Genomic Data Resources and Data Mining for Everyone

Krishna Rani Kalari Ph.D.
Associate Professor, Biomedical Informatics
College of Medicine, Mayo Clinic, Rochester, MN

Outline of my presentation

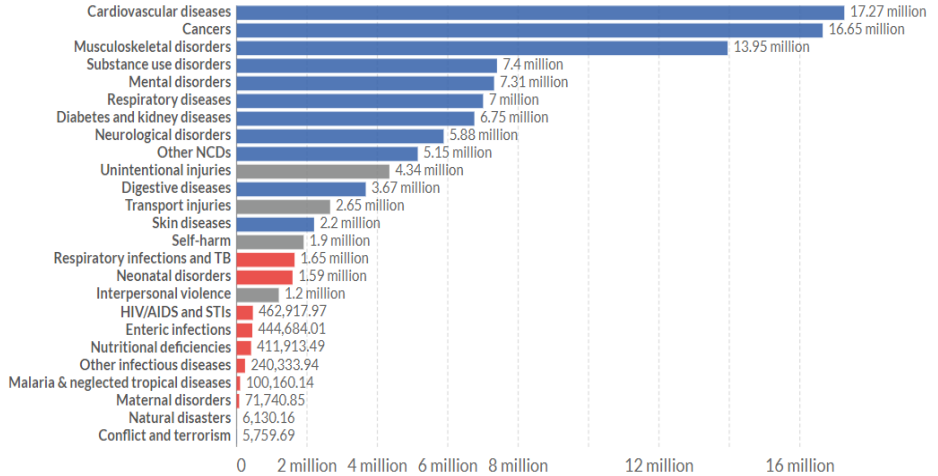
- Background
- Simple query tools for basic gene/variant search
  - UCSC, GWAS tools, and Knowledge Portals
- Genotype-Tissue Expression queries
- Cancer portals
  - TCGA, GDC, ICGC, cBioPortal, COSMIC, HTAN
- Single-cell RNA
- Proteomics
- Trans-Omics for Precision Medicine (TOPMed)

Burden of disease by cause, United States, 2019

Our World in Data

Total disease burden, measured in Disability-Adjusted Life Years (DALYs) by sub-category of disease or injury.
DALYs measure the total burden of disease – both from years of life lost due to premature death and years lived with a disability.
One DALY equals one lost year of healthy life.

⇄ Change country

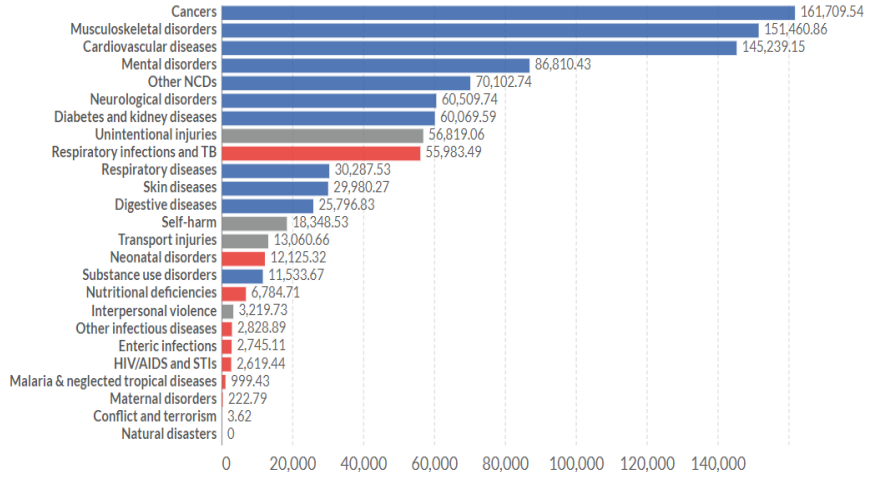


Burden of disease by cause, Singapore, 2019

Our World in Data

Total disease burden, measured in Disability-Adjusted Life Years (DALYs) by sub-category of disease or injury.
DALYs measure the total burden of disease – both from years of life lost due to premature death and years lived with a disability.
One DALY equals one lost year of healthy life.

⇄ Change country

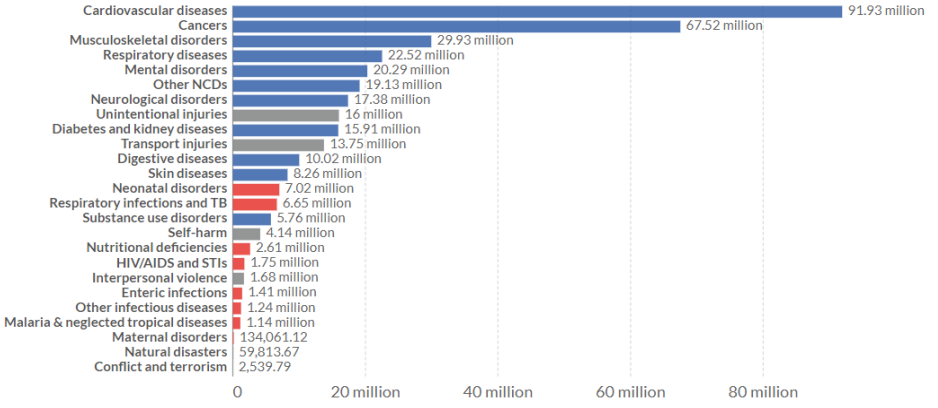


Burden of disease by cause, China, 2019

Our World in Data

Total disease burden, measured in Disability-Adjusted Life Years (DALYs) by sub-category of disease or injury.
DALYs measure the total burden of disease – both from years of life lost due to premature death and years lived with a disability.
One DALY equals one lost year of healthy life.

⇄ Change country



Source: IHME, Global Burden of Disease
OurWorldInData.org/burden-of-disease • CC BY
Note: Non-communicable diseases are shown in blue; communicable, maternal, neonatal and nutritional diseases in red; injuries in grey.

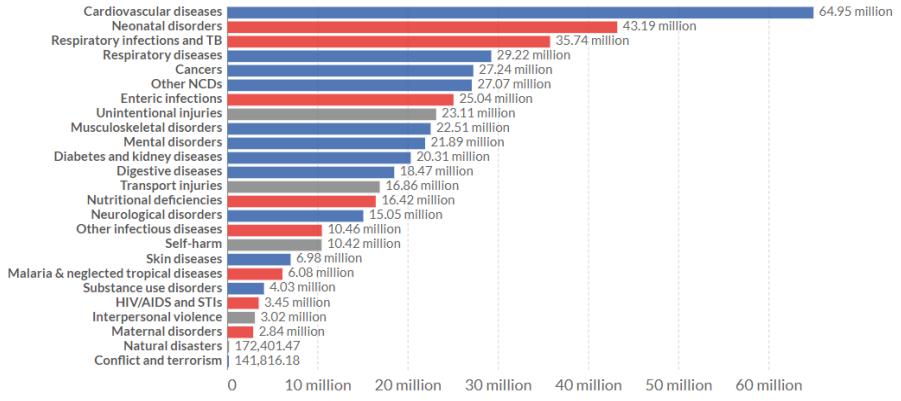
▶ 1990 ——— 2019

Burden of disease by cause, India, 2019

Our World in Data

Total disease burden, measured in Disability-Adjusted Life Years (DALYs) by sub-category of disease or injury.
DALYs measure the total burden of disease – both from years of life lost due to premature death and years lived with a disability.
One DALY equals one lost year of healthy life.

⇄ Change country



Source: IHME, Global Burden of Disease
OurWorldInData.org/burden-of-disease • CC BY
Note: Non-communicable diseases are shown in blue; communicable, maternal, neonatal and nutritional diseases in red; injuries in grey.

▶ 1990 ——— 2019

Make genomic data accessible

- Interrogate the tsunami of data
- Make genotype and phenotype data that is accessible for a relevant disease
- Software platform or a user interface that provides access of genomics data to non-experts

UCSC Genome Browser

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute | UCSC **Genome Browser**

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Projects | Help | About Us

Our tools

- **Genome Browser**
  interactively visualize genomic data
- **COVID-19 Research**
  use the SARS-CoV-2 genome browser and explore coronavirus datasets
- **BLAT**
  rapidly align sequences to the genome
- **Table Browser**
  download data from the Genome Browser database
- **Variant Annotation Integrator**
  get functional effect predictions for variant calls
- **Data Integrator**
  combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBiB)**
  run the Genome Browser on your laptop or server
- **In-Silico PCR**
  rapidly align PCR primer pairs to the genome
- **LiftOver**
  convert genome coordinates between assemblies
- **Track Hubs**
  import and view external data tracks
- **REST API**
  returns data in JSON format

More tools...

Our story

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed the first working draft of the human genome assembly, forever ensuring free public access to the genome and the information it contains. A few weeks later, on July 7, 2000, the newly

What's new

May 11, 2022 - **Ensembl 106 GENCODE models for hg19/hg38...**

May 5, 2022 - **Merged Cell Expression on hg38**

May 3, 2022 - **New GnomAD Mutation Constraint track**

UCSC Genome Browser

Genomics Institute UCSC **Genome Browser Gateway**

Genome Browser | Tools | Mirrors | Downloads | My Data | Projects | Help | About Us

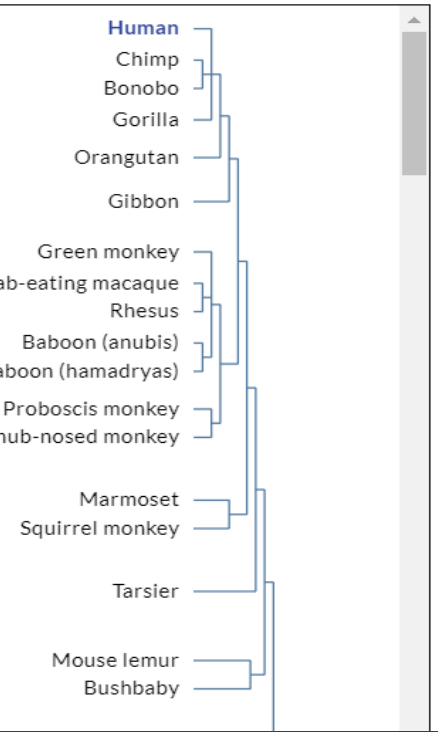
ecies

Zebrafish Fruitfly Worm Yeast

ds of genome browsers

? Send us a request.

ND CONNECTED ASSEMBLY HUBS



Find Position

Human Assembly

Dec. 2013 (GRCh38/hg38)

GO

Position/Search Term

Enter position, gene symbol or search terms

Current position: chrX:15,560,138-15,602,945

Human Genome Browser - hg38 assembly

UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38.p13 (GCA_000001405.28)
Assembly date: Dec. 2013 initial release; Dec. 2017 patch release 13
Assembly accession: GCA_000001405.28
NCBI Genome ID: 51 (Homo sapiens (human))
NCBI Assembly ID: GCF_000001405.39 (GRCh38.p13, GCA_000001405.28)
BioProject ID: PRJNA31257

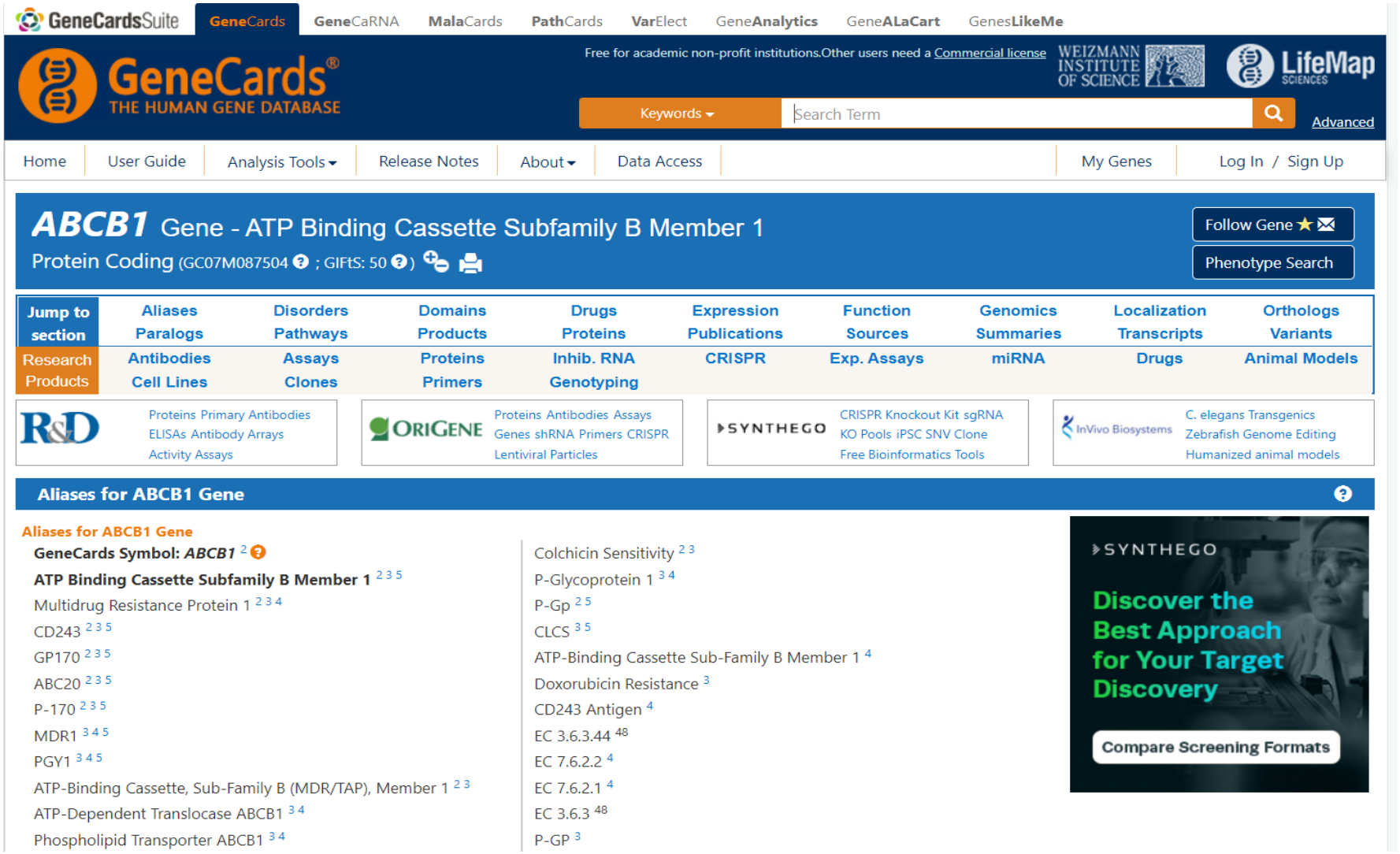
Search the assembly:

- **By position or search term:** Use the "position or search term" box to find areas of the genome associated with many
  specific chromosomal coordinate range; mRNA, EST, or STS marker names; or keywords from the GenBank descripti
  including sample queries.
- **By gene name:** Type a gene name into the "search term" box, choose your gene from the drop-down list, then press "
  assembly location associated with that gene. More information.
- **By track type:** Click the "track search" button to find Genome Browser tracks that match specific selection criteria. N

Download sequence and annotation data:

- Using rsync (recommended)
- Using HTTP
- Using FTP

Gene Cards



UCSC Genome Browser

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute — UCSC **Genome Browser Gateway**

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Projects | Help | About Us

Browse/Select Species

POPULAR SPECIES

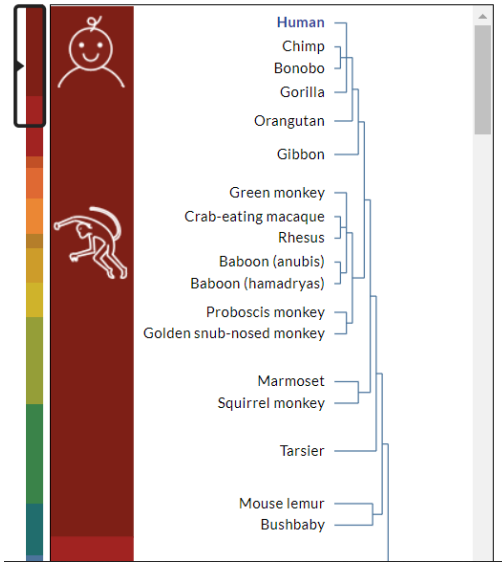
Human | Mouse | Rat | Zebrafish | Fruitfly | Worm | Yeast

Search through thousands of genome browsers

Enter species, common name or assembly ID

Unable to find a genome? Send us a request.

UCSC SPECIES TREE AND CONNECTED ASSEMBLY HUBS



Find Position

Human Assembly: Dec. 2013 (GRCh38/hg38)

GO

Position/Search Term: ABCB1

Current position: chr7:87,503,859-87,713,323

Human Genome Browser - hg38 assembly

view sequences

UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38.p13 (GCA_000001405.28)
Assembly date: Dec. 2013 initial release; Dec. 2017 patch release 13
Assembly accession: GCA_000001405.28
NCBI Genome ID: 51 (Homo sapiens (human))
NCBI Assembly ID: GCF_000001405.39 (GRCh38.p13, GCA_000001405.28)
BioProject ID: PRJNA31257

Homo sapiens
(Graphic courtesy of CBSE)

Search the assembly:

- **By position or search term:** Use the "position or search term" box to find areas of the genome associated with many different attributes, such as a specific chromosomal coordinate range; mRNA, EST, or STS marker names; or keywords from the GenBank description of an mRNA. More information, including sample queries.
- **By gene name:** Type a gene name into the "search term" box, choose your gene from the drop-down list, then press "submit" to go directly to the assembly location associated with that gene. More information.
- **By track type:** Click the "track search" button to find Genome Browser tracks that match specific selection criteria. More information.

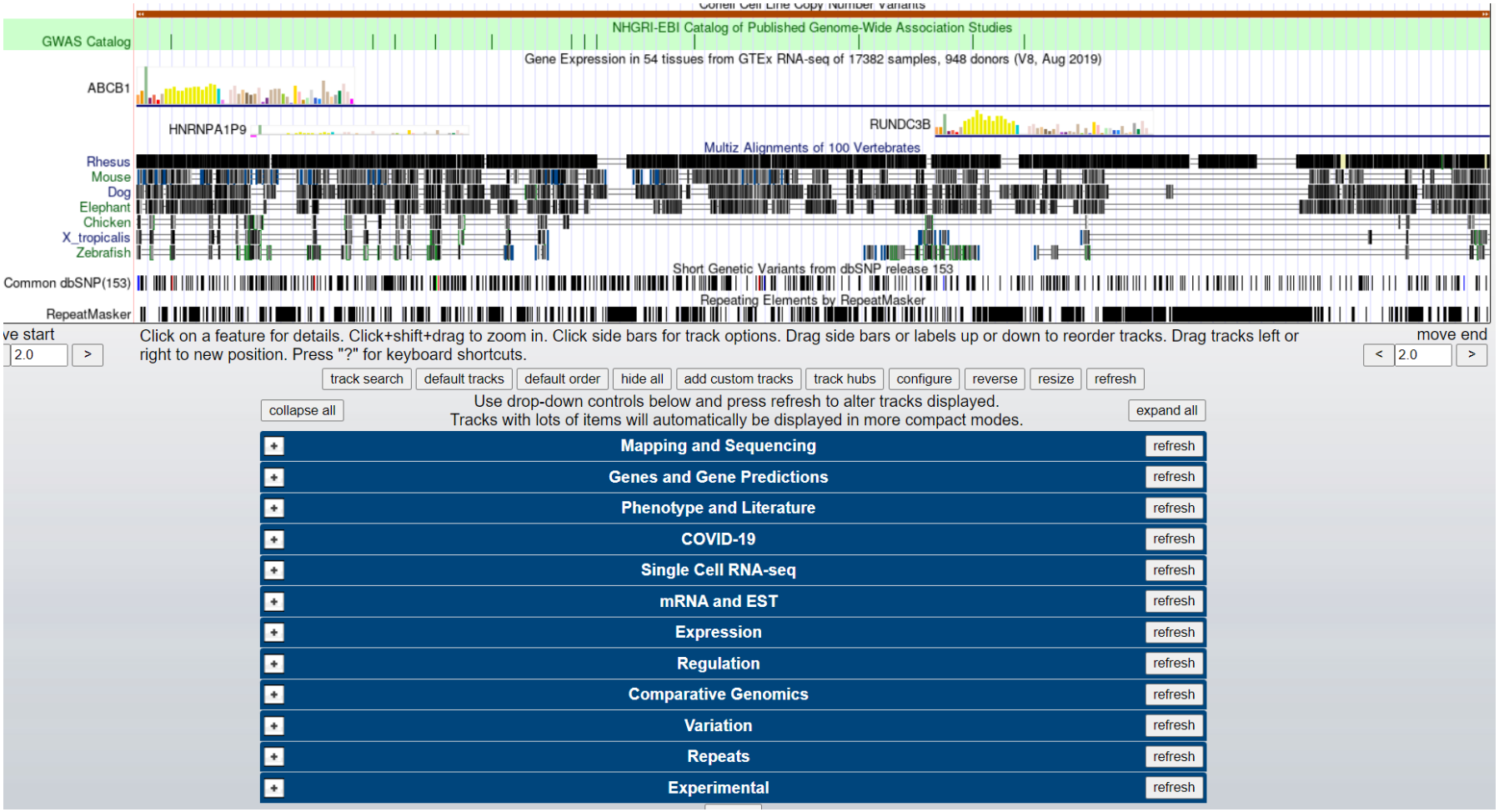
Download sequence and annotation data:

- Using rsync (recommended)
- Using HTTP
- Using FTP

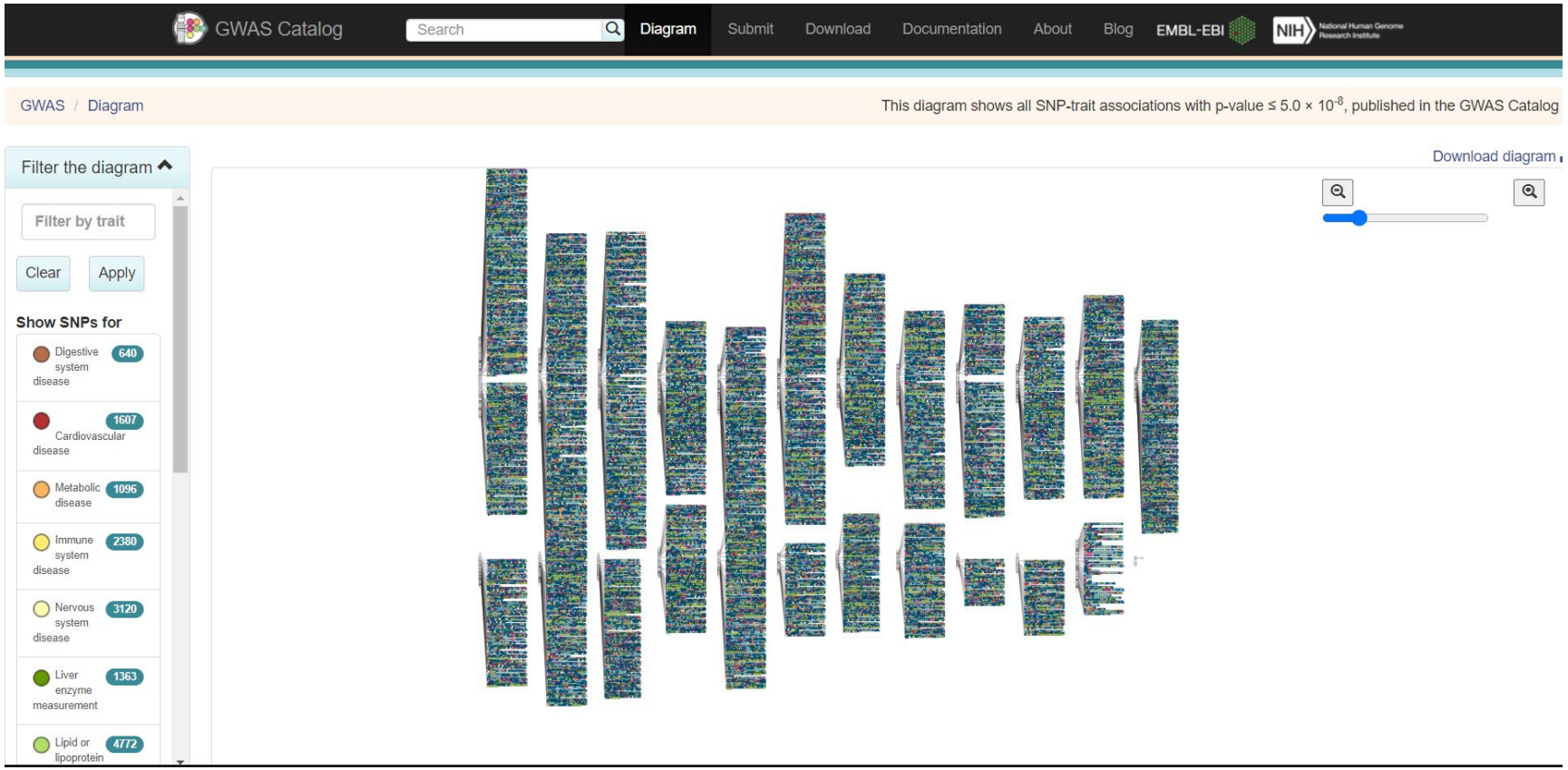
MAYO CLINIC

UCSC Genome Browser

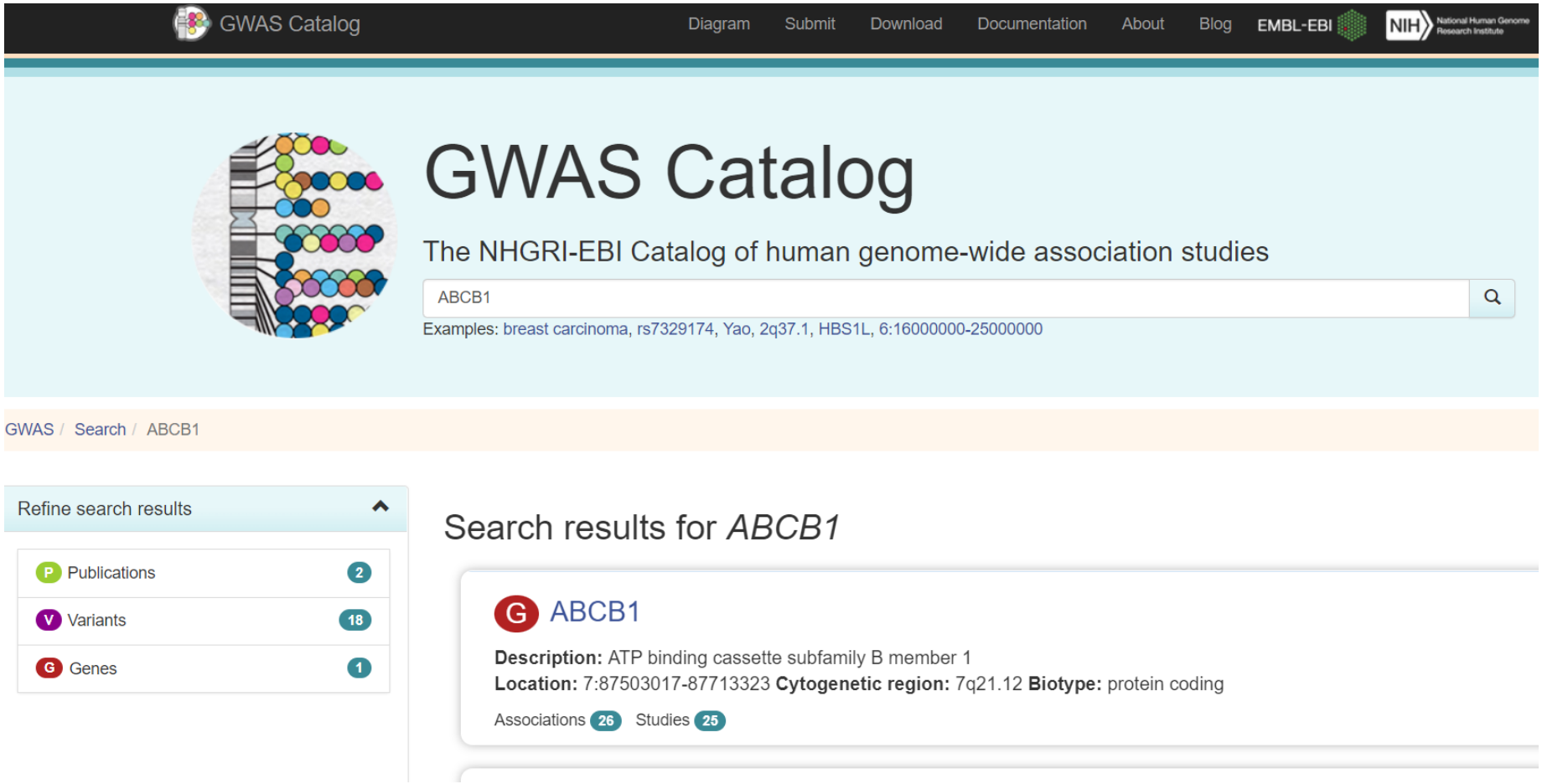
UCSC Genome Browser
Drop-down controls to display various tracks



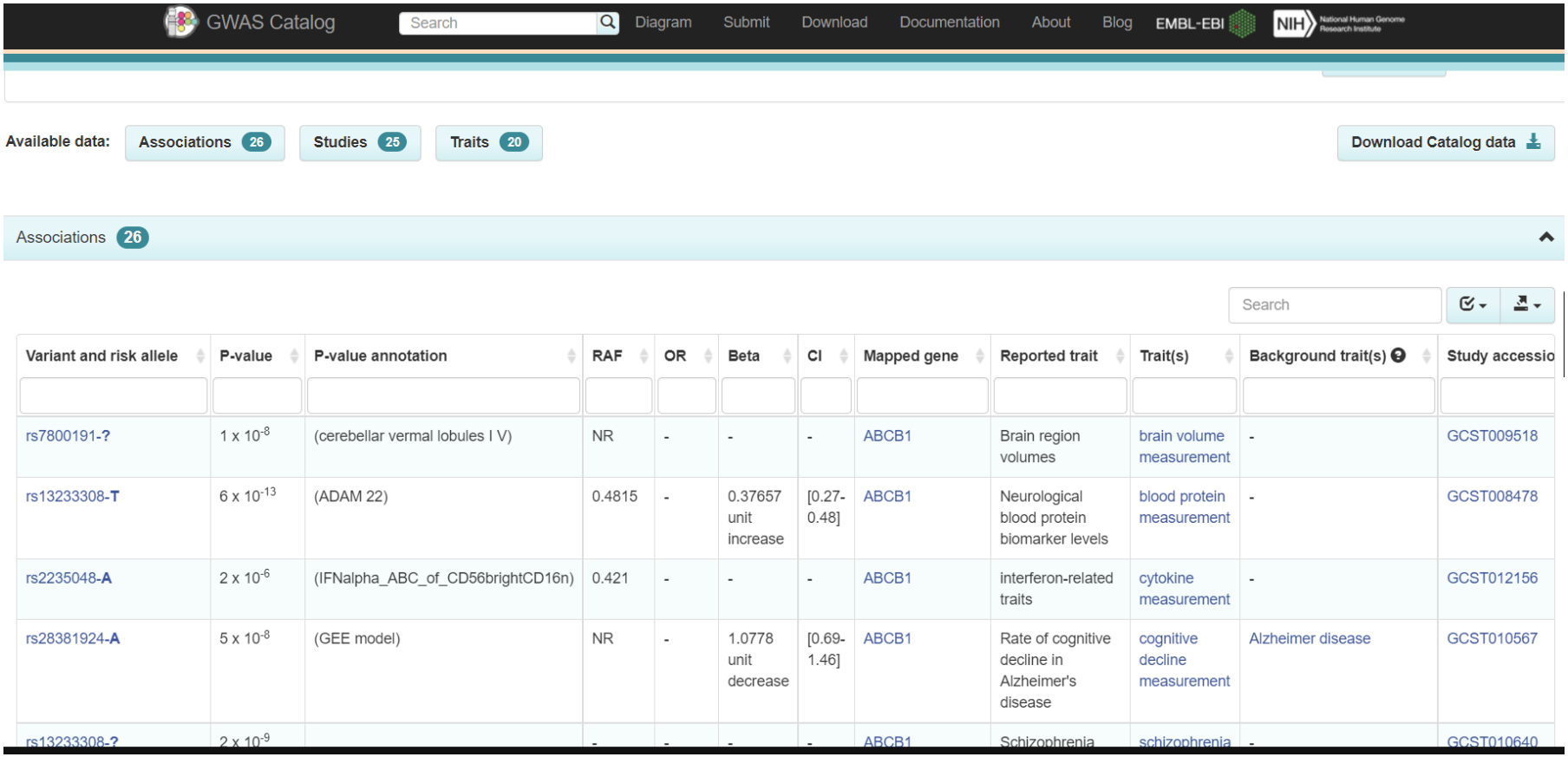
GWAS Catalog



GWAS Catalog - ABCB1 associations



GWAS Catalog - 26 ABCB1 associations



GWAS Catalog | Search | Diagram | Submit | Download | Documentation | About | Blog | EMBL-EBI | NIH National Human Genome Research Institute

