



# Cancer Informatics Lecture

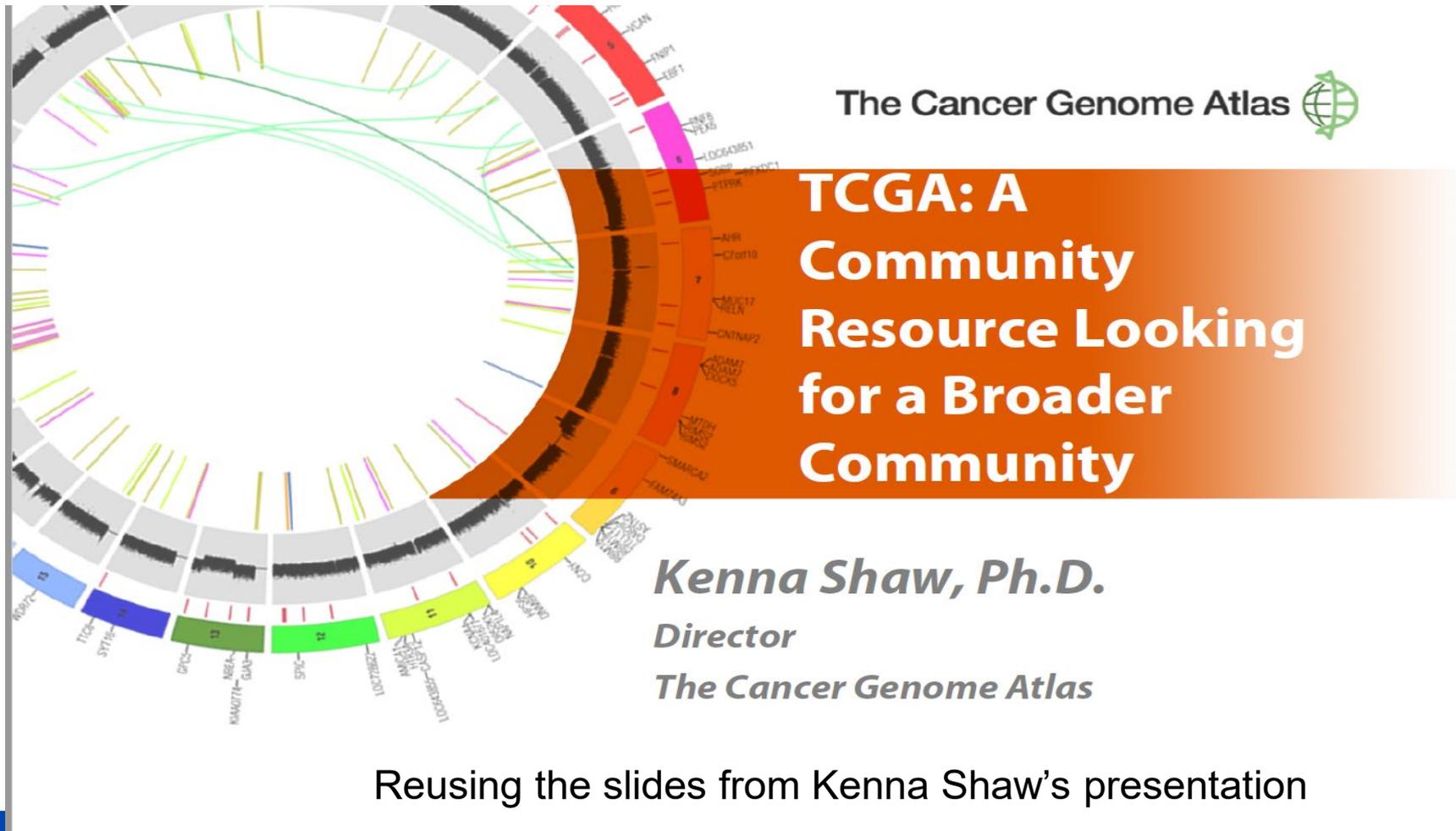
Mayo-UIUC Computational Genomics Course  
June 12, 2020

Krishna Rani Kalari Ph.D.  
Associate Professor

# Outline

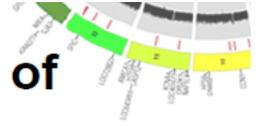
- The Cancer Genome Atlas (TCGA)
- Genomic Data Commons (GDC)
- COSMIC database (mutations database)
- cBioPortal for cancer genomics
- GTEX
- Precision Medicine in Cancer
- Single-cell RNA-Seq
- Proteomics

# The Cancer Genome Atlas (TCGA)



# TCGA core objectives

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

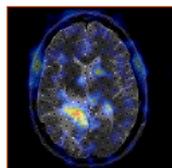


- Establish infrastructure for effective team science
- Develop a scalable “pipeline” beginning with highest quality samples
- Determine the feasibility of a large-scale, high throughput approach to identifying the molecular ‘parts-list’
- Evaluate using statistically-robust sample sets
- Make the data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

# TCGA multiple data types

## 25\* forms of cancer

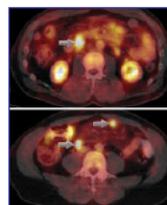
glioblastoma multiforme  
(brain)



squamous carcinoma  
(lung)



serous  
cystadenocarcinoma  
(ovarian)



Etc. Etc. Etc.

Biospecimen Core  
Resource with more  
than 150 Tissue Source  
Sites

6 Cancer Genomic  
Characterization  
Centers

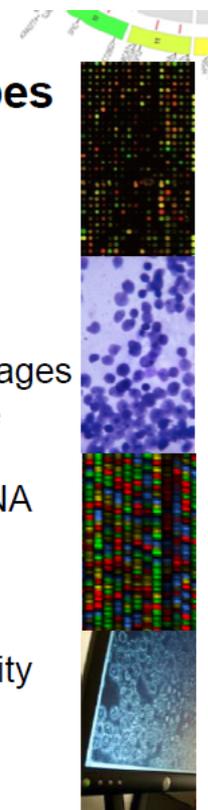
3 Genome  
Sequencing  
Centers

7 Genome Data  
Analysis Centers

Data Coordinating  
Center

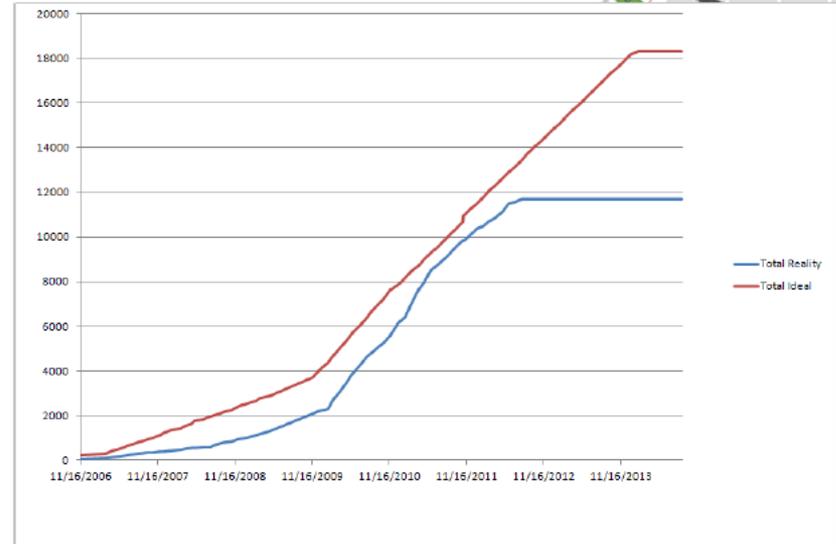
## Multiple data types

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



# 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

Harmonized Cancer Datasets

## Genomic Data Commons Data Portal

Get Started by Exploring:

Projects Exploration Analysis Repository

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

### Data Portal Summary

Data Release 24.0 - May 07, 2020

PROJECTS

65

PRIMARY SITES

67

CASES

84,031

FILES

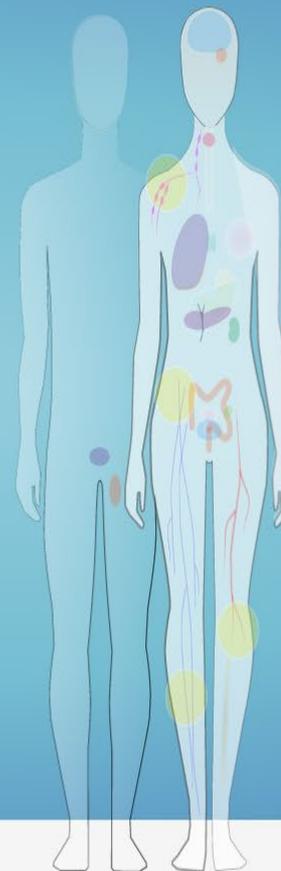
570,844

GENES

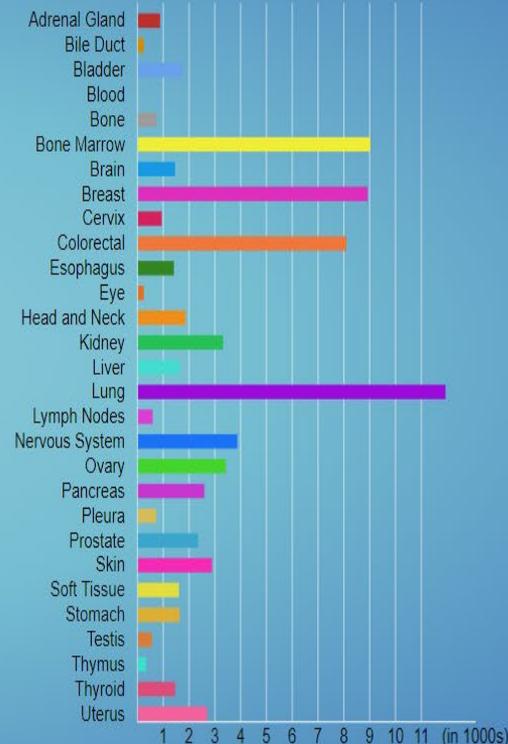
22,872

MUTATIONS

3,142,246



Cases by Major Primary Site



← 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

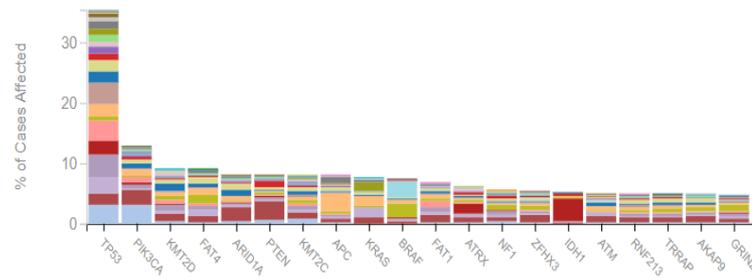
10 More...

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

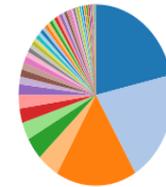


Table Graph

65Projects

JSON TSV

Available Cases per Data Category

Project	Disease Type	Primary Site	Program	Cases ↑	Seq	Exp	SNV	CNV	Meth	Clinical	Clinical		Bio		Files
											Supplement	Bio	Supplement	Bio	
<a href="#">FM-AD</a>	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	
<a href="#">GENIE-MSK</a>	49 Disease Types	49 Primary Sites	GENIE	16,824	0	0	16,823	16,823	0	16,824	0	16,824	0	36,470	

# Exploration

Cases Genes Mutations

Add a Case Filter

Case

e.g. TCGA-A5-A0G2, 432fe4a9-2...

Upload Case Set

Primary Site

- Bronchus and lung 4,977
- Breast 3,682
- Colon 2,305
- Kidney 2,201
- Unknown 1,563

63 More...

Program

- FM 18,004
- TCGA 11,315
- TARGET 3,380
- NCICCR 489
- CPTAC 322

4 More...

Project

- FM-AD 18,004
- TARGET-NBL 1,120
- TCGA-BRCA 1,098
- TARGET-AML 988
- TARGET-WT 652

42 More...

Start searching by selecting a facet

View Files in Repository

Cases (33,605) Genes (22,872) Mutations (3,142,246) OncoGrid

OncoGrid

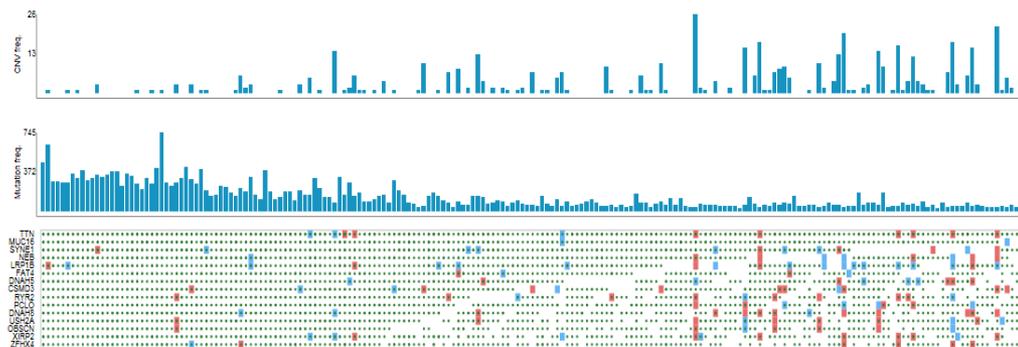
200 Most Mutated Cases and Top 50 Mutated Genes By SSM

Mutations

- Show Mutations
- Missense Start Lost Stop Gained
- Frameshift Stop Lost

CNV Changes

- Show Copy Number Variations
- Loss
- Gain



# Analysis

Launch Analysis



## Set Operations

Display Venn diagram and find intersection or union, etc. of your sets of the same type.

Select

Demo



## Cohort Comparison

Display the survival analysis of your case sets and compare characteristics such as gender, vital status and age at diagnosis.

Select

Demo



## Clinical Data Analysis

Display basic statistical analyses for the selected case set.

Select

Demo

# Repository

Files Cases

Add a File Filter

File

e.g. 142682.bam, 4f6e2e7a-b...

Data Category

- simple nucleotide variation **221,999**
- copy number variation **100,983**
- transcriptome profiling **86,875**
- sequencing reads **76,609**
- biospecimen **55,582**

4 More...

Data Type

- Annotated Somatic Mutation **91,656**
- Raw Simple Somatic Mutation **80,936**
- Aligned Reads **76,609**
- Gene Expression Quantification **52,616**
- Masked Annotated Somatic Mutation **44,755**

16 More...

Experimental Strategy

- WXS **171,975**

← Start searching by selecting a facet

Advanced Search

Add All Files to Cart Manifest View 84,031 Cases in Exploration View Images

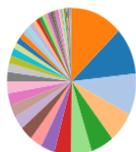
Browse Annotations

Files (570,844)

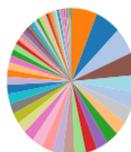
Cases (84,031)

1.43 PB

Primary Site



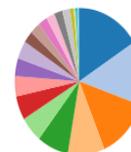
Project



Data Category



Data Type



Data Format



Show More

Showing 1 - 20 of 570,844 files

JSON TSV

Access	File Name	Cases	Project	Data Category	Data Format	File Size	Annotations
open	<a href="#">HOOLY_p_TCGAb_366_P04_NSP_GenomeWideSNP_6_D01_1390662.grch38.seg.v2.txt</a>	1	TCGA-PCPG	Copy Number Variation	TXT	27.73 KB	0
open	<a href="#">BUCKS_p_TCGA_272_273_N_GenomeWideSNP_6_C03_1320570.grch38.seg.v2.txt</a>	1	TCGA-HNSC	Copy Number Variation	TXT	35.05 KB	0
open	<a href="#">b50539eb-1a68-4e51-821d-d74ed8210b1e_mirbase21_mimas_quantification.txt</a>	1	TCGA-KIRC	Transcriptome Profiling	TXT	50.16 KB	0
open	<a href="#">418dd9f0-a87c-402e-8101-2c769e383cc5_htseq_counts.gz</a>	1	TCGA-HNSC	Transcriptome Profiling	TXT	252.09 KB	0
open	<a href="#">acc7d983-3759-489f-844f-d87543f07b26.FPKM-UQ.txt.gz</a>	1	TCGA-HNSC	Transcriptome Profiling	TXT	575.05 KB	0
controlled	<a href="#">cb8eaa7a-5584-4888-9c68-2e73eb8d4ec7.vcf.gz</a>	1	TCGA-COAD	Simple Nucleotide Variation	VCF	1.56 MB	0
open	<a href="#">TCGA_AB_2877_11A_01D_0756_21.grch38.seg.v2.txt</a>	1	TCGA-LAML	Copy Number Variation	TXT	34.46 KB	1



## 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

## Projects

---

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:



### COSMIC

The core of COSMIC, an expert-curated database of somatic mutations



### Cell Lines Project

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



### COSMIC-3D

An interactive view of cancer mutations in the context of 3D structures



### Cancer Gene Census

A catalogue of genes with mutations that are causally implicated in cancer

## COSMIC v91, released 07-APR-20

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.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below.

 **SEARCH**

### Projects

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:



**COSMIC**  
The core of COSMIC, an expert-curated database of somatic mutations



**Cell Lines Project**  
Mutation profiles of over 1,000 cell lines used in cancer research



**COSMIC-3D**  
An interactive view of cancer mutations in the context of 3D structures



**Cancer Gene Census**  
A catalogue of genes with mutations that are causally implicated in cancer

### Data curation

- [Gene Curation](#) — details of our manual curation process
- [Gene Fusion Curation](#) — details of our curation process for gene fusions
- [Genome Annotation](#) — information on the annotation of genomes
- [Drug Resistance](#) — curation of mutations conferring drug resistance

## COSMIC News

[Follow @cosmic\\_sanger](#)



### Service announcement: Normal service resumed

Following an outage 26th May 20, we can confirm our normal service has been resumed. [More...](#)



### V92 Release coming soon

We have an exciting new product, The Cancer Mutation Census, as well as expertly curated genes and drug resistance data in our July release. [More...](#)



### NEW: The Cancer Mutation Census!

We have a new product! The Cancer Mutation Census allows for the prioritisation of somatic mutations that introduce biologically relevant changes to protein function, and participate in the development of cancer. [More...](#)

### Tools

- [Cancer Browser](#) — browse COSMIC data by tissue type and histology
- [Genome Browser](#) — browse the human genome with COSMIC annotations
- [GA4GH Beacon](#) — access COSMIC data through the [GA4GH Beacon Project](#)
- [COSMIC in BigQuery](#) — search COSMIC via the [ISB Cancer Genomics Cloud](#)

### Help

- [Downloads](#) — data that you can download from our SFTP site
- [Documentation](#) — view our help documentation
- [FAQ](#) — a compilation of our Frequently Asked Questions
- [Release Notes](#) — information about the latest COSMIC release

## Gene

### KRAS

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

[Reset page](#)

### Search

### Filters

[Show advanced filters](#)

**Range** [Show input fields](#)

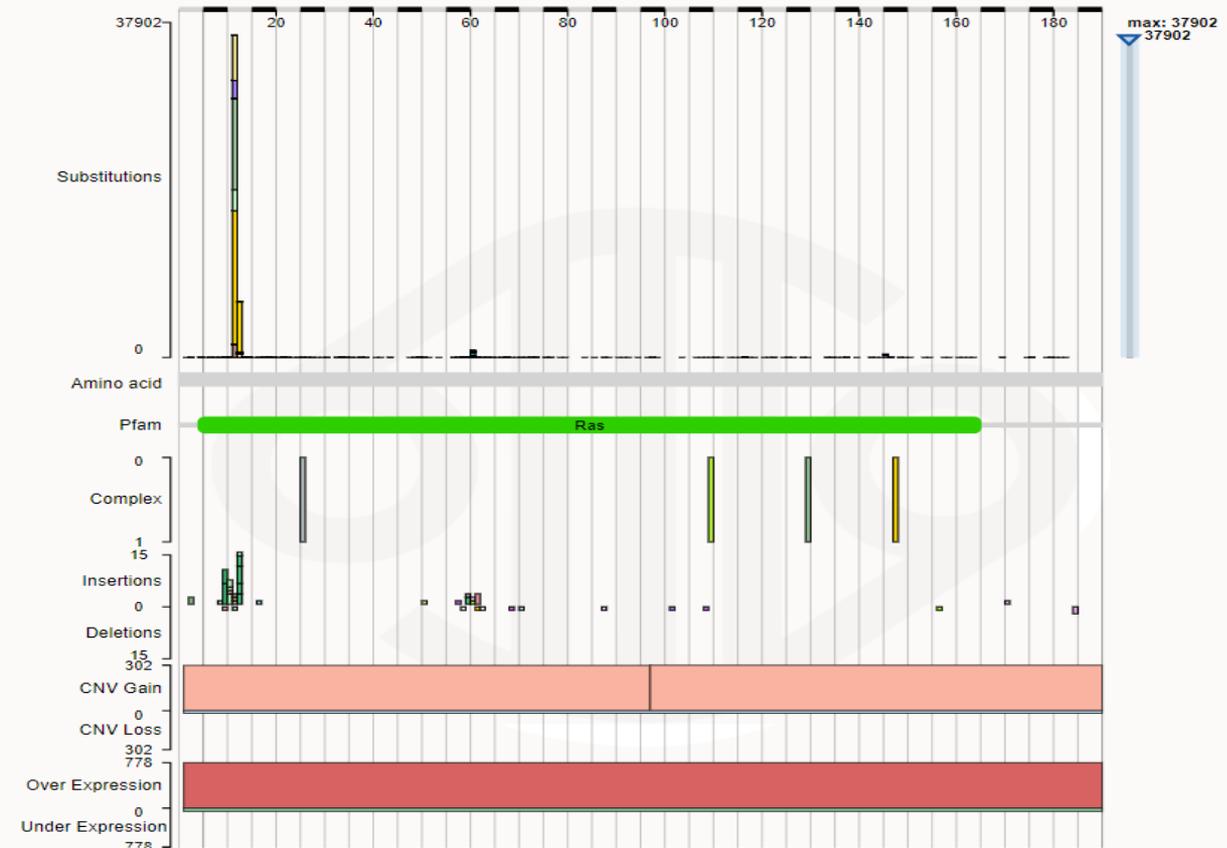


### Coordinate system

- Amino-acid
- cDNA

## Gene view

The gene view histogram is a graphical view of mutations across KRAS. These mutations are displayed at the amino acid 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](#)

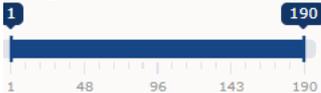
## Search

Search COSMIC...

