalla KK, Johnson S, Yao JZ, Harrington SC, Suman VJ, Wang L, Weinshilboum RL, Boughey JC, Kocher JP, Nelson H, Goetz MP, Kalari KR.
PLoS One. 2017 Nov 30;12(11):e0188873. doi: 10.1371/journal.pone.0188873. eCollection 2017.

20. Circular RNAs and their associations with breast cancer subtypes.
Nair AA, Niu N, Tang X, Thompson KJ, Wang L, Kocher JP, Subramanian S, Kalari KR.
Oncotarget. 2016 Dec 6;7(49):80967-80979. doi: 10.18632/oncotarget.13134.

HGT-ID: An efficient and sensitive workflow to detect human-viral insertion sites using next-generation sequencing data

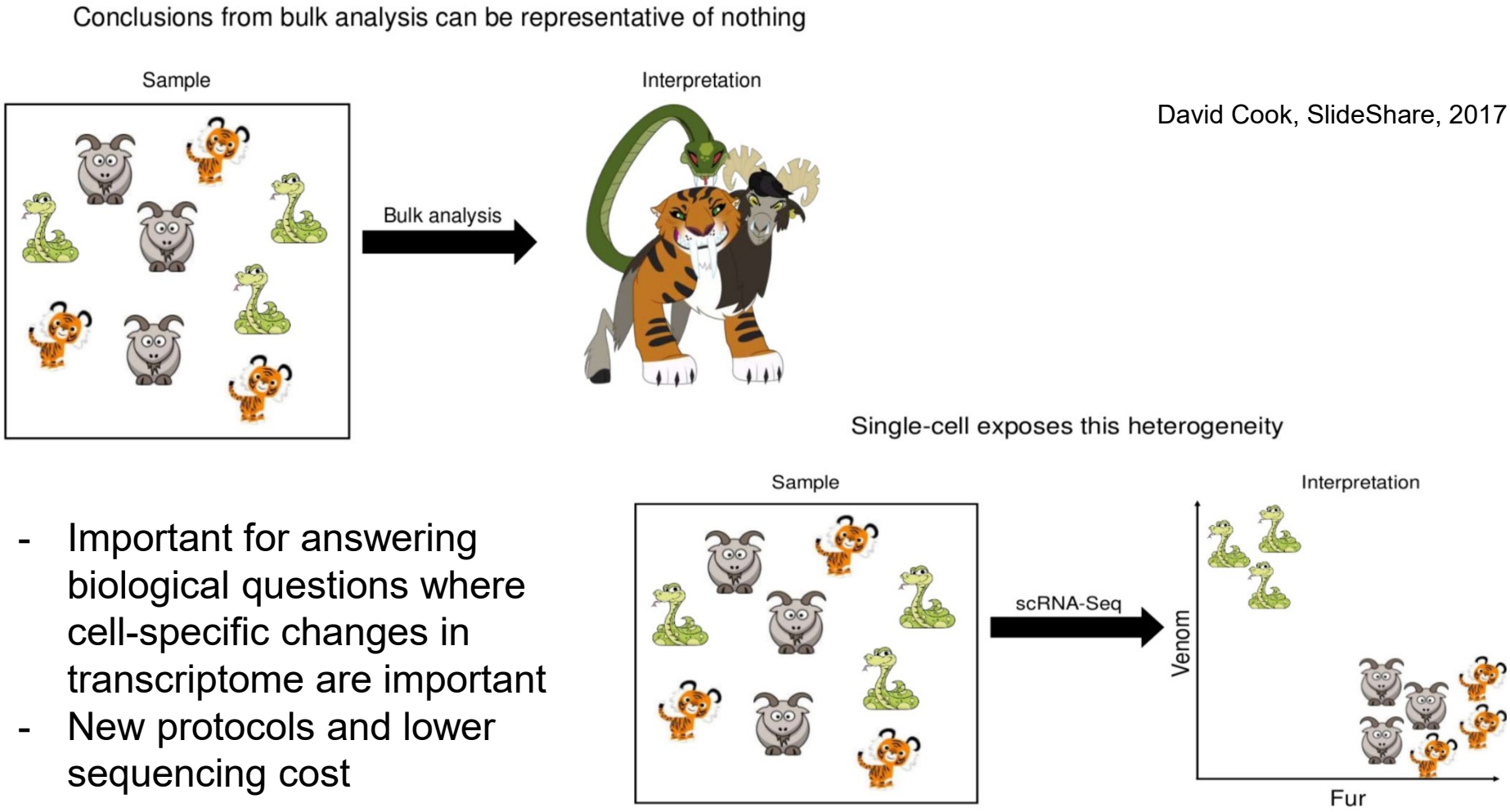
Saurabh Baheti†, Xiaojia Tang†, Daniel R. O'Brien, Nicholas Chia, Lewis R. Roberts, Heidi Nelson, Judy C. Boughey, Liewei Wang, Matthew P. Goetz, Jean-Pierre A. Kocher, Krishna R. Kalari*