Available data: Associations 26 | Studies 25 | Traits 20 | Download Catalog data

Associations 26

Search

Variant and risk allele	P-value	P-value annotation	RAF	OR	Beta	CI	Mapped gene	Reported trait	Trait(s)	Background trait(s)	Study accessio
rs7800191-?	1 x 10^{-8}	(cerebellar vermal lobules I V)	NR	-	-	-	ABCB1	Brain region volumes	brain volume measurement	-	GCST009518
rs13233308-T	6 x 10^{-13}	(ADAM 22)	0.4815	-	0.37657 unit increase	[0.27-0.48]	ABCB1	Neurological blood protein biomarker levels	blood protein measurement	-	GCST008478
rs2235048-A	2 x 10^{-6}	(IFNalpha_ABC_of_CD56brightCD16n)	0.421	-	-	-	ABCB1	interferon-related traits	cytokine measurement	-	GCST012156
rs28381924-A	5 x 10^{-8}	(GEE model)	NR	-	1.0778 unit decrease	[0.69-1.46]	ABCB1	Rate of cognitive decline in Alzheimer's disease	cognitive decline measurement	Alzheimer disease	GCST010567
rs13233308-?	2 x 10^{-9}		-	-	-	-	ABCB1	Schizophrenia	schizophrenia	-	GCST010640

Nature 2021 publication

ARTICLES
https://doi.org/10.1038/s41588-021-00921-z
nature genetics
Check for updates

A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease

Douglas P. Wightman[1], Iris E. Jansen[1], Jeanne E. Savage[1], Alexey A. Shadrin[2,3], Shahram Bahrami[2,3,4], Dominic Holland[5], Arvid Rongve[6,7], Sigrid Børte[3,8,9], Bendik S. Winsvold[9,10,11], Ole Kristian Drange[12,13], Amy E. Martinsen[3,9,10], Anne Heidi Skogholt[9,14], Cristen Willer[15], Geir Bråthen[16,17,18], Ingunn Bosnes[12,19], Jonas Bille Nielsen[9,15,20], Lars G. Fritsche[21], Laurent F. Thomas[9,14], Linda M. Pedersen[10], Maiken E. Gabrielsen[9], Marianne Bakke Johnsen[3,8,9], Tore Wergeland Meisingset[16,17], Wei Zhou[22,23], Petroula Proitsi[24], Angela Hodges[24], Richard Dobson[25,26,27,28,29], Latha Velayudhan[24], Karl Heilbron[30], Adam Auton[30], 23andMe Research Team*, Julia M. Sealock[31,32], Lea K. Davis[31,32], Nancy L. Pedersen[33], Chandra A. Reynolds[34], Ida K. Karlsson[33,35], Sigurdur Magnusson[36], Hreinn Stefansson[36], Steinunn Thordardottir[37], Palmi V. Jonsson[37,38], Jon Snaedal[37], Anna Zettergren[39], Ingmar Skoog[39,40], Silke Kern[39,40], Margda Waern[39,41], Henrik Zetterberg[42,43,44,45], Kaj Blennow[44,45], Eystein Stordal[12,19], Kristian Hveem[9,46], John-Anker Zwart[3,9,10], Lavinia Athanasiu[2,4], Per Selnes[47], Ingvild Saltvedt[16,18], Sigrid B. Sando[16,17], Ingun Ulstein[48], Srdjan Djurovic[49,50], Tormod Fladby[3,47], Dag Aarsland[24,51], Geir Selbæk[3,48,52], Stephan Ripke[23,53,54], Kari Stefansson[36], Ole A. Andreassen[2,3,4,56] and Danielle Posthuma[1,55,56] ✉

Late-onset Alzheimer's disease is a prevalent age-related polygenic disease that accounts for 50–70% of dementia cases. Currently, only a fraction of the genetic variants underlying Alzheimer's disease have been identified. Here we show that increased sample sizes allowed identification of seven previously unidentified genetic loci contributing to Alzheimer's disease. This study highlights microglia, immune cells and protein catabolism as relevant to late-onset Alzheimer's disease, while identifying and prioritizing previously unidentified genes of potential interest. We anticipate that these results can be included in larger meta-analyses of Alzheimer's disease to identify further genetic variants that contribute to Alzheimer's pathology.

Dementia has an age- and sex-standardized prevalence of ~7.1% in Europeans[1], with Alzheimer's disease (AD) being the most common form of dementia (50–70% of cases)[2]. identify the missing causal variants and may highlight additional disease mechanisms. In combination with increasing the number of samples, it is beneficial to use different approaches to identify rare

Nature 2021 publication

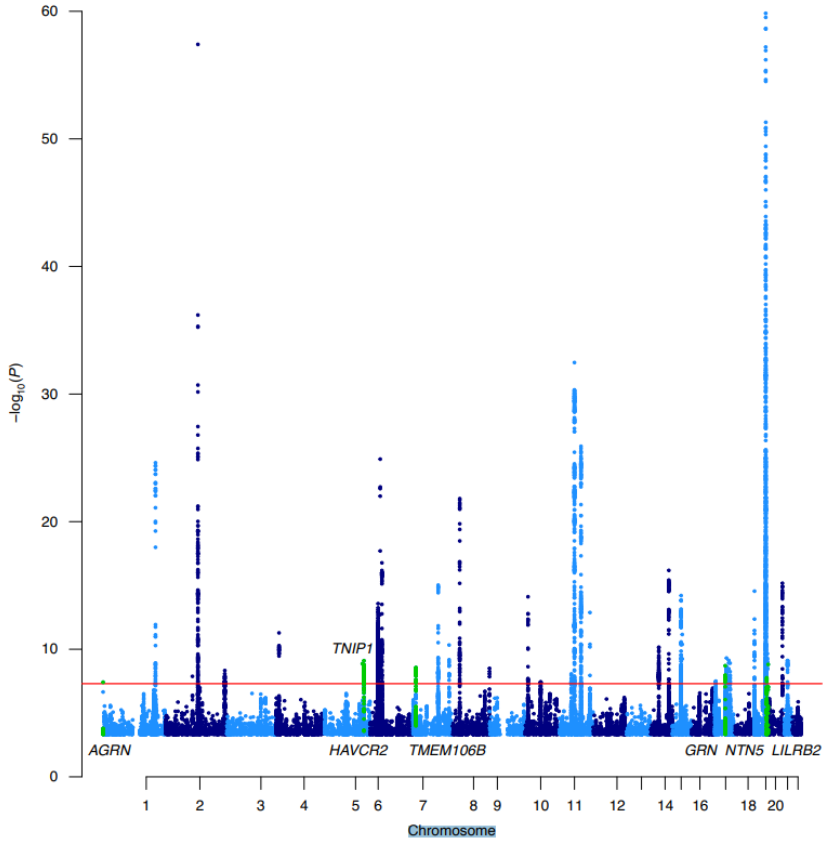
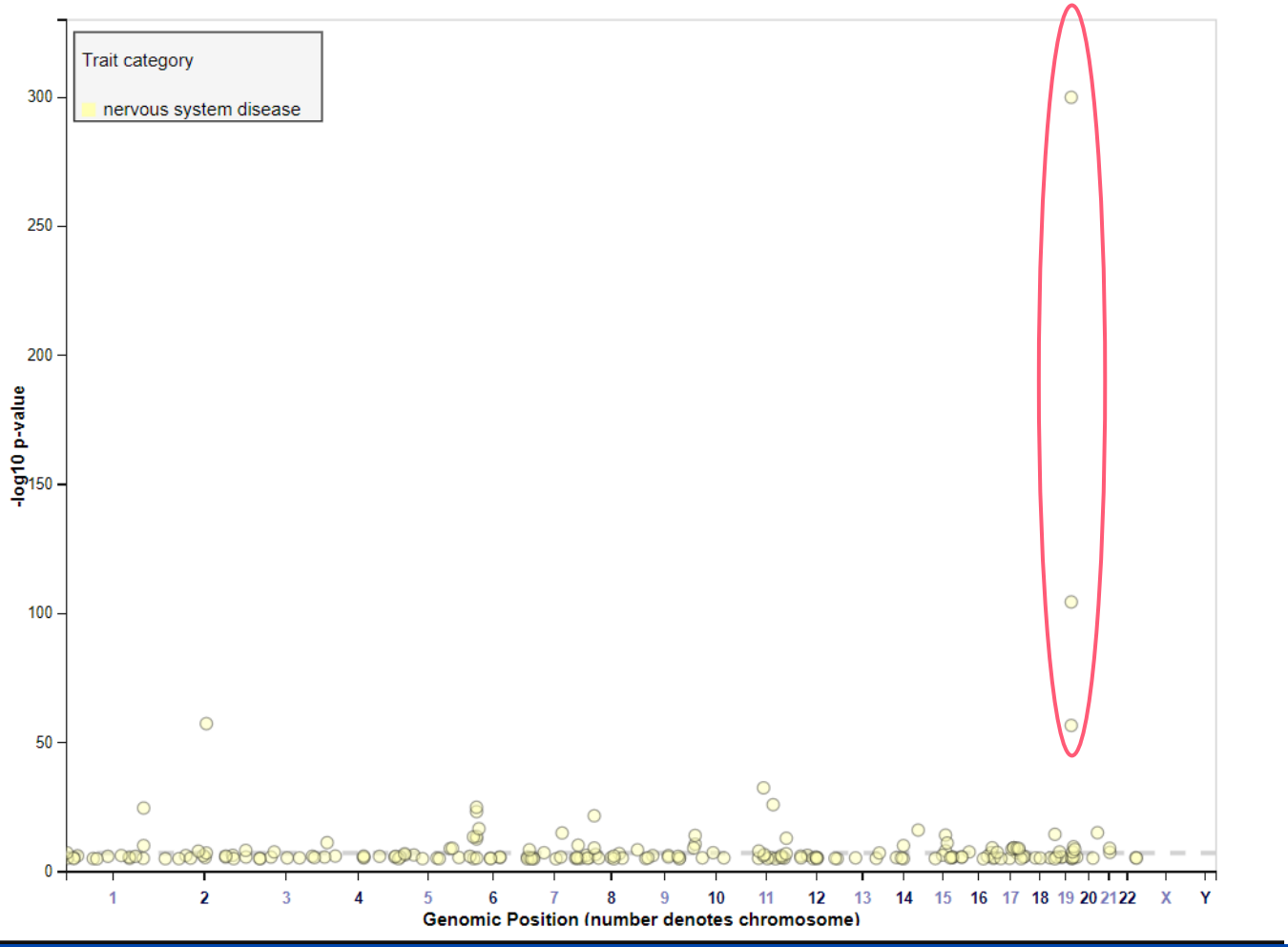


Fig. 1 | A Manhattan plot of the meta-analysis results highlighting 38 loci, including seven previously unidentified regions. Only variants with $P < 0.0005$ are displayed. The *APOE* region cannot be fully observed because the *y* axis is limited to the top variant in the second most significant locus, $-\log_{10}(1\times10^{-60})$, to display the less significant variants. The red line represents genome-wide significance (5×10^{-8}). The *P* values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing. The previously unidentified loci are highlighted in green and indicated by the assigned gene name. The *TNIP1/HAVCR2* regions and the *NTN5/LILRB2*

GWAS – Catalog Alzheimer's disease associations



Knowledge Portal Network

Common Metabolic Diseases Knowledge Portal
Learn more

Cardiovascular Disease Knowledge Portal
Learn more

Cerebrovascular Disease Knowledge Portal
Learn more

Sleep Disorder Knowledge Portal
Learn more

Type 1 Diabetes Knowledge Portal
Learn more

Type 2 Diabetes Knowledge Portal
Learn more

Musculoskeletal Knowledge Portal
Learn more

ALS Knowledge Portal
Learn more

Lung Disease Knowledge Portal
Learn more

Lipid Droplet Knowledge Portal
Learn more

Non-Additive Genetic Effects Knowledge Portal
Learn more

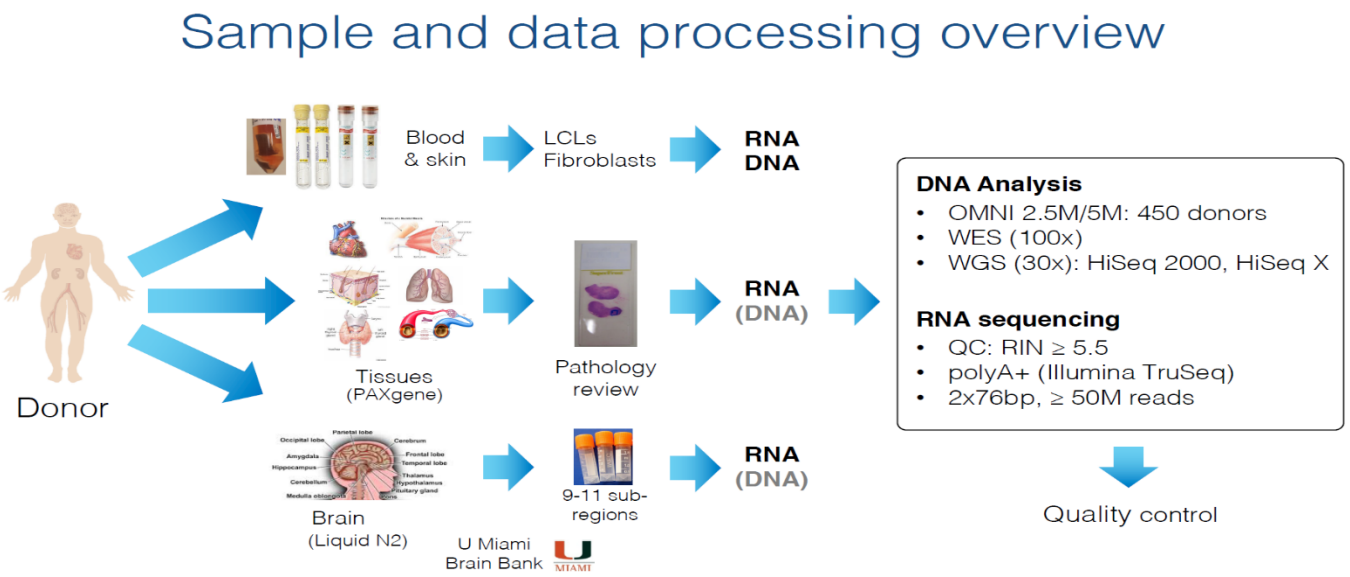
Collaborate on methods
Learn more

Collaborate to create a new Knowledge Portal
Learn more

MAYO CLINIC

Genotype Tissue-Expression Project (GTEX)

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



Overview of GTEx resources: open-access data

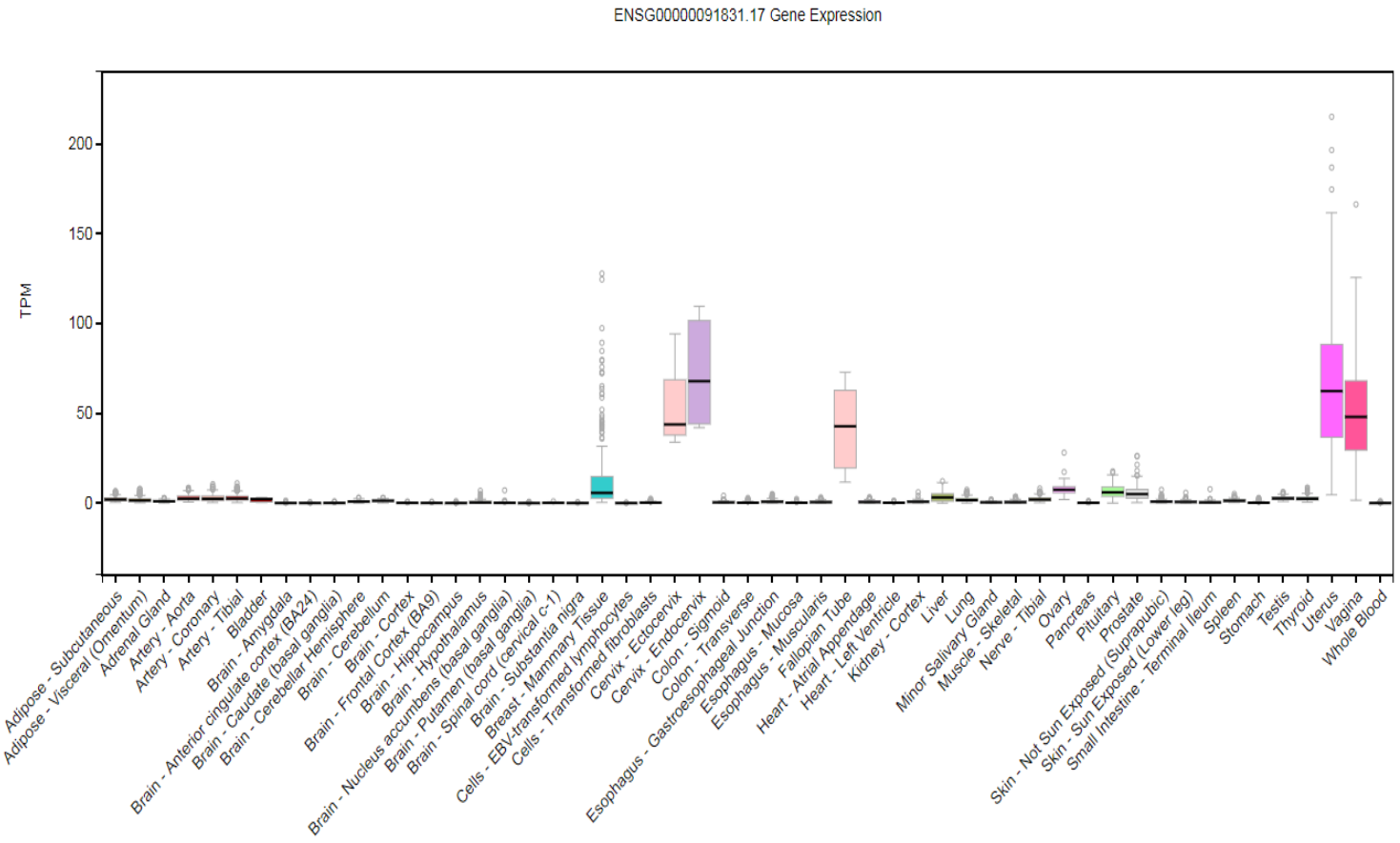
- Expression
  - Gene-level expression (TPM, counts)
  - Transcript-level expression (TPM, counts, isoform proportions)
  - Exon read counts
- QTLs
  - Single-tissue eQTLs (*cis*- and *trans*-)
  - Multi-tissue eQTLs
  - Future: splicing QTLs
- Histology images
- De-identified public access sample and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

All open-access data is available at gtexportal.org

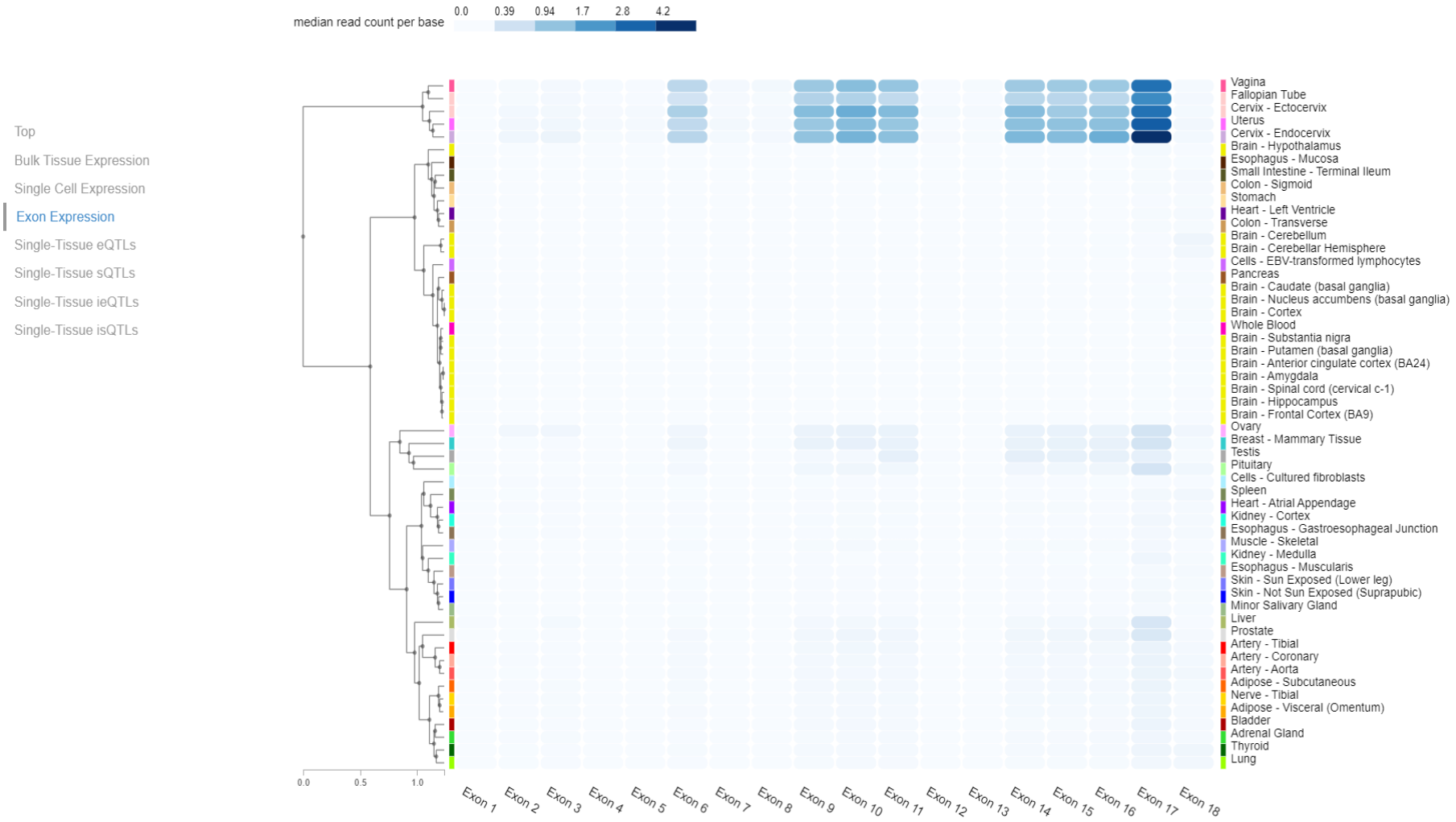
ESR1 query

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

Plot: Gene | Isoform
Differentiation: None | Sex
Scale: Log | Linear
Sort: ABC | ▲ | ▼
Outliers: On | Off
Display: Boxes | Medians
Filter: Tissue



Exon expression



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

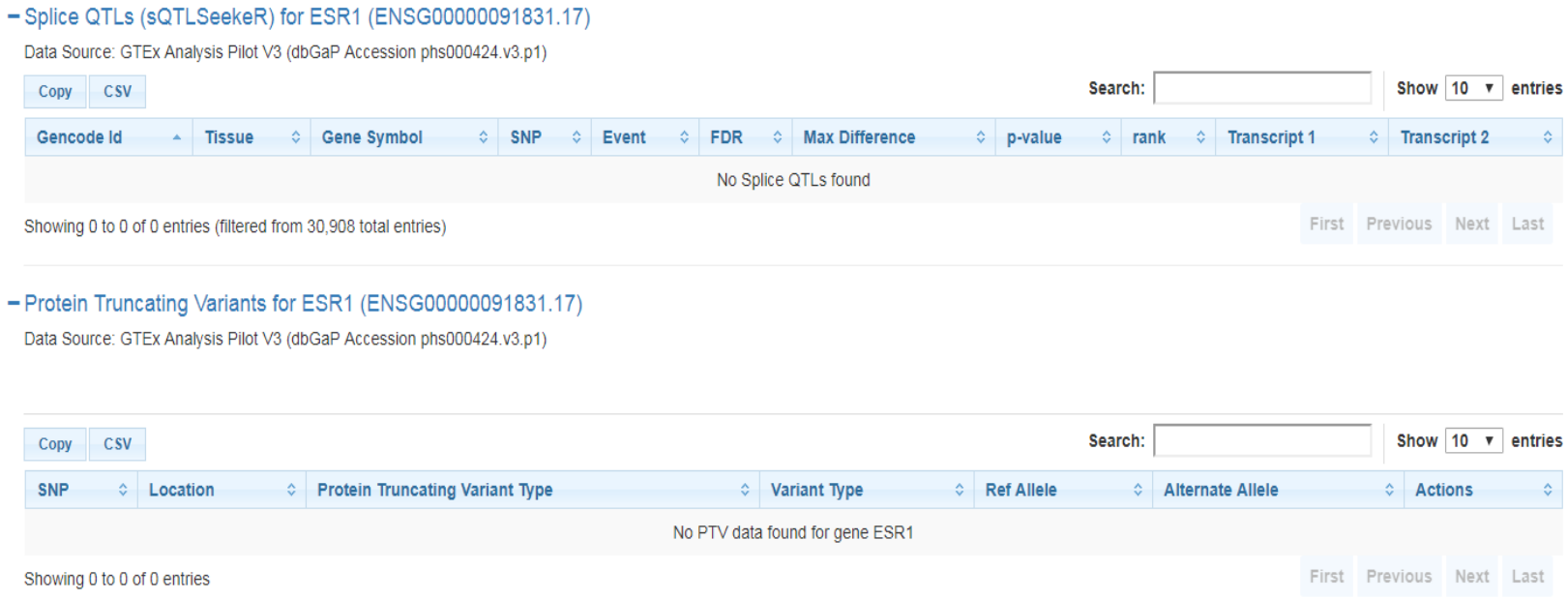
Copy CSV Search: Show 10 entries

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

The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA)

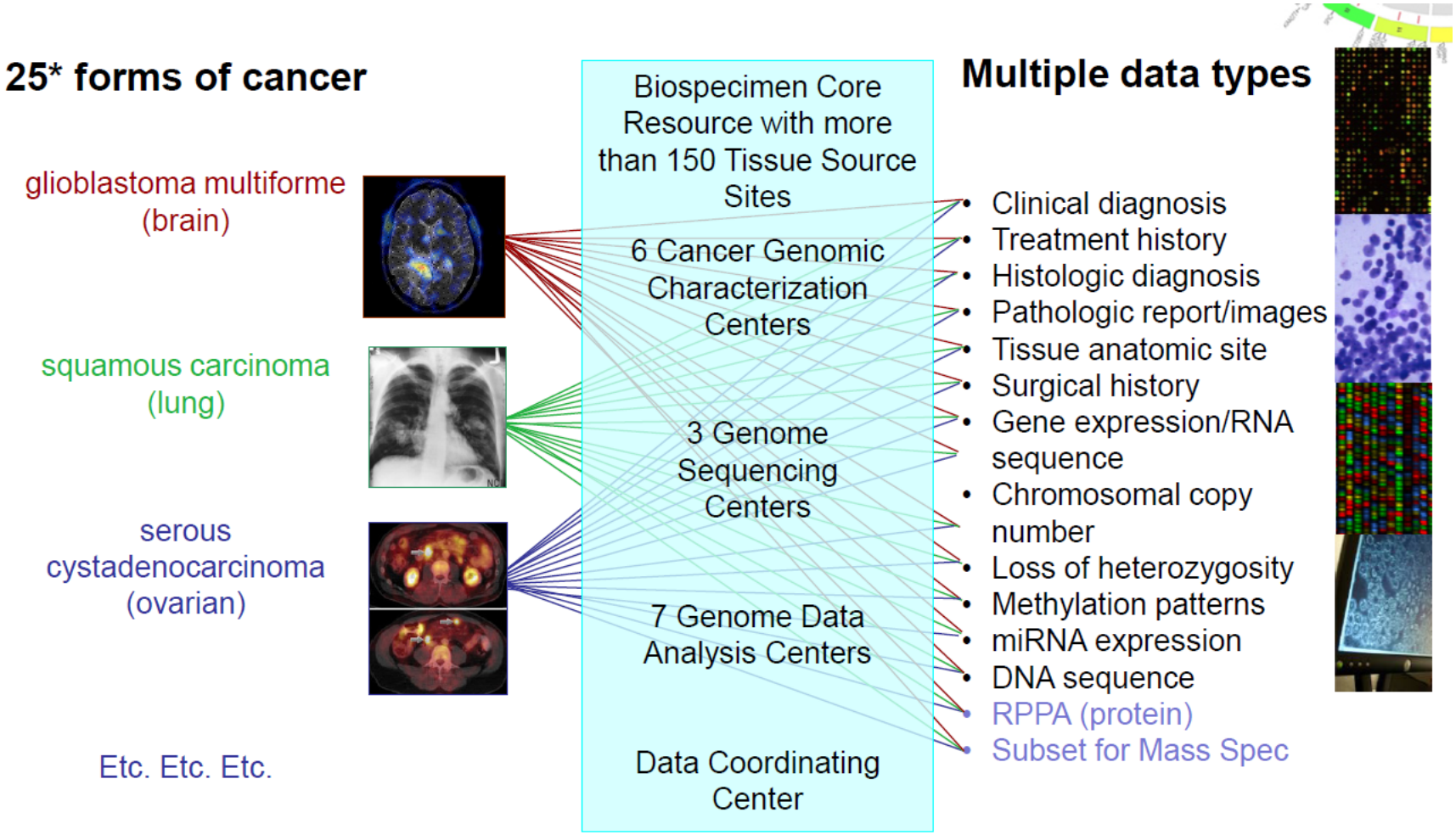
A comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing.