## Filters

[Show advanced filters](#)

Range [Show input fields](#)



## 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.

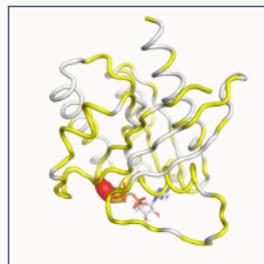
- [Census gene](#)
- [Curated gene](#)
- [Mouse gene](#)
- [Hallmark gene](#)

**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

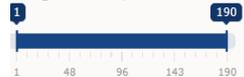
[Reset page](#)

## Search

## Filters

[Show advanced filters](#)

Range [Show input fields](#)



## Coordinate system

- Amino-acid
- cDNA

## 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 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  entries

Search:

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

# Cell Lines Project

## Projects

---

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:



### COSMIC

The core of COSMIC, an expert-curated database of somatic mutations



### Cell Lines Project

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



### COSMIC-3D

An interactive view of cancer mutations in the context of 3D structures



### Cancer Gene Census

A catalogue of genes with mutations that are causally implicated in cancer

# Cell lines project

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

The screenshot shows the COSMIC Cell Lines v91 web interface. The top navigation bar includes 'Projects', 'Data', 'Tools', 'News', 'Help', 'About', 'Genome Version', a search bar with 'SEARCH', and a 'Login' button. The main header displays 'GRCh38 · CELL LINES v91'. On the left, a sidebar lists navigation options: Overview, Circos, Genome browser, Variants, Mutation spectrum, Sequence context, Heatmap, Non-mutant genes, and References. The main content area is titled 'Overview' and contains the following information:

This tab shows an overview of the data that we have for this sample. You can read more about these data on our [help pages](#).

**Sample information**

- Sample name:** MCFZ
- COSMIC sample ID:** COSS905946
- Tumour location:** Breast (Carcinoma)  
View this tissue/histology in the [Cancer Browser](#)
- Screening method:** Whole exome screening
- Source:**
  - Sample type:** Cultured Sample
  - Cell line source:** primary
  - Sample source:** cell-line

**Curated features**

- Sample details:** n/a
- Tumour details:** n/a
- Individual details:**
  - Ethnicity:** Caucasian
  - Normal tissue tested:** No
  - Gender:** Unknown
  - Age:** 69

**Additional information**

- Microsatellite instability (MSI):** Stable
- Other samples linked to the same individual:** n/a
- Supplier:**
  - Institute:** Developmental Therapeutics Program
  - Address:** National Cancer Institute, Frederick, MD 21701

# COSMIC-3D

## Projects

---

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:



### COSMIC

The core of COSMIC, an expert-curated database of somatic mutations



### Cell Lines Project

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



### COSMIC-3D

An interactive view of cancer mutations in the context of 3D structures



### Cancer Gene Census

A catalogue of genes with mutations that are causally implicated in cancer

# COSMIC-3D

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

COSMIC-3D Settings Help

Back to COSMIC Report Bugs and Feedback

Search for a gene or protein, e.g. BRAF, EGFR, or BRCA2...

## EGFR

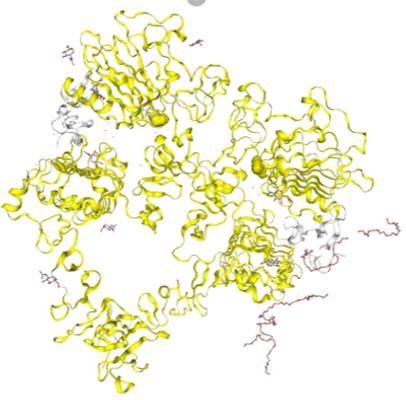
P00533 In Census 183 structures

Single Structure Mode

Receptor tyrosine kinase binding ligands of the EGF family and activating several signaling cascades to convert extracellular cues into appropriate cellular responses (PubMed:2790960, PubMed:10805725, PubMed:27153536).

External Links Toggle Usage Hints Share

Structures 183



Information

Click on a mutation in the Sequence Feature Viewer below to view more information.

# COSMIC

## Projects

---

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:



### COSMIC

The core of COSMIC, an expert-curated database of somatic mutations



### Cell Lines Project

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



### COSMIC-3D

An interactive view of cancer mutations in the context of 3D structures



### Cancer Gene Census

A catalogue of genes with mutations that are causally implicated in cancer

# 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  entries

Search:

Sorted By ▲	Number ▼
<a href="#">Amplifications</a>	24
<a href="#">Chromosome</a>	576
<a href="#">Frameshift Mutations</a>	156
<a href="#">Gene Symbol</a>	576
<a href="#">Germline Mutations</a>	102
<a href="#">Large Deletions</a>	41
<a href="#">Missense Mutations</a>	253
<a href="#">Nonsense Mutations</a>	155
<a href="#">Other Mutations</a>	37
<a href="#">Somatic Mutations</a>	536
<a href="#">Splicing Mutations</a>	73
<a href="#">Translocations</a>	314

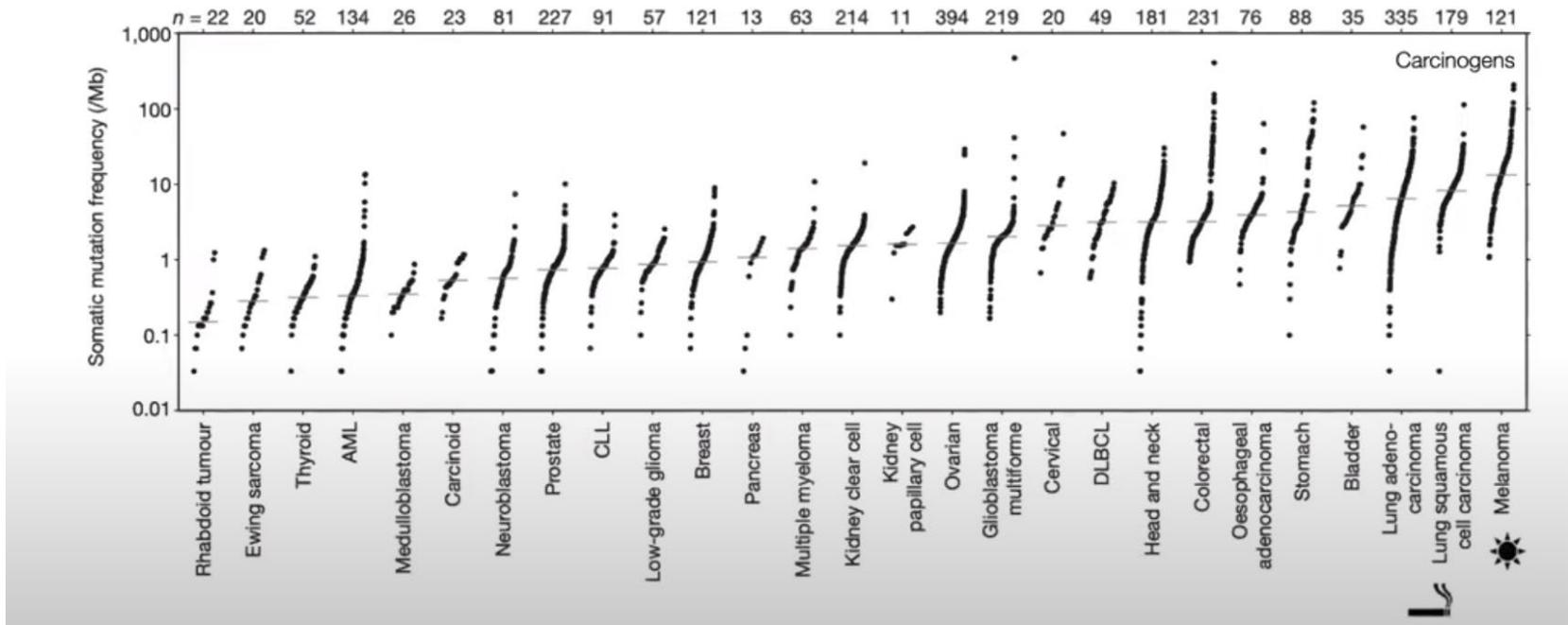
Showing 1 to 12 of 12 entries

Previous  Next



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

# 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**

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

**List of all studies, organized by organ system**

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

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

# Single study query

1. Filter the list of studies (optional)

**Select Studies for Visualization & Analysis:** 1 study selected (530 samples) Deselect all glioma X

**CNS/Brain** 15  Select all listed studies matching filter (16)

**Soft Tissue** 1

**Diffuse Glioma**

- Brain Lower Grade Glioma (TCGA, Firehose Legacy) 530 samples
- Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 514 samples
- Glioma (MSK, 2018) 91 samples
- Glioma (MSKCC, Clin Cancer Res 2019) 1004 samples
- Low-Grade Gliomas (UCSF, Science 2014) 61 samples
- Merged Cohort of LGG and GBM (TCGA, Cell 2016) 1102 samples

**GLIOBLASTOMA**

- Brain Tumor PDXs (Mayo Clinic, 2019) 95 samples
- Glioblastoma (TCGA, Cell 2013) 543 samples
- Glioblastoma (TCGA, Nature 2008) 206 samples
- Glioblastoma Multiforme (TCGA, Firehose Legacy) 604 samples
- Glioblastoma Multiforme (TCGA, PanCancer Atlas) 592 samples

**OLIGODENDROGLIOMA**

- Anaplastic Oligodendroglioma and Anaplastic Oligoastrocytoma (MSK...) 22 samples

**Encapsulated Glioma**

**PILOCYTIC ASTROCYTOMA**

- Pilocytic Astrocytoma (ICGC, Nature Genetics 2013) 96 samples

**Miscellaneous Neuroepithelial Tumor**

1 study selected (530 samples) Deselect all **Query By Gene** OR **Explore Selected Studies**

2. Check the box for study of interest.

3. Select "Query By Gene"

Query

Quick Search **Beta!**

Download

Selected Studies:

Brain Lower Grade Glioma (TCGA, Firehose Legacy) (530 total samples)

Select Genomic Profiles:

- Mutations 
- Putative copy-number alterations from GISTIC 
- mRNA Expression. Select one of the profiles below:
  - mRNA expression z-scores relative to diploid samples (microarray) 
  - mRNA expression z-scores relative to diploid samples (RNA Seq V2 RSEM) 
  - mRNA expression z-scores relative to all samples (log RNA Seq V2 RSEM) 
  - mRNA expression z-scores relative to all samples (microarray) 
- Protein expression Z-scores (RPPA) 

Select Patient/Case Set:

To build your own case set,  
try out our enhanced Study View.

Samples with mutation and CNA data (283)  

Enter Genes:

Hint: Learn Onco Query Language (OQL)  
to write more powerful queries 

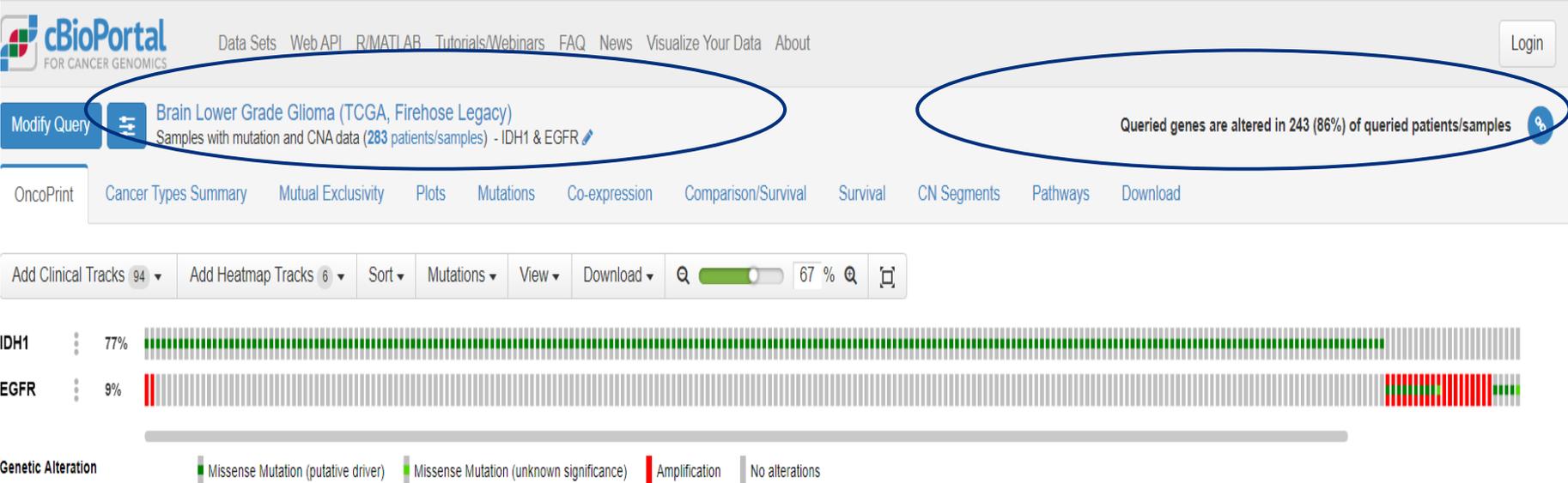
User-defined List  

IDH1  
EGFR

 All gene symbols are valid.

Submit Query

# Glioma Query



# Annotations and Filtering

**cBioPortal** FOR CANCER GENOMICS

Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data About Login

**Modify Query** **Brain Lower Grade Glioma (TCGA, Firehose Legacy)** Queried genes are altered in 243 (86%) of queried patients/samples

Samples with mutation and CNA data (283 patients/samples) - IDH1 & EGFR

OncoPrint **Annotate Data** Plots Mutations Co-expression Comparison/Survival Survival CN Segments Pathways Download

Add Clinical Track

Putative drivers vs VUS

OncoKB driver annotation

Hotspots

cBioPortal  $\geq 10$

COSMIC  $\geq 10$

**Genetic Alteration** **Filter Data**

Exclude mutations and copy number alterations of unknown significance

Exclude germline mutations

Mutations View Download 67%

Missense Mutation (unknown significance) Amplification No alterations

# 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.



The percentage of samples with an alteration in each query gene.

To change the order, click on a gene name and drag, or click on the . Samples will re-sort based on this new order.

# OncoPrint: Features

The image shows a screenshot of the OncoPrint web interface. At the top, there are navigation tabs: OncoPrint, Cancer Types Summary, Mutual Exclusivity, Plots, Mutations, Core-expression, Comparison, Survival, CN Segments, Pathways, and Download. Below these is a toolbar with buttons for 'Add Clinical Tracks' (94), 'Add Heatmap Tracks' (4), 'Sort', 'Mutations', 'View', and 'Download'. A search bar shows '49 %' and a zoom slider. On the left, a sidebar shows a table of clinical tracks with columns for 'Name' and 'Freq'. The table lists 'Diagnosis Age' (100.0%), 'Animal Insect Allergy Age' (2.8%), and 'Age of Food Allergy' (1.4%). The main area displays a genomic track with a green bar at the top and a red bar below it. Callout boxes with arrows point to various features: 'Add clinical tracks (options will vary depending on the data available for each study)' points to the 'Add Clinical Tracks' button; 'Add a heatmap with RNA or protein levels or treatment response (when available)' points to the 'Add Heatmap Tracks' button; 'Change the sample sorting order' points to the 'Sort' button; 'Customize visualization' points to the 'View' button; 'Change the rules by which mutations are colored.' points to the 'Mutations' button; and 'Download figure as PNG, PDF or SVG. Download patient/sample IDs in same order as OncoPrint.' points to the 'Download' button.

**Add clinical tracks (options will vary depending on the data available for each study)**

**Add a heatmap with RNA or protein levels or treatment response (when available)**

**Change the sample sorting order**

**Customize visualization**

**Change the rules by which mutations are colored.**

**Download figure as PNG, PDF or SVG. Download patient/sample IDs in same order as OncoPrint.**

Name	Freq
<input checked="" type="checkbox"/> Diagnosis Age	100.0%
<input type="checkbox"/> Animal Insect Allergy Age	2.8%
<input type="checkbox"/> Age of Food Allergy	1.4%

# 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.

OncoPrint

The query contains 3 gene pairs with mutually exclusive alterations (2 significant), and no gene

Mutual exclusivity  Co-occurrence  Significant only

Gene A	Gene B	Neither	A Not B	B Not A	Both	Log Odds Ratio	p-Value ▲	Tendency ←
EGFR	IDH1	40	24	217	2	<-3	<0.001	Mutual exclusivity <b>Significant</b>
IDH1	IDH2	52	218	12	1	<-3	<0.001	Mutual exclusivity <b>Significant</b>
EGFR	IDH2	244	26	13	0	<-3	0.278	Mutual exclusivity

Showing 1-3 of 3

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.