Under review

MAYO CLINIC

Single-cell RNA-Sequencing

Single Cell vs. Bulk Samples

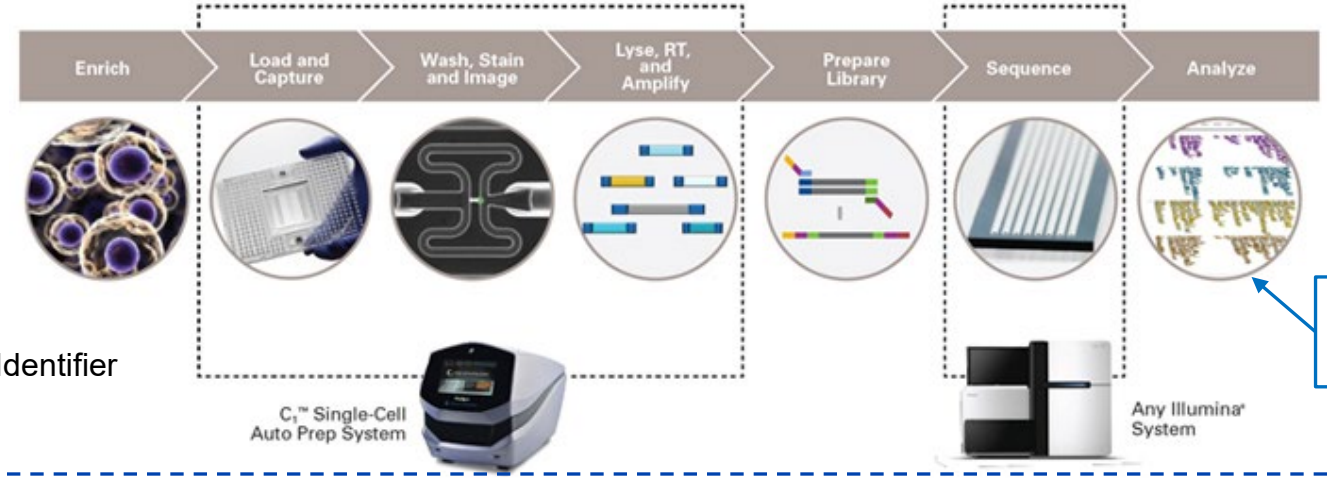


- Important for answering biological questions where cell-specific changes in transcriptome are important
- New protocols and lower sequencing cost