Cancer.gov

The Cancer Genome Atlas (TCGA; https://cancergenome. nih.gov/)

- Multi-omics data sets for > 33 cancer types
- For more than 30000 individual tumor samples
- RNA-Seq, DNA-Seq, miRNA-Seq, single-nucleotide variant (SNV), copy number variation (CNV), DNA methylation, and reverse phase protein array (RPPA) data
- The biospecimens from TCGA are analyzed by mass spectrometry technique, and the cancer cohort proteomics data are available at Clinical Proteomic Tumor Analysis Consortium (CPTAC) (https://cptac-data-portal. georgetown.edu/cptacPublic/)

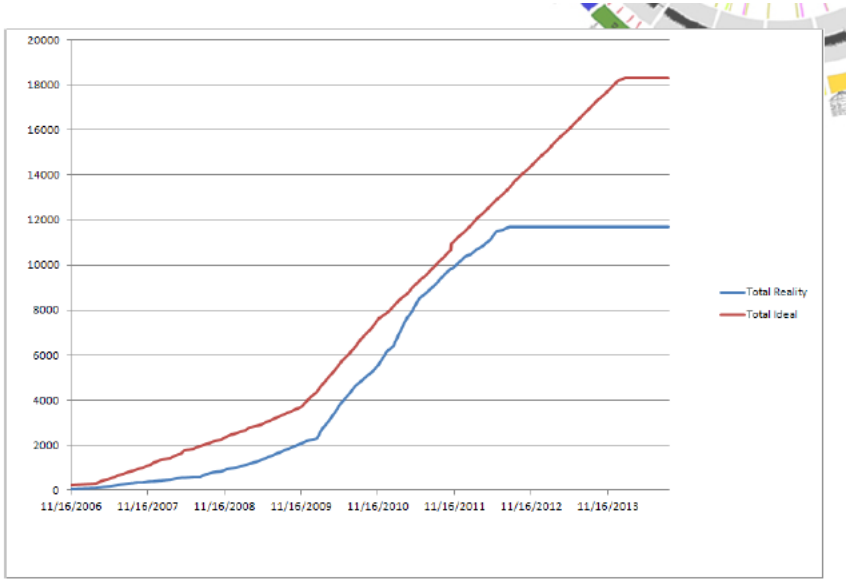
TCGA multiple data types



Reusing the slides from Kenna Shaw's presentation

Rare tumors projects

- 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??

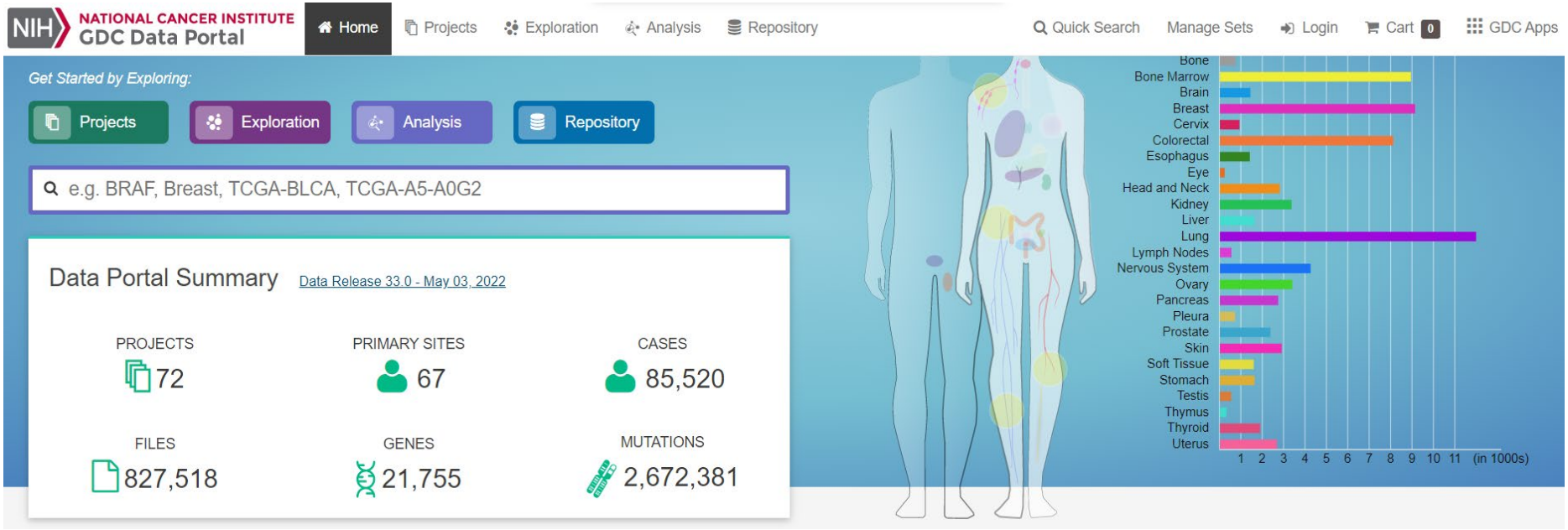


Genomic Data Commons - GDC

Genomic Data Commons

- A NCI repository for The Cancer Genome Atlas and Genomics data.
- It consists of data from 72 projects
- 67 primary sites
- >85K cases
- >2.6 million mutations
- 827,518 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

Search Projects

e.g. TCGA-GBM, Brain

Primary Site

	# Projects
kidney	20
hematopoietic and reticuloendothelial syst…	18
bronchus and lung	17
breast	16
colon	16

62 More...

Program

	# Projects
TCGA	33
TARGET	9
GENIE	8
CMI	3
BEATAML1.0	2

15 More...

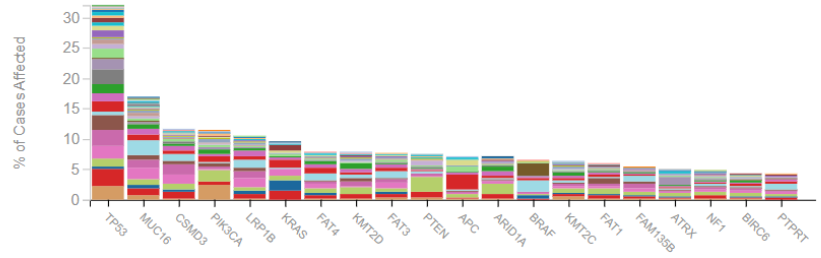
Disease Type

	# Projects
adenomas and adenocarcinomas	36
cystic, mucinous and serous neoplasms	22
squamous cell neoplasms	20
epithelial neoplasms, nos	18
ductal and lobular neoplasms	17

54 More...

Top Mutated Cancer Genes in Selected Projects

% of Cases Affected | # of Cases Affected



Case Distribution per Project

85,520 Cases across 72 Projects

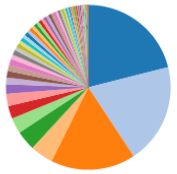


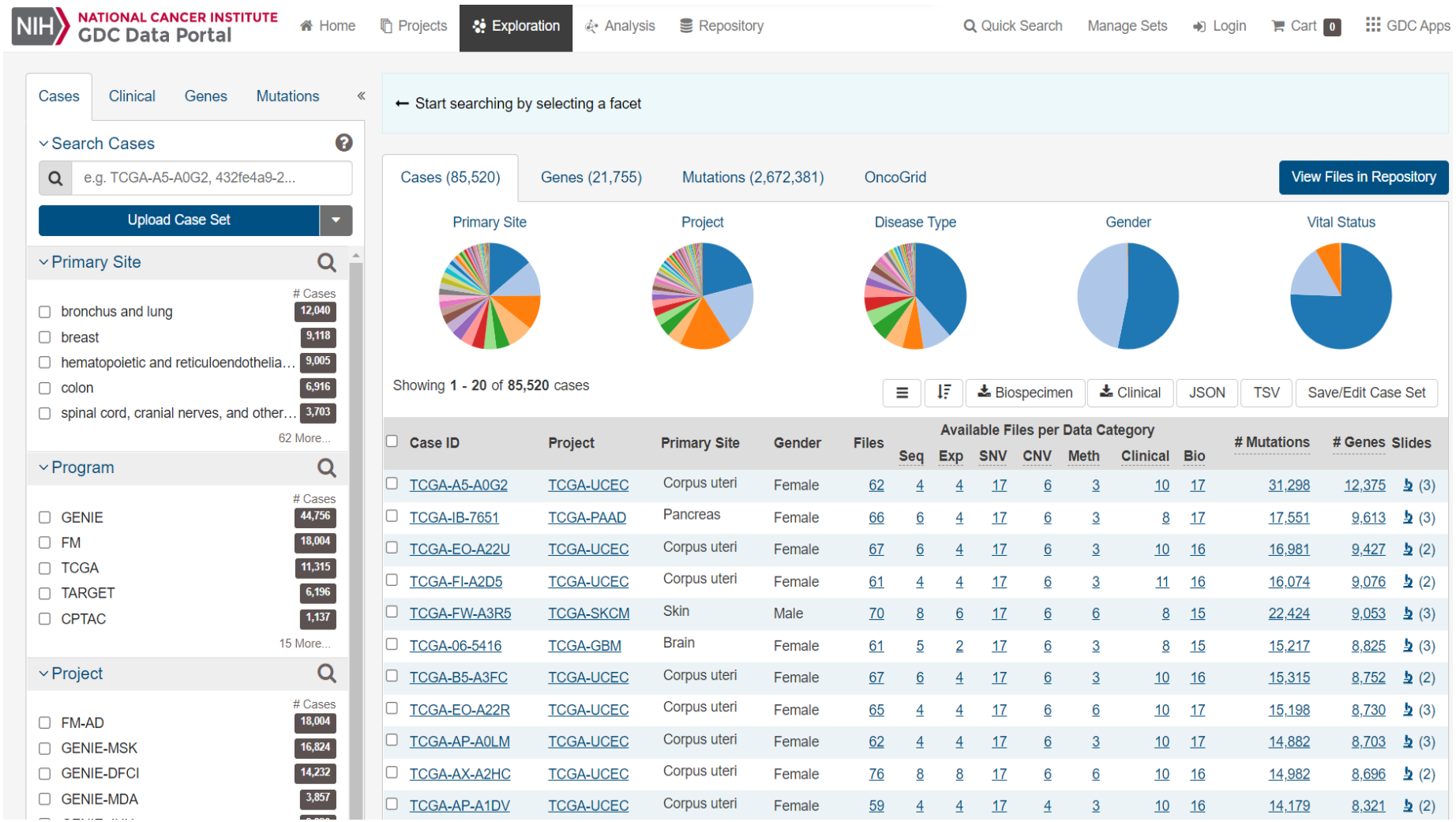
Table | Graph

72 Projects

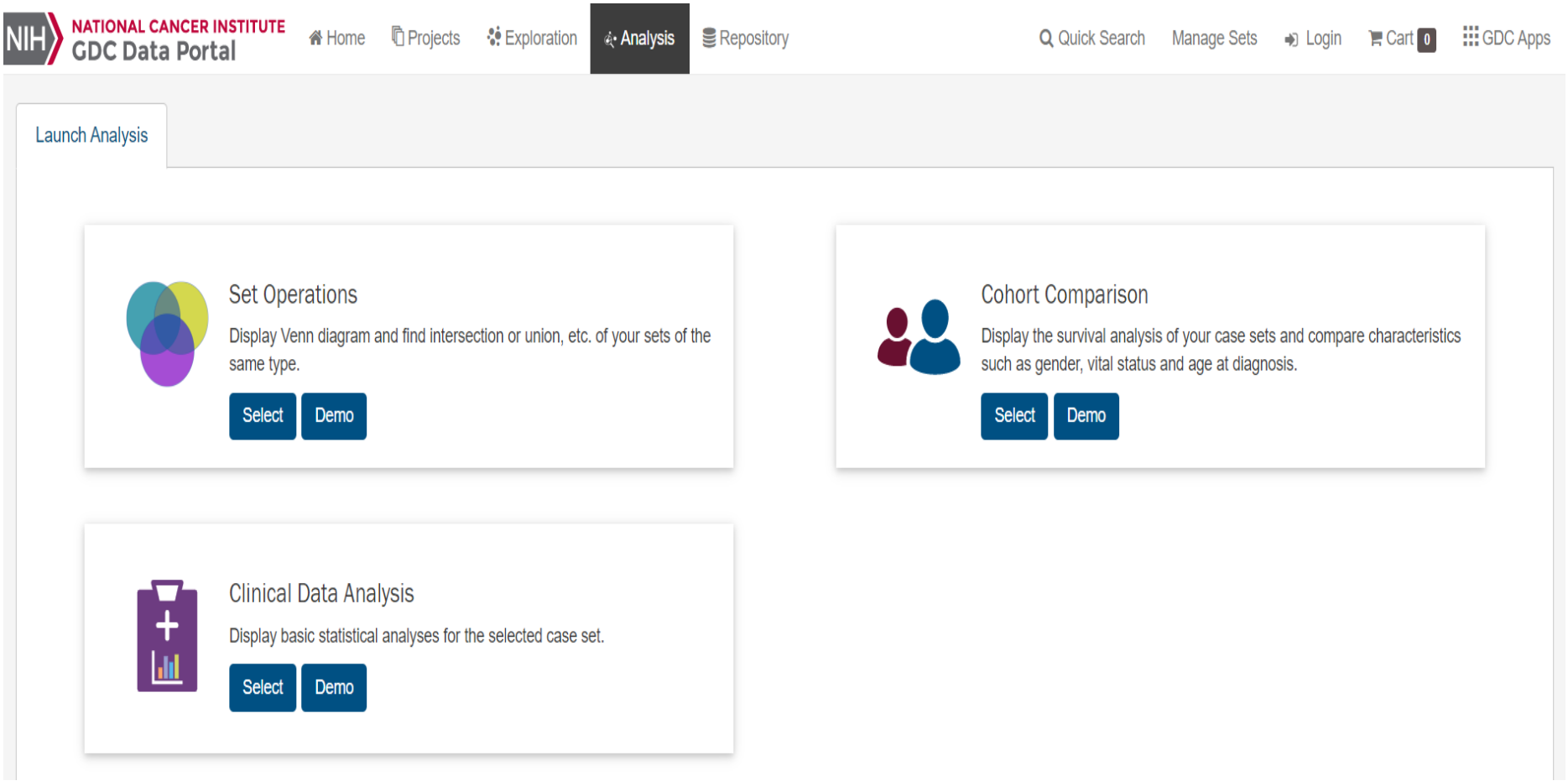
JSON | TSV

					Available Cases per Data Category									
Project	Disease Type	Primary Site	Program	Cases ↑	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	54,096
GENIE-MSK	▾ 49 Disease Types	▾ 49 Primary Sites	GENIE	16,824	0	0	16,823	16,823	0	16,824	0	16,824	0	36,470
GENIE-DFCI	▾ 52 Disease	▾ 49 Primary	GENIE	14,232	0	0	14,232	14,232	0	14,232	0	14,232	0	28,464

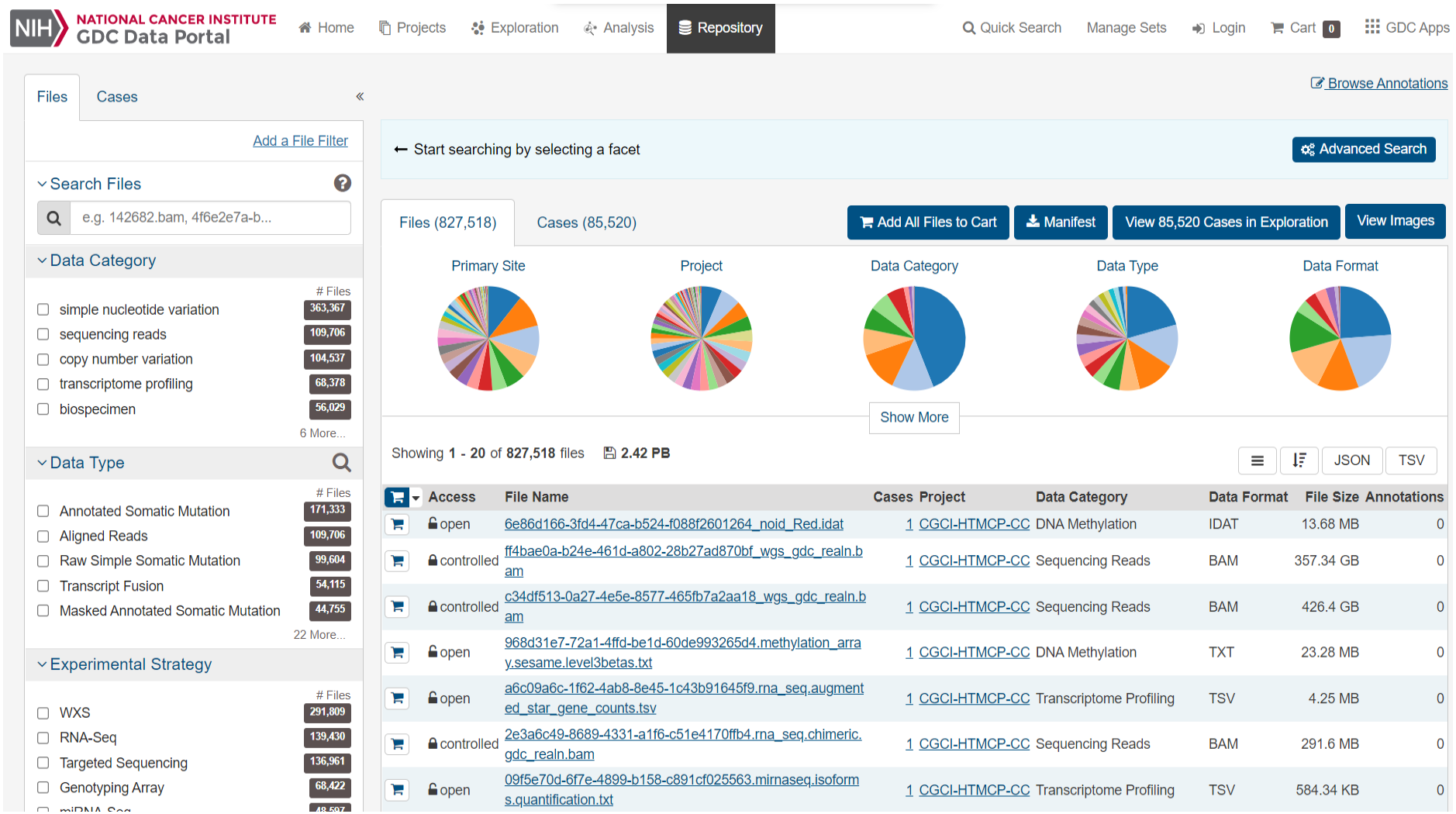
Exploration



Analysis



Repository



International Cancer Genomics Consortium (ICGC; https:// icgc.org/)

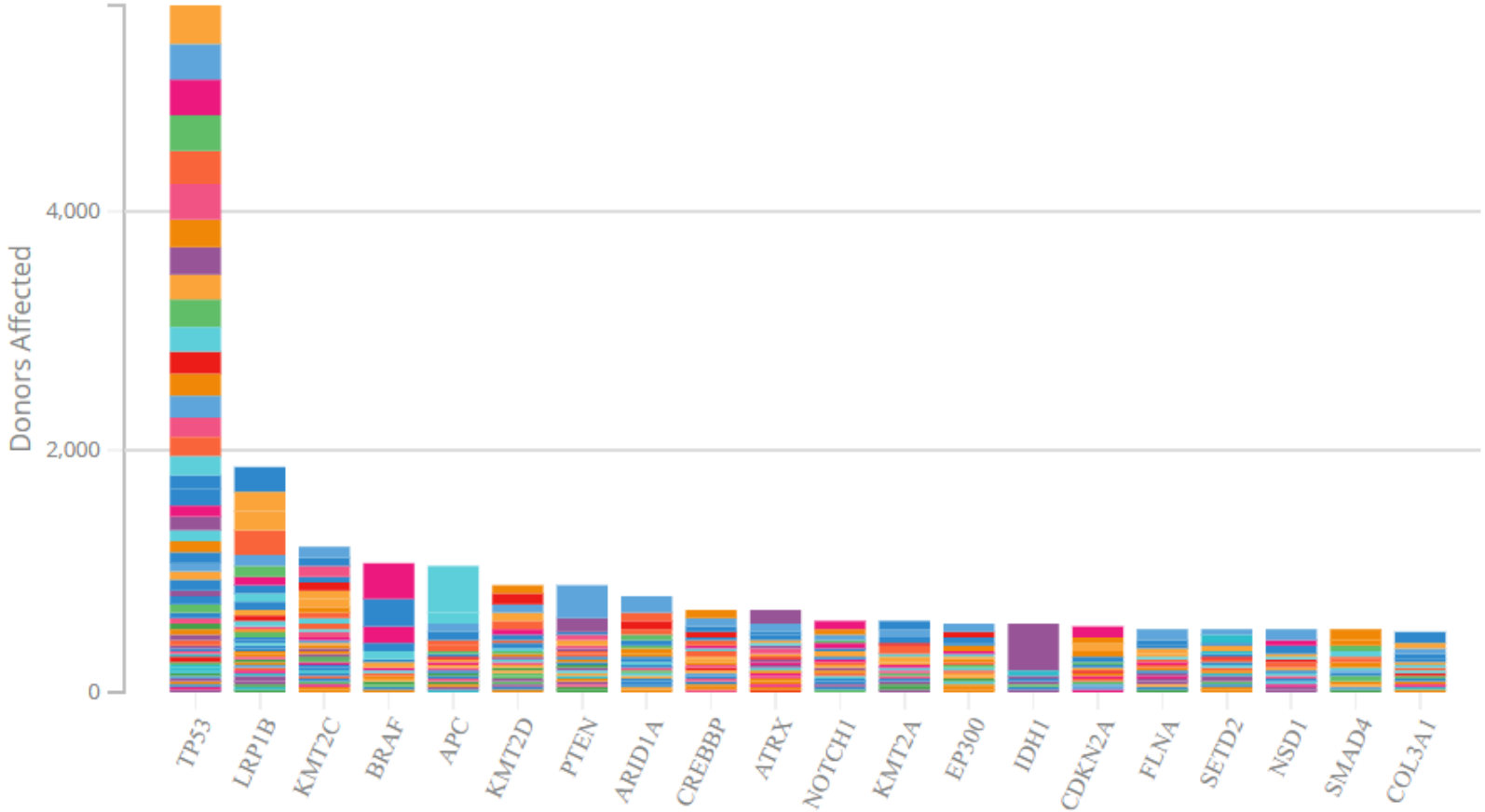
- Large-scale generation of genome studies from 86 cancer projects in 22 primary cancer sites from 22,330 donors (RELEASE 28)
- This project mainly contains mutation-related genomic alteration data (both germline and somatic) across cancer types from various ethnicities
- The Pan-cancer analysis of whole genomes (PCAWG; https://dcc.icgc.org/pcawg) allows the exploration and analysis of more than 2800 whole genomes from ICGC

International Cancer Genome Consortium

• Cancer Projects	86
• Cancer Primary Sites	22
• Patients with molecular data	22,330
• Total Donors	24,289
• Somatic mutations identified	81.7million

https://dcc.icgc.org/projects

Top 20 mutated cancer genes with high functional impact somatic mutations



https://dcc.icgc.org/projects

Cancer Cell Line Encyclopedia (CCLE; (https://portals.broadinstitute.org/ccle)

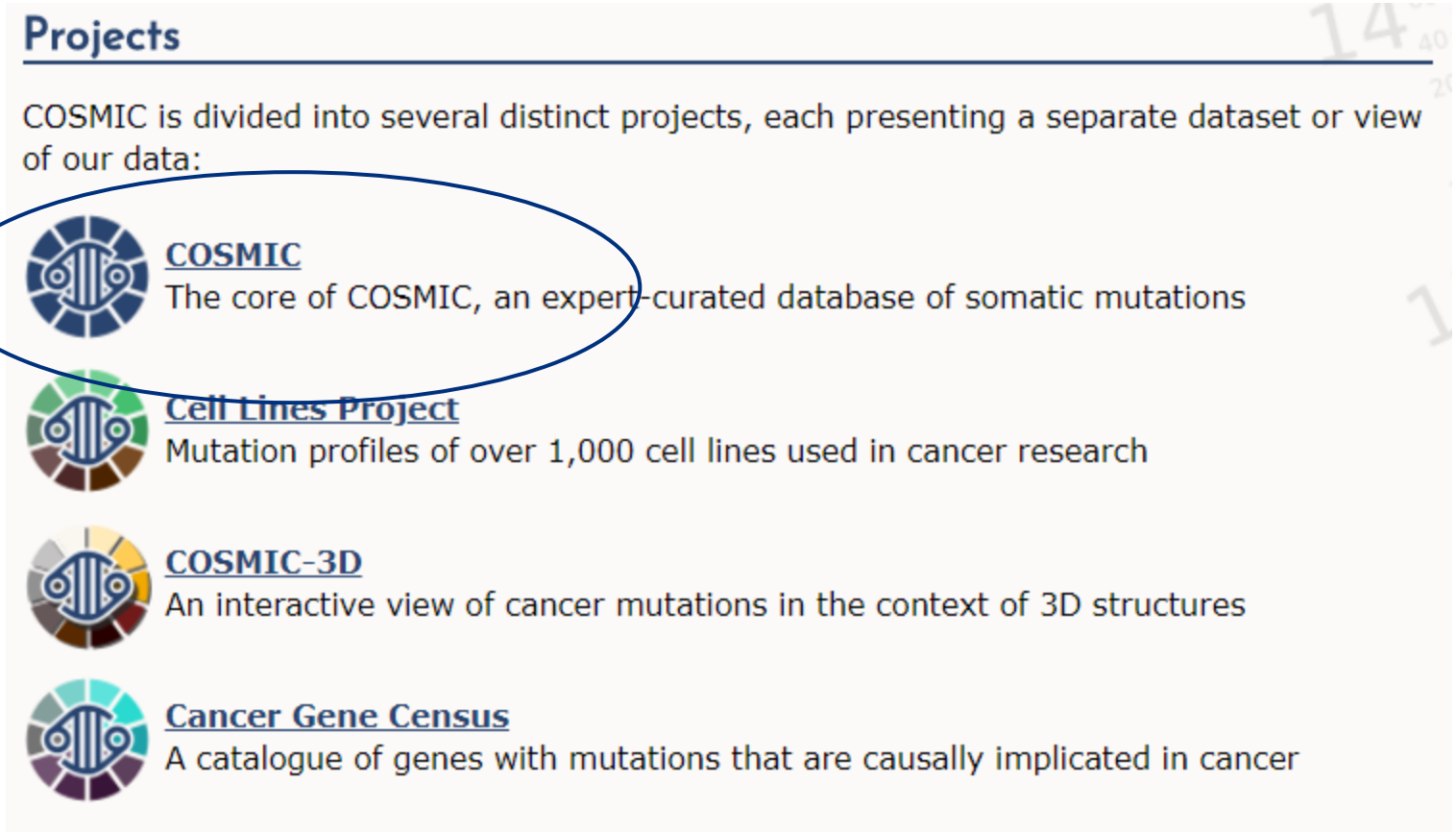
- Hosted by Broad institute is a compilation of gene expression, copy number, and sequencing data from 947 human cell lines and for 36 tumor types.
- It also houses the pharmacological profiles of 24 anticancer drugs across the cancer cell lines.
- https://depmap.org/portal/interactive/

MAYO CLINIC

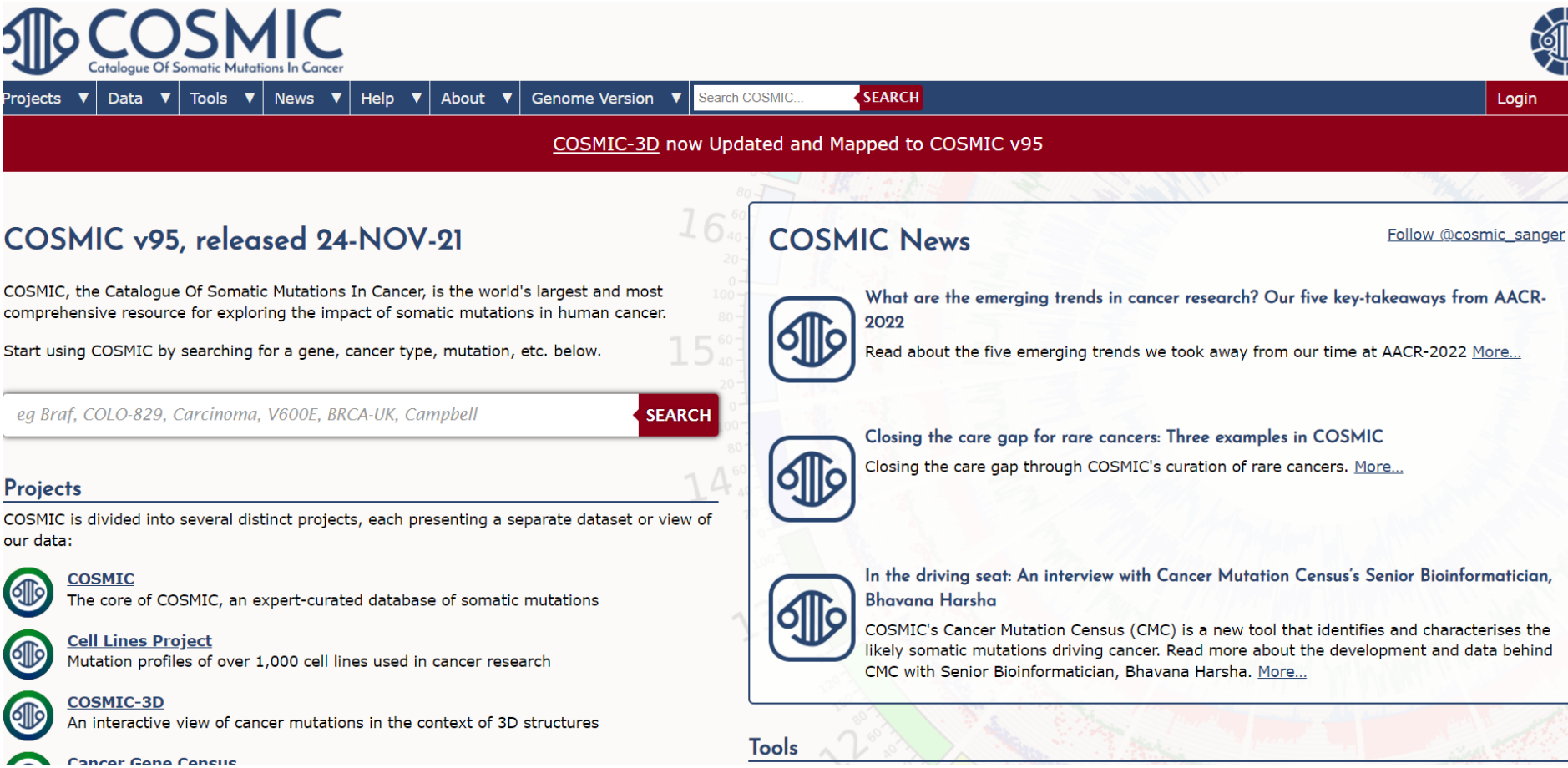
COSMIC

“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



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

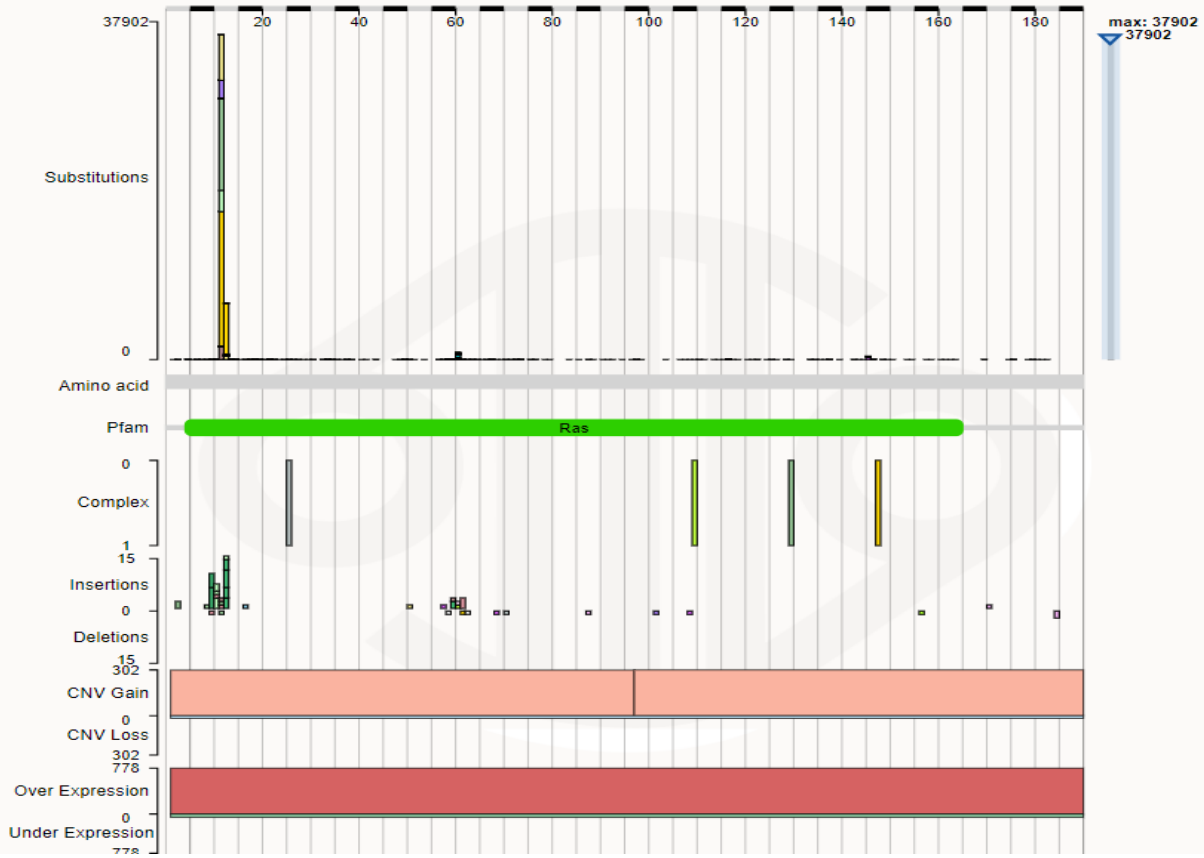
Coordinate system

- (•) 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_ENST00000311...