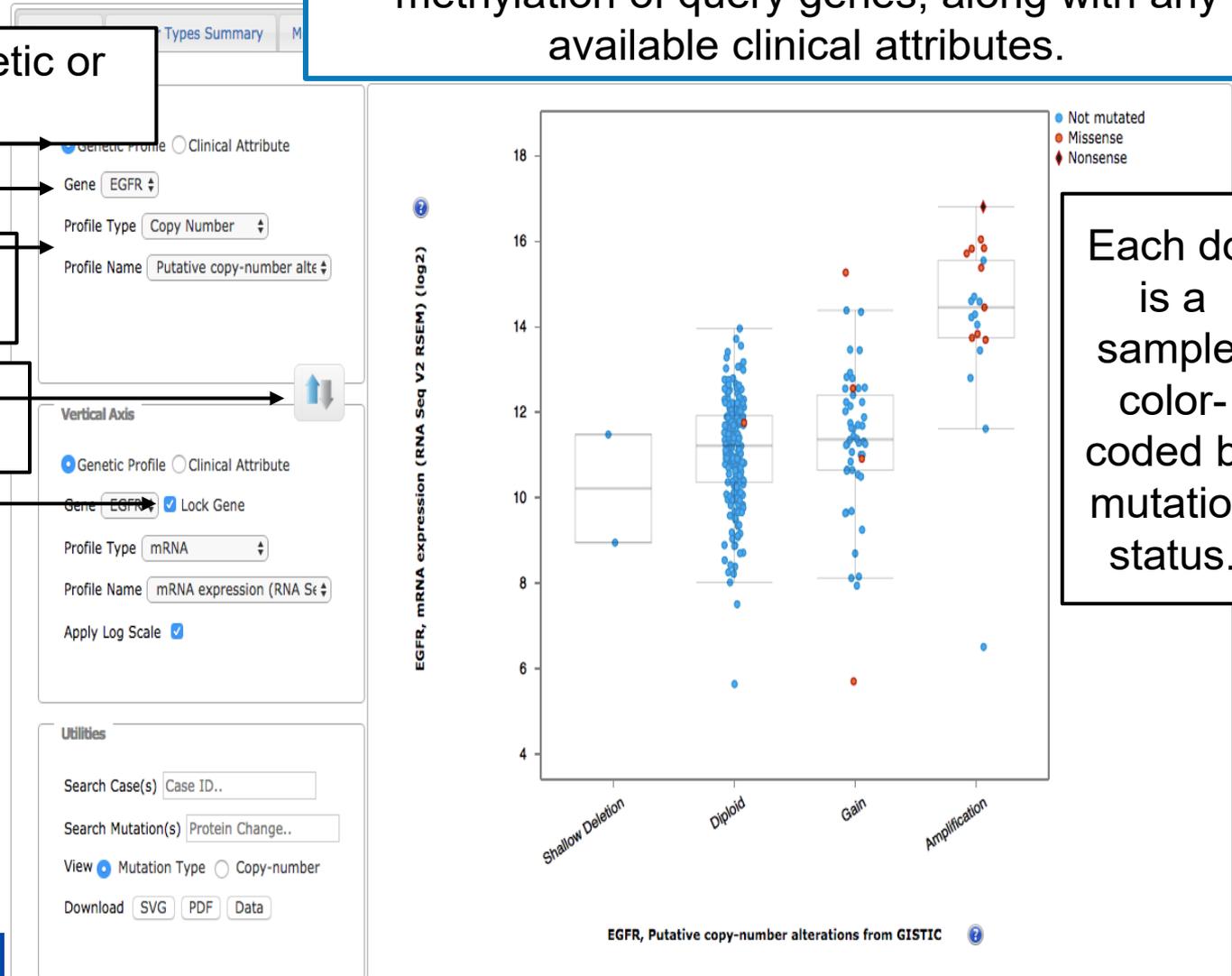
Choose genetic or clinical

Select a query gene

Select data type and processing

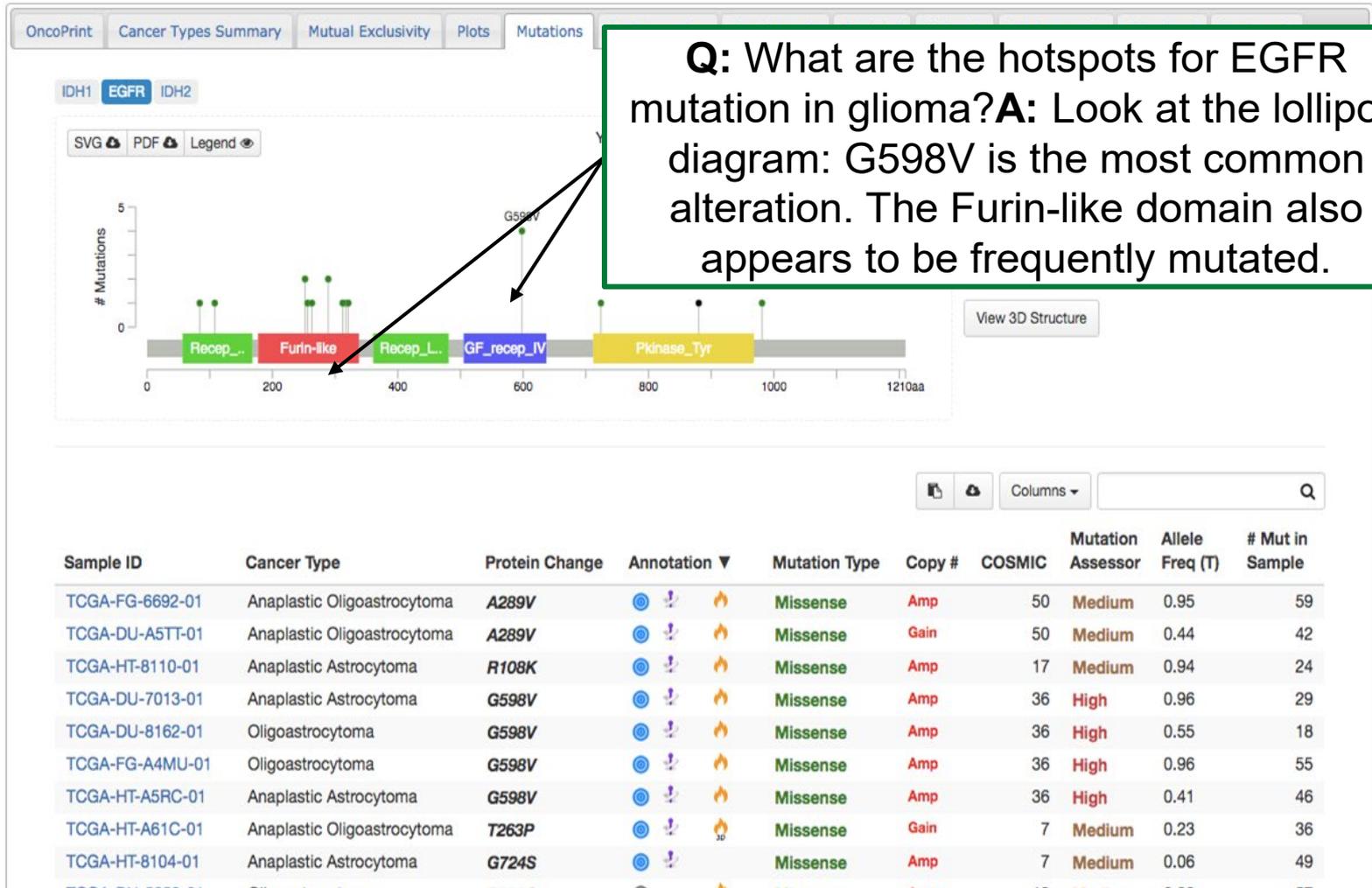
Swap horizontal & vertical axis

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

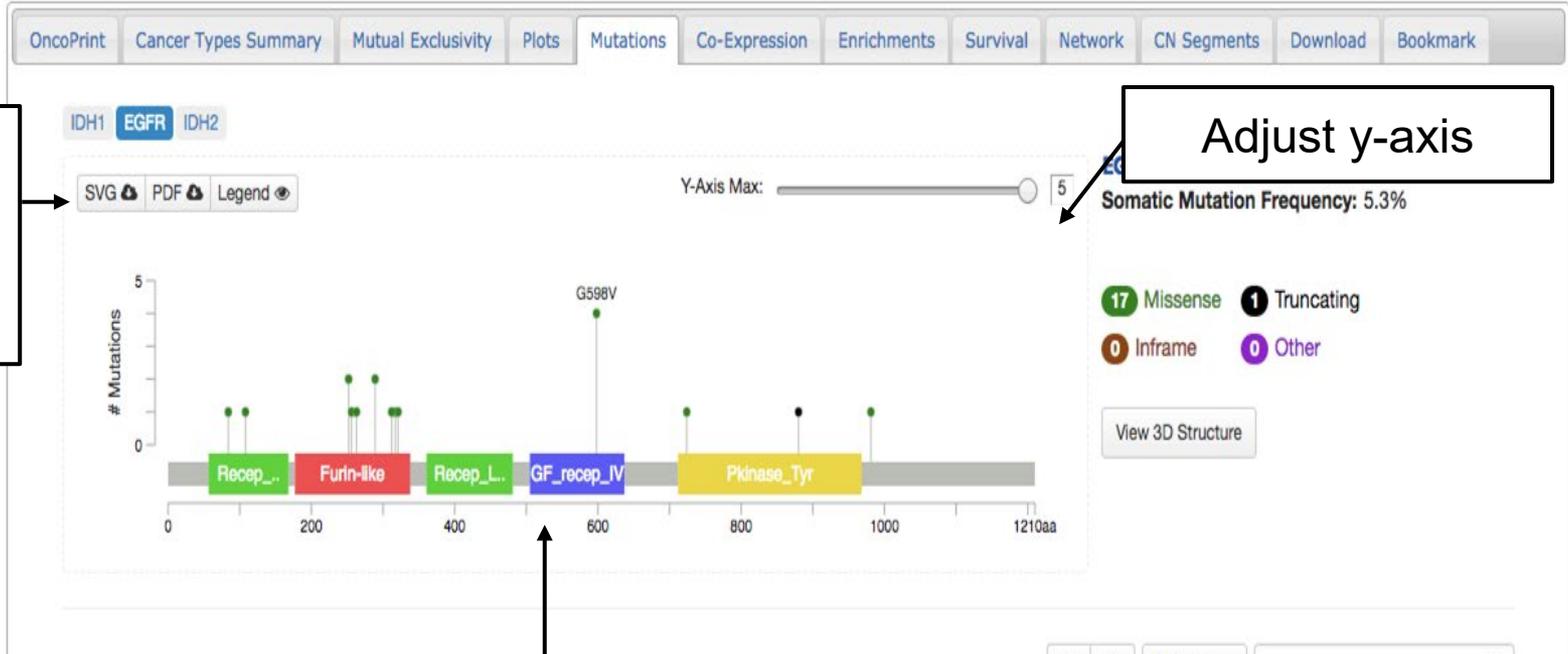


Each dot is a sample, color-coded by mutation status.

# Mutations



# 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.

Copy #	COSMIC	Mutation Assessor	Allele Freq (T)	# Mut in Sample
Amp	50	Medium	0.95	59
Gain	50	Medium	0.44	42
Amp	17	Medium	0.94	24
Amp	36	High	0.96	29
Amp	36	High	0.55	18
Amp	36	High	0.96	55
Amp	36	High	0.41	46
Gain	7	Medium	0.23	36
Amp	7	Medium	0.06	49
Amp	10	Medium	0.22	57

TCGA ID	Cancer Type	Mutation	Assessor	Assessor	Assessor	Assessor	Assessor	Assessor
TCGA-HT-A5RC-01	Anaplastic Astrocytoma	G598V	High	High	High	High	High	High
TCGA-HT-A61C-01	Anaplastic Oligoastrocytoma	T263P	High	High	High	High	High	High
TCGA-HT-8104-01	Anaplastic Astrocytoma	G724S	High	High	High	High	High	High
TCGA-DU-5959-01	Oligoastrocytoma	R265G	High	High	High	High	High	High

# 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.

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

# Breast cancer example

Modify Query

# Breast Invasive Carcinoma (TCGA, PanCancer Atlas)

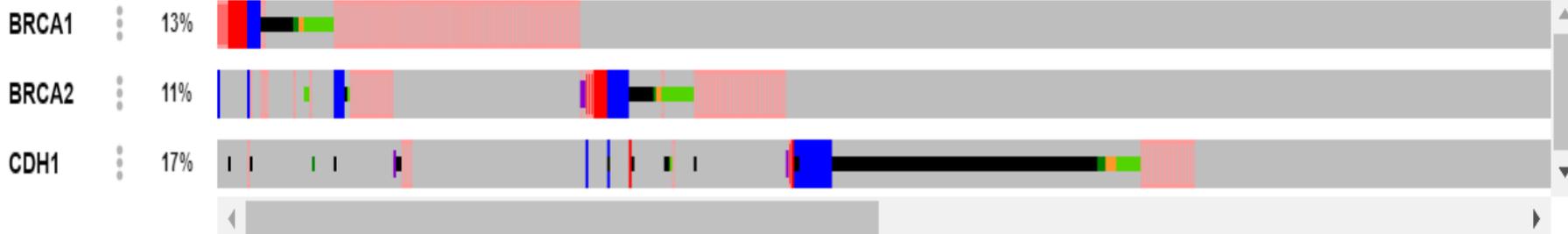
All Complete Tumors (993 samples) / 3 Genes

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

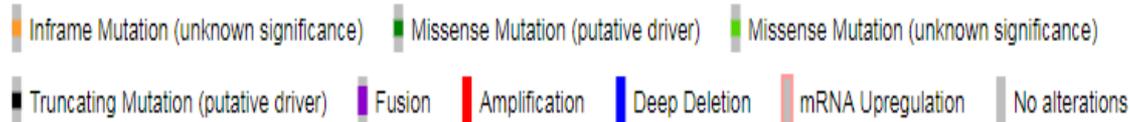
OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-Expression Enrichments Network CN Segments Download Bookmark

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

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



Genetic Alteration



# CNV

Modify Query

## Breast Invasive Carcinoma (TCGA, PanCancer Atlas)

All Complete Tumors (993 samples) / 4 Genes

Gene Set / Pathway is altered in 420 (42.3%) of queried samples



# mRNA overexpressed



# Co-Expression

Select from available data types

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-Expression

Data Set mRNA expression (RNA Seq V2 RSEM) +

This table lists the genes with the highest expression correlation with the query genes. Click on a row

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.

Each gene appears on a separate tab

IDH1 EGFR IDH2

Search Gene:

Show All

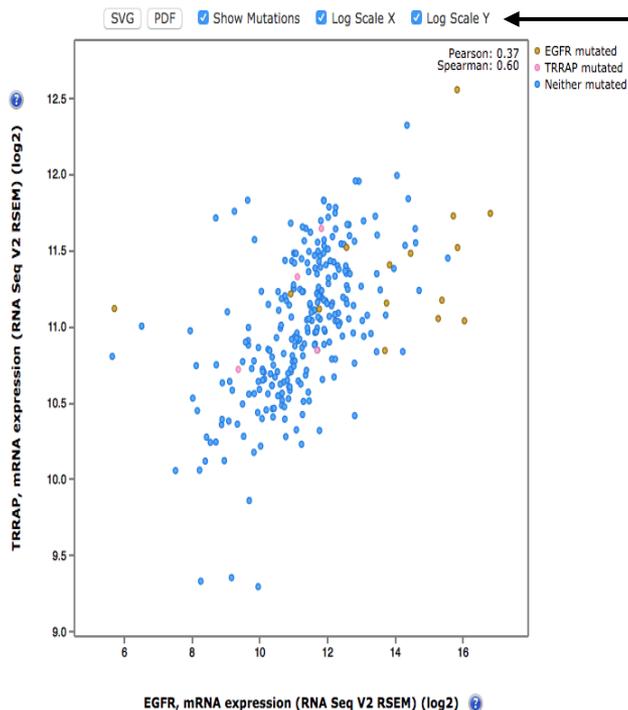
Correlated Gene	Cytoband	Pearson's Correlation	Spearman's Correlation
<b>TRRAP</b>	7q21.2-q22.1	<b>0.37</b>	0.60
ZNF107	7q11.2	<b>0.36</b>	0.60
RBL1	20q11.2	<b>0.37</b>	0.59
BAZ1B	7q11.23	<b>0.40</b>	0.58
ZNF713	7p11.2	<b>0.42</b>	0.57
ILDR2	1q24.1	<b>0.33</b>	0.57
PRKDC	8q11	<b>0.31</b>	0.57
UEVLD	11p15.1	<b>0.30</b>	0.57
XRCC2	7q36.1	<b>0.41</b>	0.56
TNPO1	5q13.2	<b>0.33</b>	0.55
FKBP14	7p14.3	<b>0.30</b>	0.55
TMEM131	2q11.2	<b>0.31</b>	0.54
RAD18	3p25.3	<b>0.40</b>	0.52
SKP2	5p13	<b>0.31</b>	0.51
KIAA1524	3q13.13	<b>0.37</b>	0.50
CENPO	2p23.3	<b>0.35</b>	0.50
NUP205	7q33	<b>0.33</b>	0.50
TRIM24	7q32-q34	<b>0.36</b>	0.49
ZNF558	19p13.2	<b>0.36</b>	0.49
PHF14	7p21.3	<b>0.32</b>	0.48
MCM4	8q11.2	<b>0.38</b>	0.47
CDC23	5q31	<b>0.35</b>	0.47
NCAPG2	7q36.3	<b>0.42</b>	0.46
CHCHD1	10q22.2	<b>-0.34</b>	<b>-0.46</b>
ZNF829	19q13.12	<b>0.31</b>	0.46
SLC25A28	10q24.2	<b>-0.30</b>	<b>-0.46</b>
PRR11	17q22	<b>0.36</b>	0.45
UHRF1	19p13.3	<b>0.34</b>	0.45

1 to 30 of 127

Download Full Results

Click on a gene name to see correlation plot

mRNA co-expression: EGFR vs. TRRAP



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

BRCA1 BRCA2

Show Any Correlation

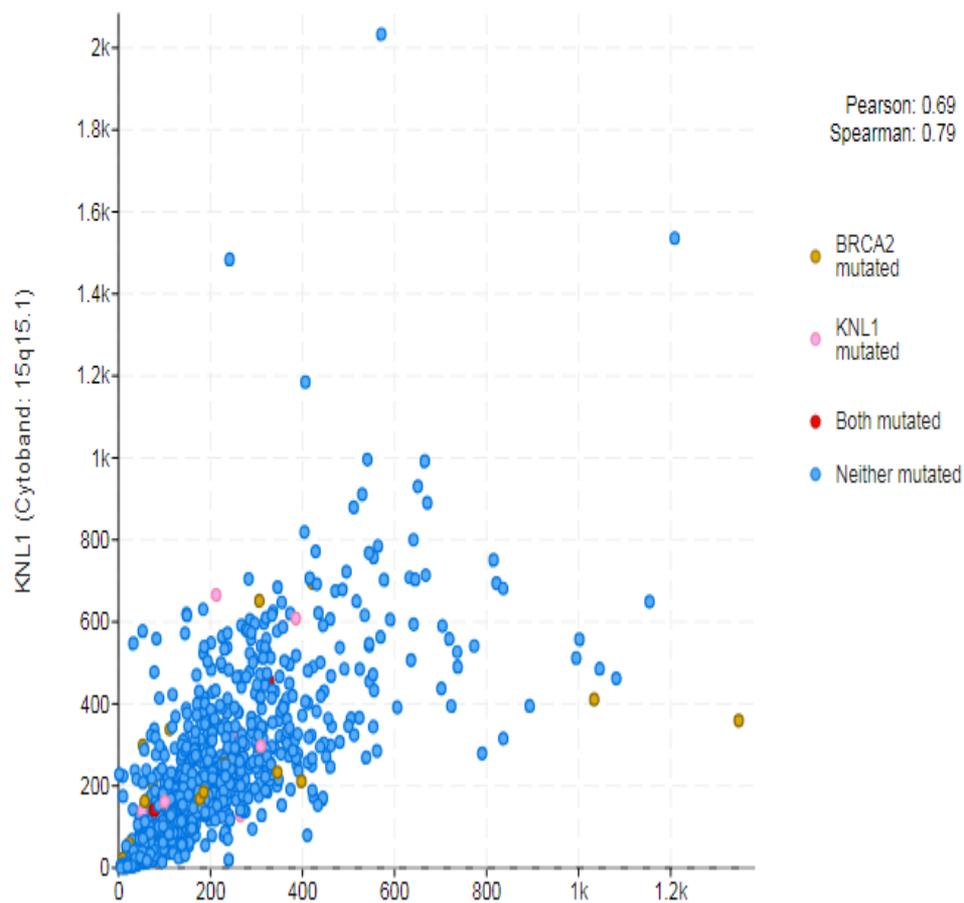
Enter gene or cytoband. Q

Correlated Gene	Cytoband	Pearson's Correlation	Spearman's Correlation
<b>KNL1</b>	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

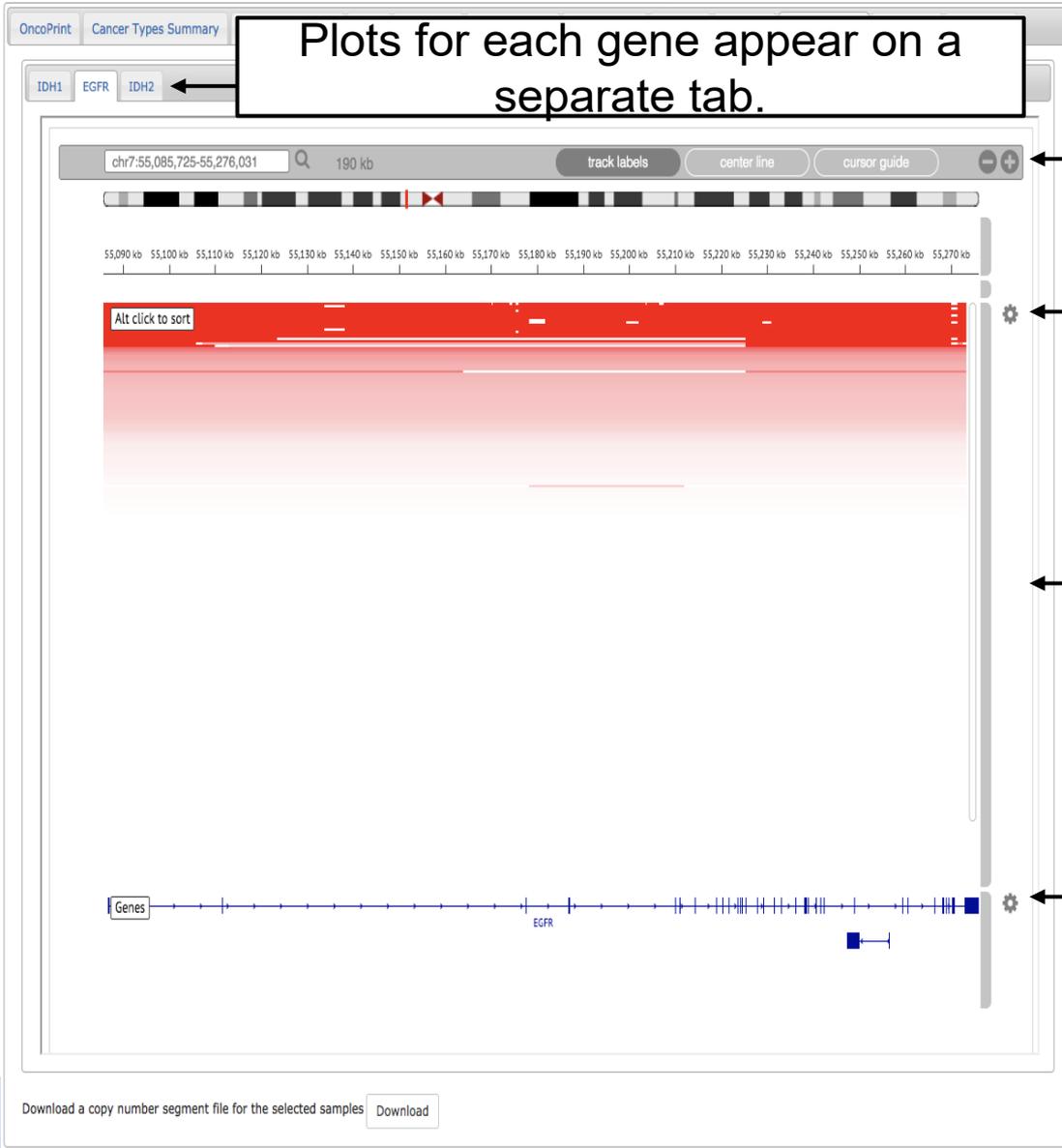
SVG PDF

mRNA Expression Batch Normalized/Merged from Illumina HiSeq\_RNASeqV2 syn4976369: BRCA2 vs



# 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.

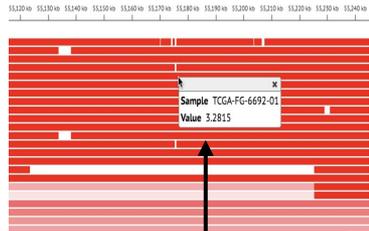
Toggle track labels, a vertical line marking the center of the viewing screen, and a vertical line that moves with your cursor. Use to zoom in or out.

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<sup>52</sup>





# 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. <sup>54</sup>

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

The following are downloadable data files (click to download)

- Mutations: [ Tab-delimited Format ] [ Transposed Matrix ]
- Putative copy-number alterations from GISTIC: [ Tab-delimited Format ] [ Transposed Matrix ]

Click to download data with other genetic profiles ...

Contents below can be copied and pasted into Excel

Frequency of Gene Alteration:

GENE_SYMBOL	NUM_CASES_ALTERED	PERCENT_CASES_ALTERED
IDH1: MUT FUSION AMP HOMDEL;	219	77%
EGFR: MUT FUSION AMP HOMDEL;	26	9%
IDH2: MUT FUSION AMP HOMDEL;	13	5%

Type of Genetic alterations across all cases: (Alterations are summarized as MUT, Gain, HetLoss, etc.)