Two scRNA-Seq Platforms at MGF

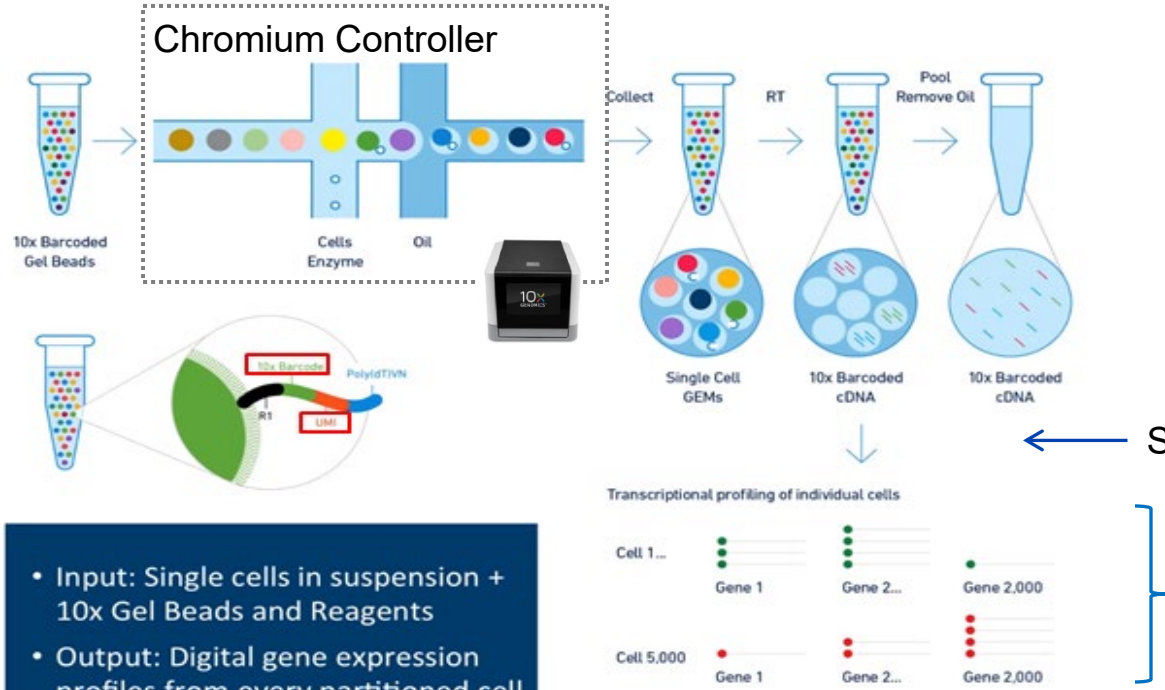
Fluidigm C1 (fluidic circuits) (non-UMI)

UMI=Unique Molecular Identifier



MAPRSeq pipeline

Chromium 10X Genomics Single cell gene expression (Droplet) (UMI)



- Input: Single cells in suspension + 10x Gel Beads and Reagents
- Output: Digital gene expression profiles from every partitioned cell