- 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 — 189

1 48 95 142 189

Coordinate system

- (•) Amino-acid
- ( ) cDNA

Apply filters | Reset filters

Overview

This section gives an overview of KRAS_ENST00000311936, along with links to any related data and resources.

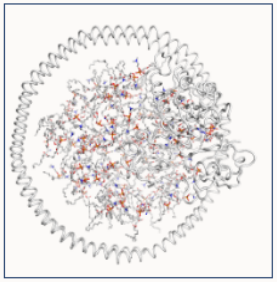
Census gene | Curated gene | **Mouse gene** | Hallmark gene

COSMIC gene KRAS_ENST00000311936 (COSG4)

Genomic coordinates 12:25204789..25250931 (negative strand)

Synonyms KRAS, KRAS1, KRAS2, CCDS8702.1, P01116, ENSG00000133703.11, NM_004985.4, NP_004976

COSMIC-3D There are **201** structures for **KRAS_ENST00000311936**. View them in COSMIC-3D.



Number of samples 275946 unique samples
49449 unique samples with mutations

Alternative transcripts KRAS, KRAS_ENST00000556131, KRAS_ENST00000557334

Sequences You can see various sequences for this gene:
cDNA (ENST00000311936.7)
Protein (KRAS_ENST00000311936)
Transcript and protein aligned (ENST00000311936.7+KRAS_ENST00000311936)

Gene fusions KRAS_ENST00000311936 is involved in 1 fusion, with the following gene:
UBE2L3_ENST00000342192 (1 mutation in 1 sample)

Drug sensitivity data n/a

Drug Resistance and tissue distribution

Gene

KRAS_ENST00000311...

- 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

1 189

1 48 95 142 189

Coordinate system

- (•) Amino-acid
- ( ) cDNA

Apply filters | Reset filters

Drug resistance

This section shows the drugs associated with **KRAS_ENST00000311936** 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

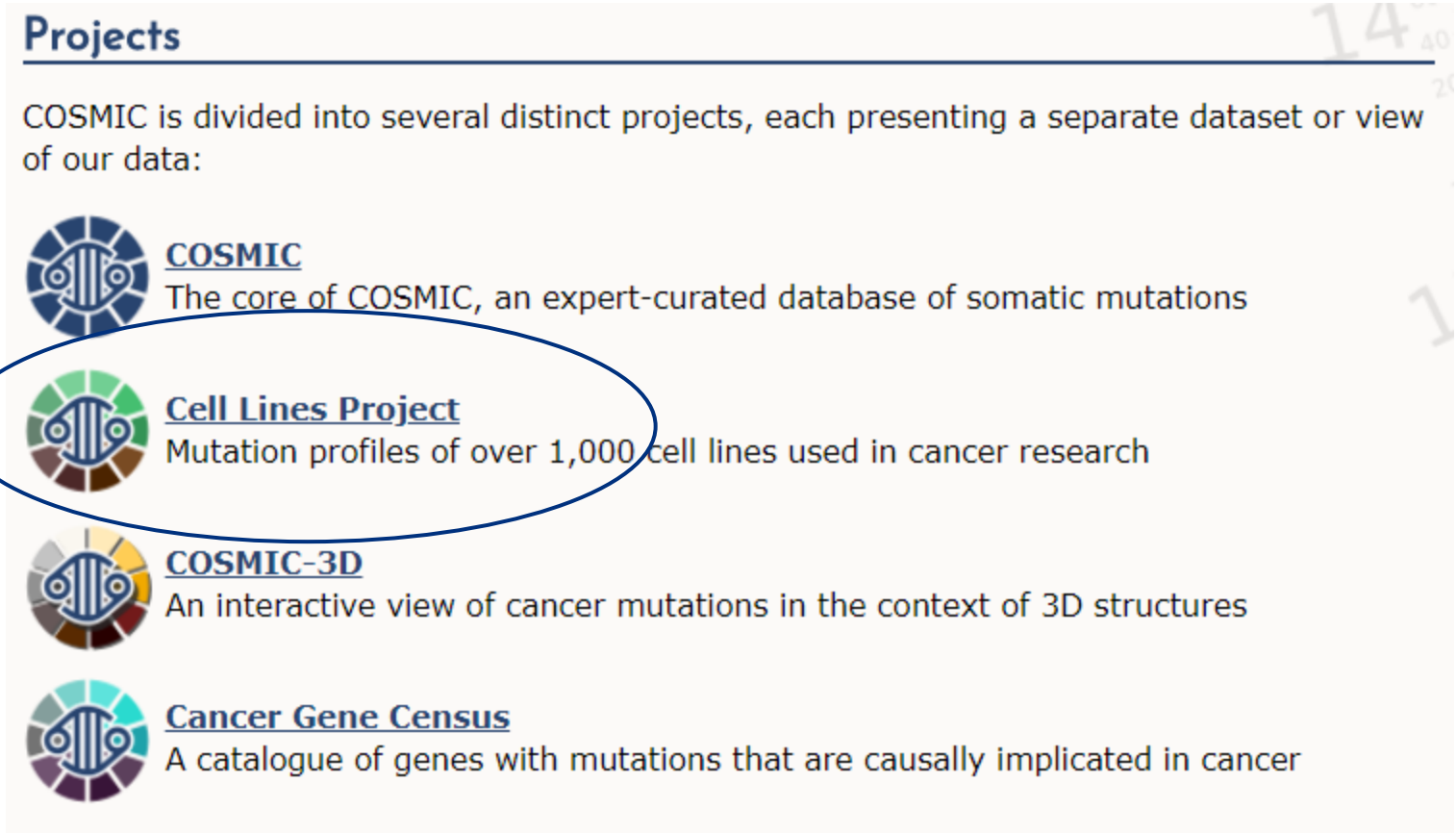
The table shows the distribution of mutations across the primary tissue types that are curated by COSMIC. Histograms show the percentage of mutated samples for point mutations, CNV data and gene expression data. Moving your mouse over the histograms will show additional data. The number of samples tested 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:

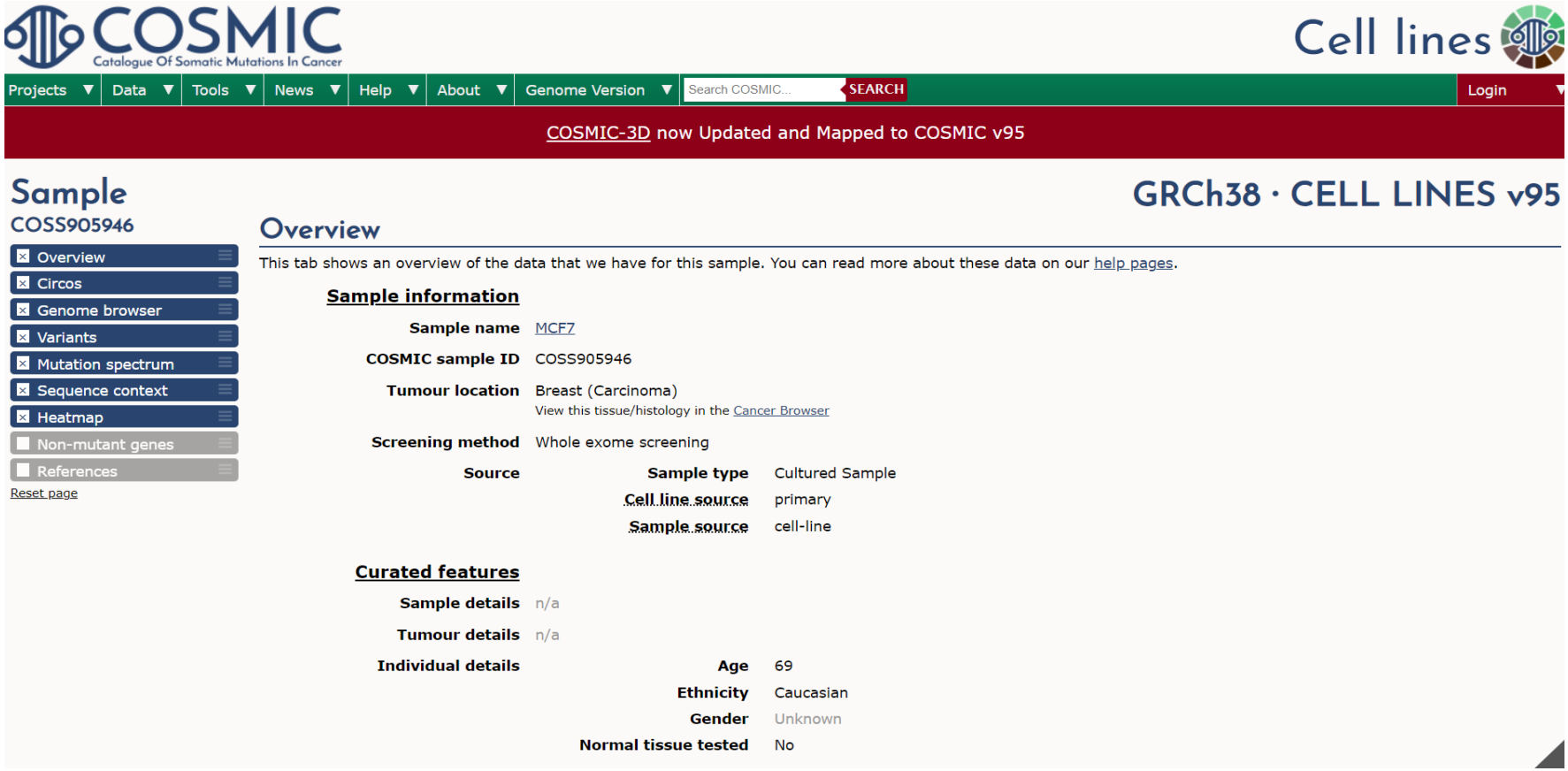
Tissue	Point Mutations		Copy Number Variation		Gene Expression		Methylation	
	% Mutated	Tested	Variant %	Tested	% Regulated	Tested	% Diff. Methylated	Tested
Adrenal gland		1330		267		-		-
Autonomic ganglia		1591		-		-		-
Biliary tract		5583		-		-		-
Bone		1202		-		-		-
Breast		11611		1492		-		-
Central nervous system		5051		1035		-		-
Cervix		2865		299		-		-
Endometrium		4817		586		-		-

Cell Lines Project

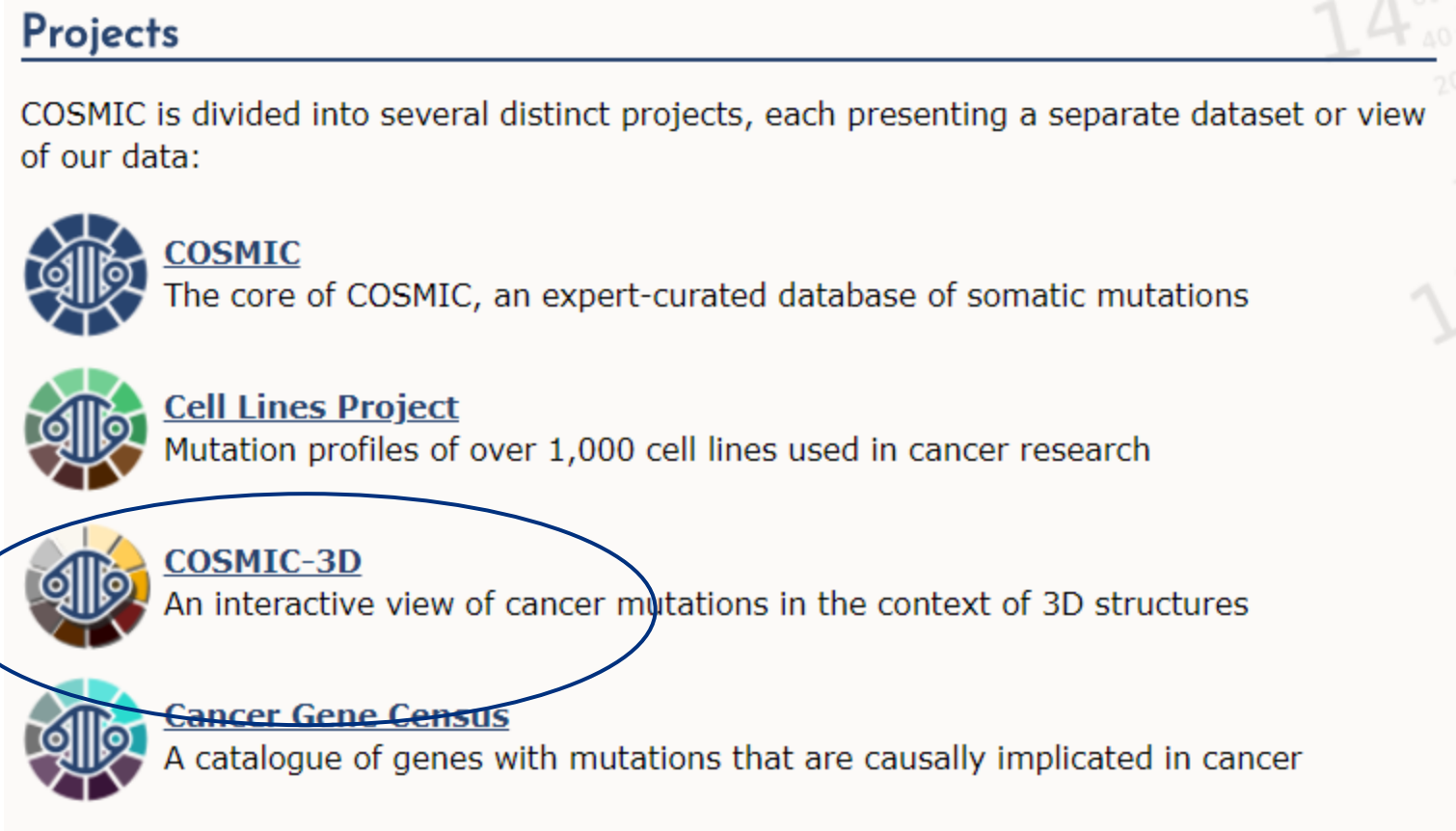


Cell lines project

- Mutation profiles of over 1,000 cell lines used in cancer research (e.g.MCF7)

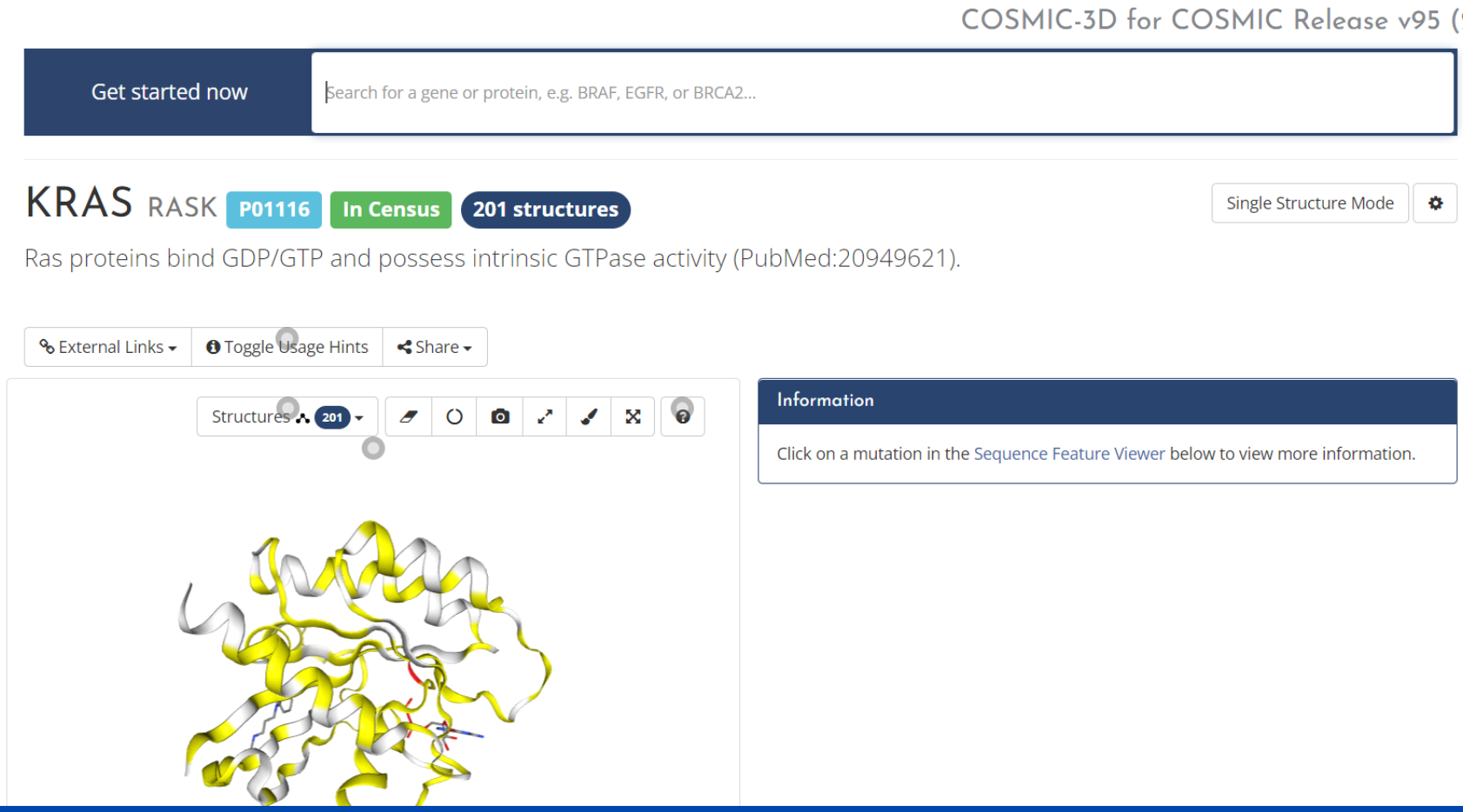


COSMIC-3D

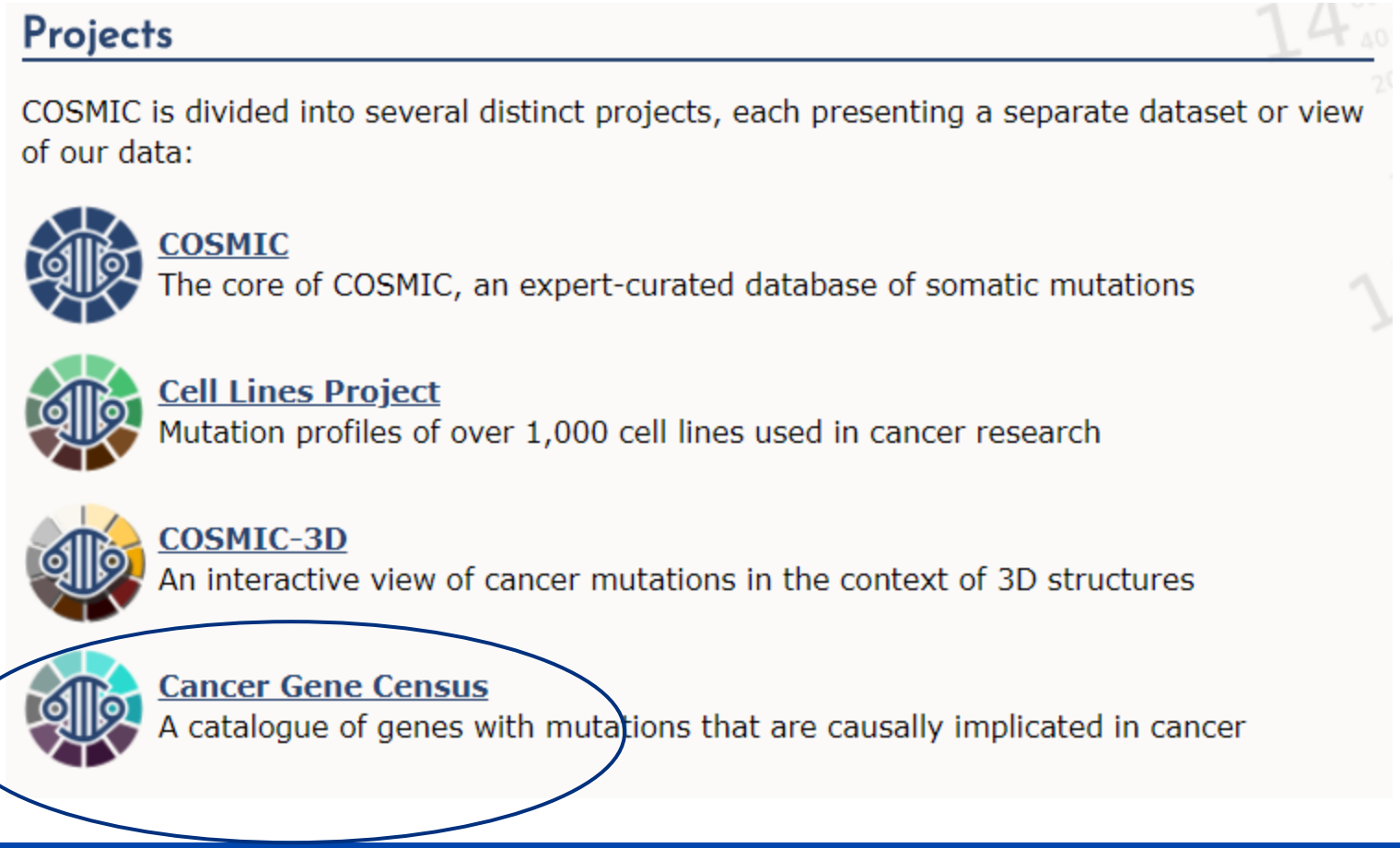


COSMIC-3D

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



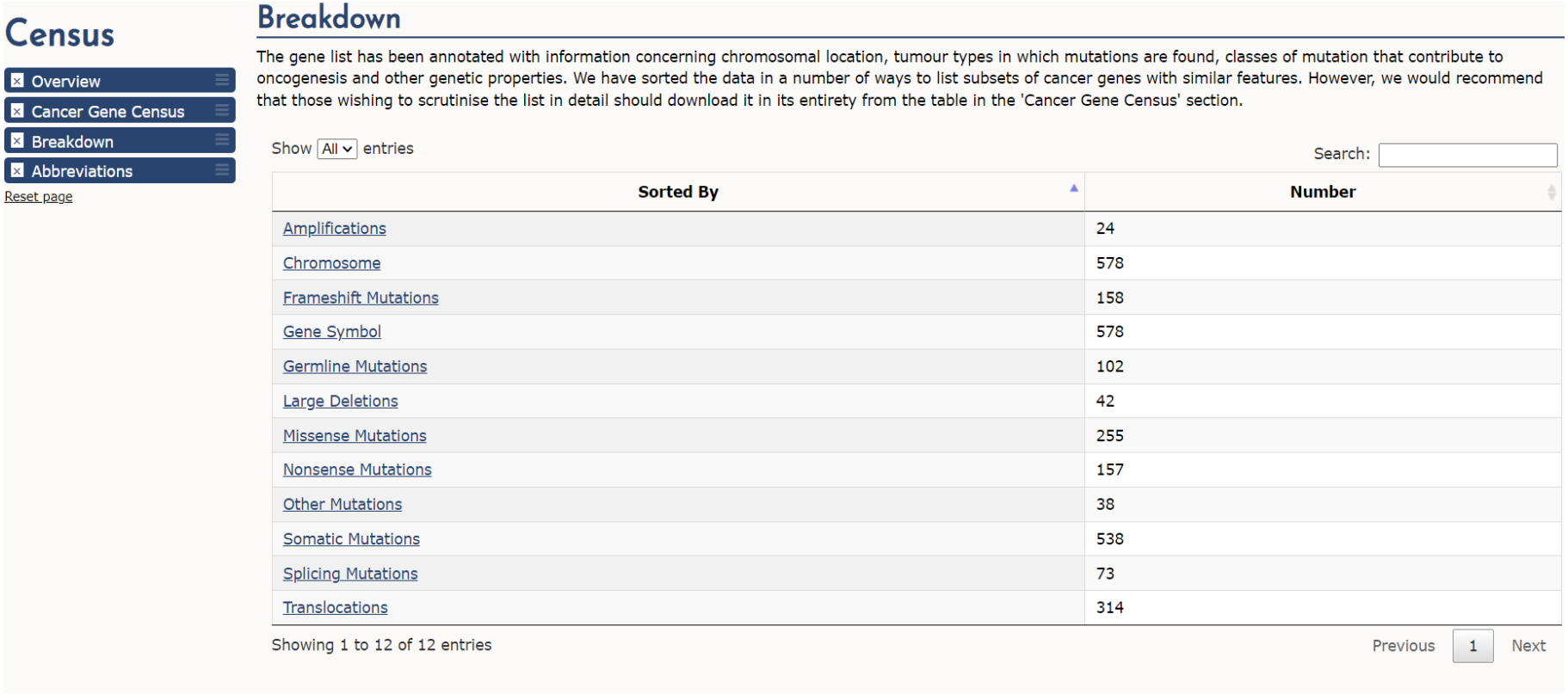
COSMIC



Gene Tiers in Cancer Gene Census

- **Census tiers – 578 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:

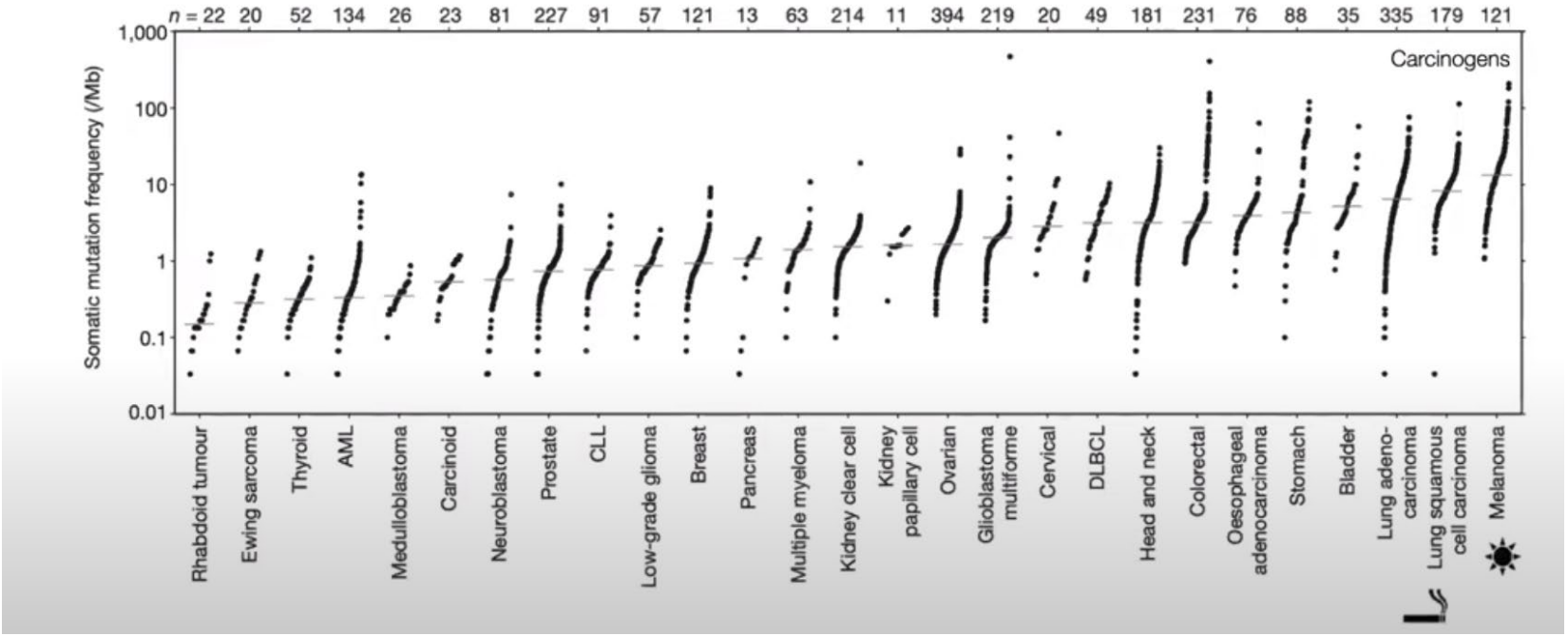
Sorted By	Number
Amplifications	24
Chromosome	578
Frameshift Mutations	158
Gene Symbol	578
Germline Mutations	102
Large Deletions	42
Missense Mutations	255
Nonsense Mutations	157
Other Mutations	38
Somatic Mutations	538
Splicing Mutations	73
Translocations	314