Case ID	IDH1: MUT FUSION AMP HOMDEL;	EGFR: MUT FUSION AMP HOMDEL;	IDH2: MUT FUSION AMP HOMDEL;
TCGA-CS-4938-01	MUT: R132H;		
TCGA-CS-4941-01	AMP;		
TCGA-CS-4942-01	MUT: R132H;		
TCGA-CS-4943-01	MUT: R132H;		
TCGA-CS-4944-01	MUT: R132H;		
TCGA-CS-5390-01	MUT: R132H;		
TCGA-CS-5393-01	MUT: R132H;	AMP;	
TCGA-CS-5394-01	MUT: R132H;		
TCGA-CS-5395-01			
TCGA-CS-5396-01	MUT: R132H;		
TCGA-CS-5397-01			

Cases affected: (Only cases with an alteration are included)

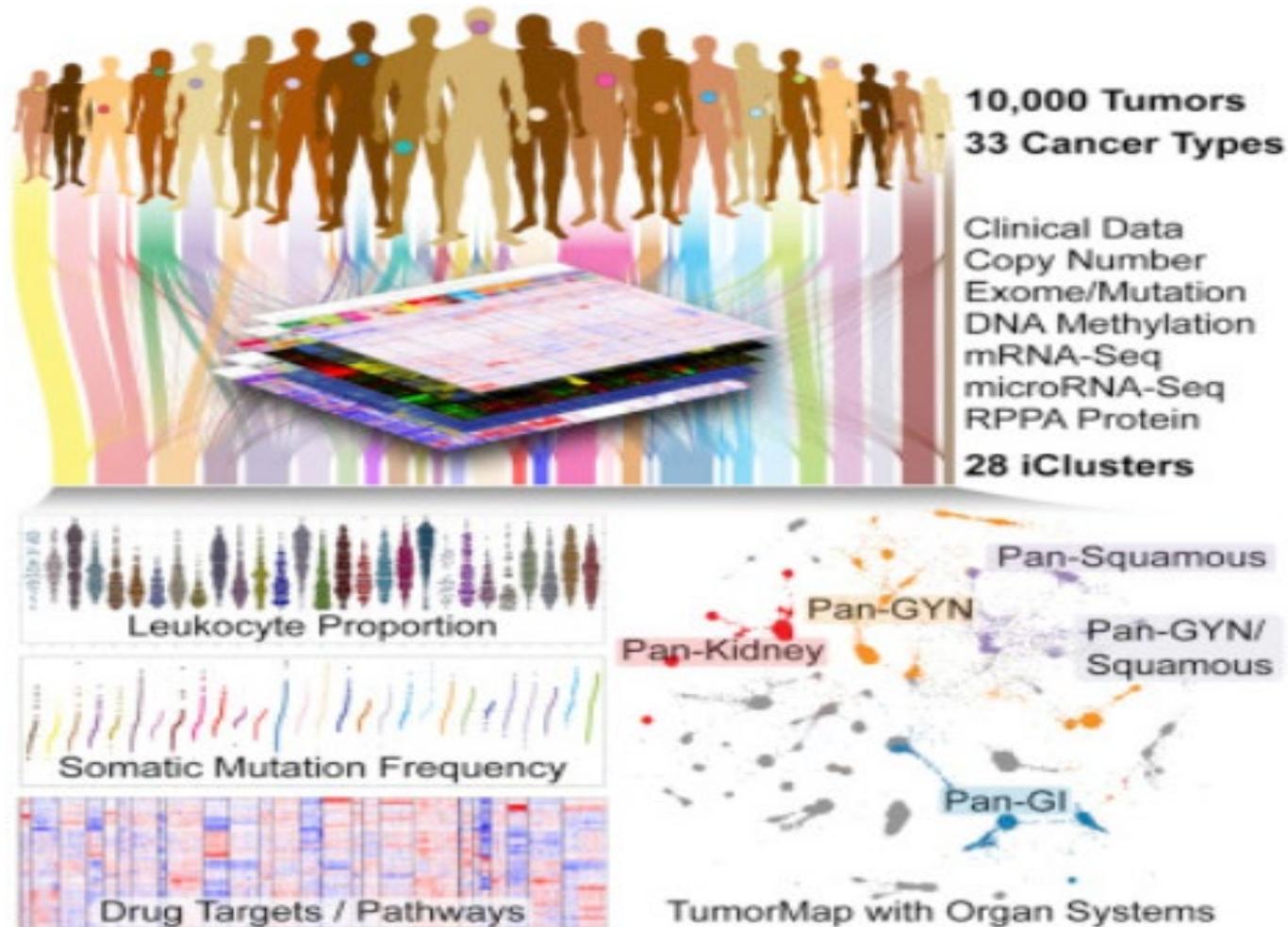
TCGA-CS-4938-01
TCGA-CS-4941-01
TCGA-CS-4942-01
TCGA-CS-4943-01
TCGA-CS-4944-01
TCGA-CS-5390-01
TCGA-CS-5393-01
TCGA-CS-5394-01
TCGA-CS-5396-01
TCGA-CS-6188-01
TCGA-CS-6290-01
TCGA-CS-6665-01

Case matrix: (1= Case harbors alteration in one of the input genes)

TCGA-CS-4938-01	1
TCGA-CS-4941-01	1
TCGA-CS-4942-01	1

# PAN-can example

# PAN-Can datasets and analyses



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

# PAN-Can datasets and analyses



Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data About

Query

Quick Search **Beta!**

Download

Please cite: [Cerami et al., 2012](#) & [Gao et al., 2013](#)

Select Studies for Visualization & Analysis:

0 studies selected (0 samples)

Search...

PanCancer Studies	6
Pediatric Cancer Studies	13
Cell lines	3
Adrenal Gland	3
Ampulla of Vater	1
Biliary Tract	9
Bladder/Urinary Tract	15
Bone	2
Bowel	10
Breast	16

Quick select:

TCGA PanCancer Atlas Studies

Curated set of non-redundant studies

## PanCancer Studies

- MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017) 10945 samples
- Metastatic Solid Cancers (UMich, Nature 2017) 500 samples
- MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018) 249 samples
- SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018) 141 samples
- TMB and Immunotherapy (MSKCC, Nat Genet 2019) 1661 samples
- Tumors with TRK fusions (MSK, 2019) 106 samples

## Pediatric Cancer Studies

- Pediatric Preclinical Testing Consortium (Maris, 2019) 261 samples
- Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018) 1978 samples
- Pediatric Rhabdoid Tumor (TARGET, 2018) 72 samples
- Pediatric Wilms' Tumor (TARGET, 2018) 657 samples
- Pediatric Acute Myeloid Leukemia (TARGET, 2018) 1025 samples
- Pediatric Neuroblastoma (TARGET, 2018) 1089 samples

0 studies selected (0 samples)

Query By Gene

OR

Explore Selected Studies

# PAN-Can datasets and analyses



Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data About

Query

Quick Search **Beta!**

Download

Please cite: Cerami et al., 2012 & Gao et al., 2013

Selected Studies:

Acute Myeloid Leukemia (TCGA, PanCancer Atlas) Adrenocortical Carcinoma (TCGA, PanCancer Atlas) Bladder Urothelial Carcinoma (TCGA, PanCancer Atlas)  
Brain Lower Grade Glioma (TCGA, PanCancer Atlas) [and 28 more \(10967 total samples\)](#)

Select Molecular Profiles:

Mutation  Copy number alterations

Select Patient/Case Set:

To build your own case set,  
try out our enhanced Study View.

All (10967)

Enter Genes:

User-defined List

Hint: Learn Onco Query Language (OQL)  
to write more powerful queries [↗](#)

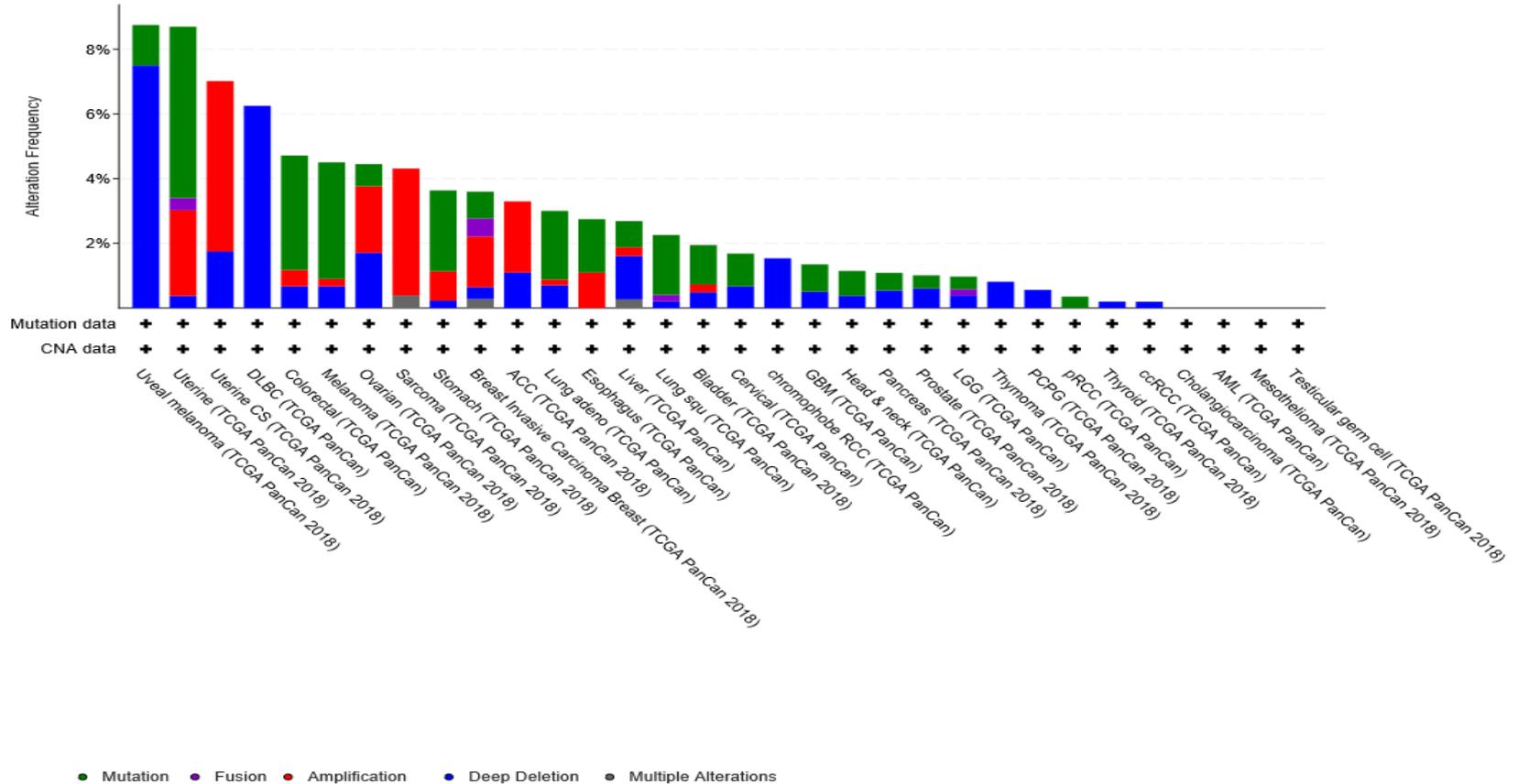
ESR|

All gene symbols are valid.

Submit Query

# PAN-Can datasets and analyses

32 of 32 categories (Cancer Study) are shown based on filtering.

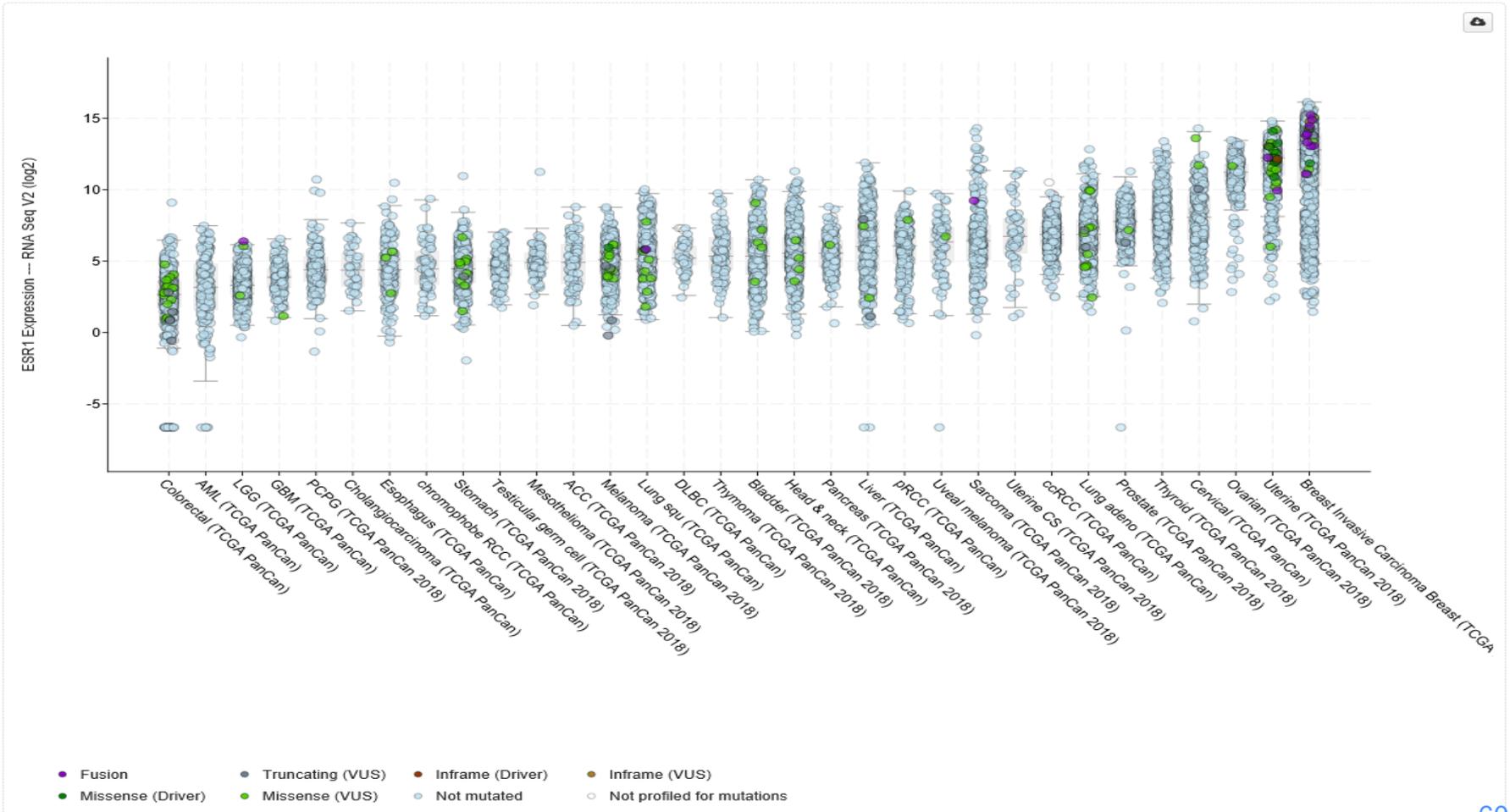


# PAN-Can datasets and analyses

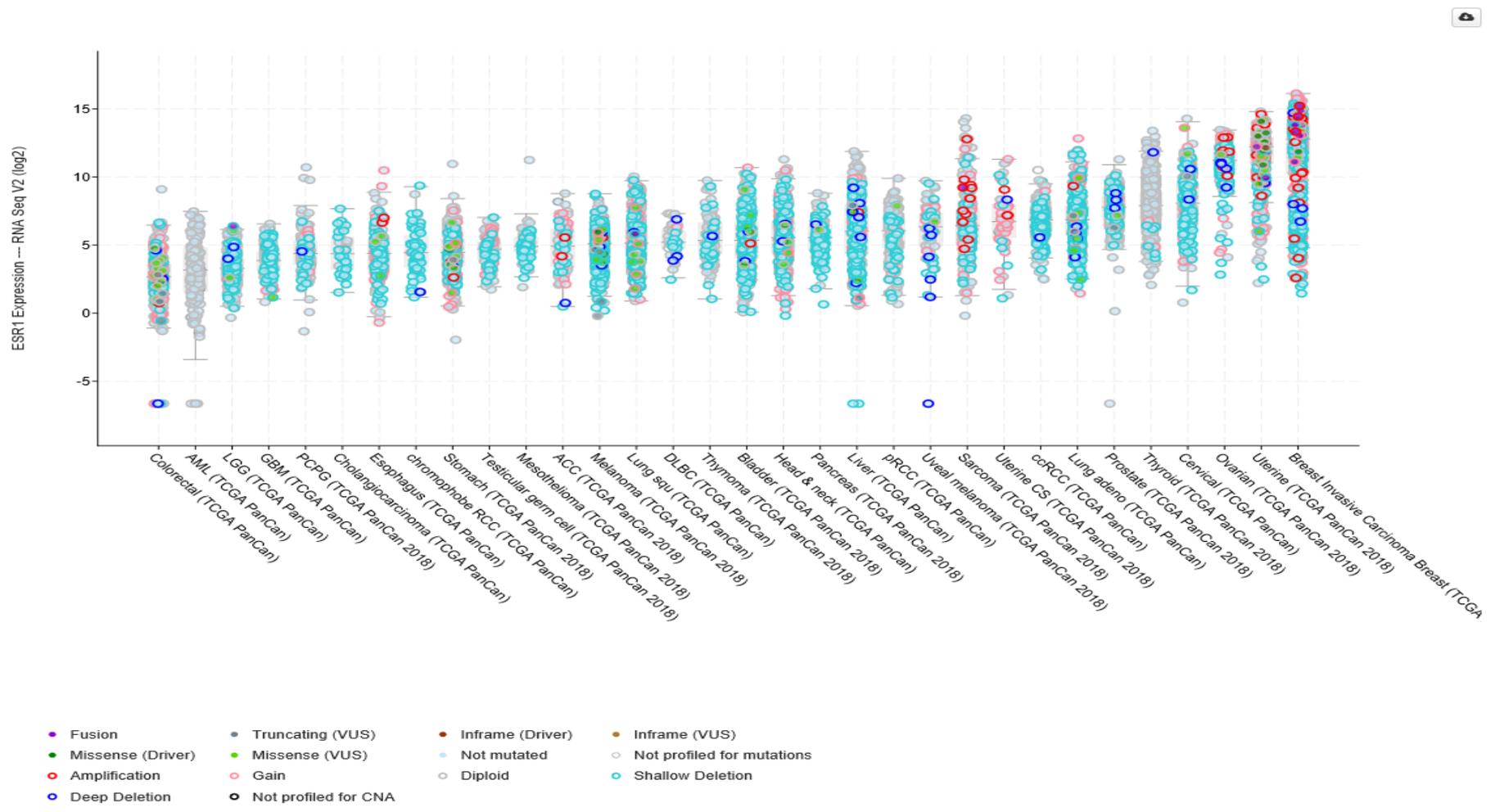
ESR1

Profile: RNA Seq V2 Sort By: Median  Log scale  Show mutations \*  Show copy number alterations

Select studies: TCGA Pan-Can Atlas (32) Custom list



# PAN-Can datasets and analyses



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

# Clustering example

 **cBioPortal**  
FOR CANCER GENOMICS

Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data About

Query Quick Search **Beta!** Download

---

Selected Studies:  Metastatic Prostate Adenocarcinoma (SU2C/PCF Dream Team, PNAS 2019) (444 total samples)

---

Select Genomic Profiles:

- Mutations 
- Putative copy-number alterations 
- mRNA Expression. Select one of the profiles below:
  - mRNA expression Z-scores relative to diploid samples (FPKM capture) 
  - mRNA expression Z-scores relative to diploid samples (FPKM polyA) 
  - mRNA expression z-scores relative to all samples (log FPKM capture) 
  - mRNA expression z-scores relative to all samples (log FPKM polyA) 

Enter a z-score threshold  $\pm$

---

Select Patient/Case Set:  
To build your own case set,  
try out our enhanced Study View.