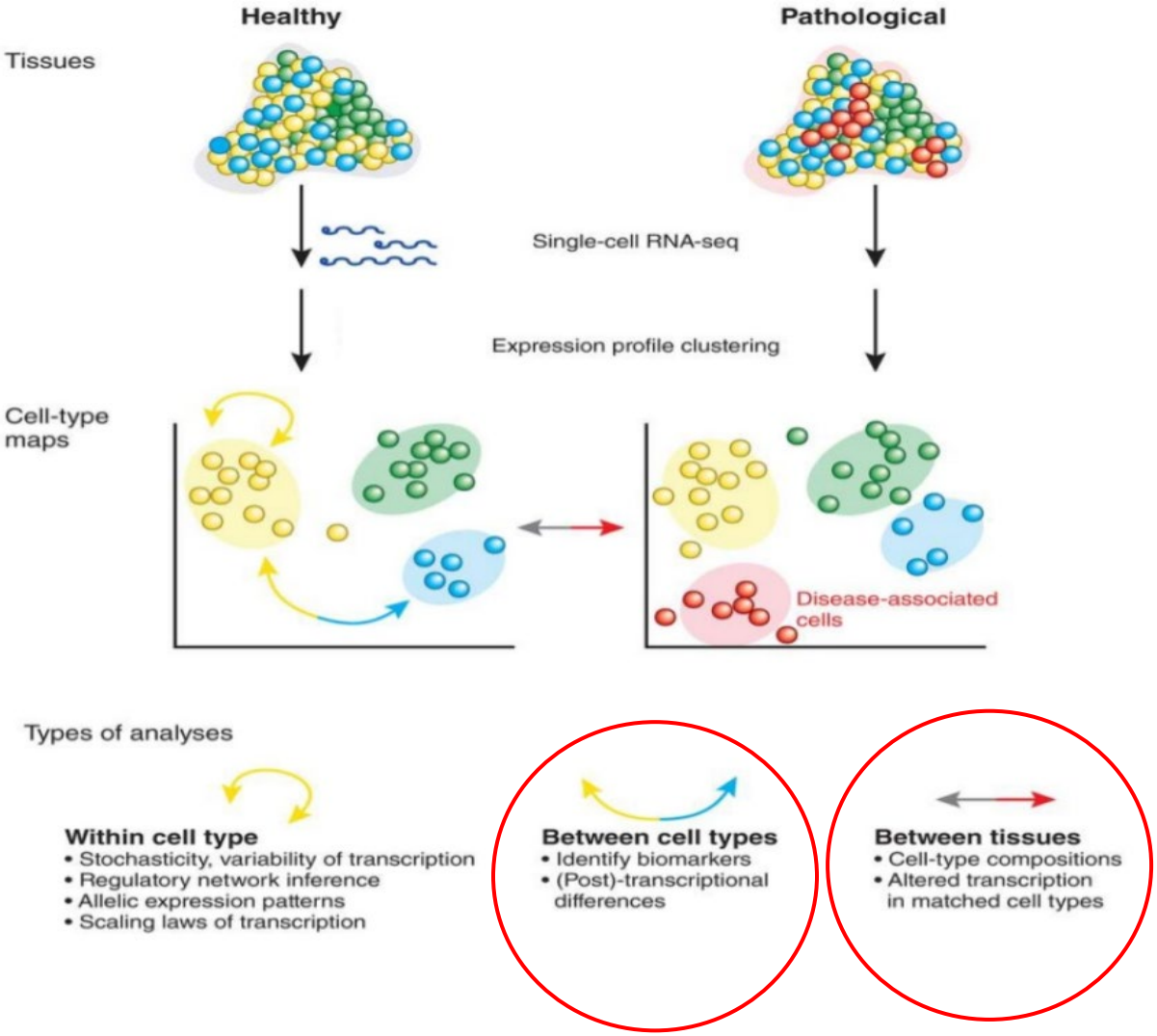
Tertiary analysis

10X Genomics Pipeline

Fluidigm C1 vs. Chromium 10X Genomics

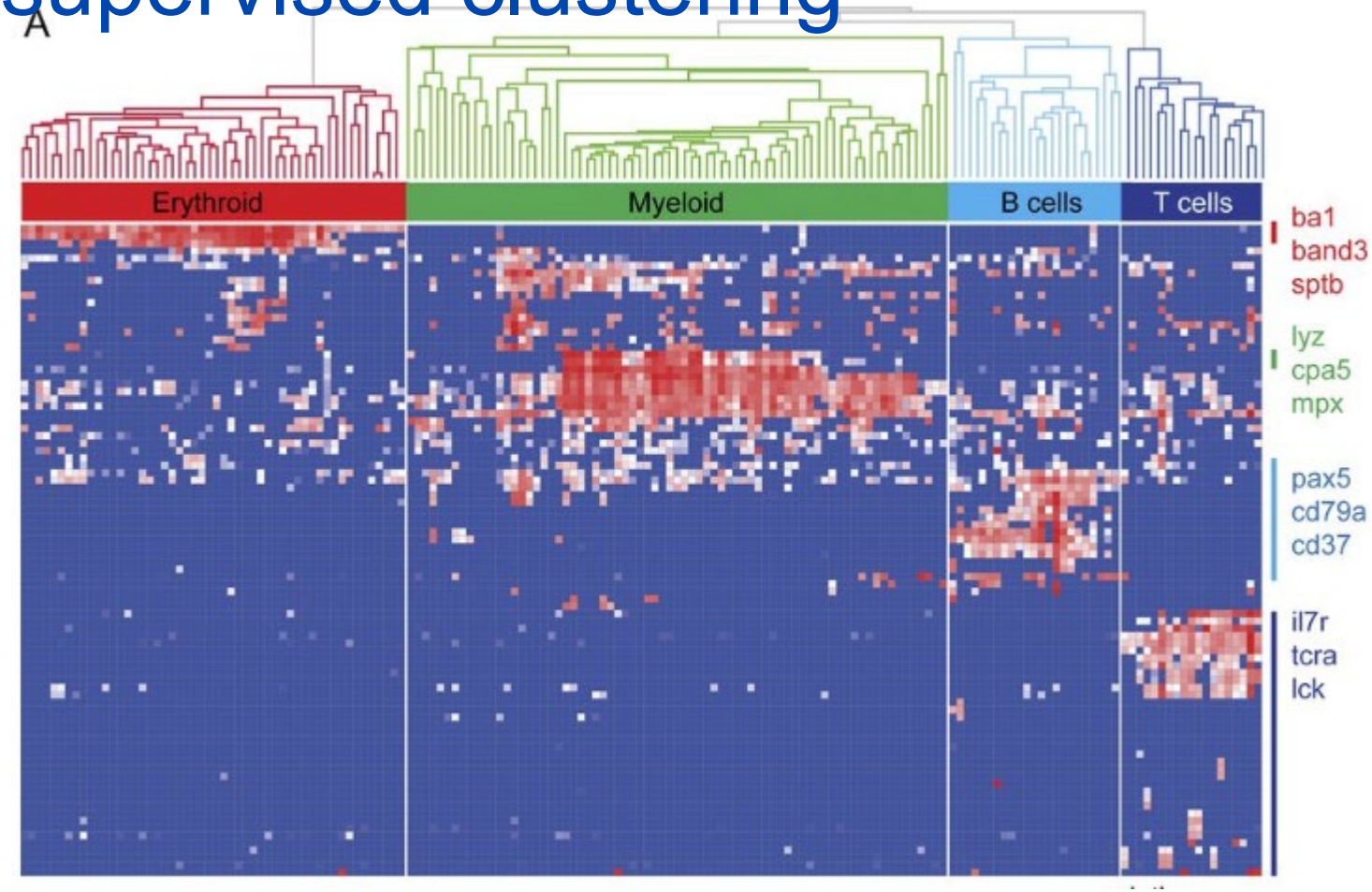
Instrument	Fluidigm C1	Chromium 10x Genomics
Launched in	2012	10/2016
Principles (Reference)	Integrated fluidic circuits	Droplet-based
RNA-Seq solution	Full transcript	3'-tag
Throughput (# of cells analyzed)	Low-medium (48-800)	High (100-10,000+)
Visual Inspection	Yes	No
Cell Selection	Yes (C1 size based)	No
Starting Amount of Cells	Medium-low	High
Flexibility (Own Protocols)	Yes	No
Advantage	Allows visual inspection of captured cells customizable protocols	**High cell capture efficiency, cell size <50µm, nuclei suspensions can be studied, lower system cost**
Limitation	Size-based cell selection (C1) (5-10, 10-17, 17-25 µm)	High initial cell concentration required, no users modification possible