Showing 1 to 12 of 12 entries Previous 1 Next

MAYO CLINIC

cBioPortal

TCGA - 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

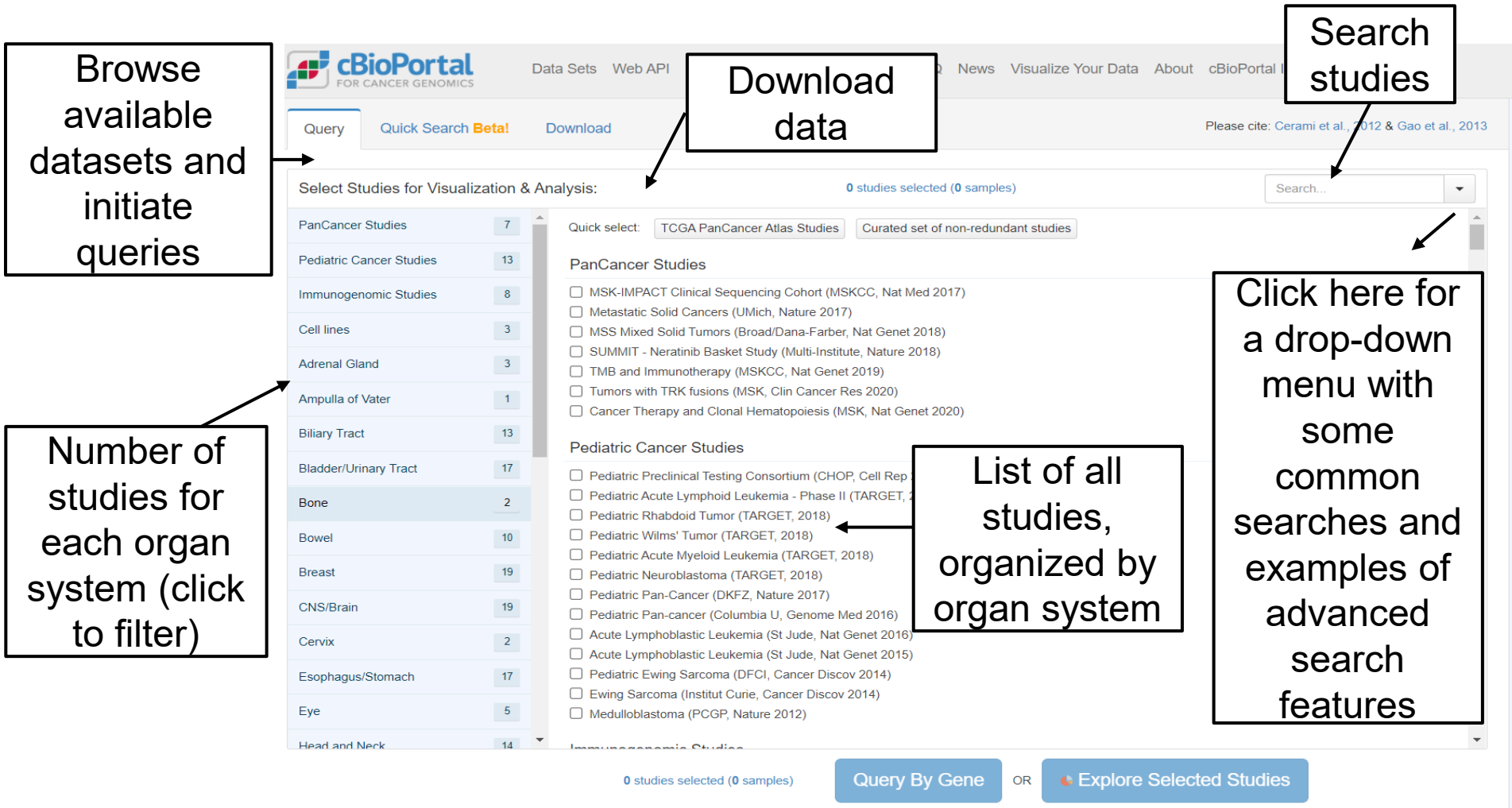
Overview of Tabs in a Single Study Query

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)

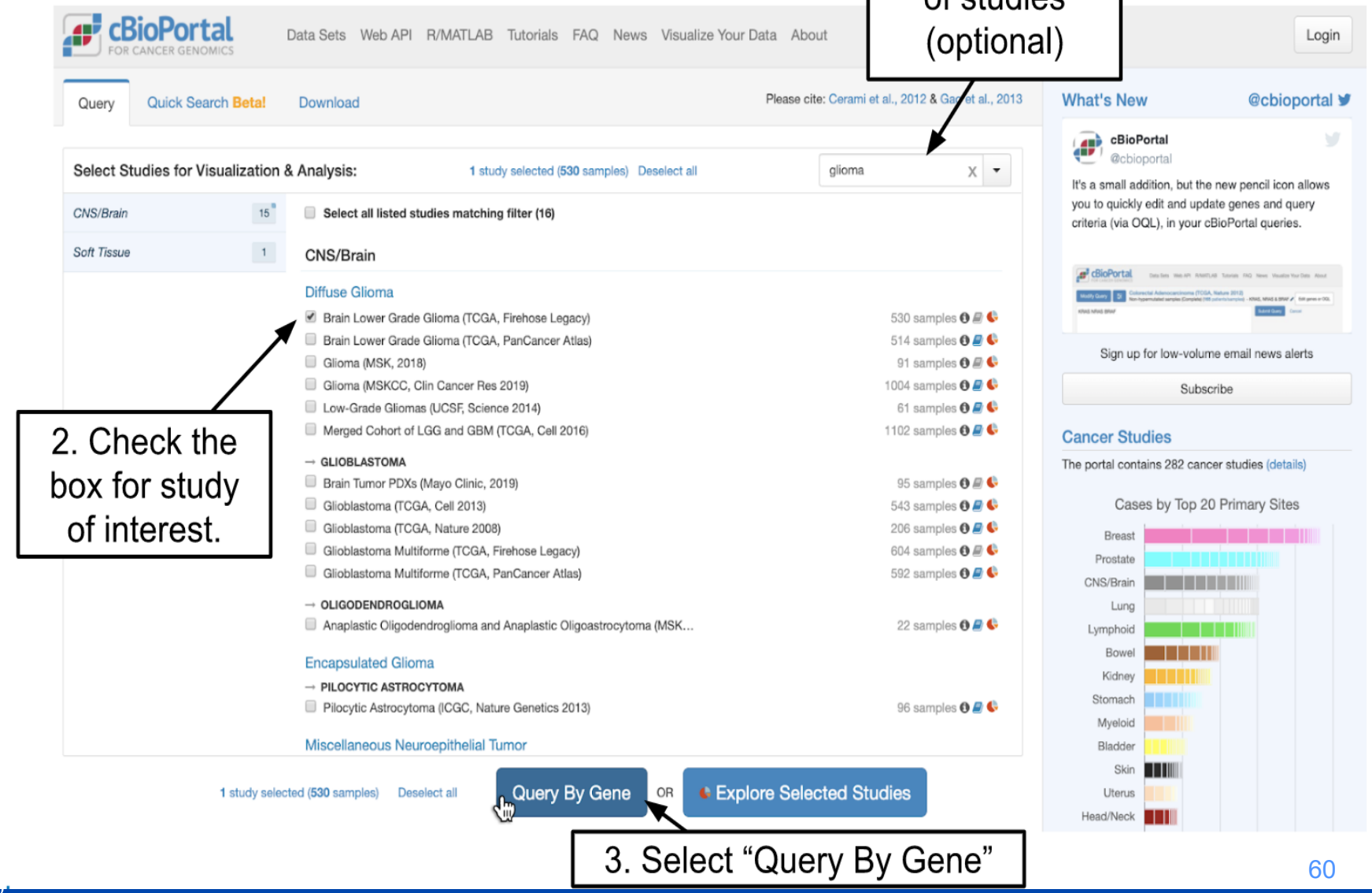
- **OncoPrint:** Overview of genetic alterations per sample in each query gene
- **Cancer Types Summary:** Frequency of alteration in each query gene in the detailed cancer types included in this study
- **Mutual Exclusivity:** Statistical analysis to determine if query genes are mutually exclusively 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

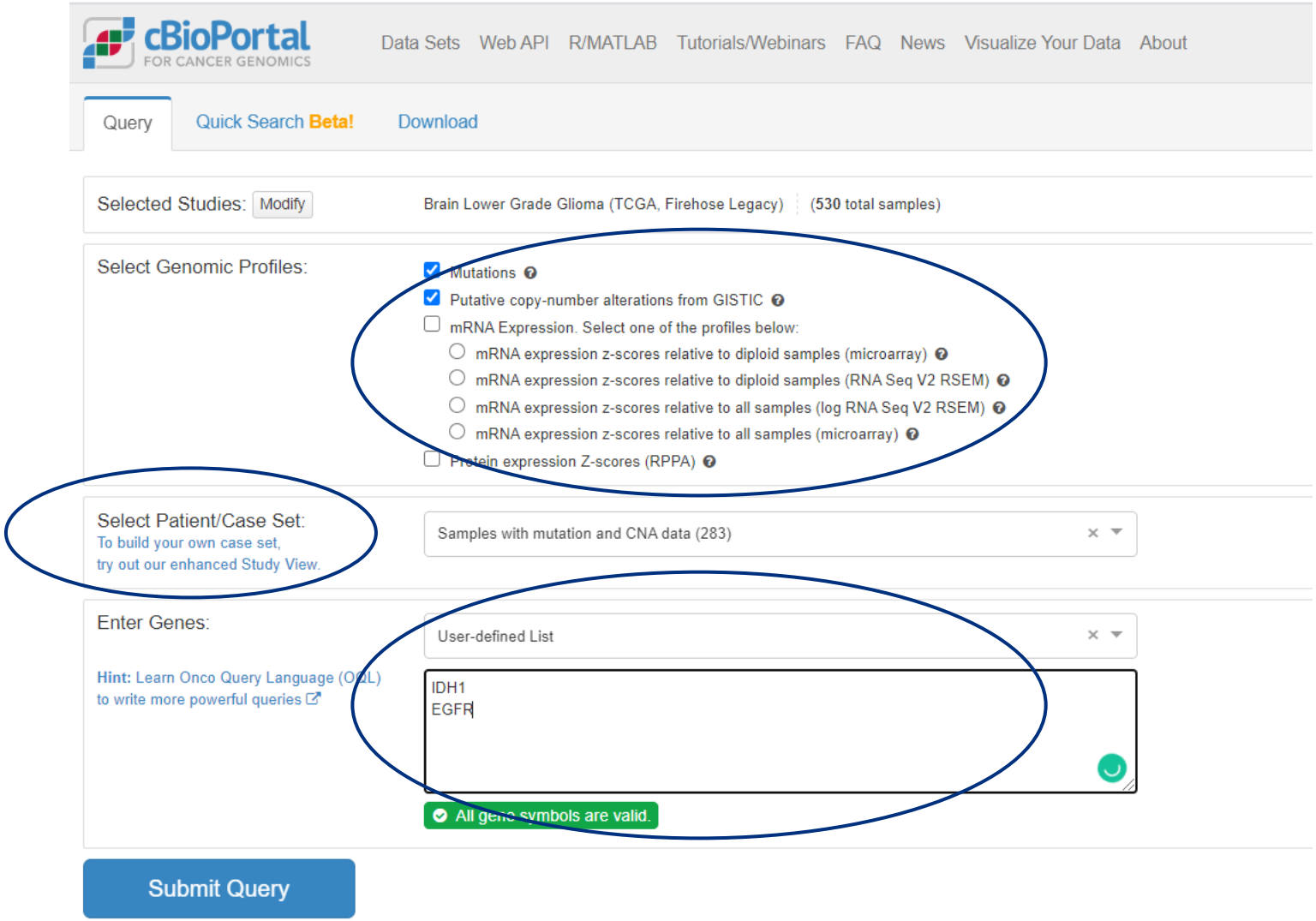
Glioma example

Query overview

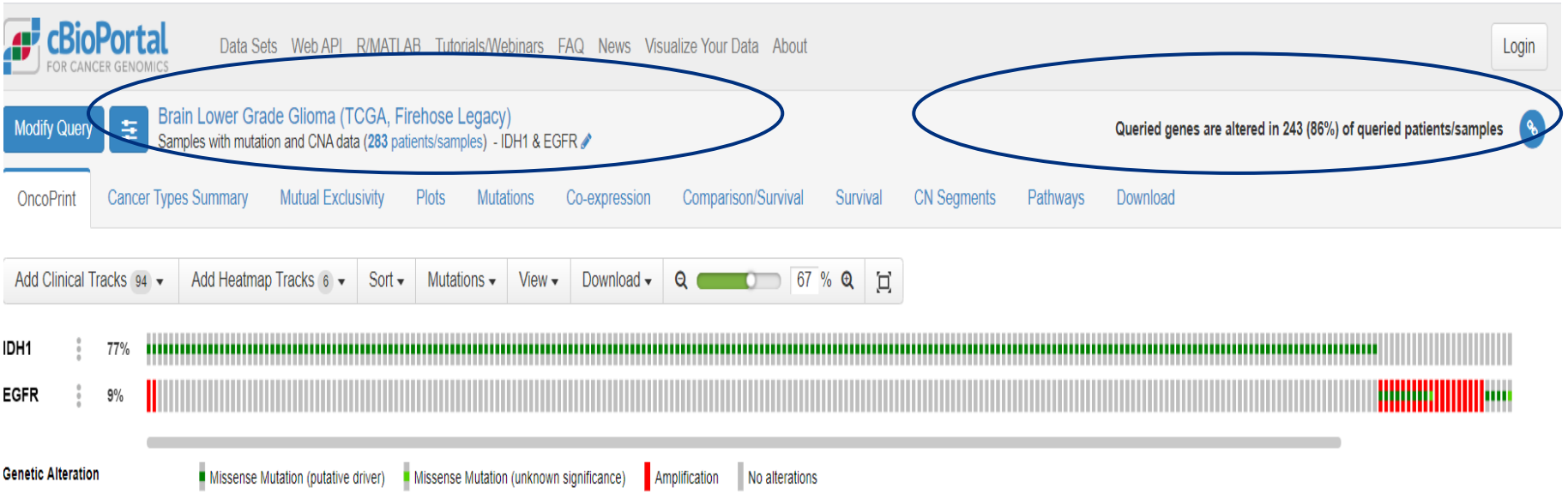


Single study query

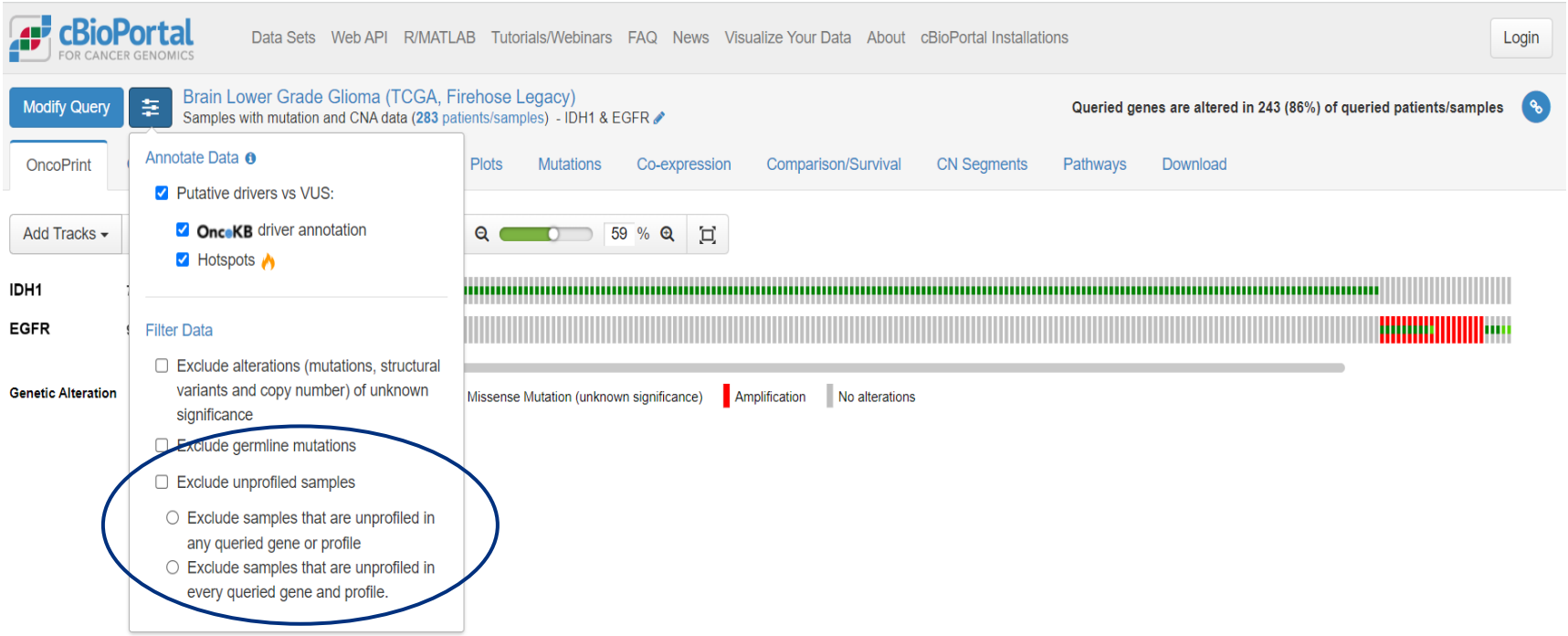




Glioma Query

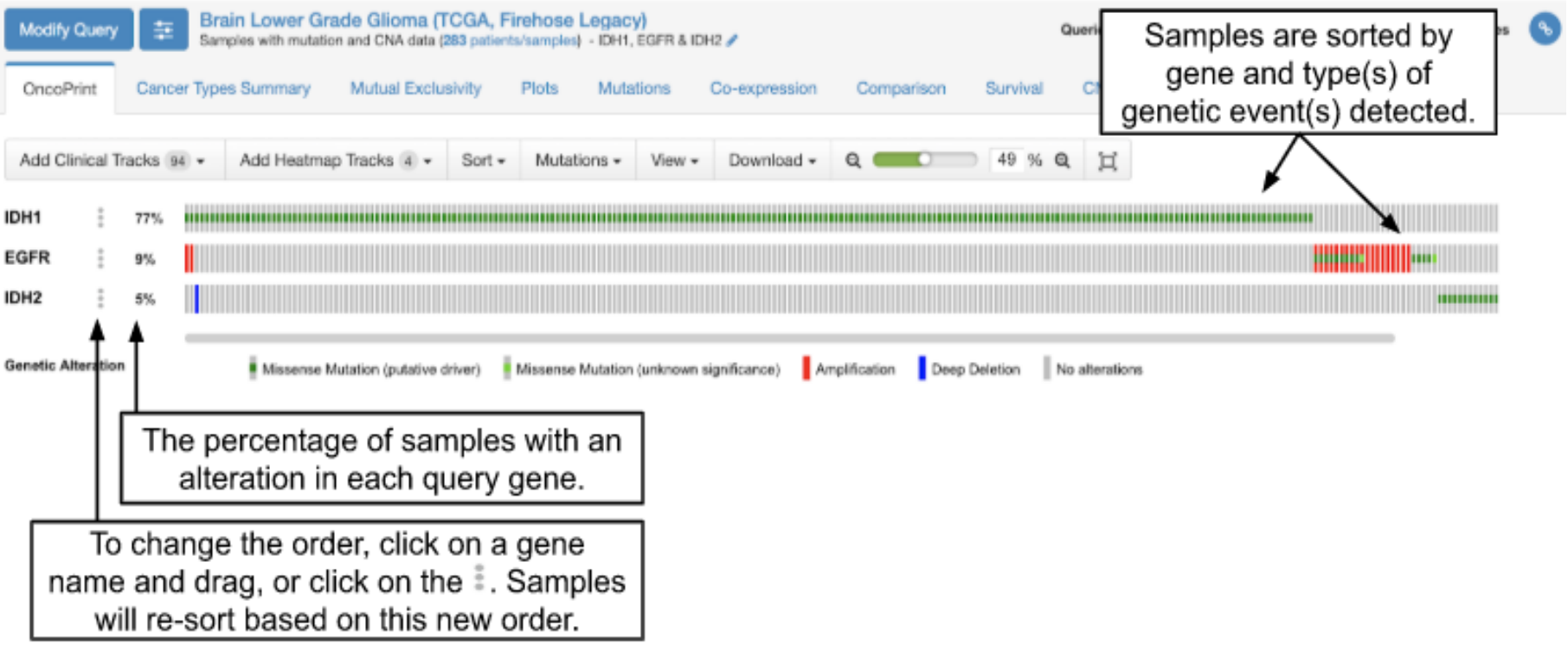


Annotations and Filtering

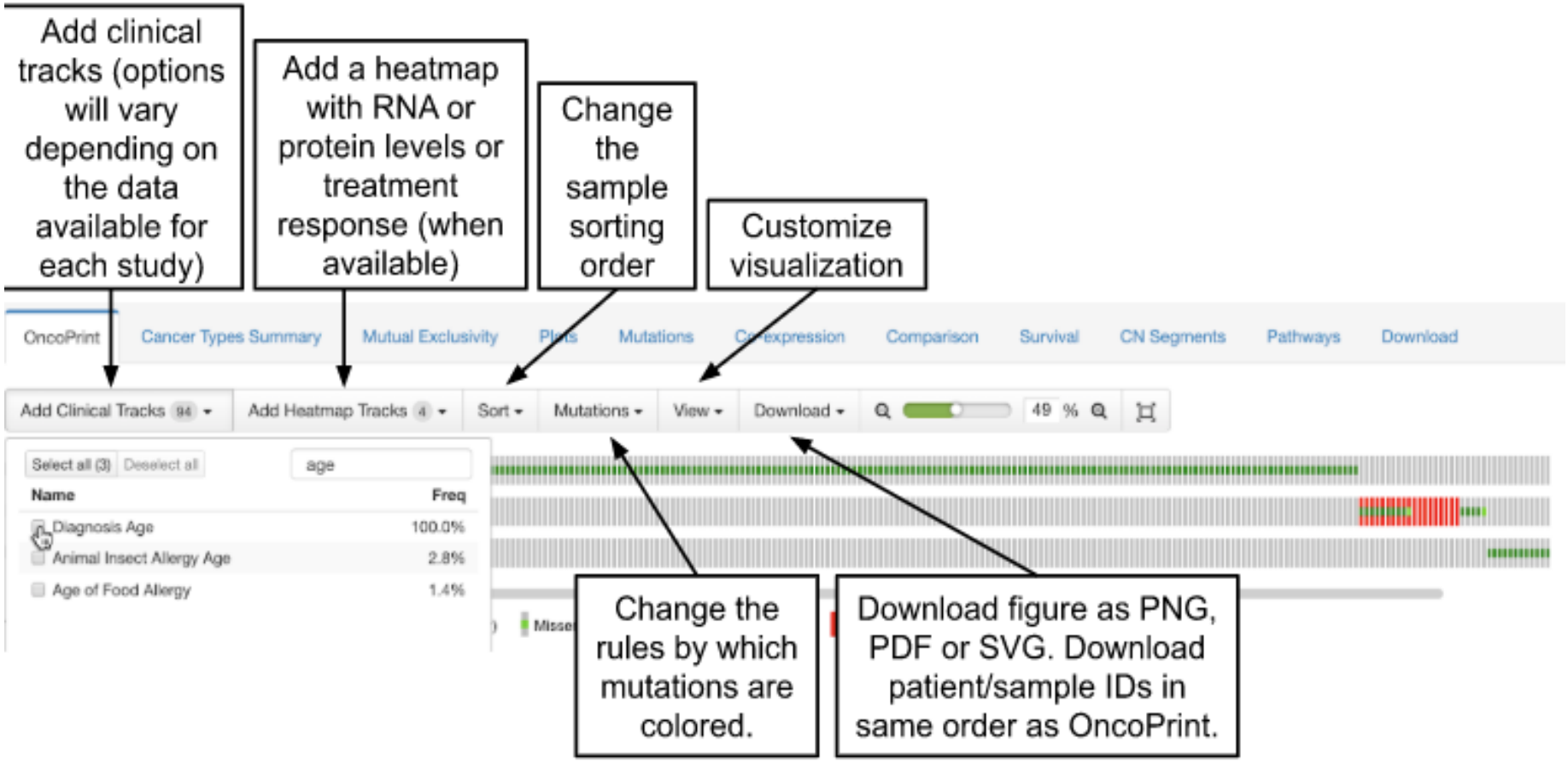


OncoPrint

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.

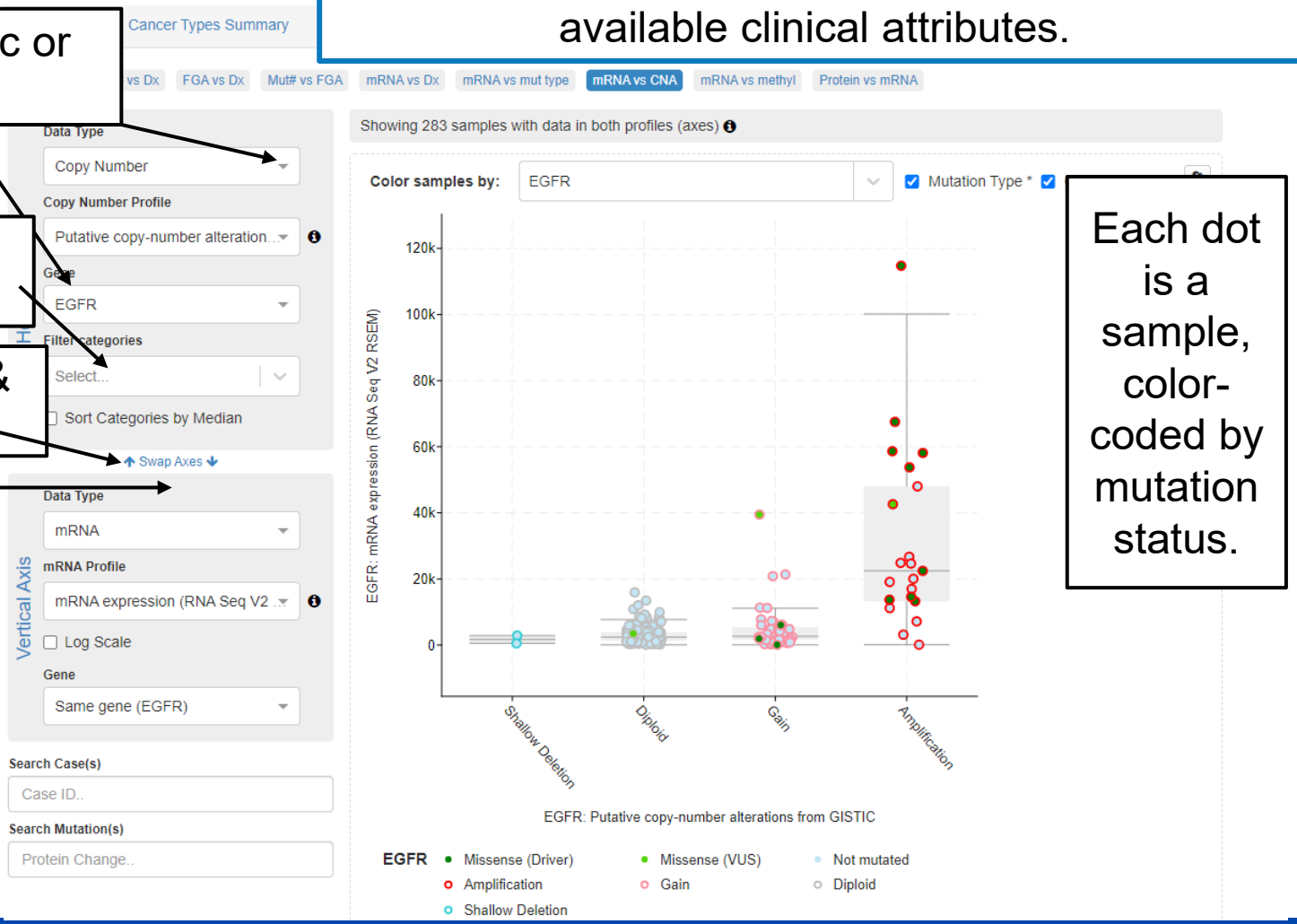
Choose genetic or clinical

Select a query gene

Select data type and processing

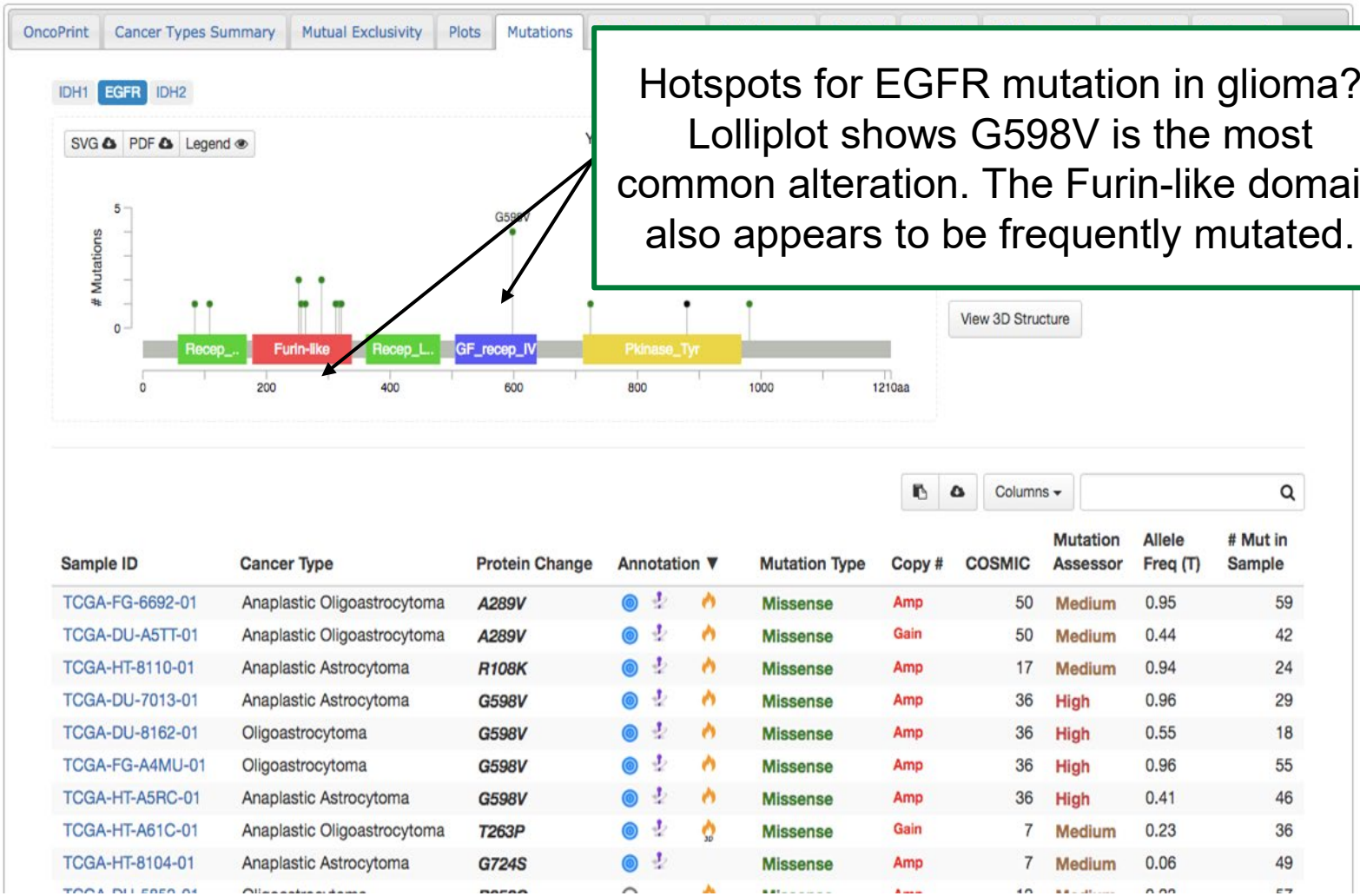
Swap horizontal & vertical axis

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



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

Mutations

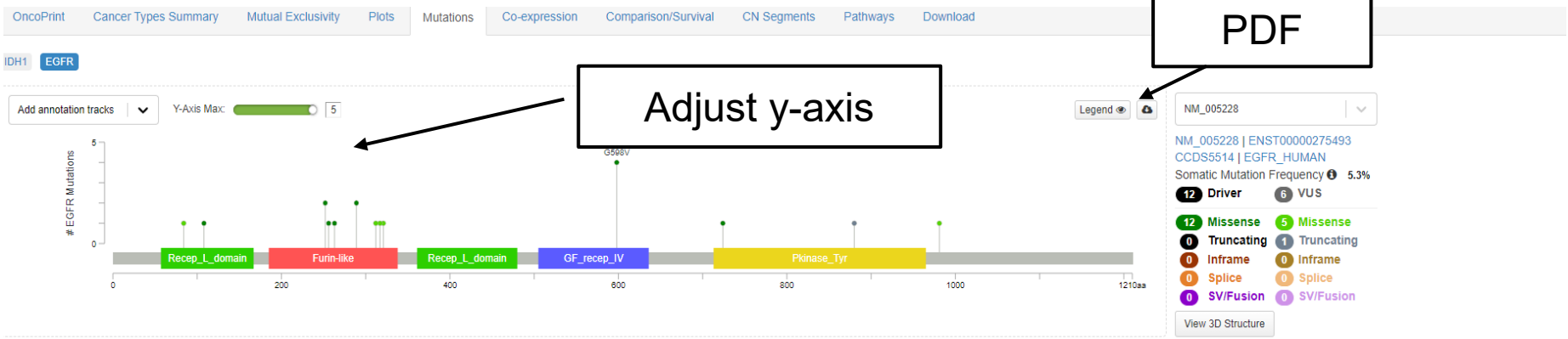


Hotspots for EGFR mutation in glioma? Lolliplot shows G598V is the most common alteration. The Furin-like domain also appears to be frequently mutated.