All Tumors (444)

---

Enter Genes:

Prostate Cancer: AR Signaling (10 genes)

**Hint:** Learn Onco Query Language (OQL) to write more powerful queries 

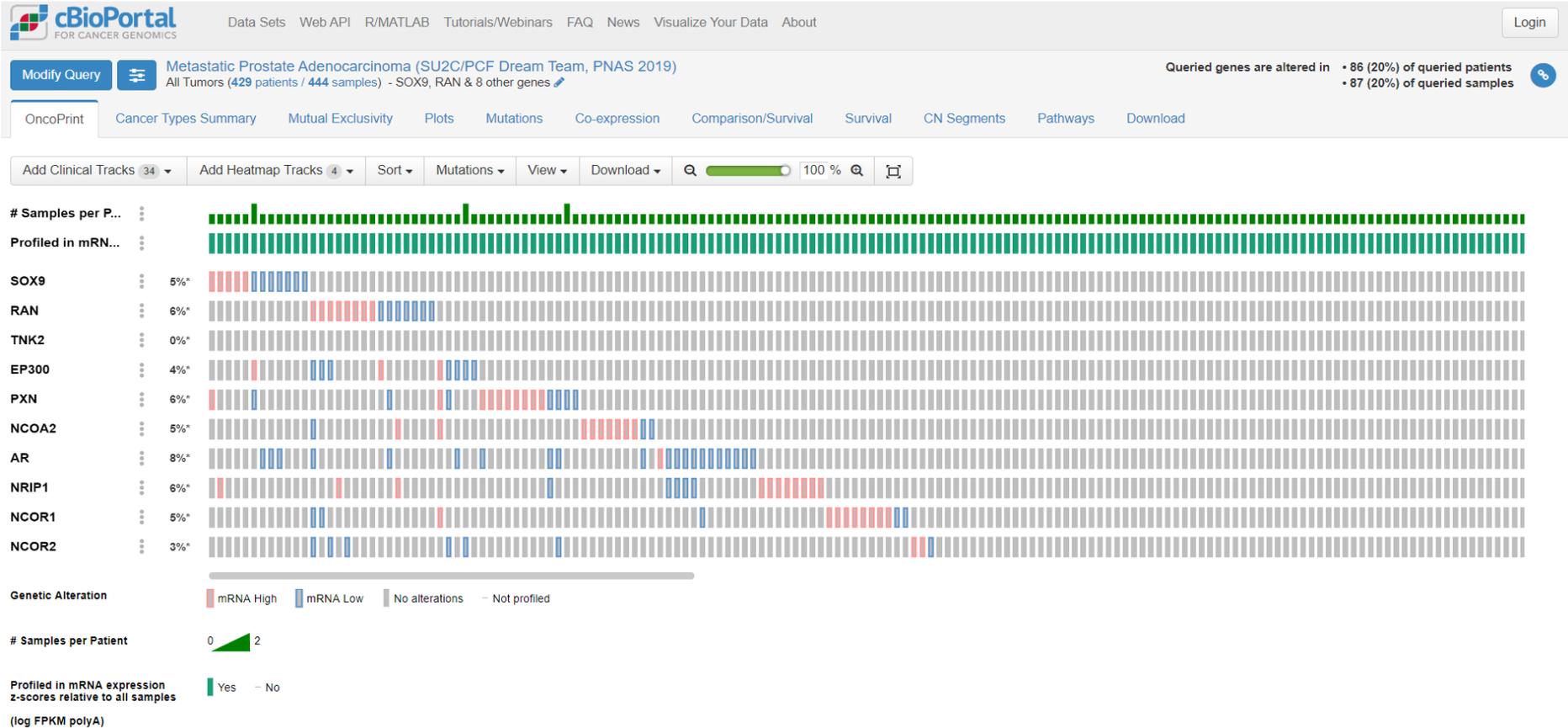
SOX9 RAN TNK2 EP300 PXN NCOA2 AR NRIP1 NCOR1 NCOR2



All gene symbols are valid.

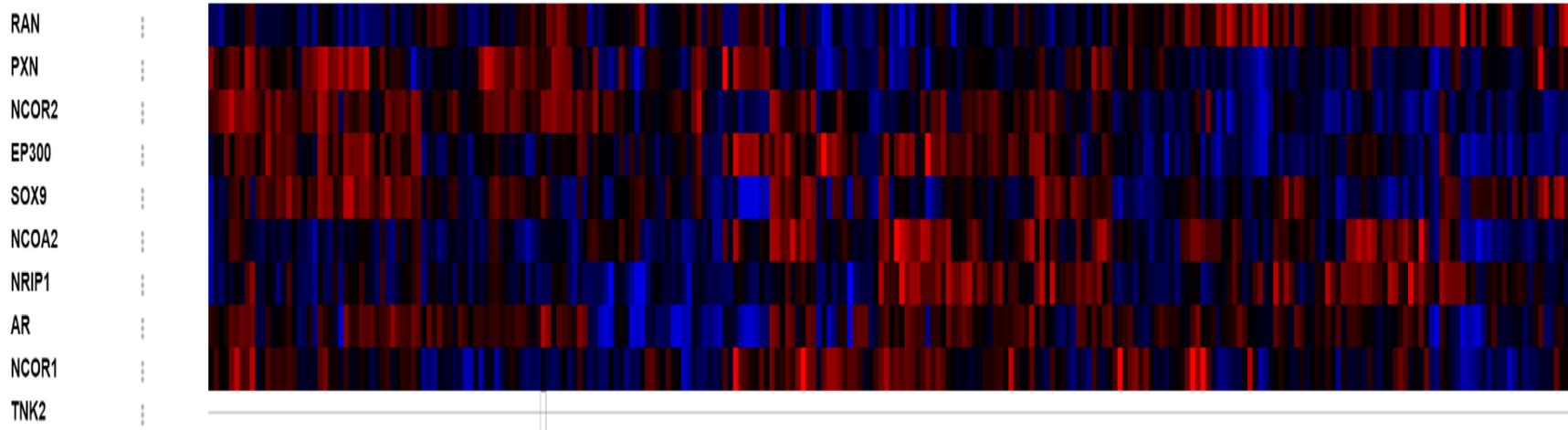
---

# Clustering example



# Clustering example

mRNA expression z-scores relative to all samples (log FPKM polyA) :



Genetic Alteration mRNA High mRNA Low No alterations Not profiled

# Samples per Patient 0 2

Profiled in mRNA expression z-scores relative to all samples Yes No

(log FPKM polyA)

Expression Heatmap -3 3 No data

# Ovarian – multi-omics example

Ovary/Fallopian Tube

Ovarian Epithelial Tumor

→ **SEROUS OVARIAN CANCER**

- Ovarian Serous Cystadenocarcinoma (TCGA, Firehose Legacy)
- Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011)
- Ovarian Serous Cystadenocarcinoma (TCGA, PanCancer Atlas)

→ **SMALL CELL CARCINOMA OF THE OVARY**

- Small Cell Carcinoma of the Ovary (MSKCC, Nat Genet 2014)

---

Select Genomic Profiles:

- Mutations ⓘ
- Putative copy-number alterations from GISTIC ⓘ
- mRNA Expression. Select one of the profiles below:
  - mRNA Expression z-Scores (U133 microarray only) ⓘ
  - mRNA Expression z-Scores (microarray) ⓘ
  - mRNA Expression z-Scores (RNA Seq V2 RSEM) ⓘ
- Protein/phosphoprotein level. Select one of the profiles below:
  - Protein expression Z-scores (RPPA) ⓘ
  - Protein level Z-scores (mass spectrometry by CPTAC) ⓘ

---

Select Patient/Case Set:

To build your own case set, try out our enhanced Study View.

Samples with mutation and CNA data (311) × ▾

---

Enter Genes:

Hint: Learn Onco Query Language (OQL) to write more powerful queries [↗](#)

User-defined List × ▾

PIK3CA  
TP53

All gene symbols are valid.

Submit Query

# Ovarian – CNV (GISTIC) -mRNA

Modify Query + **Ovarian Serous Cystadenocarcinoma (TCGA, Firehose Legacy)** Queried g  
Samples with mutation and CNA data (311 patients/samples) - PIK3CA & TP53

OncoPrint Cancer Types Summary Mutual Exclusivity **Plots** Mutations Co-expression Comparison CN Segments Pathways Download

Examples: Mut# vs FGA mRNA vs mut type **mRNA vs CNA** Protein vs mRNA

**Horizontal Axis**

Data Type: Copy Number

Copy Number Profile: Putative copy-number alteration..

Gene: PIK3CA

↑ Swap Axes ↓

**Vertical Axis**

Data Type: mRNA

mRNA Profile: mRNA Expression z-Scores (RN..)

Gene: Same gene (PIK3CA)

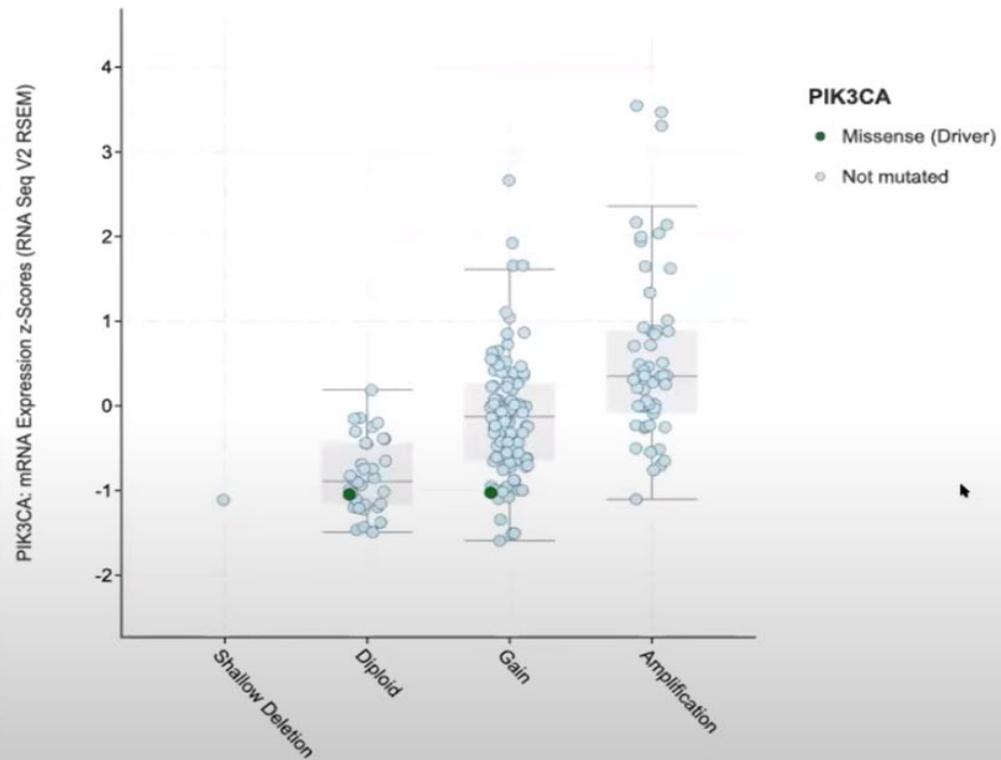
Search Case(s)

Case ID..

Search Mutation(s)

Protein Change..

Color samples by: Gene: PIK3CA  Mutation Type \*  Copy Number Alteration  Both



PIK3CA: Putative copy-number alterations from GISTIC

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

Modify Query ☰ **Ovarian Serous Cystadenocarcinoma (TCGA, Firehose Legacy)**  
 Samples with mutation and CNA data (311 patients/samples) - PIK3CA & TP53 [↗](#)
Queried genes are altered in 284 (91%) of queried patients/s

OncoPrint Cancer Types Summary Mutual Exclusivity **Plots** Mutations Co-expression Comparison CN Segments Pathways Download

**Examples:** Mut# vs FGA mRNA vs mut type **mRNA vs CNA** Protein vs mRNA

**Horizontal Axis**

Data Type: Copy Number

Copy Number Profile: Relative linear copy-number val... **i**

Gene: TP53

↑ Swap Axes ↓

**Vertical Axis**

Data Type: mRNA

mRNA Profile: mRNA Expression z-Scores (RN... **i**

Gene: Same gene (TP53)

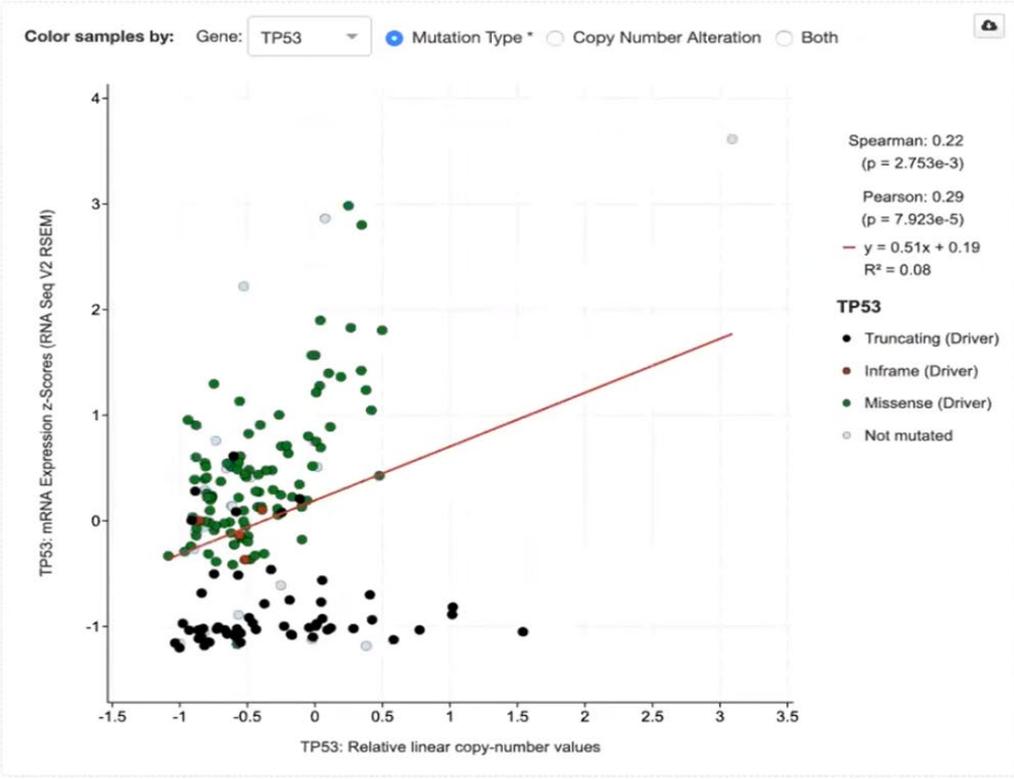
**Search Case(s)**

Case ID..

**Search Mutation(s)**

Protein Change..

Show Regression Line



\* Driver annotation settings are located in the settings menu ☰ at the top of the page.

# Ovarian – TP53 mRNA, CNV and mutations

Modify Query Queried genes

Ovarian Serous Cystadenocarcinoma (TCGA, Firehose Legacy)  
 Samples with mutation and CNA data (311 patients/samples) - PIK3CA & TP53

OncoPrint Cancer Types Summary Mutual Exclusivity **Plots** Mutations Co-expression Comparison CN Segments Pathways Download

Examples: Mut# vs FGA mRNA vs mut type **mRNA vs CNA** Protein vs mRNA

**Horizontal Axis**

Data Type: Copy Number

Copy Number Profile: Putative copy-number alteration..

Gene: TP53

↑ Swap Axes ↓

---

**Vertical Axis**

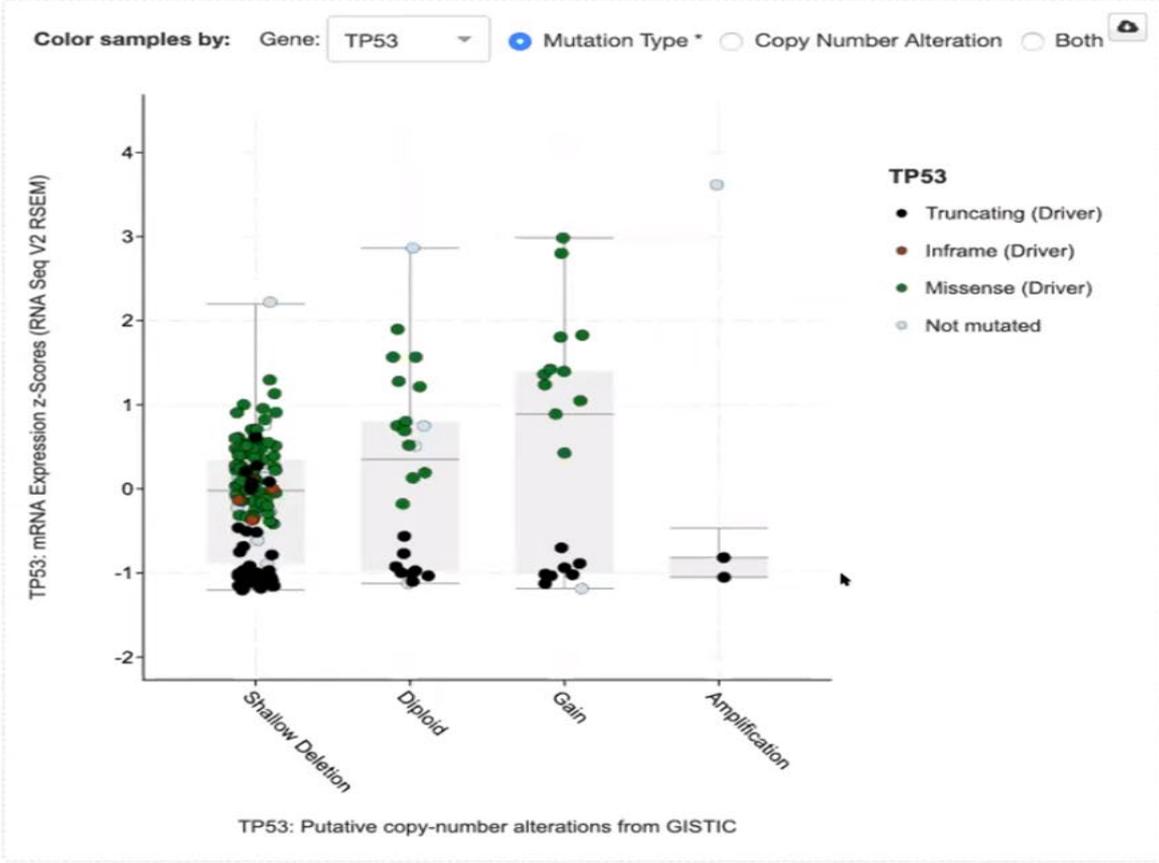
Data Type: mRNA

mRNA Profile: mRNA Expression z-Scores (RN..

Gene: Same gene (TP53)

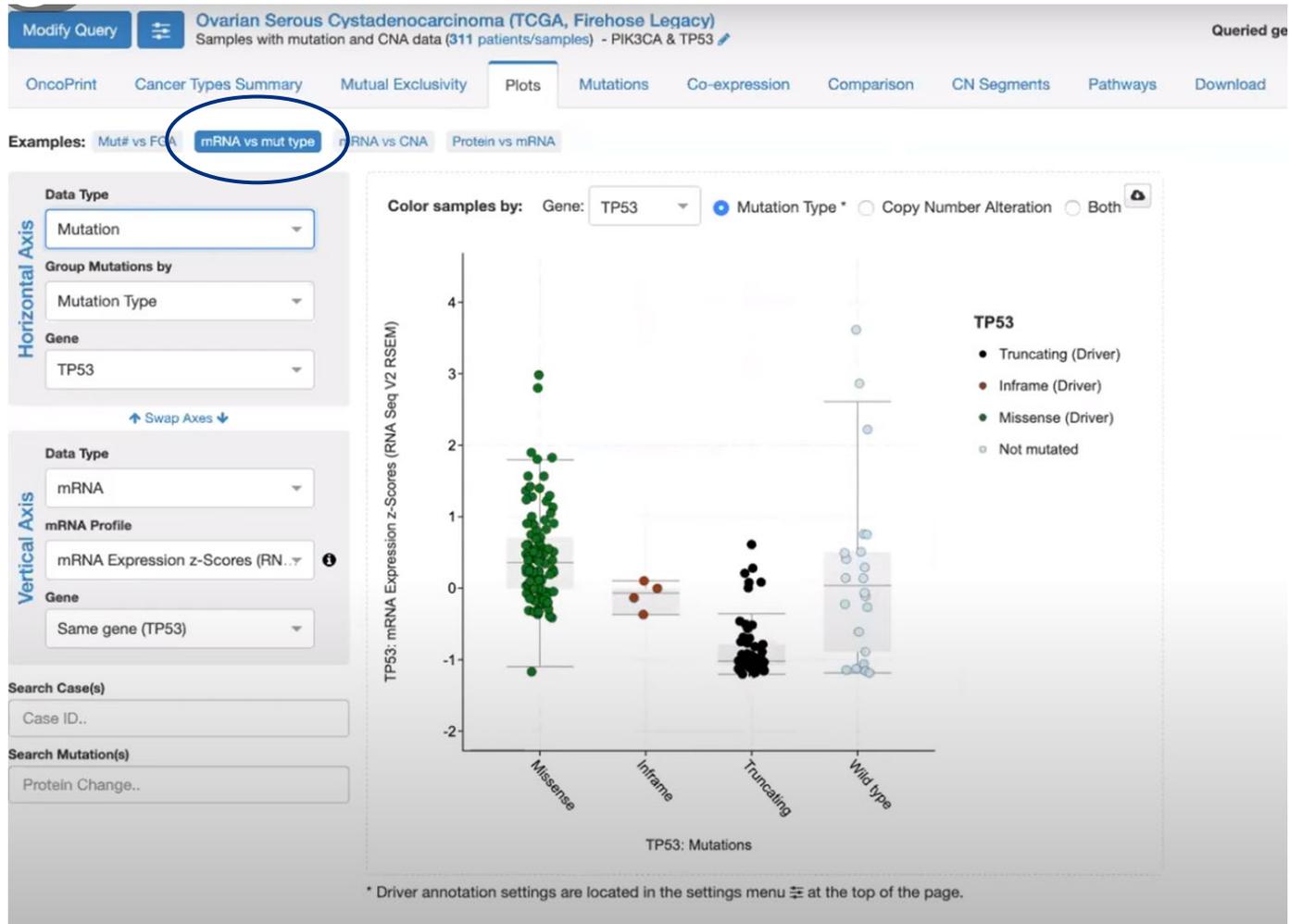
Search Case(s): Case ID..

Search Mutation(s): Protein Change..