Customized Tertiary Analysis



Bushra Arif, SlideShare, 2016

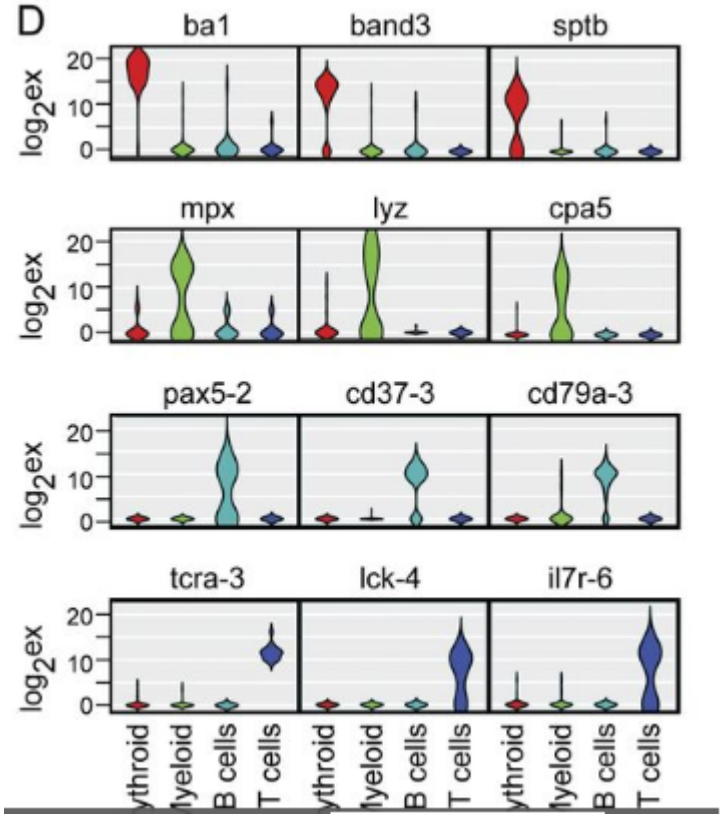
Unsupervised clustering



Unsupervised hierarchical clustering after gene expression analysis of single blood cells isolated from the whole kidney marrow. Heat map shows high transcript expression in red and low/absent expression in blue. Four major clusters were identified, including the following: erythroid (red), myeloid (green), B cells (light blue), and T cells (dark blue).