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

8 Mutations (page 1 of 1)

Columns

Sample ID	Cancer Type	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-FG-6692-01	Anaplastic Oligoastrocytoma	*A289V*		Missense	Amp	50	0.95	59
TCGA-DU-A5TT-01	Anaplastic Oligoastrocytoma	*A289V*		Missense	Gain	50	0.44	42
					Amp	17	0.94	24
					Gain	7	0.23	36
					Amp	36	0.96	29
					Amp	36	0.55	18
					Amp	36	0.96	55
					Amp	36	0.41	46
					Amp	7	0.06	49
					Amp	12	0.79	34
					Gain	5	0.38	39
					Amp	12	0.23	57
					Diploid	1	0.26	33
					Amp	1	0.02	49
					Gain	1	0.06	40

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.

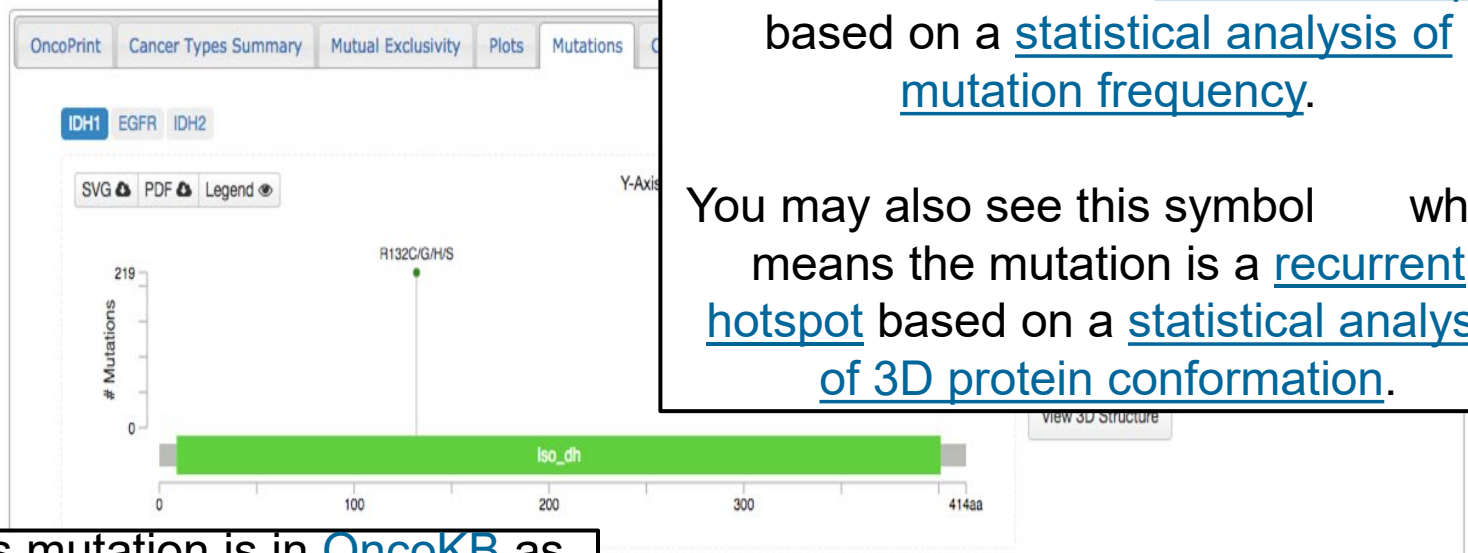
http://www.cbioportal.org/index.do

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.

Columns

			Annotation	Mutation Type	Copy #	COSMIC	Mutation Assessor	Allele Freq (T)	# Mut in Sample
TCGA-HT-7479-01	Anaplastic Astrocytoma	R132C		Missense	Diploid	4964	High	0.41	26
TCGA-FG-8185-01	Anaplastic Astrocytoma	R132C		Missense	Diploid	4964	High	0.24	17
TCGA-HT-7693-01	Oligodendroglioma	R132C		Missense	Diploid	4964	High	0.39	29
					ShallowDel	4964	High	0.48	24
									30
									28
									18
									25
									25
				Missense	Diploid	4964	High	0.31	10
TCGA-DB-5276-01	Oligoastrocytoma	R132C		Missense	Diploid	4964	High	0.20	13

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

This mutation is in My Cancer Genome.

Co-Expression

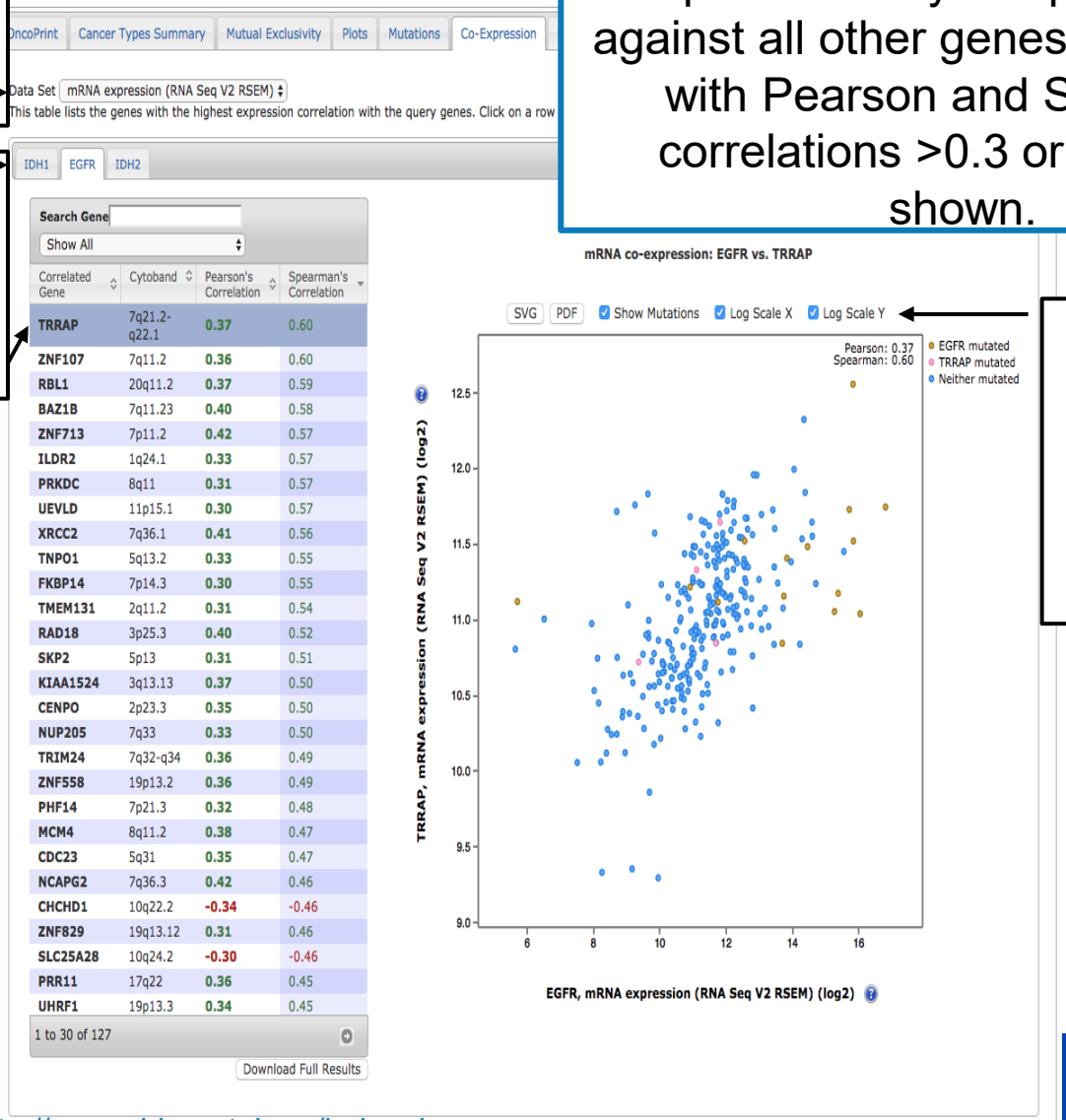
Select from available data types

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.

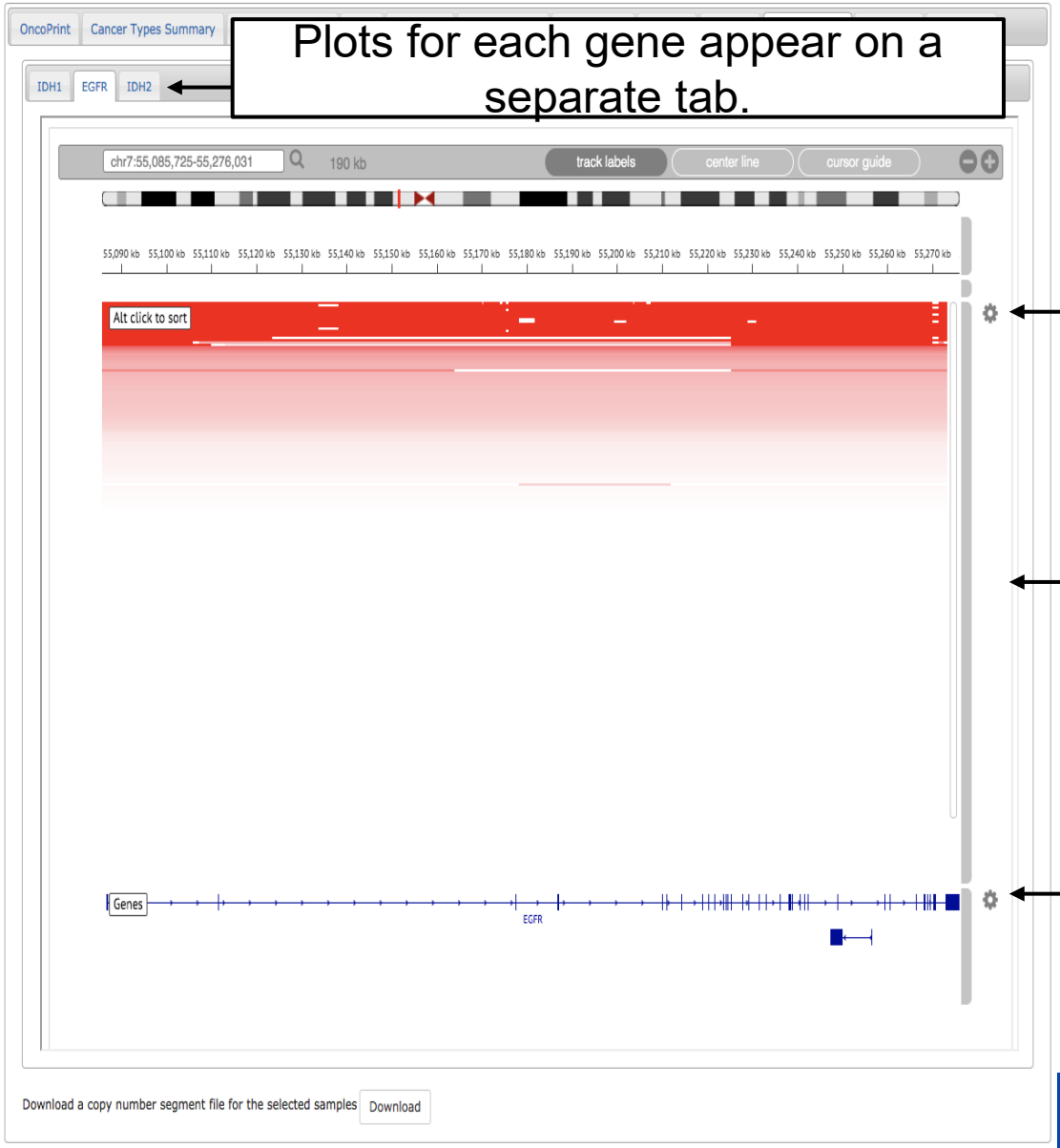
Check boxes to color-code sample dots by mutation status or change x- or y-axis to log scale



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

CN Segments

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

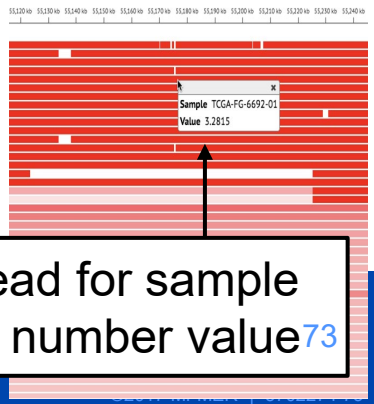


Plots for each gene appear on a separate tab.

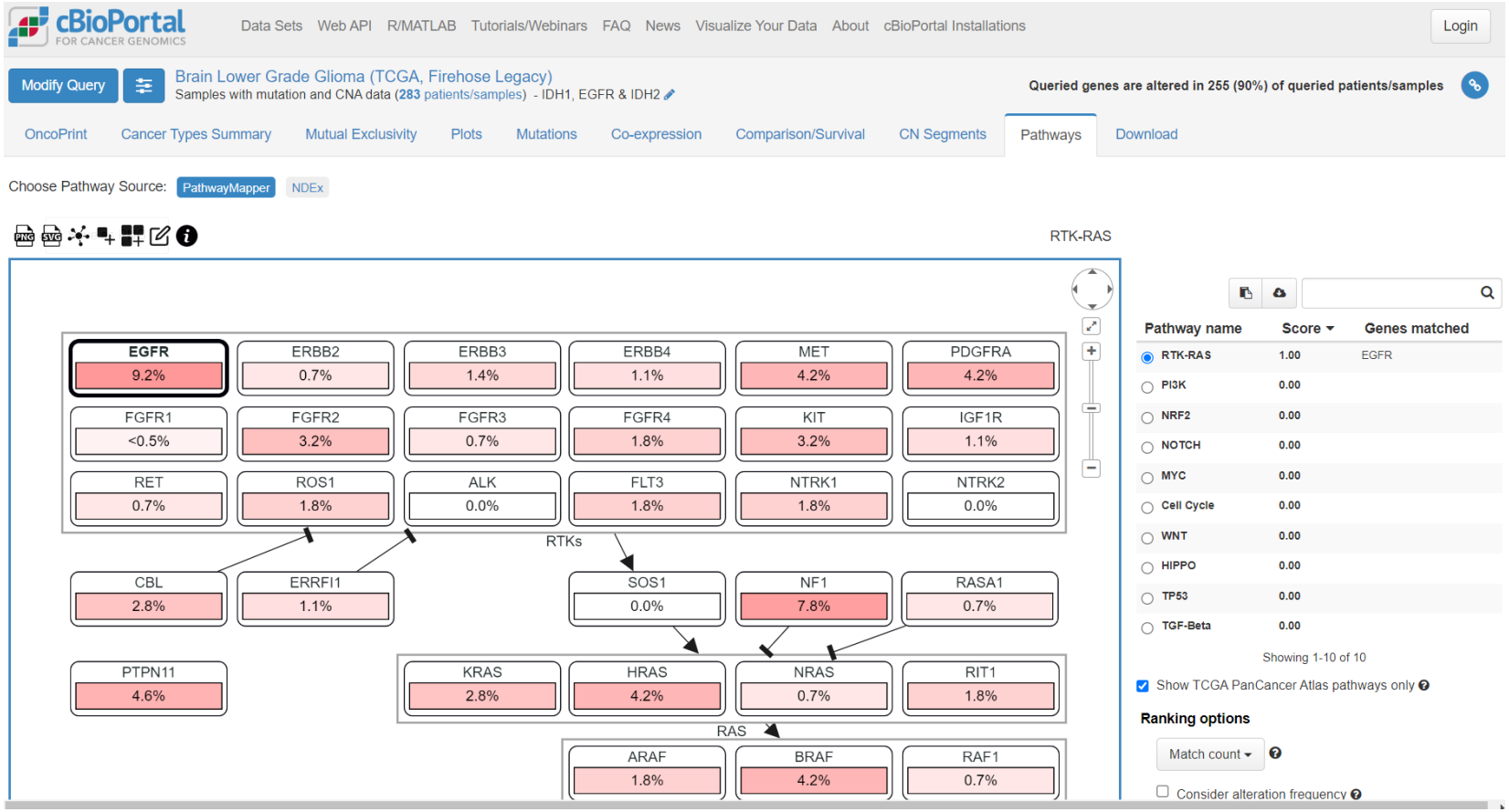
Click ⚙ for track settings, including expanding the height of each sample (see below)

Each row is a single sample

Gene structures

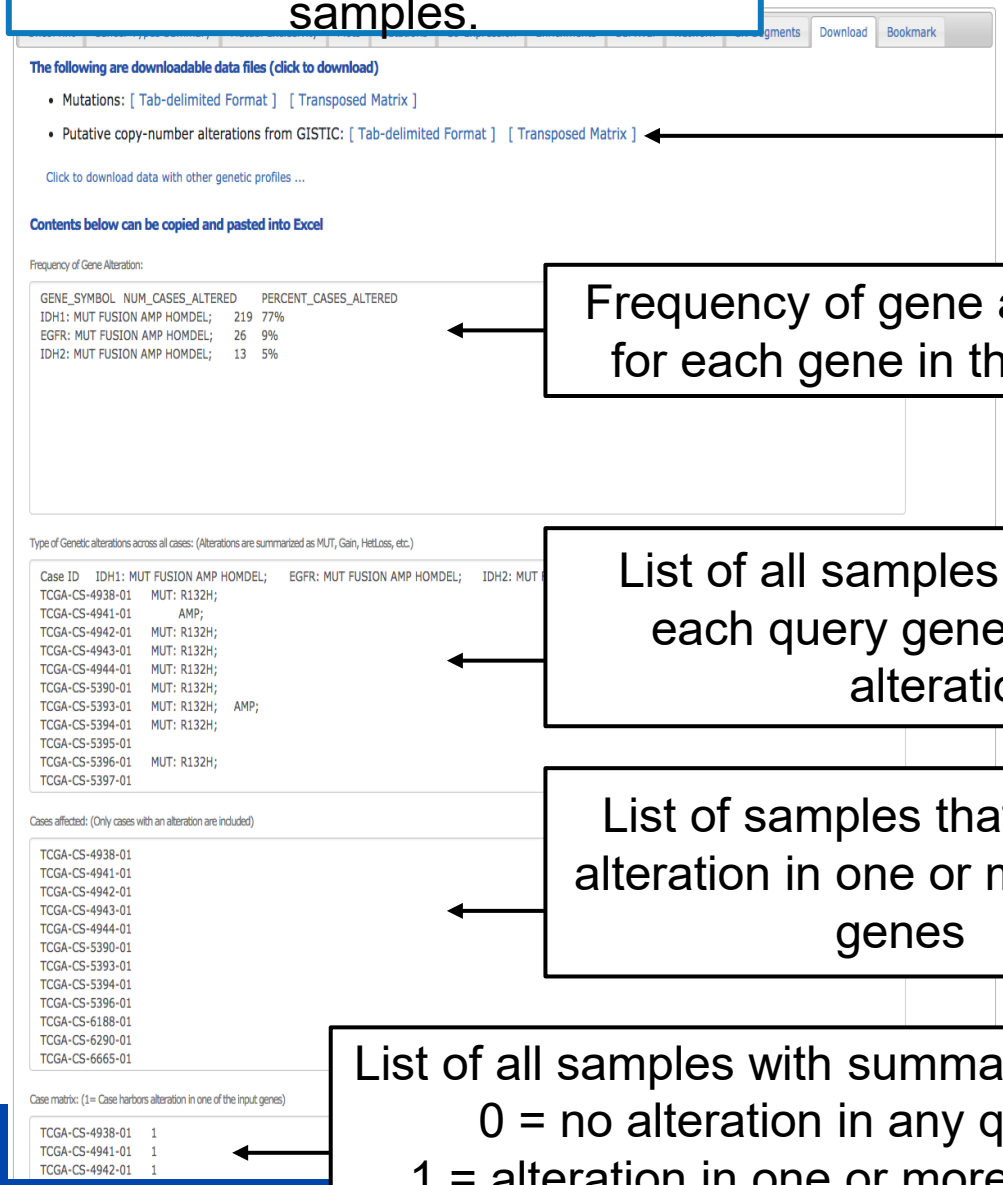


Click on a read for sample ID and copy number value



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

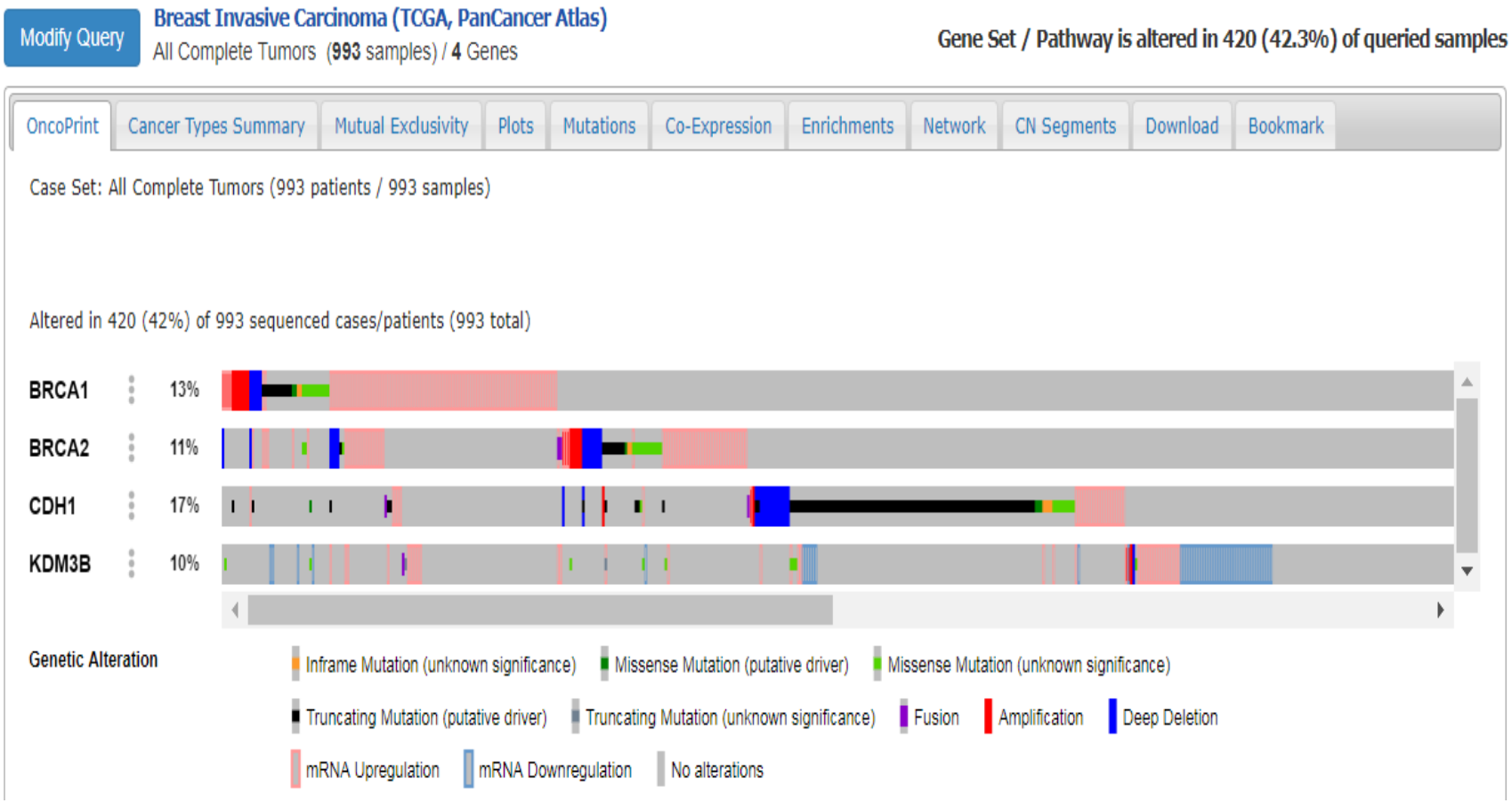
List of all samples with summary classification:
0 = no alteration in any query gene
1 = alteration in one or more query genes

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.