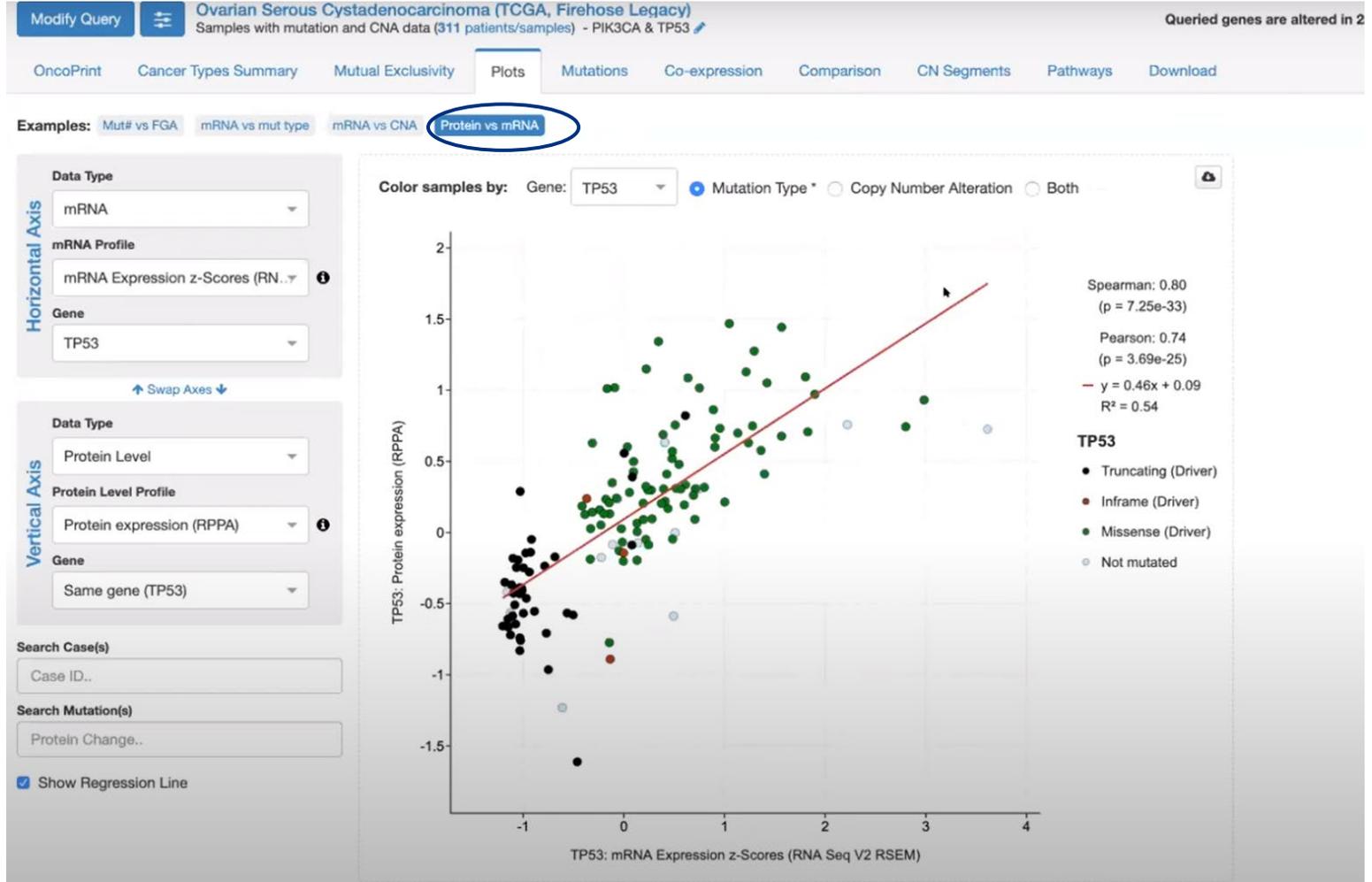


\* Driver annotation settings are located in the settings menu ☰ at the top of the page.

# Ovarian – mRNA-mutation



# Ovarian – mRNA-protein



# Breast Cancer (TP53 + and TP53 - analysis)

 **cBioPortal**  
FOR CANCER GENOMICS

Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data About

Query [Quick Search Beta!](#) [Download](#)

---

Selected Studies: [Modify](#) Breast Invasive Carcinoma (TCGA, PanCancer Atlas) (1084 total samples)

---

Select Genomic Profiles:

- Mutations [?](#)
- Putative copy-number alterations from GISTIC [?](#)
- mRNA Expression. Select one of the profiles below:
  - mRNA expression z-scores relative to diploid samples (RNA Seq V2 RSEM) [?](#)
  - mRNA expression z-scores relative to all samples (log RNA Seq V2 RSEM) [?](#)

Enter a z-score threshold  $\pm$

---

Select Patient/Case Set:  [x](#) [v](#)  
[To build your own case set,](#)  
[try out our enhanced Study View.](#)

---

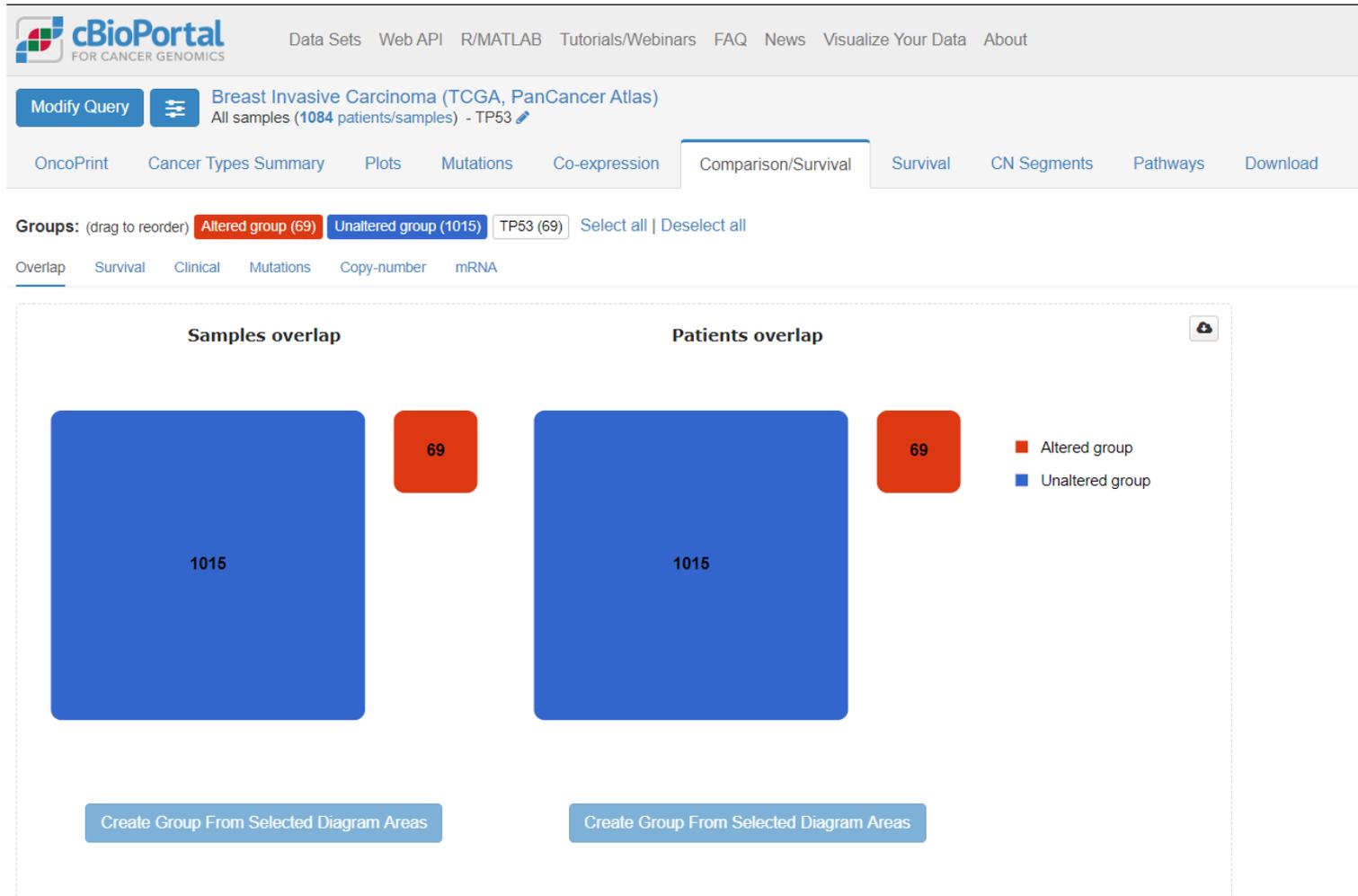
Enter Genes:  [x](#) [v](#)

[Hint: Learn Onco Query Language \(OQL\)](#)  
[to write more powerful queries](#) [?](#)

All gene symbols are valid.

[Submit Query](#)

# Breast Cancer (TP53+/TP53-) analysis



# Breast Cancer (TP53 high and TP53 low)



Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data About

Login

Modify Query

Breast Invasive Carcinoma (TCGA, PanCancer Atlas)  
All samples (1084 patients/samples) - TP53

Queried gene is altered in 285 (26%) of queried patients/samples

OncoPrint

Cancer Types Summary

Mutual Exclusivity

Plots

Mutations

Co-expression

Comparison/Survival

Survival

CN Segments

Pathways

Download

✓ The results below reflect the OQL specification from your query.

Add Clinical Tracks 95

Add Heatmap Tracks 2

Sort

Mutations

View

Download

Q



55 %

Q



Profiled in mRN...



TP53: EXP>1

12%\*



TP53: EXP<-1

14%\*



Genetic Alteration

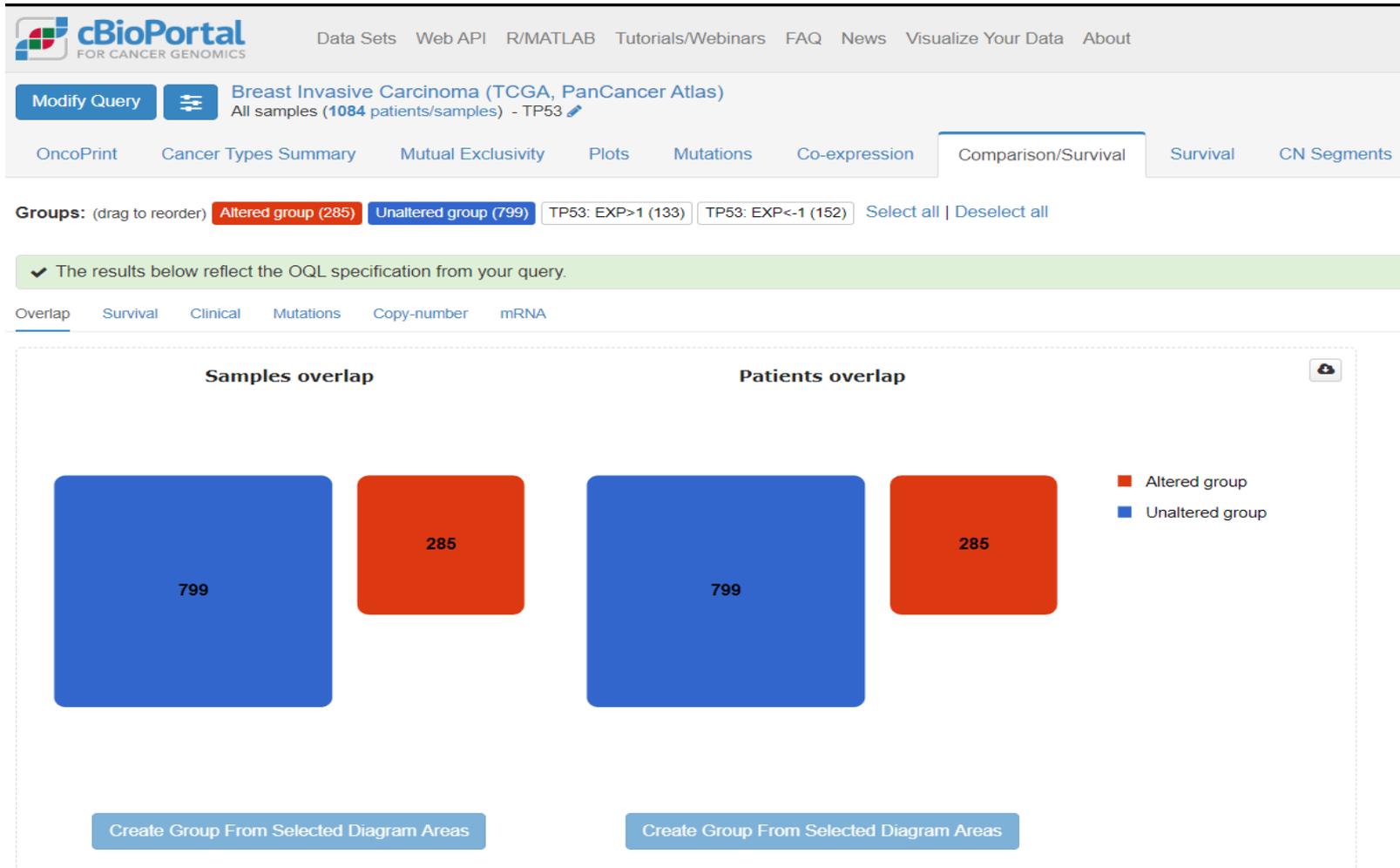
mRNA High mRNA Low No alterations Not profiled

Profiled in mRNA expression  
z-scores relative to all samples

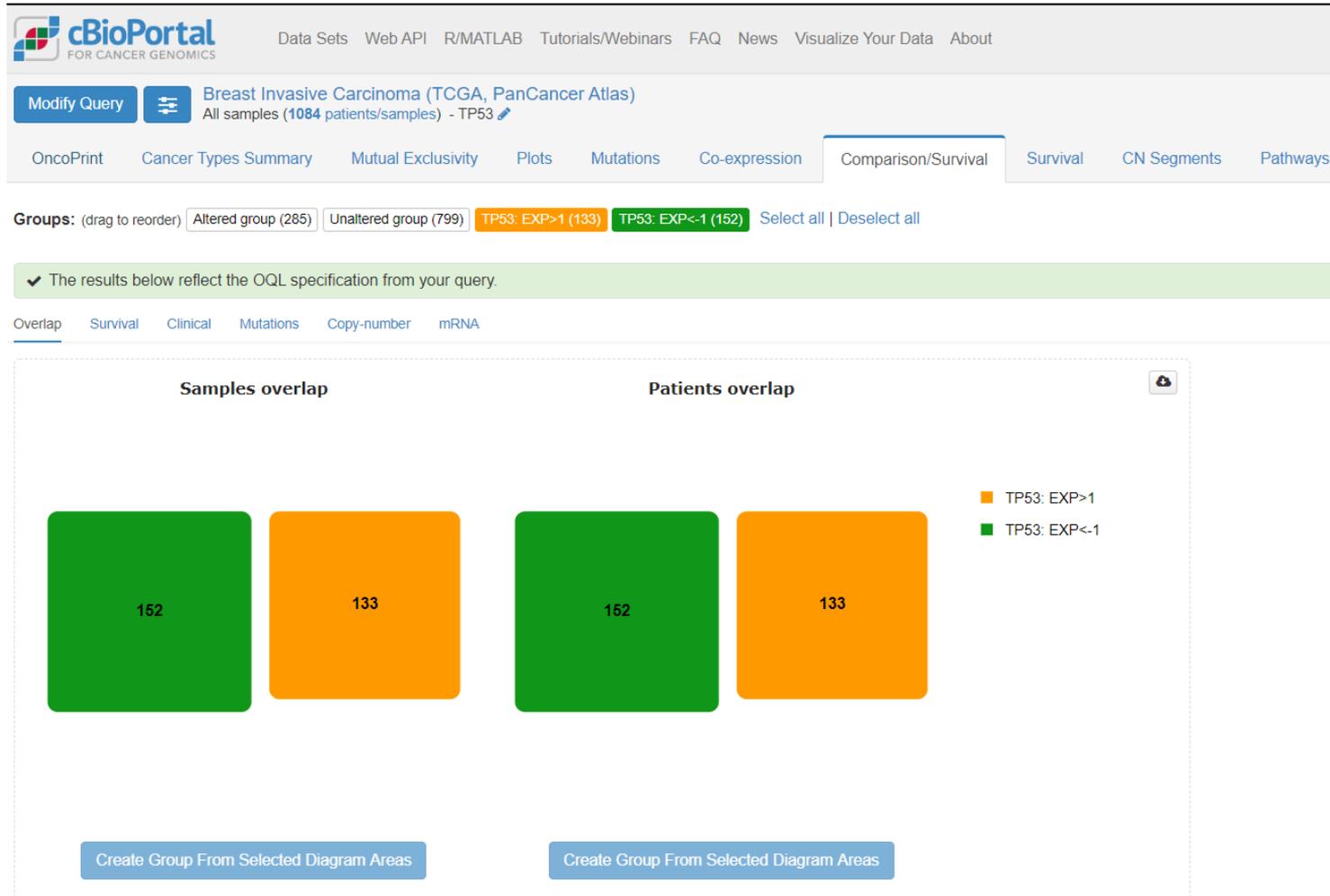
Yes No

(log RNA Seq V2 RSEM)

# Breast Cancer (TP53 high and TP53 low)



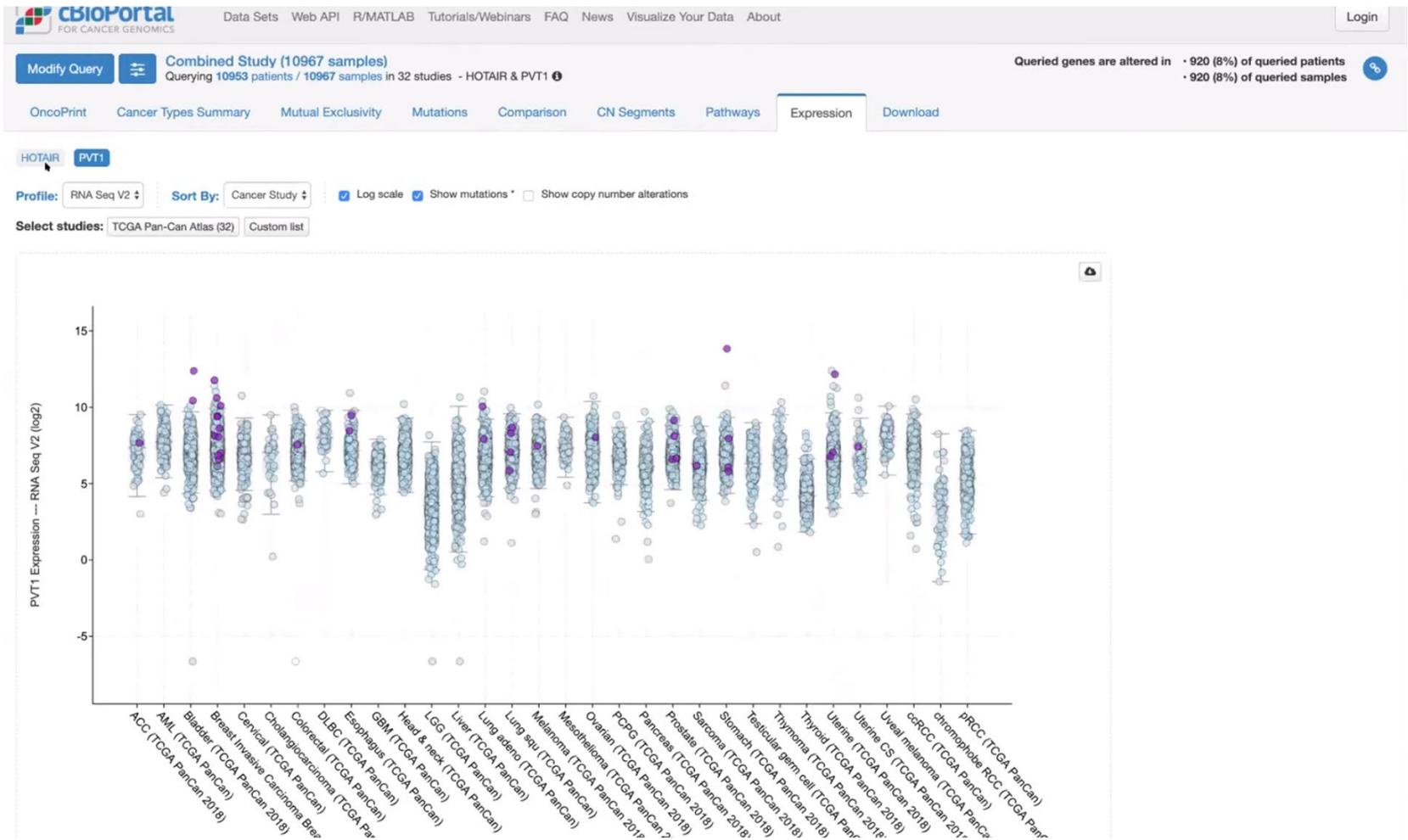
# Breast Cancer (TP53 high and TP53 low)



# Long-noncoding RNA example



# Long-noncoding RNA example



# miRNA example

**cbioPortal** Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data About

cbioPortal Webinar 3: Expression Data Analysis

Watch later Share

## Datasets

The table below lists the number of available samples per cancer study and data type.