http://jem.rupress.org/content/213/6/979

Unsupervised clustering

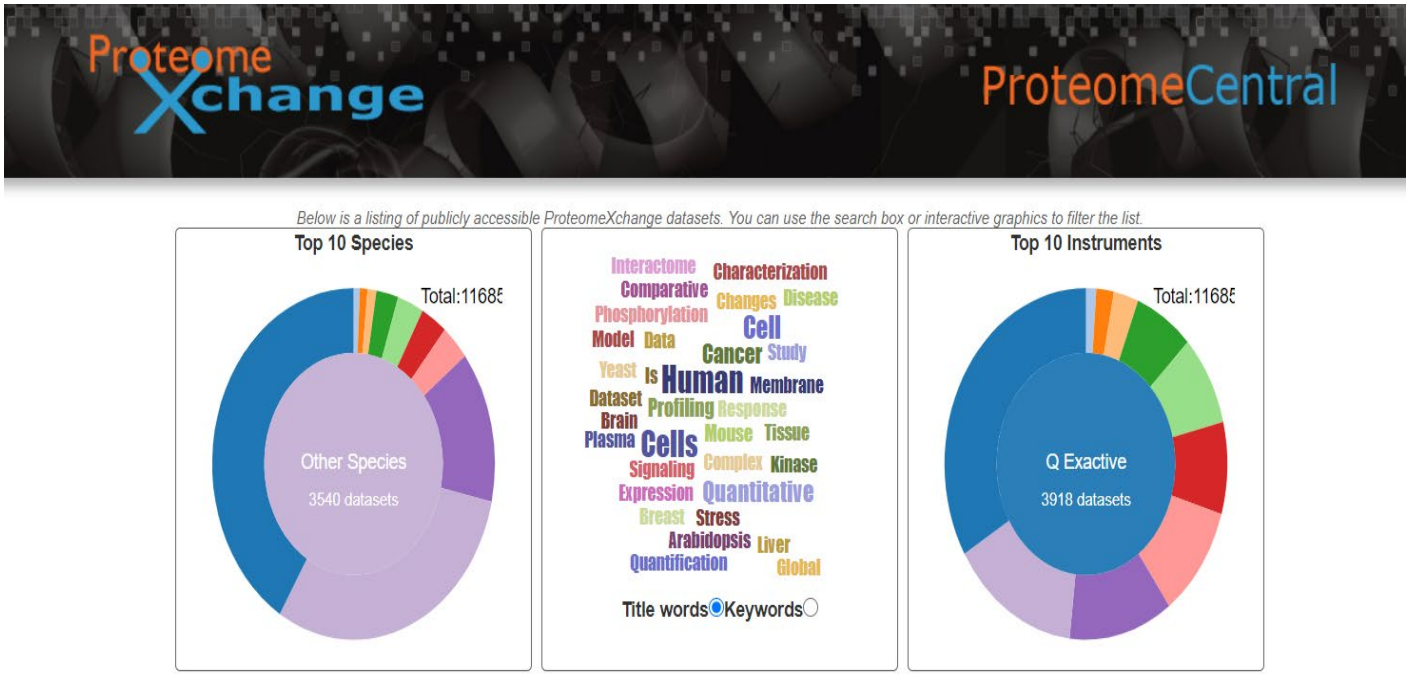


- Violin plots show the distribution of gene expression of single cells. Cells types were assigned based on hierarchical clustering and assessed for transcript expression of well-known blood cell lineage genes

http://jem.rupress.org/content/213/6/979

Proteomics datasets

- http://proteomecentral.proteomexchange.org/cgi/GetDataset



Impact

Prophylactic oophorectomy

- In women who have a known *BRCA* mutation, prophylactic oophorectomy can decrease breast cancer incidence by 50%
  - Rebbeck et al. Breast cancer risk after bilateral prophylactic oophorectomy in BRCA1 mutation carriers, J Natl cancer Inst 1999;91(17):1475-9.

Questions & Discussion