Link to this page

Breast cancer example

CNV



mRNA overexpressed

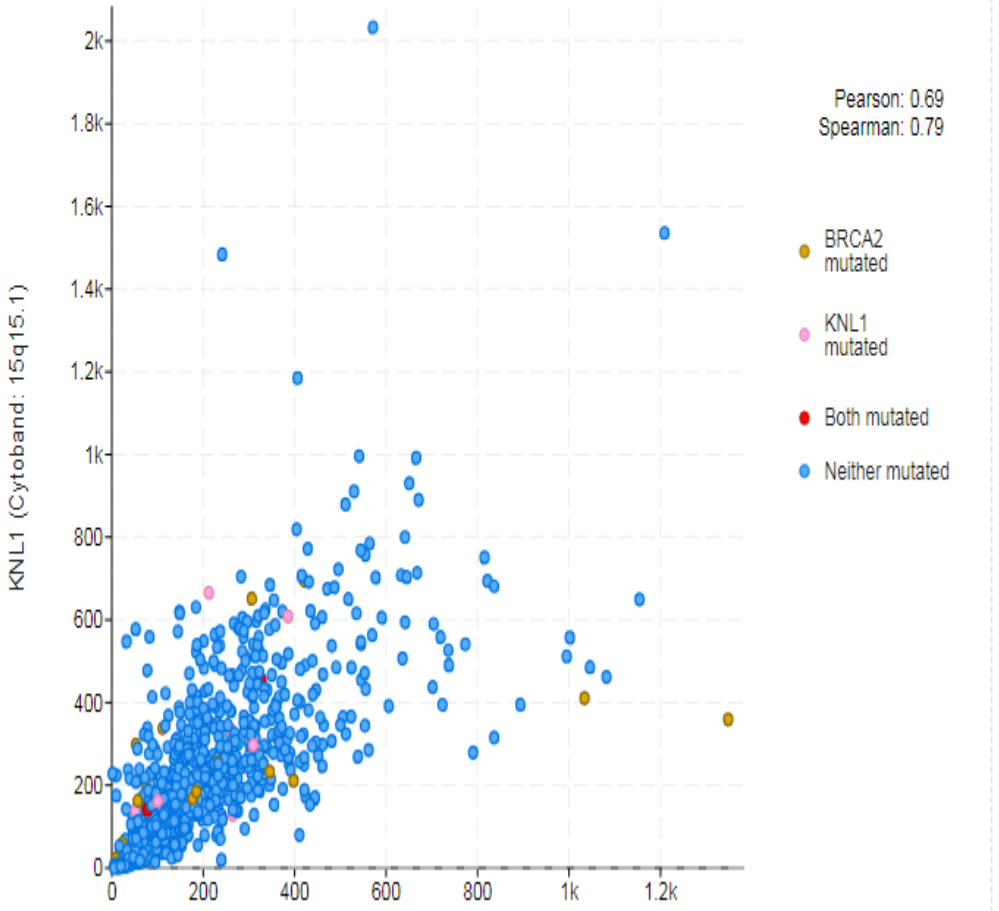
BRCA1 BRCA2

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

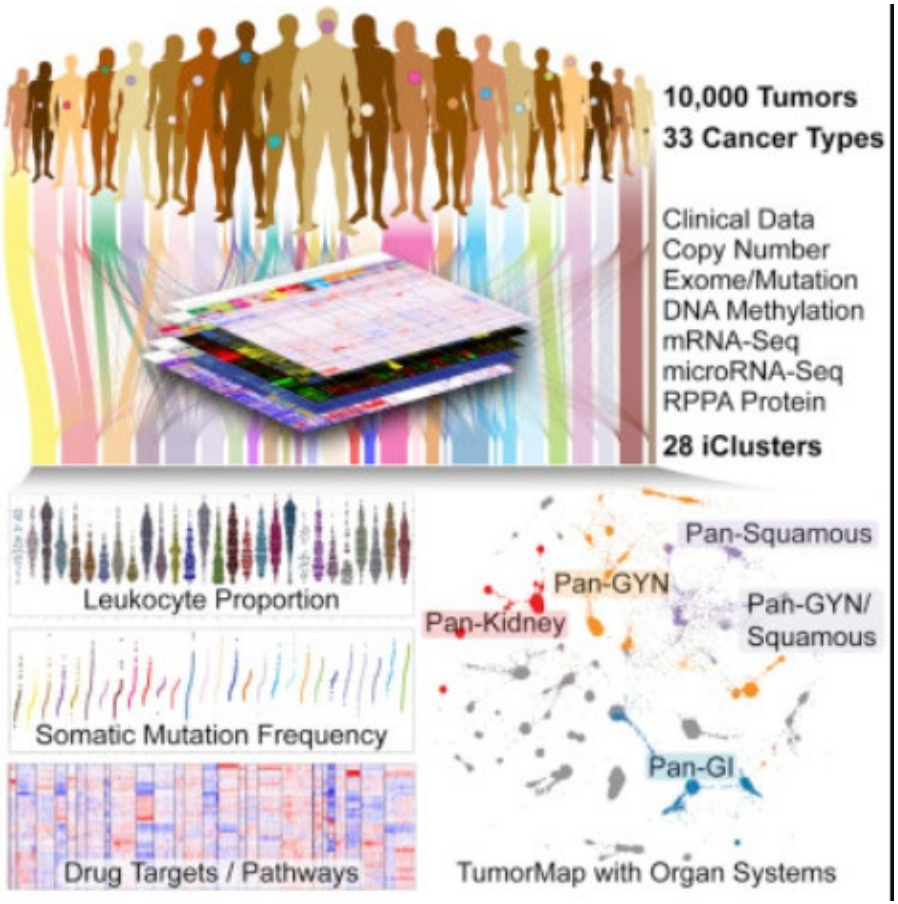
mRNA Expression Batch Normalized/Merged from Illumina HiSeq_RNASeqV2 syn4976369: BRCA2 vs



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

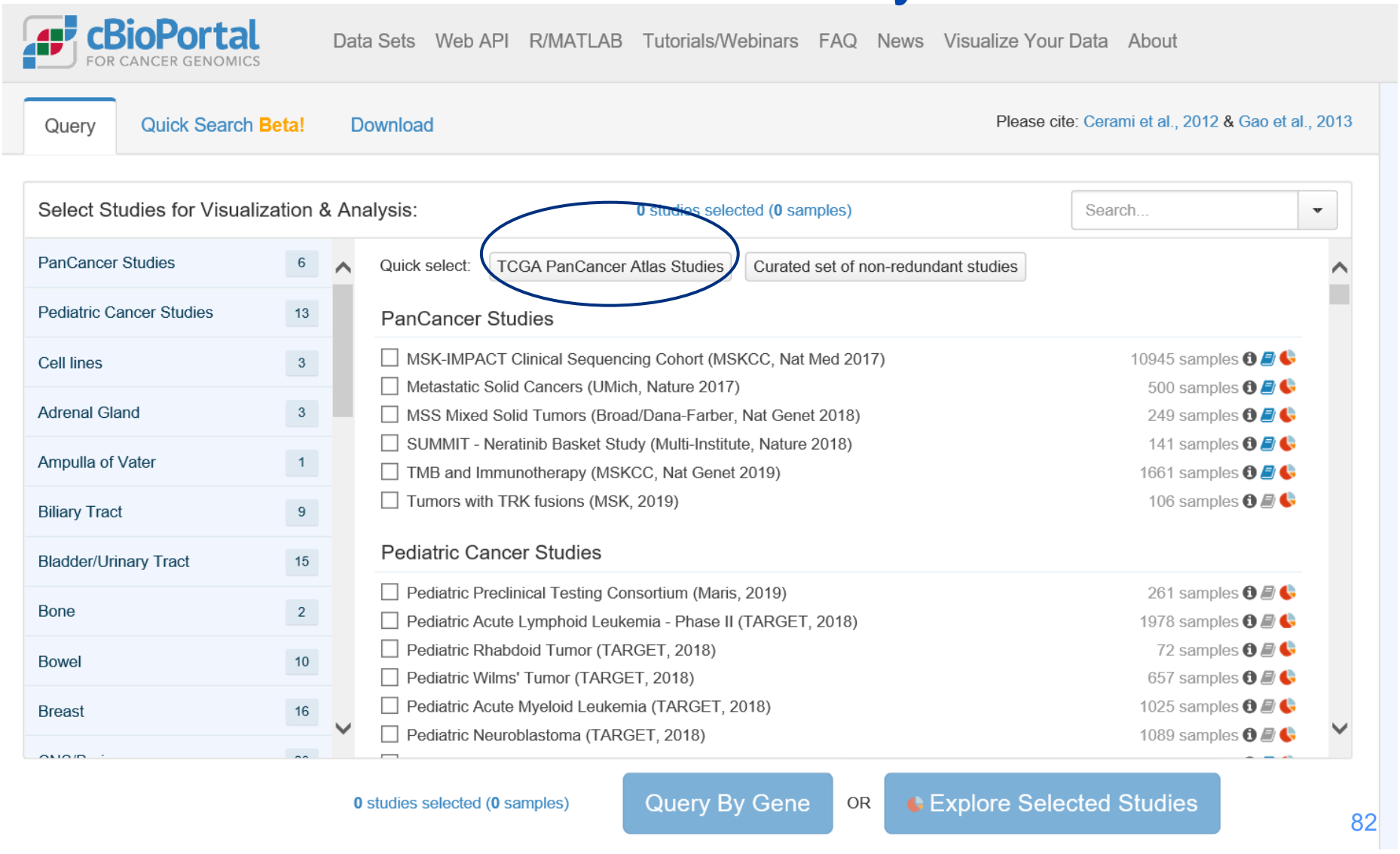
PAN-can example

PAN-Can datasets and analyses

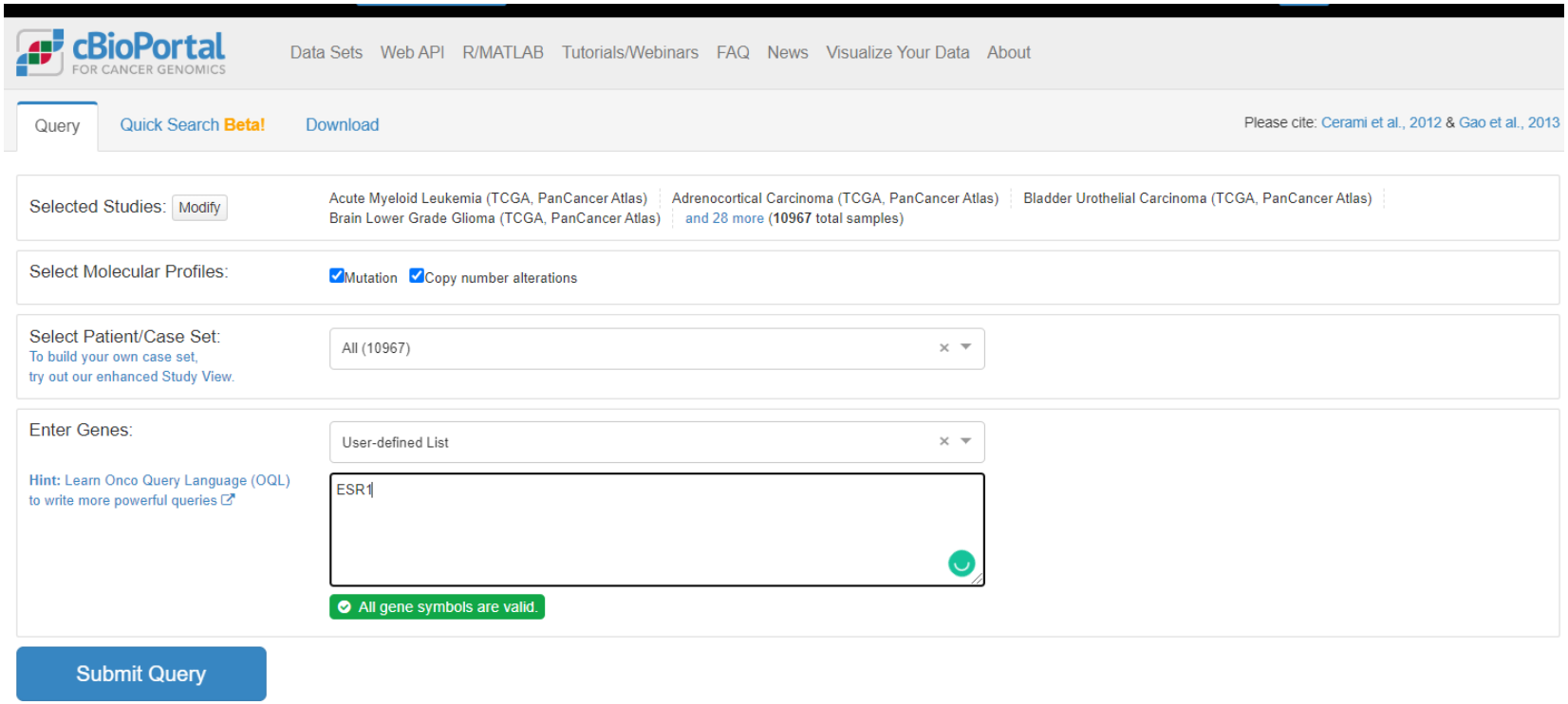


https://www.sciencedirect.com/science/article/pii/S0092867418303027

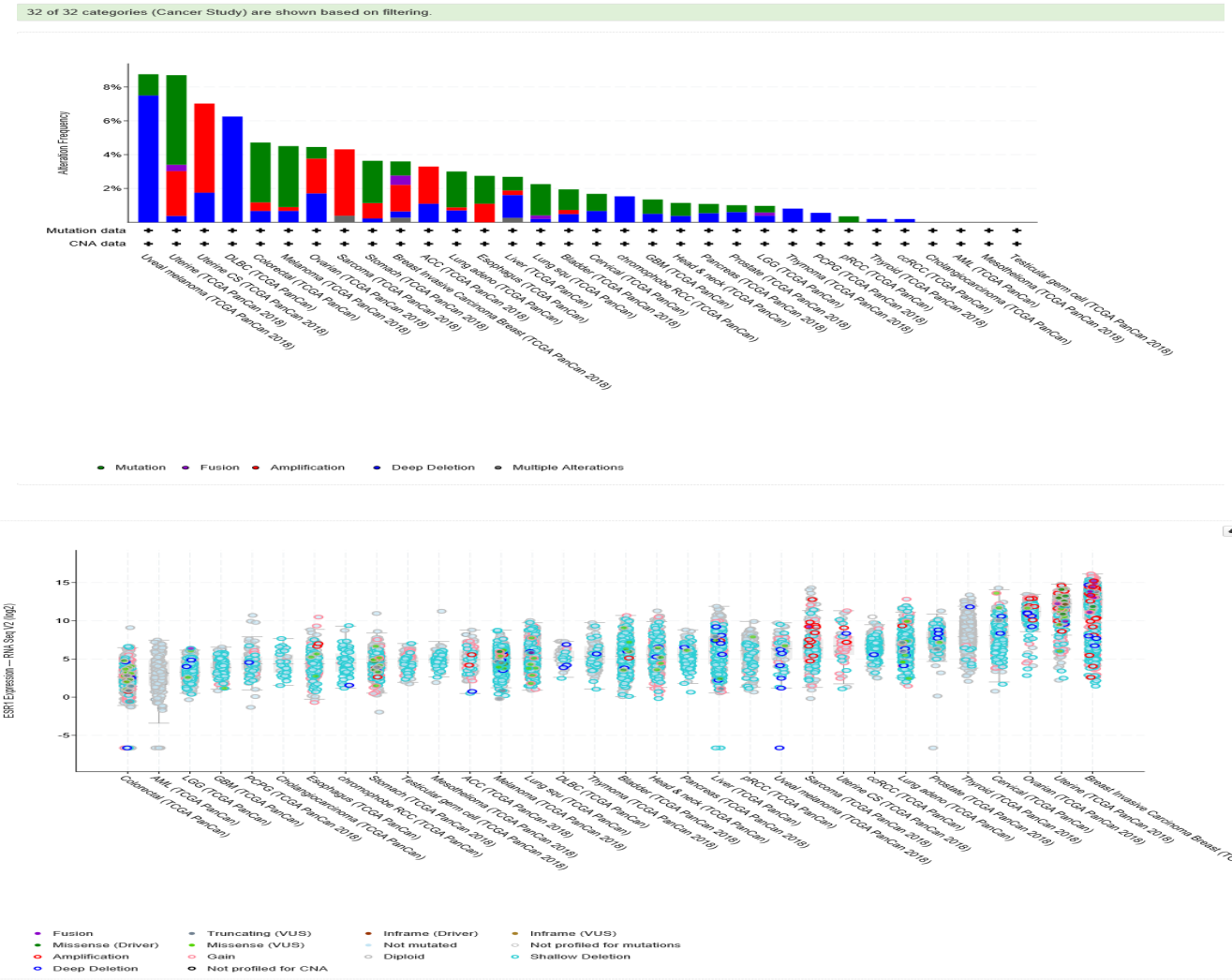
PAN-Can datasets and analyses



PAN-Can datasets and analyses – ESR1

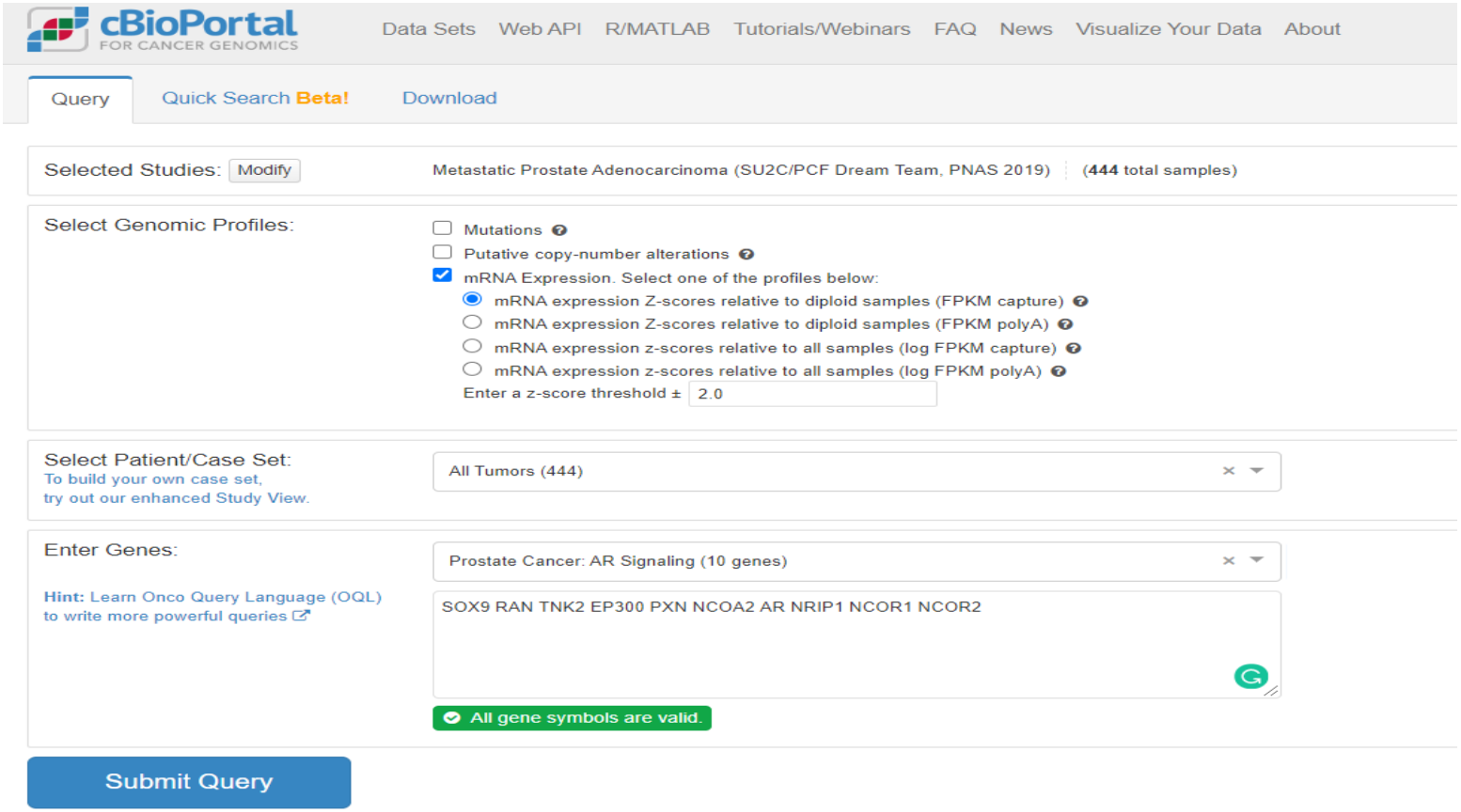


PAN-Can datasets and alterations – ESR1

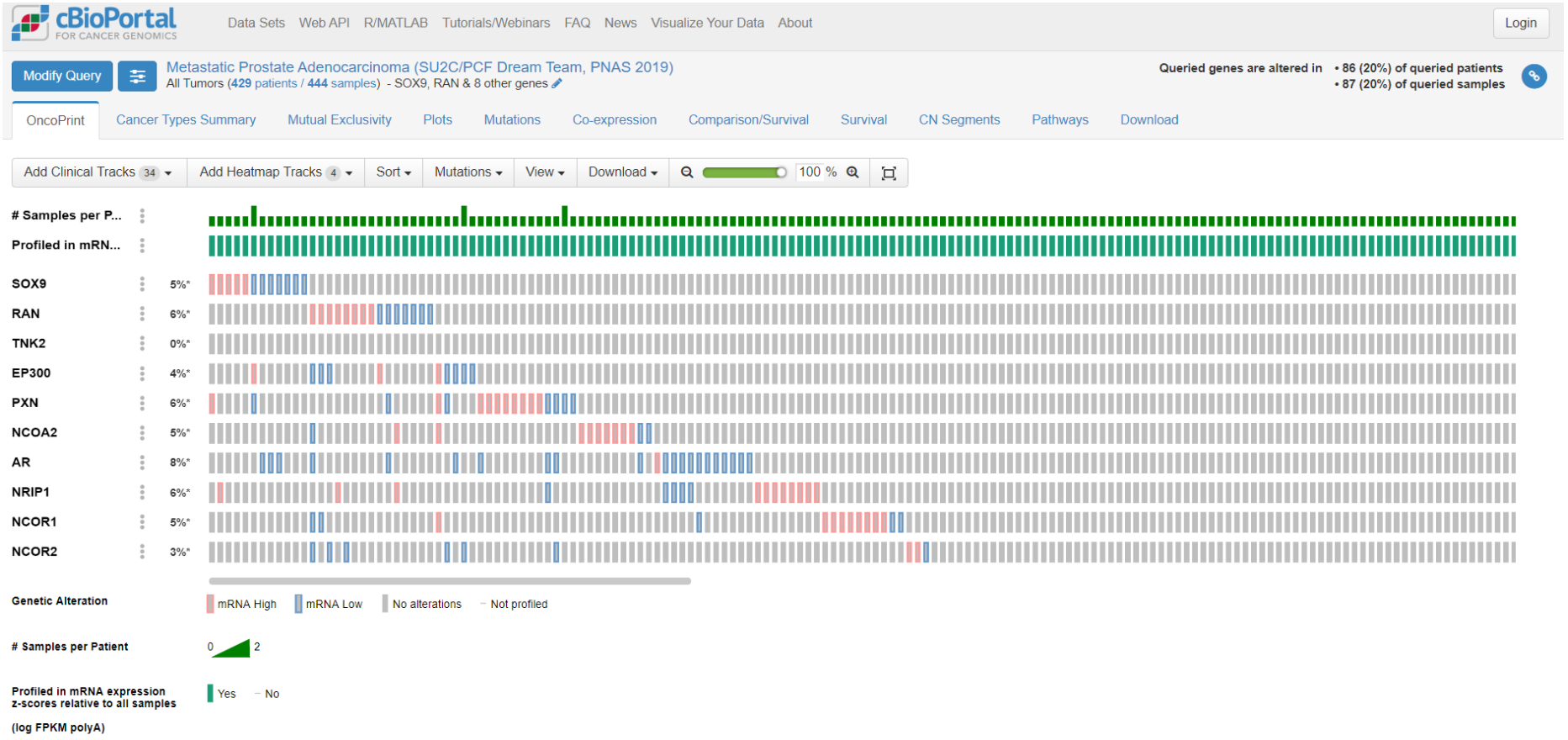


* Driver annotation settings are located in the Mutation Color menu of the Oncoprint.

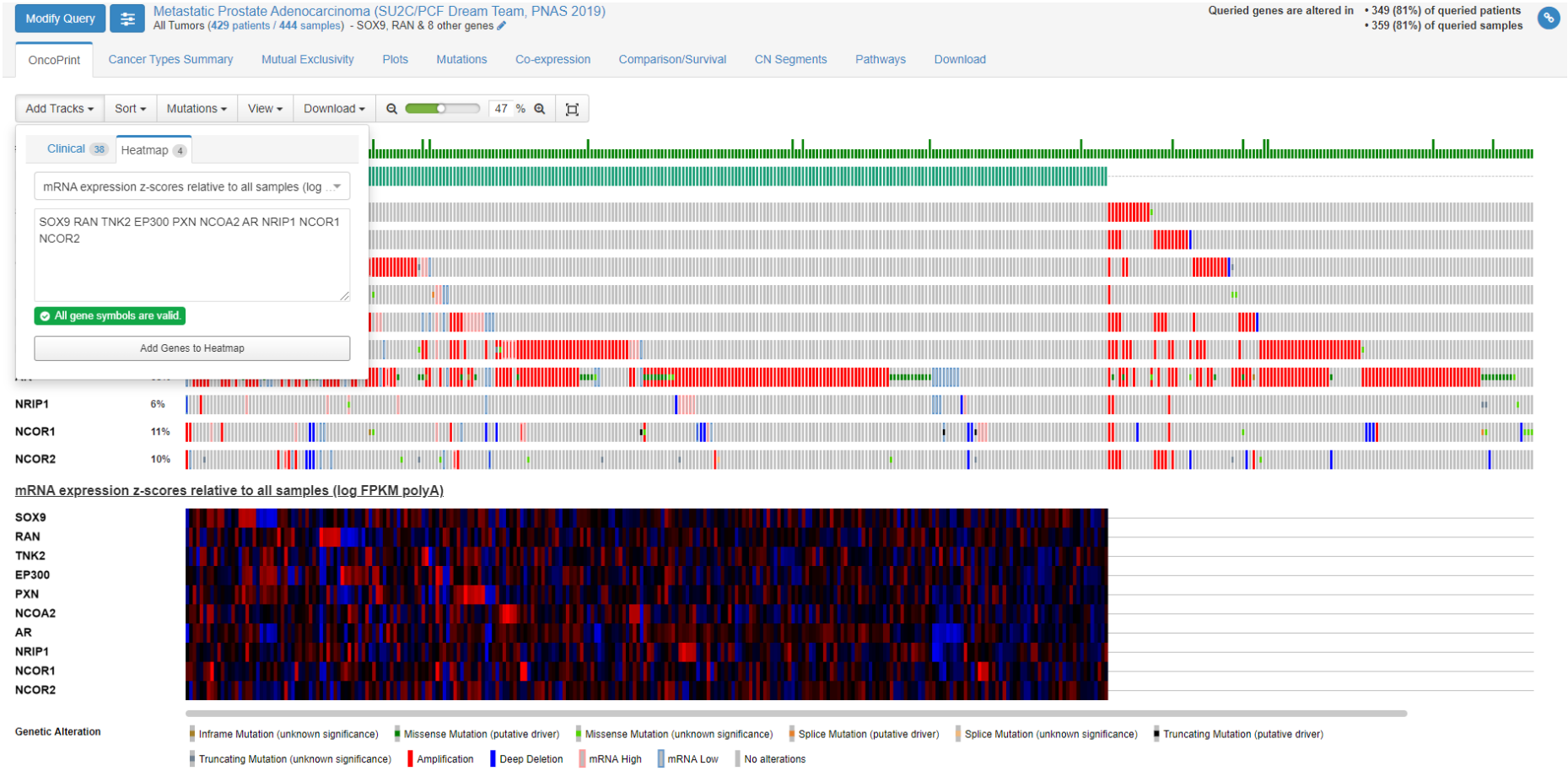
Clustering example



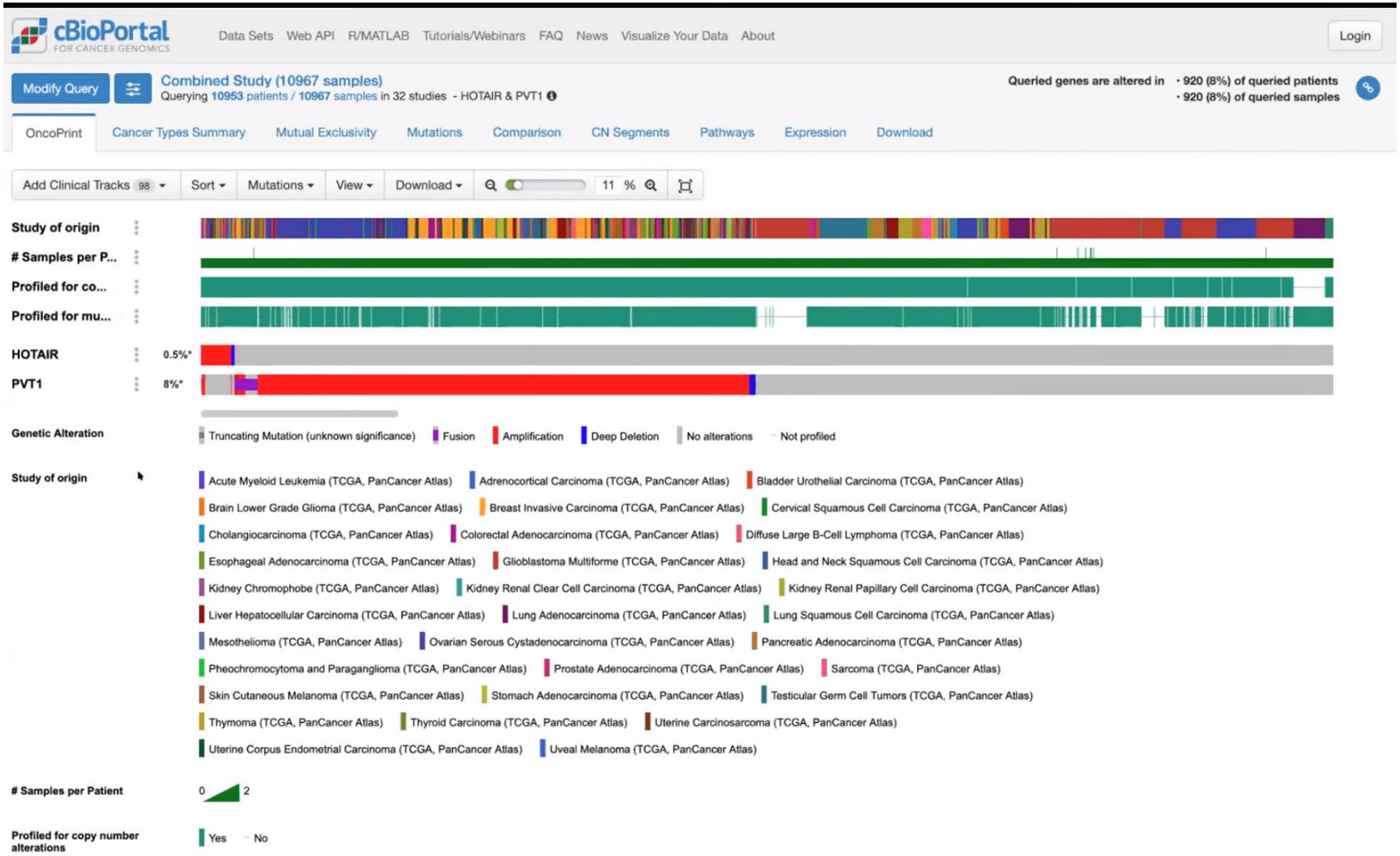
Clustering example



Clustering example

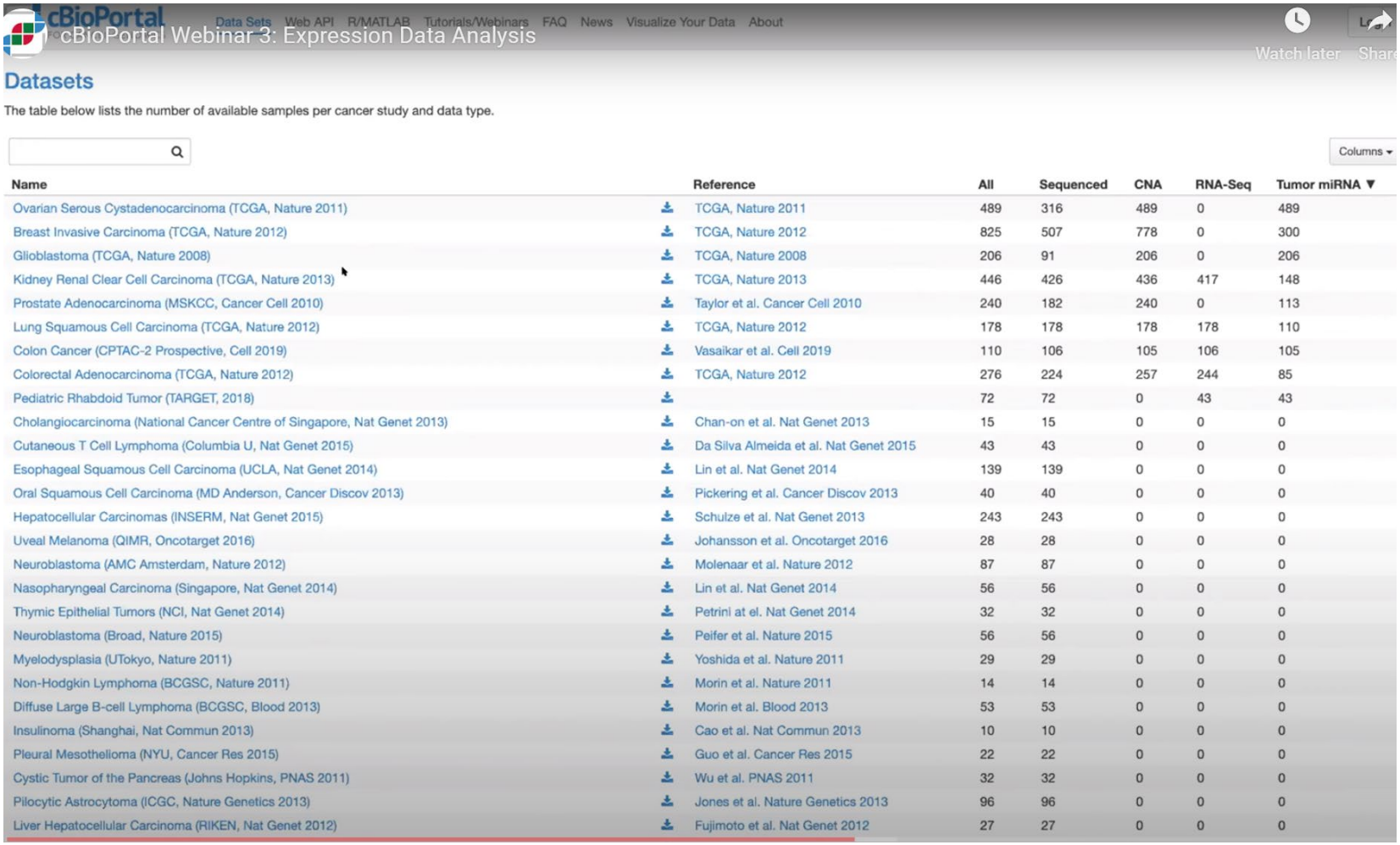


Long-noncoding RNA example



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

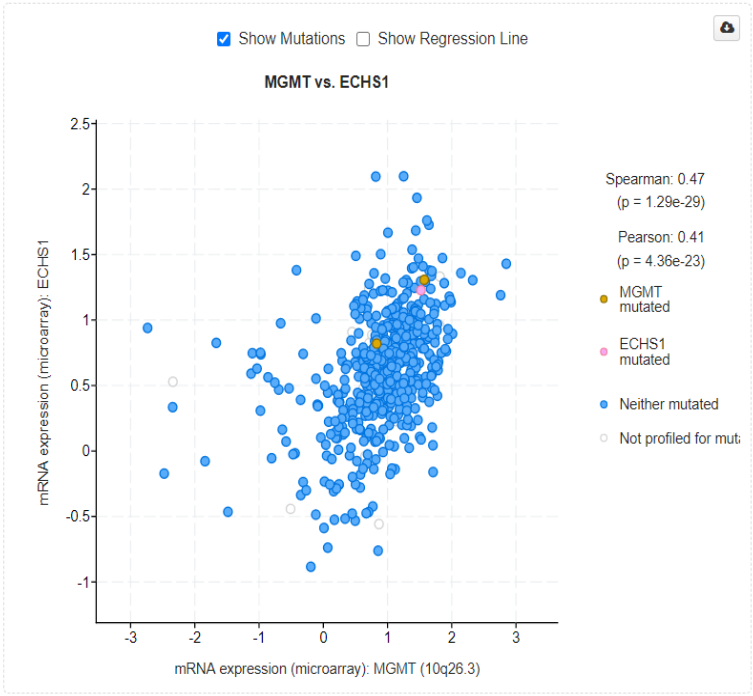
OncoPrint | Cancer Types Summary | Plots | Mutations | Co-expression | Comparison/Survival | Survival | CN Segments | Pathways | Download