Columns ▾

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

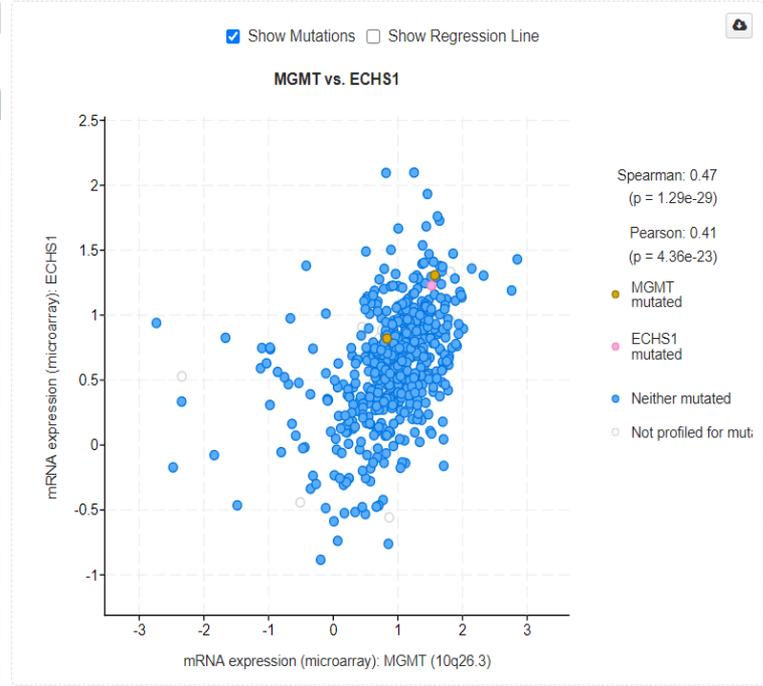
OncoPrint | Cancer Types Summary | Plots | Mutations | **Co-expression** | Comparison/Survival | Survival | CN Segments | Pathways | Download

MGMT

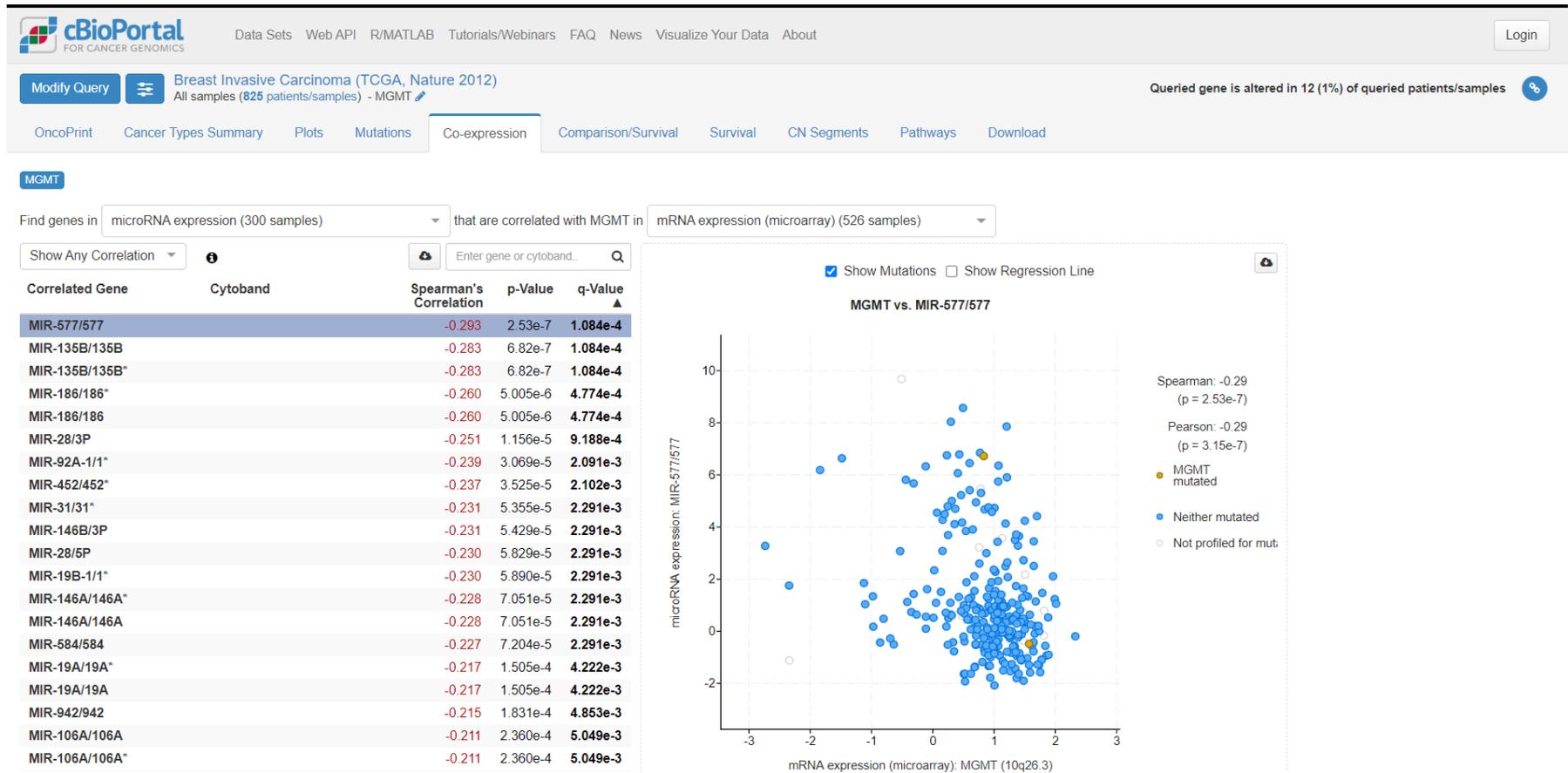
Find genes in mRNA expression (microarray) (526 samples) that are correlated with MGMT in mRNA expression (microarray) (526 samples)

Show Any Correlation | Enter gene or cytoband..

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



# RNA and miRNA example



# RNA and miRNA example

**cBioPortal** FOR CANCER GENOMICS | Data Sets | Web API | R/MATLAB | Tutorials/Webinars | FAQ | News | Visualize Your Data | About | [Login](#)

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

MGMT  
MIR-577/577

All gene symbols are valid.

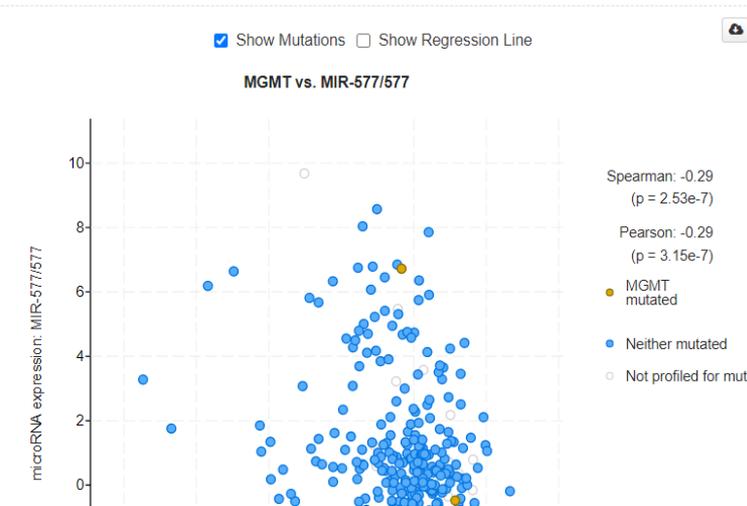
OncoPrint | Cancer Types Summary | Plots | Mutations | **Co-expression** | Comparison/Survival | Survival | CN Segments | Pathways | Download

MGMT

Find genes in **microRNA expression (300 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
MIR-577/577		-0.293	2.53e-7	1.084e-4
MIR-135B/135B		-0.283	6.82e-7	1.084e-4
MIR-135B/135B*		-0.283	6.82e-7	1.084e-4
MIR-186/186*		-0.260	5.005e-6	4.774e-4
MIR-186/186		-0.260	5.005e-6	4.774e-4
MIR-28/3P		-0.251	1.156e-5	9.188e-4
MIR-92A-1/1*		-0.239	3.069e-5	2.091e-3
MIR-452/452*		-0.237	3.525e-5	2.102e-3
MIR-31/31*		-0.231	5.355e-5	2.291e-3
MIR-146B/3P		-0.231	5.429e-5	2.291e-3
MIR-28/5P		-0.230	5.829e-5	2.291e-3
MIR-19B-1/1*		-0.230	5.890e-5	2.291e-3
MIR-146A/146A*		-0.228	7.051e-5	2.291e-3
MIR-146A/146A		-0.228	7.051e-5	2.291e-3
MIR-584/584		-0.227	7.204e-5	2.291e-3



# miRNA correlation with mRNA example

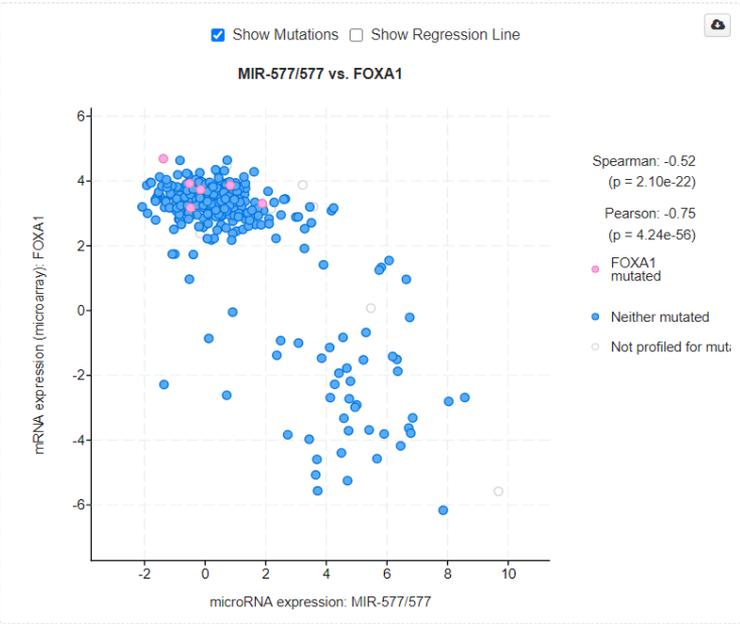
MGMT **MIR-577/577**

Find genes in  that are correlated with MIR-577/577 in

Show Any Correlation   Enter gene or cytoband.

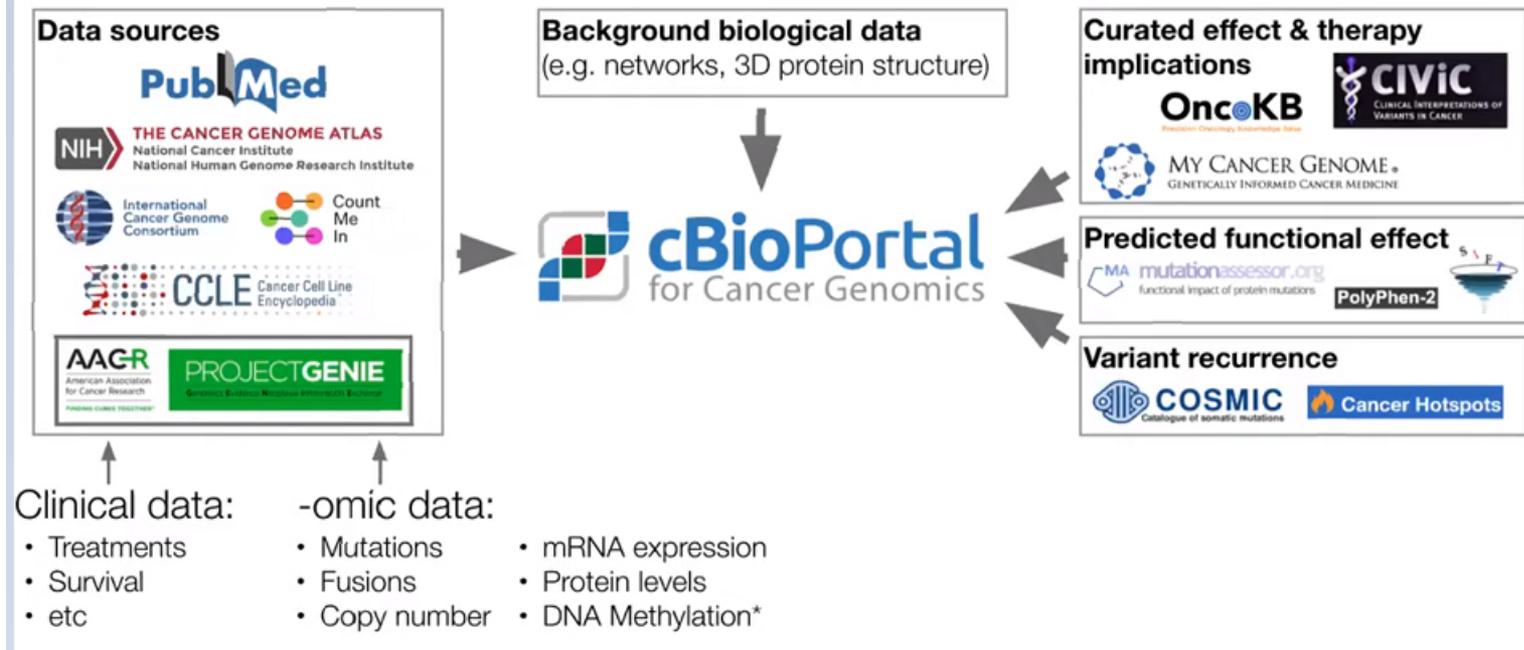
Correlated Gene	Cytoband	Spearman's Correlation	p-Value	q-Value ▲
<b>FOXA1</b>	14q21.1	<b>-0.523</b>	2.10e-22	<b>1.90e-18</b>
UGT8	4q26	0.523	2.21e-22	<b>1.90e-18</b>
BCL11A	2p16.1	0.515	1.13e-21	<b>6.47e-18</b>
GATA3	10p14	<b>-0.511</b>	2.74e-21	<b>1.18e-17</b>
CDC20	1p34.2	0.510	3.62e-21	<b>1.25e-17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

Showing 1-25 of 17237



# Cbioportal.org

What data is in cBioPortal?



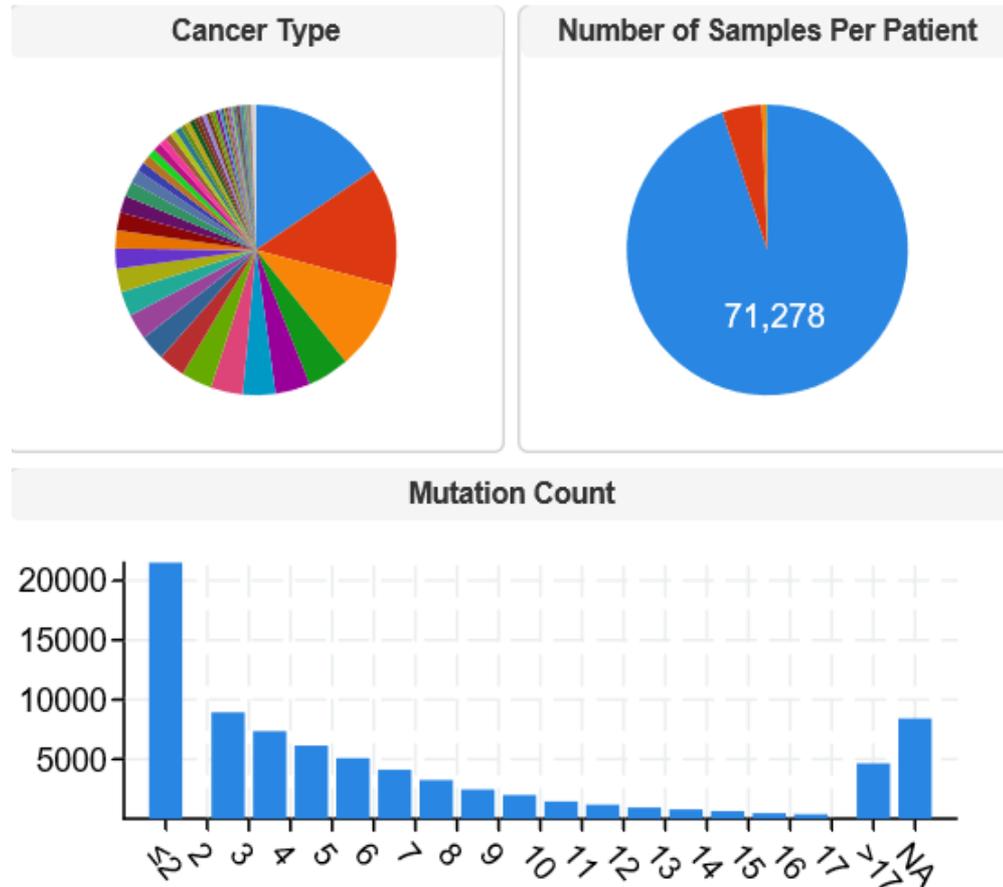
Slide borrowed from cbioportal website

# 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 cbiportal

- <https://genie.cbiportal.org/>
- It consists of more 79K samples



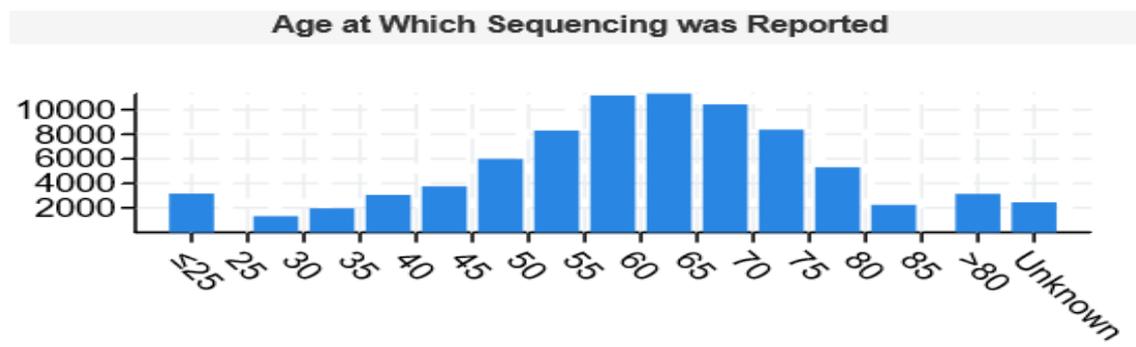
# Search cancer types

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

Search...

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

Search...



MAYO  
CLINIC



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 Sample Info

Sample Counts by Tissues  
e/sGenes vs Sample Size

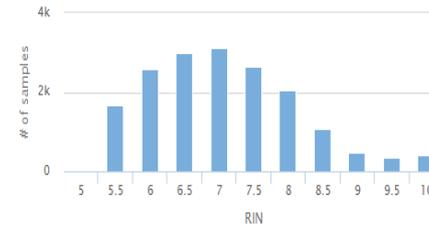
### V8 Donor Info

Sex, Race & Age  
Cause of Death  
Tissue Counts Per Donor  
Age By Tissues

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

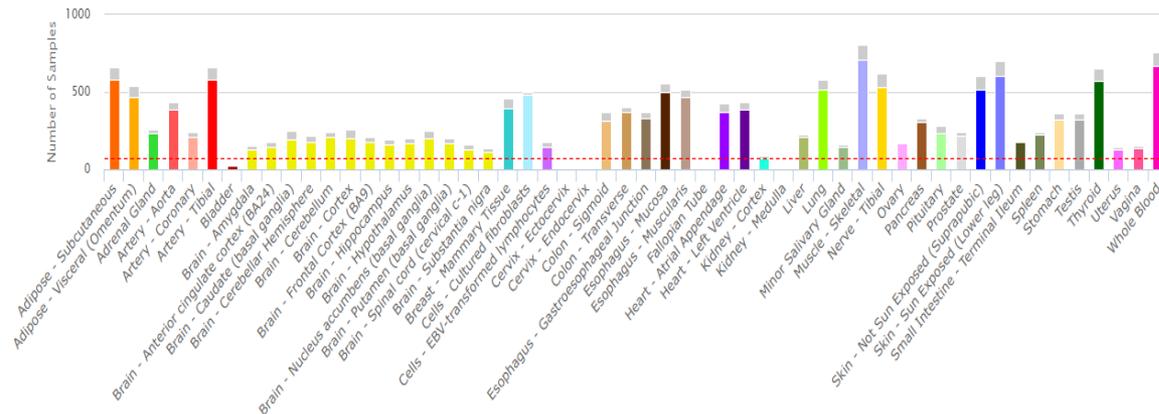


### V8 Sample Counts by Tissues

Sort tissues by:

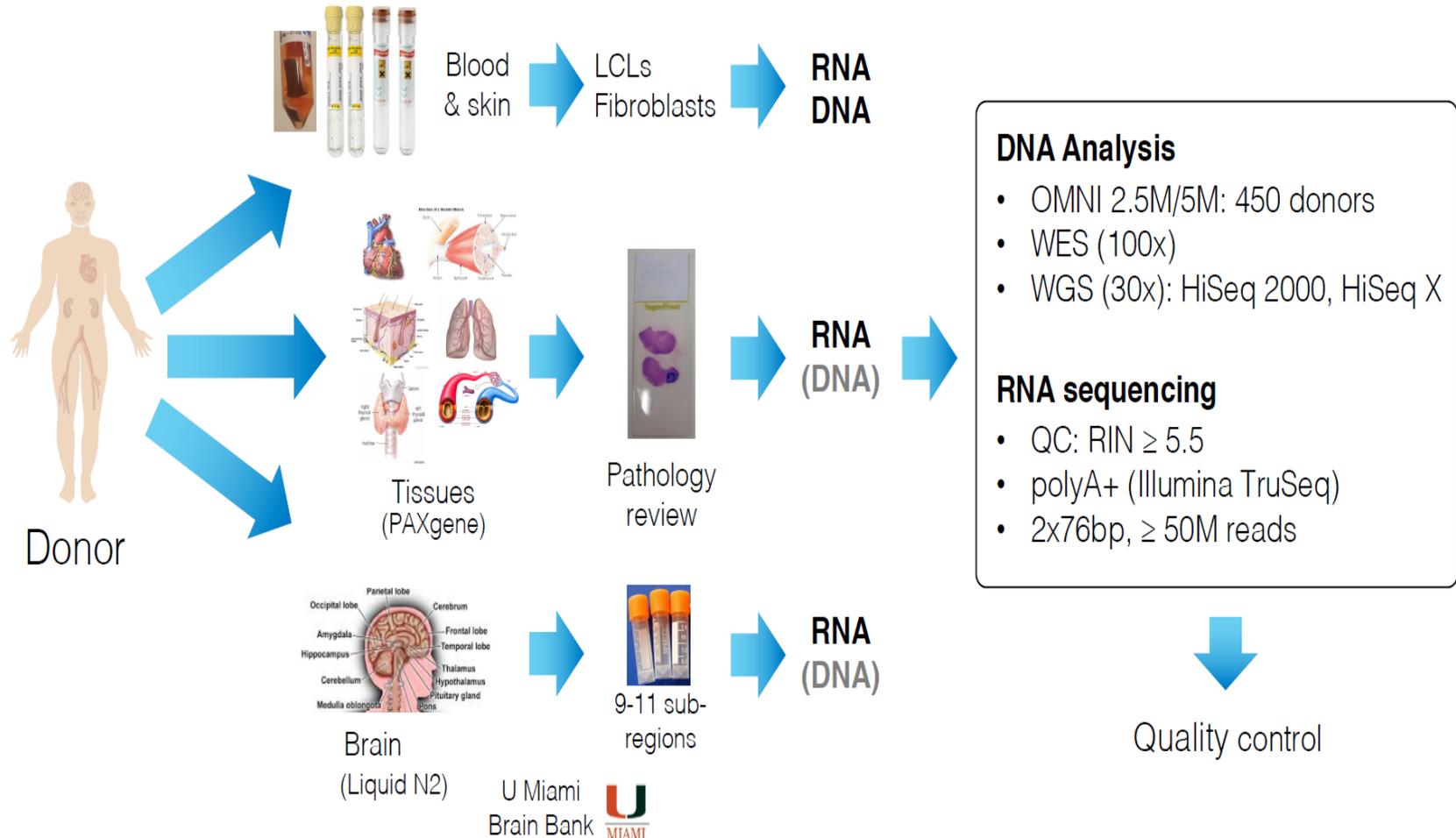
[Download](#)