MGMT

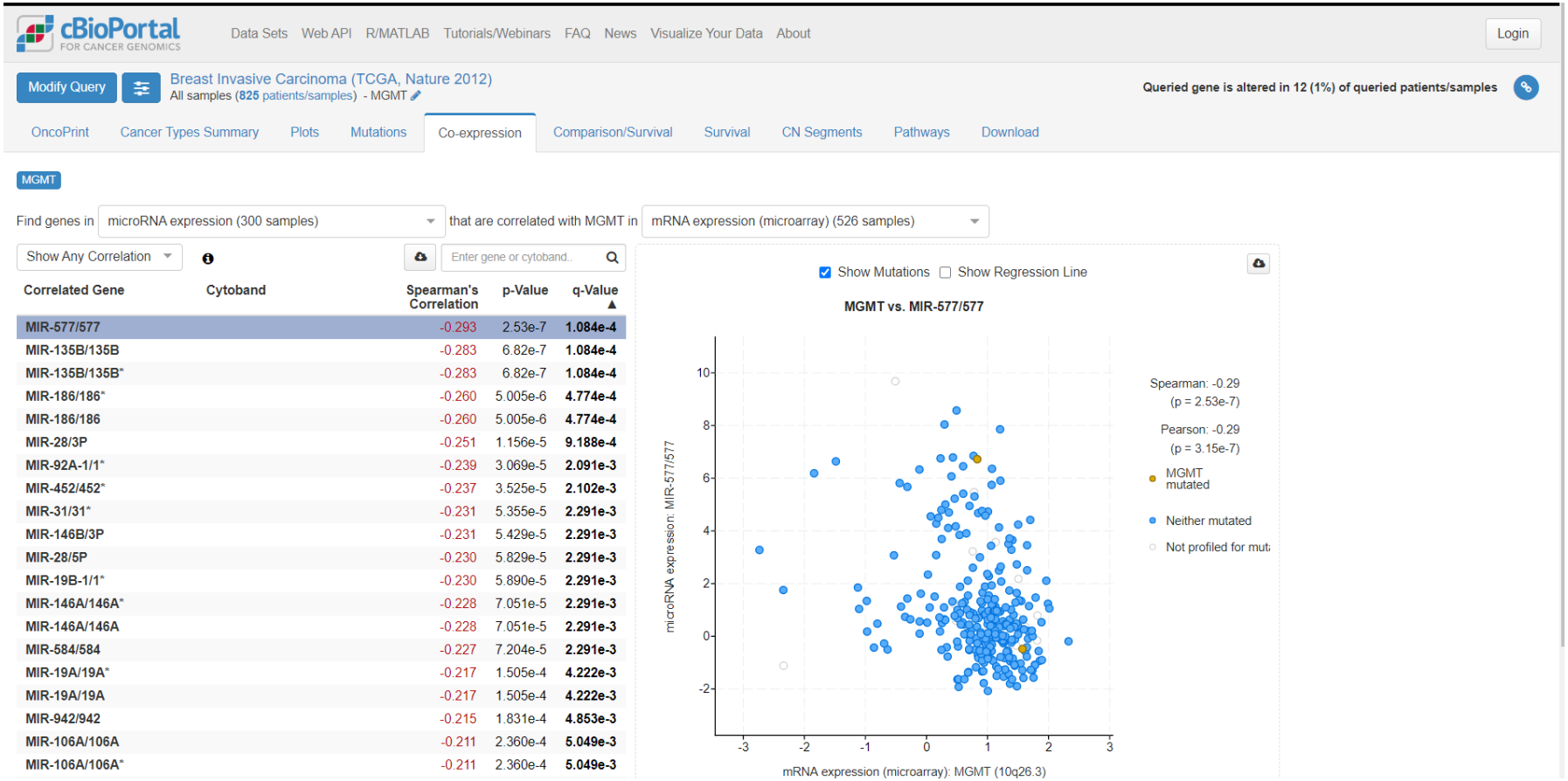
Find genes in mRNA expression (microarray) (526 samples) that are correlated with MGMT in mRNA expression (microarray) (526 samples)

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



miRNA correlation with mRNA example

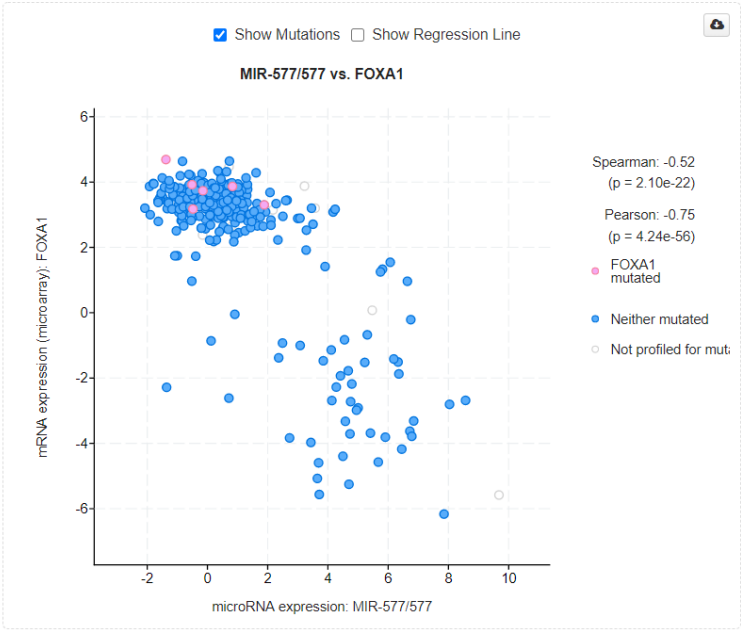
MGMT MIR-577/577

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 ›

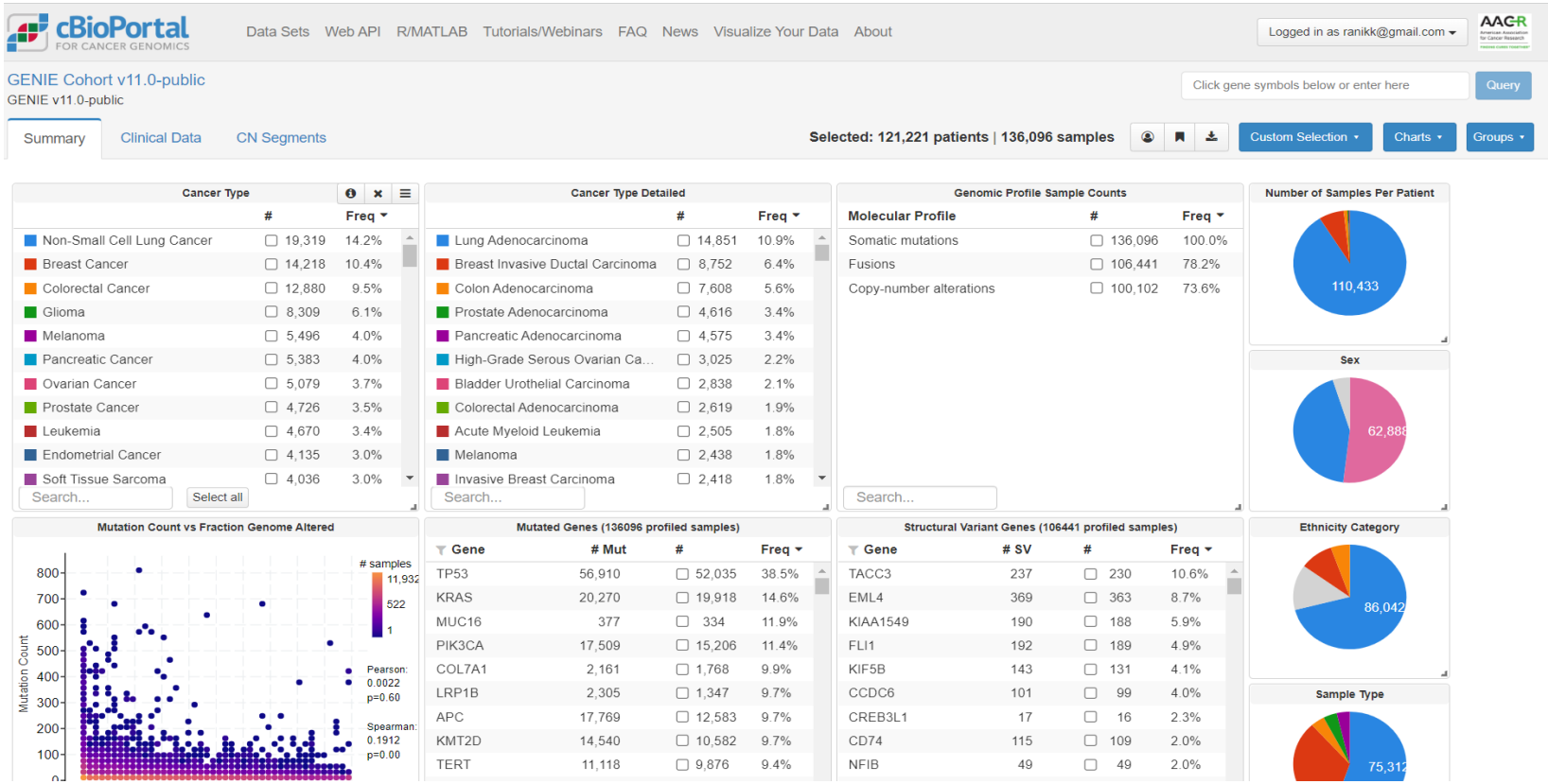


AACR Project GENIE

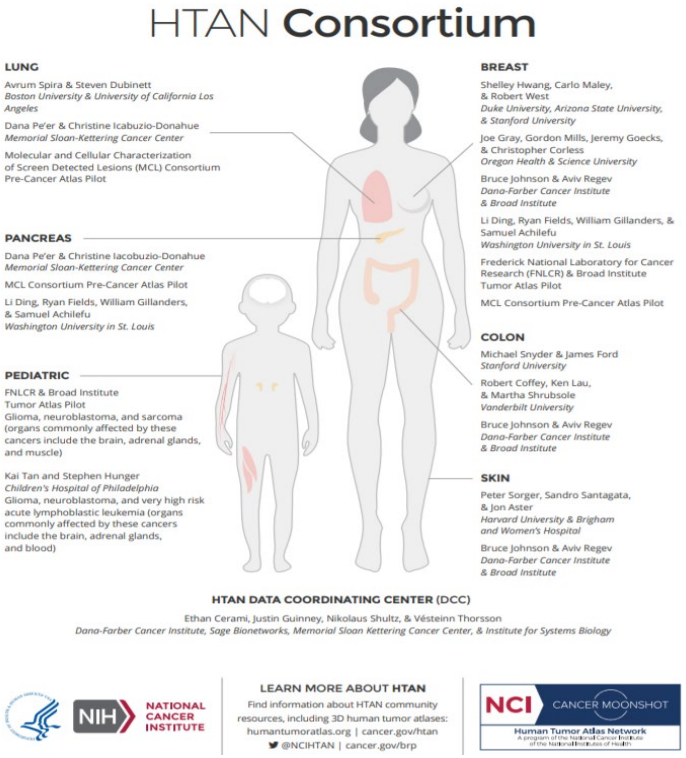
- Clinical sequencing data from 19 cancer centers worldwide.
- It consists of primary and metastatic tumor unlike TCGA where they only have primary and untreated tumors
- For some samples, GENIE also consists of pre and post treatment.
- Targeted gene panels (# of genes targeted varied across the cancer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

GENIE cbioportal

- https://genie.cbioportal.org/
- It consists of more 136 K samples (121K patients)



HTAN – Human Tumor Atlas Network

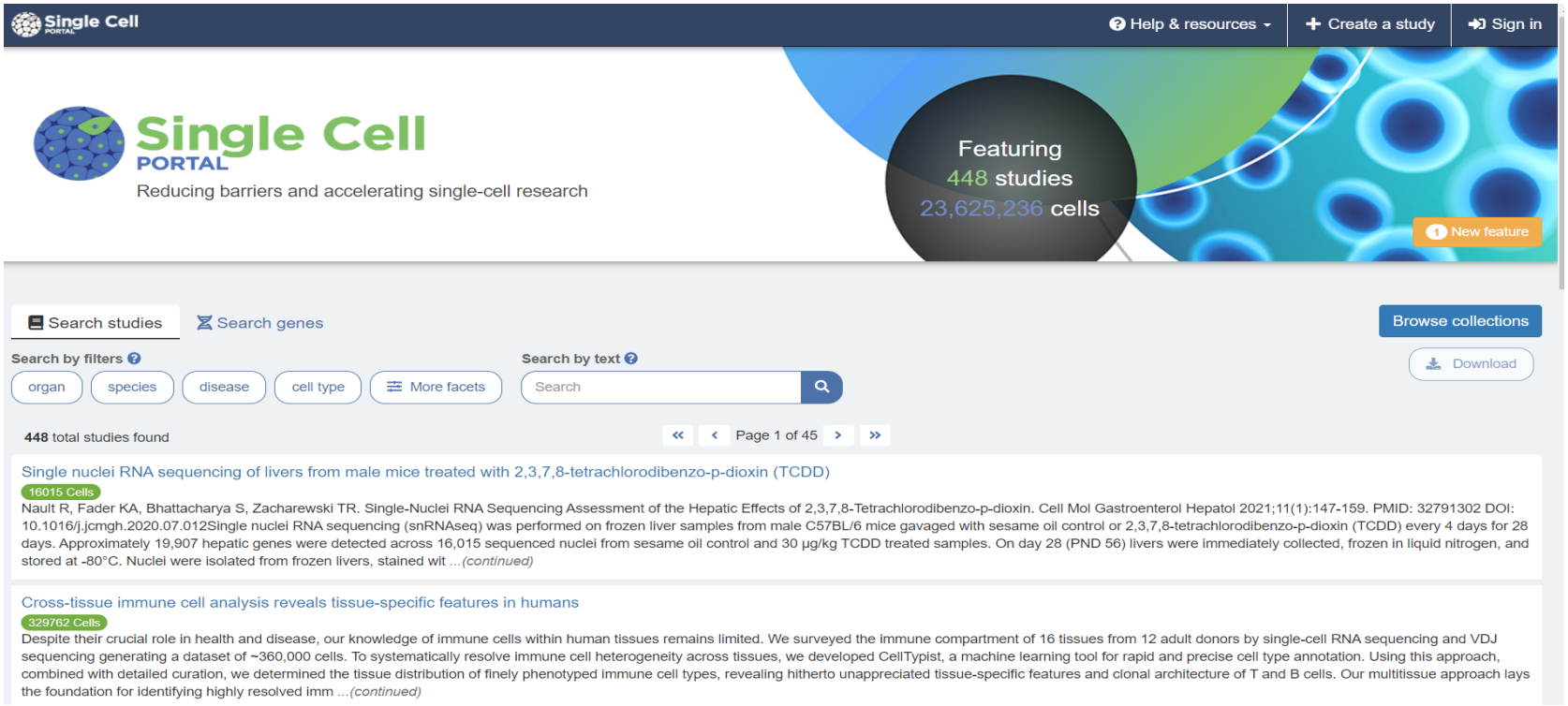


The Human Tumor Atlas Network (HTAN) is a National Cancer Institute (NCI)-funded Cancer Moonshot℠ initiative to construct 3-dimensional atlases of the dynamic cellular, morphological, and molecular features of human cancers as they evolve from precancerous lesions to advanced disease.

10	45	1223	2752
Atlases	Organs	Cases	Biospecimens

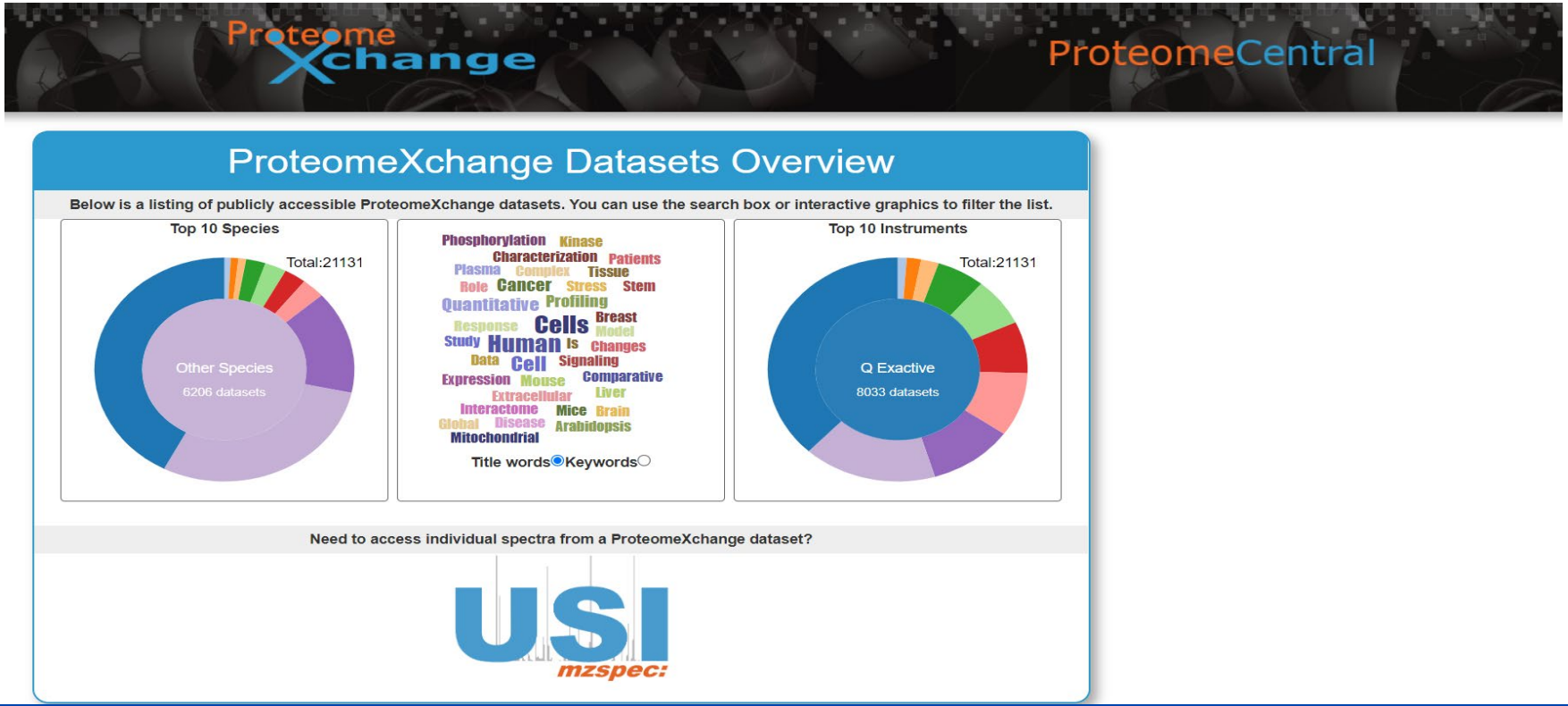
Single-cell Sequencing

- Lots happening in this area.
- Exciting area of research



Proteomics datasets

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



Trans Omics Precision MEDicine (TOPMed)

- Large omics data resource to support discovery science and precision medicine.
- Build a genomic database for more 20 or more diseases or conditions
- 170K Whole genome sequencing data, 32K RNA-Seq, 58K methylome, 16K metabolome and 3K proteome (75% of data released by dbGAP)

About TOPMed

Updated 10/12/2021

Contents

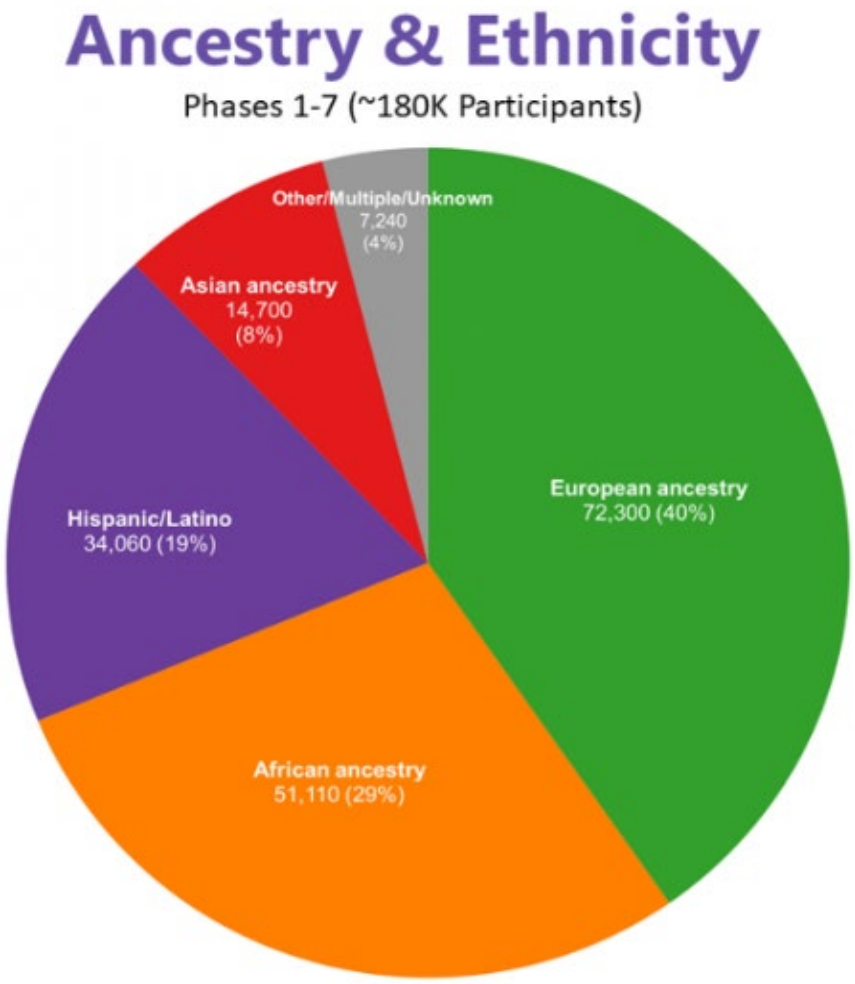
Overview

The Trans-Omics for Precision Medicine (TOPMed) program, sponsored by the National Institutes of Health (NIH) National Heart, Lung and Blood Institute (NHLBI), is part of a broader Precision Medicine Initiative, which aims to provide disease treatments tailored to an individual's unique genes and environment. TOPMed contributes to this Initiative through the integration of whole-genome sequencing (WGS) and other omics (e.g., metabolic profiles, epigenomics, protein and RNA expression patterns) data with molecular, behavioral, imaging, environmental, and clinical data.

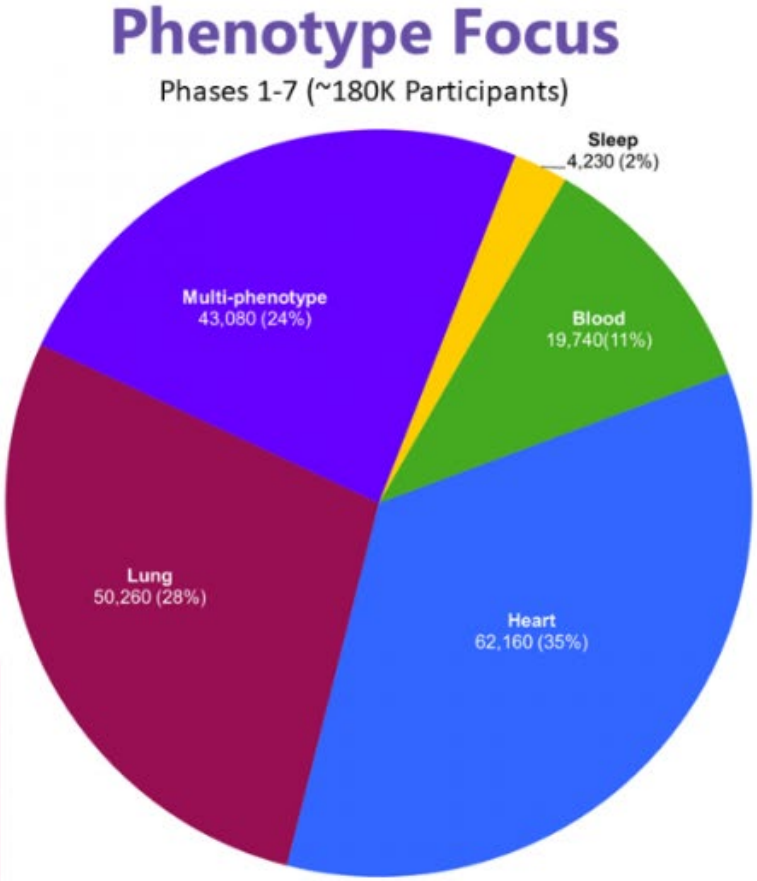
Study Characteristics

A primary goal of the TOPMed program is to improve scientific understanding of the fundamental biological processes that underlie heart, lung, blood, and sleep (HLBS) disorders. TOPMed is providing deep WGS and other omics data to pre-existing 'parent' studies having large samples of human

TOPMed



TOPMed



Blood:

- Hemophilia
- Sickle Cell Disease
- Platelets
- Lipids
- Blood Cancers

Heart:

- Hypertension
- Myocardial Infarction
- Coronary Artery Disease
- Stroke
- Small Vessel Disease
- Venous Thromboembolism
- Congenital Heart Disease
- Atrial Fibrillation
- Coronary Artery Calcification
- Adiposity
- Congestive Heart Failure

Lung:

- Asthma
- Chronic Obstructive Pulmonary Disease
- Idiopathic Pulmonary Fibrosis
- Sarcoidosis
- Interstitial Lung Disease

My Research

Kalari Lab

- Team
- Graduate Students
- Blog
- Softwares
- Projects
- Resources
- Visualization
- Digest Club
- Collaborators
- Photos

Krishna Rani Kalari Ph.D.

Associate Professor, Biomedical Informatics,
College of Medicine, Mayo Clinic, Rochester, MN

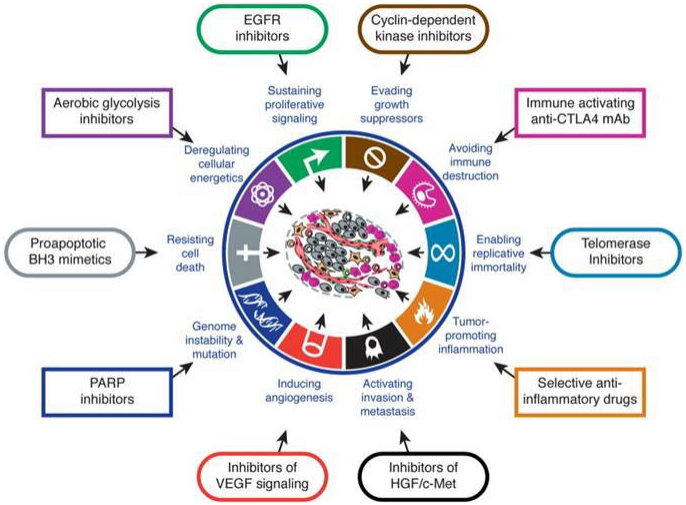
Google Scholar
Krishna Rani Kalari

Email

Dr. Kalari's research focuses on:

1. Pharmacogenomics
2. Biostatistics and Computational genomics
3. Biological systems modeling

COMPUTATIONAL BIOLOGY METHODS



HALLMARKS OF CANCER

PANOPLY
Precision Cancer Genomic Report: Single Sample Inventory Read more.

CIRC-SEQ
A comprehensive bioinformatics workflow for detecting circular RNAs Read more.

MAP-RSEQ
A comprehensive system for RNA-Sequencing data analysis Read more.

eSNV-Detect
Reliable Identification of Variants Using RNA-seq Data Read more.

HGT-ID
A program for detecting viral insertion sequences in the genome of human cancers Read more.

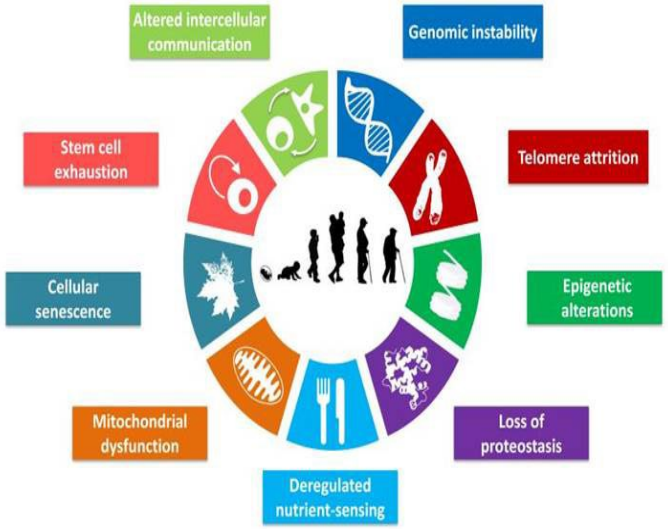
IM-TORNADO
A tool for comparison of 16S reads from paired-end libraries. Read more.

Onco-MATCH
A web-based tool for molecular profile matching of individual cancer patient

CALAR
Clustering Among Luminal Androgen Receptor and basal subtypes. Read more.

ReMIX
Integration of mRNA and microRNA data

CIC
Clone Initiating Calculator.



HALLMARKS OF AGING

Thank you all and email me if you have any questions at Kalari.Krishna@mayo.edu