View as:



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](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	<b>GRCh37</b>	<b>GRCh38</b>	GRCh38
GENCODE annotation	v19	<b>v19</b>	<b>v26</b>	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/GTEEx\\_ASHG\\_workshop\\_101817\\_final.pdf](https://storage.googleapis.com/ashg-workshop-files/GTEEx_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](https://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](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

Plot:  Gene  Isoform

Differentiation:  None  Sex

Scale:  Log  Linear

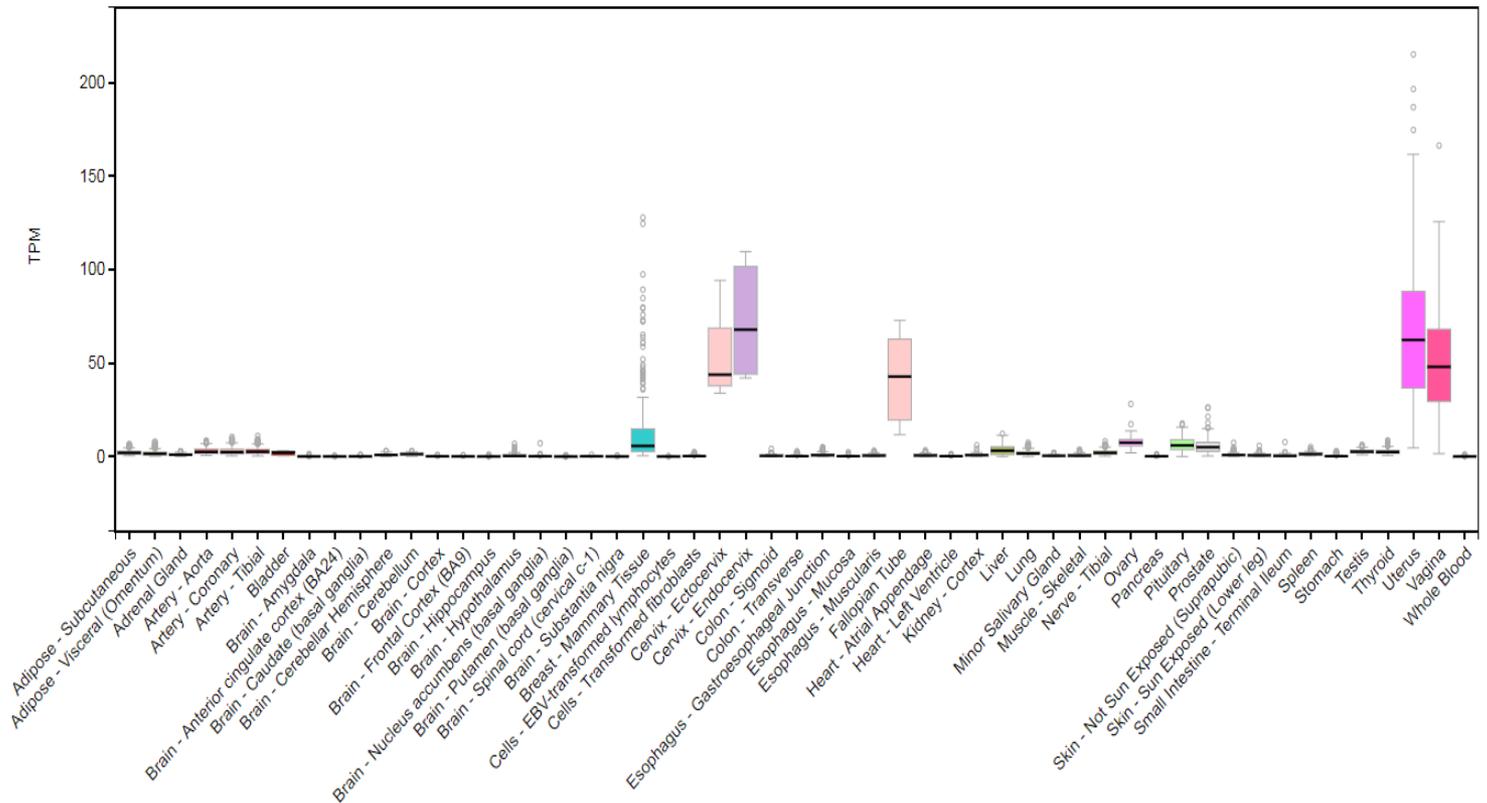
Sort:  ABC  ▲  ▼

Outliers:  On  Off

Display:  Boxes  Medians

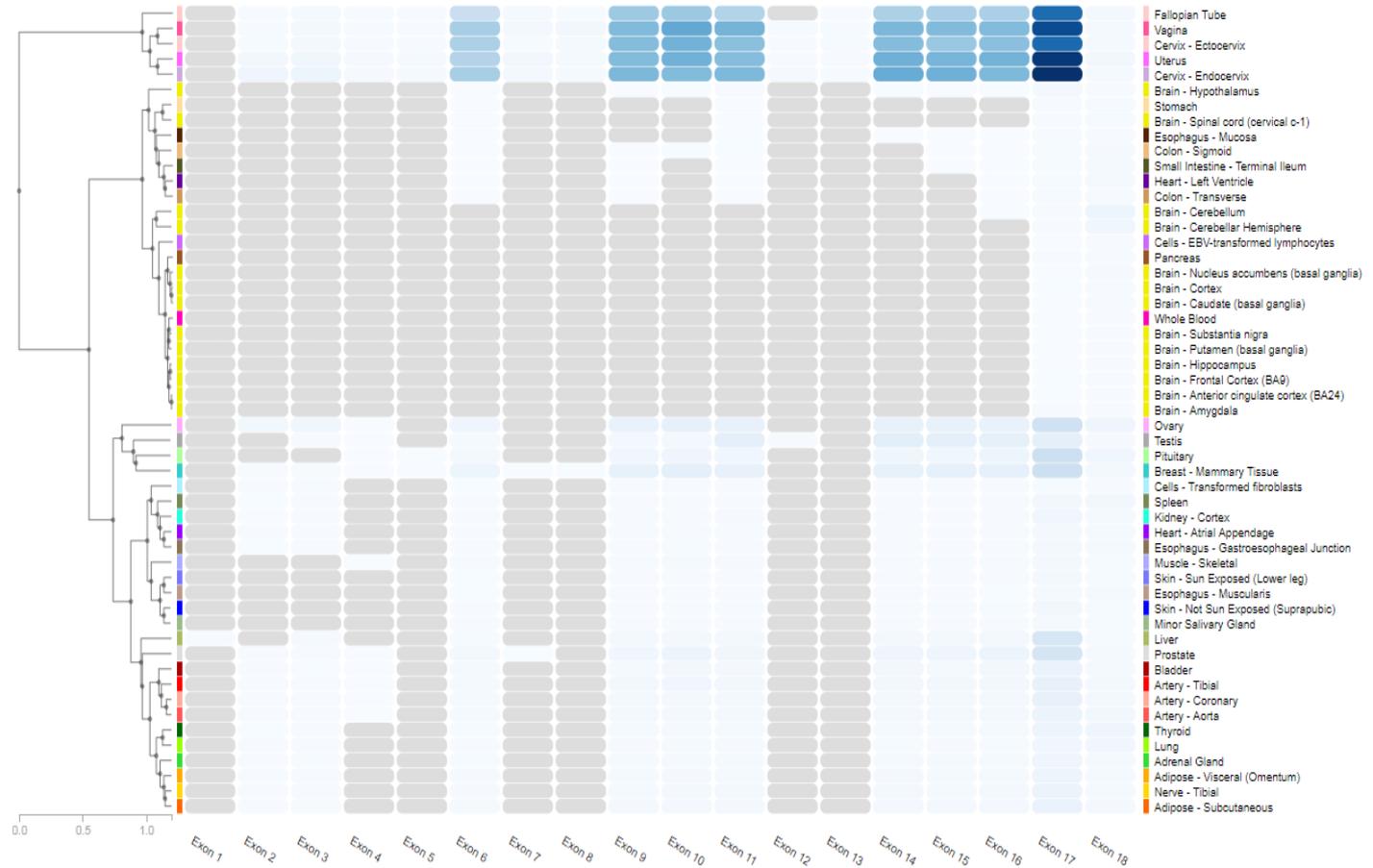
Filter:

ENSG00000091831.17 Gene Expression



# 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



# WebQTL software



## Select and Search

**Species:**

**Group:**  [Info](#)

**Type:**

**Data Set:**  [Info](#)

Databases marked with \*\* suffix are not public yet.  
 Access requires [user login](#).

**Get Any:**

Enter terms, genes, ID numbers in the **Get Any** field.  
 Use \* or ? wildcards (Cyp\*a?, synap\*).  
 Use **Combined** for terms such as *tyrosine kinase*.

**Combined:**

[Search](#) [Make Default](#) [Advanced Search](#)

## Websites Affiliated with GeneNetwork

- GeneNetwork **Time Machine**: Full versions from 2009 to 2016 (mm9)
- UTHSC Genome Browser **Classic** and **Newest**
- UTHSC **Galaxy** Service
- UTHSC **Bayesian Network Web Server**
- GeneNetwork Classic on **Amazon Cloud**
- GeneNetwork Classic Code on **GitHub**
- GeneNetwork 2.0 Development Code on **GitHub**
- **GeneNetwork 2.0** Development

## Getting Started

1. Select **Species** (or select All)
2. Select **Group** (a specific sample)
3. Select **Type** of data:
  - Phenotype (traits)
  - Genotype (markers)
  - Expression (mRNAs)
4. Select a **Database**
5. Enter search terms in the **Get Any** or **Combined** field: words, genes, ID numbers, probes, advanced search commands
6. Click on the **Search** button
7. Optional: Use the **Make Default** button to save your preferences

GeneNetwork Intro Help Tools Collections **0** Source Code Sign in

Genes / Molecules Search All

**Search Results:** We searched GTExv8 Human Breast Mammary Tissue RNA-Seq (Feb20) TPM log2 to find all records that match:

**"esr1"**

1 records are shown below.

[Correlations](#) [Networks](#) [WebGestalt](#) [GeneWeaver](#) [BNW](#) [WGCNA](#) [CTL Maps](#) [MultiMap](#) [Comparison Bar Chart](#)

Select  Add     Deselect

Show/Hide Columns:

Index  Record  Symbol  Description  Location  Mean  Max LRS?  Max LRS Location  Additive Effect?

Showing 1 to 1 of 1 entries

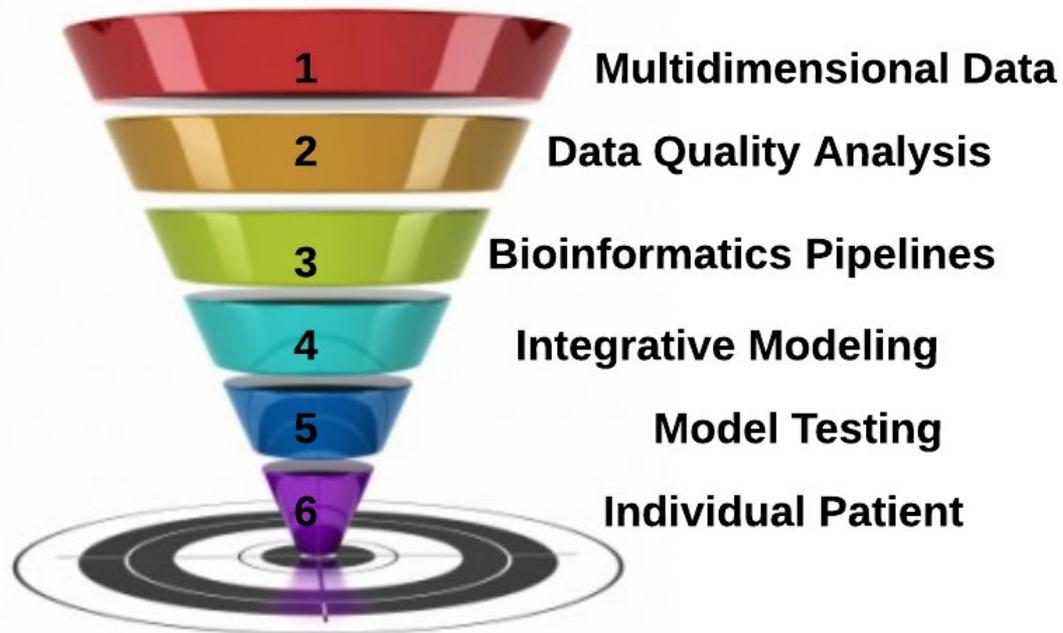
<input type="checkbox"/>	Index	Record	Symbol	Description	Location	Mean	Max LRS?	Max LRS Location	Additive Effect?
<input checked="" type="checkbox"/>	1	ENSG00000091831	ESR1	estrogen receptor 1	Chr6: 151.656691	3.078	N/A	N/A	N/A

Showing 1 to 1 of 1 entries



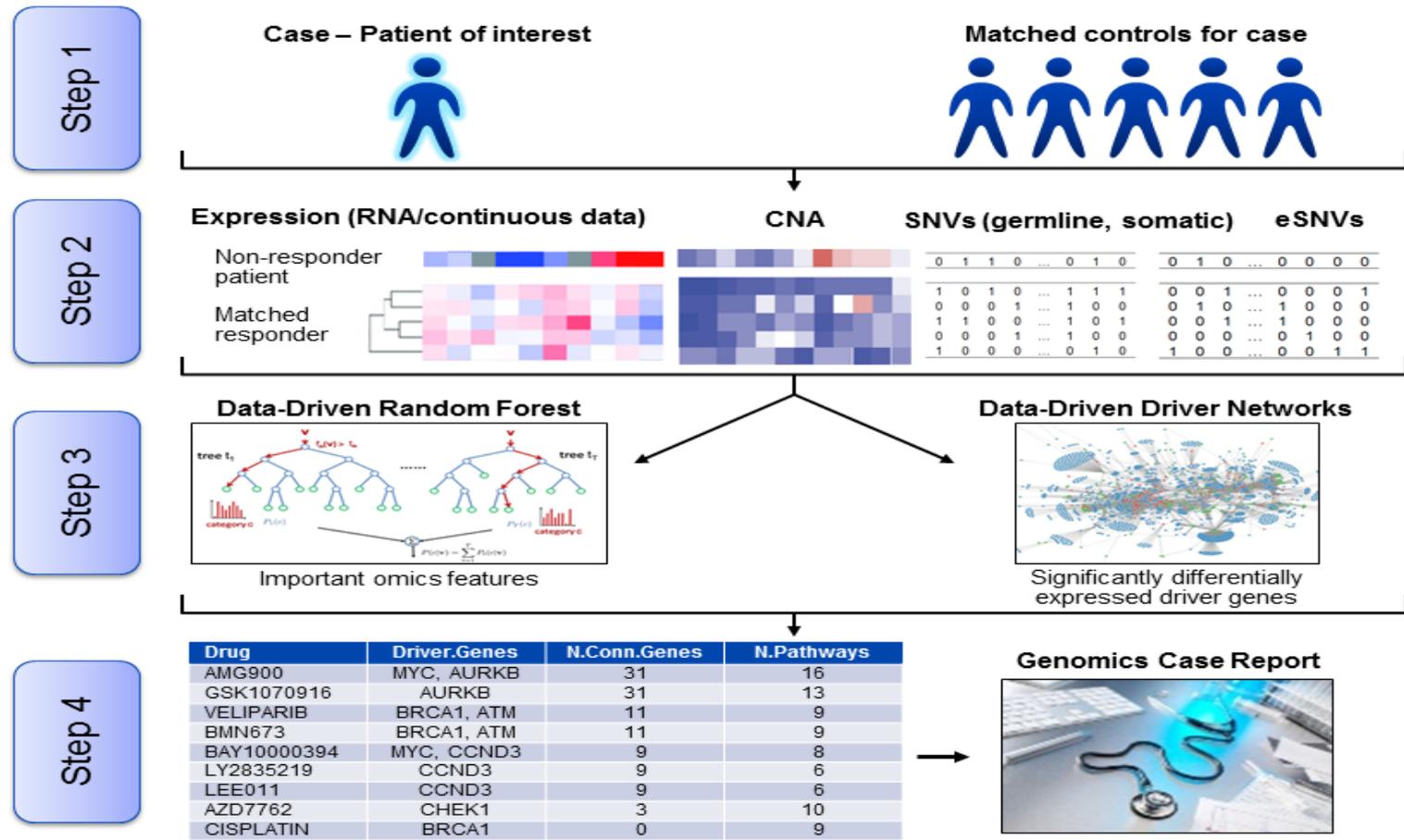
# Precision medicine for cancer patients using clinical and molecular data

# Multi-dimensional data to individual patient



# PANOPLY – Precision cancer genomics report: single sample inventory

## Patient-Specific Multi-Omics Integration Analysis



# Integration of multi-omics data for precision medicine

- **PANOPLY- Precision Cancer Genomic Report: Single Sample Inventory**
- [PANOPLY: Omics-Guided Drug Prioritization Method](#)

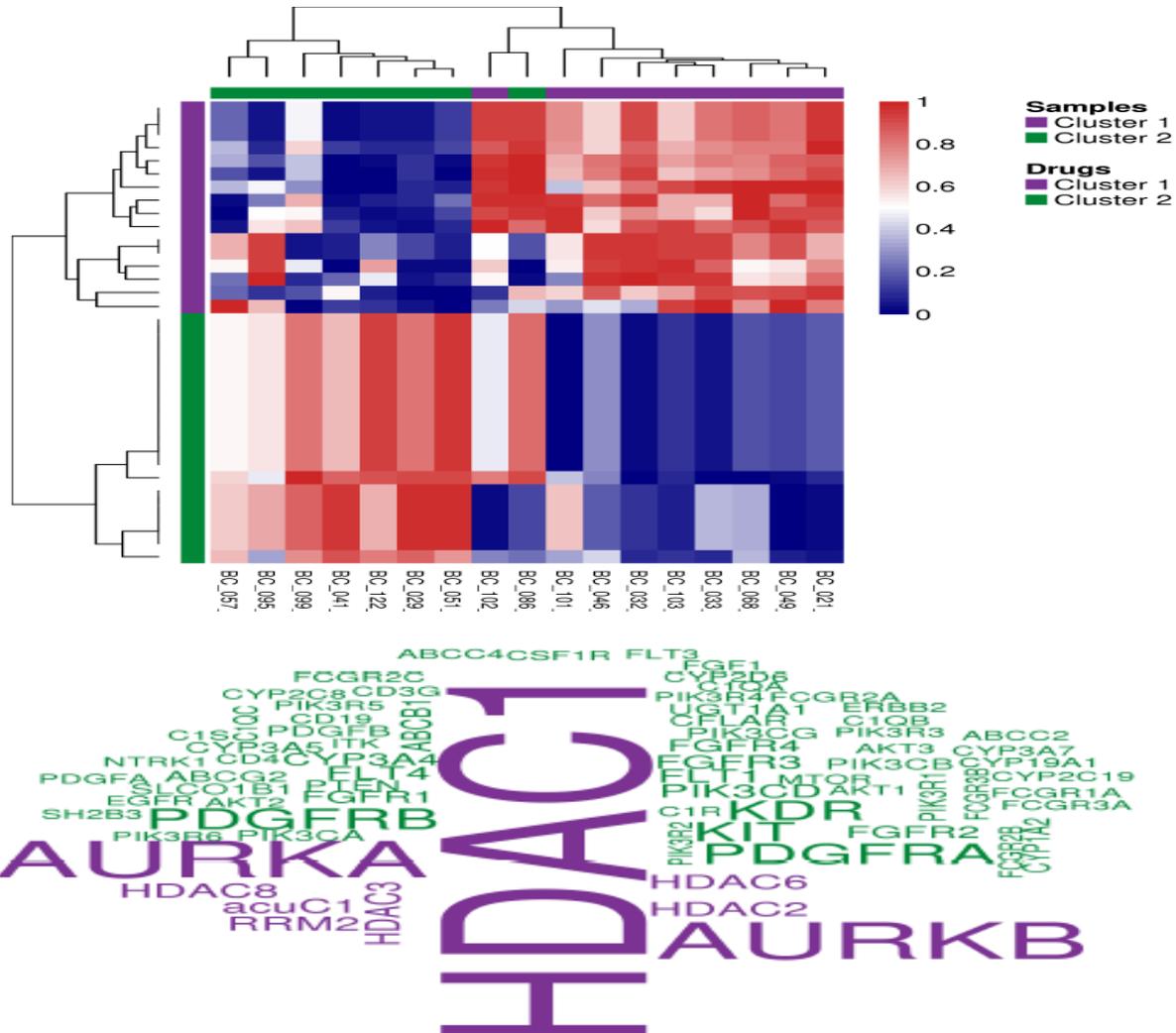
Tailored to an Individual Patient.

## 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

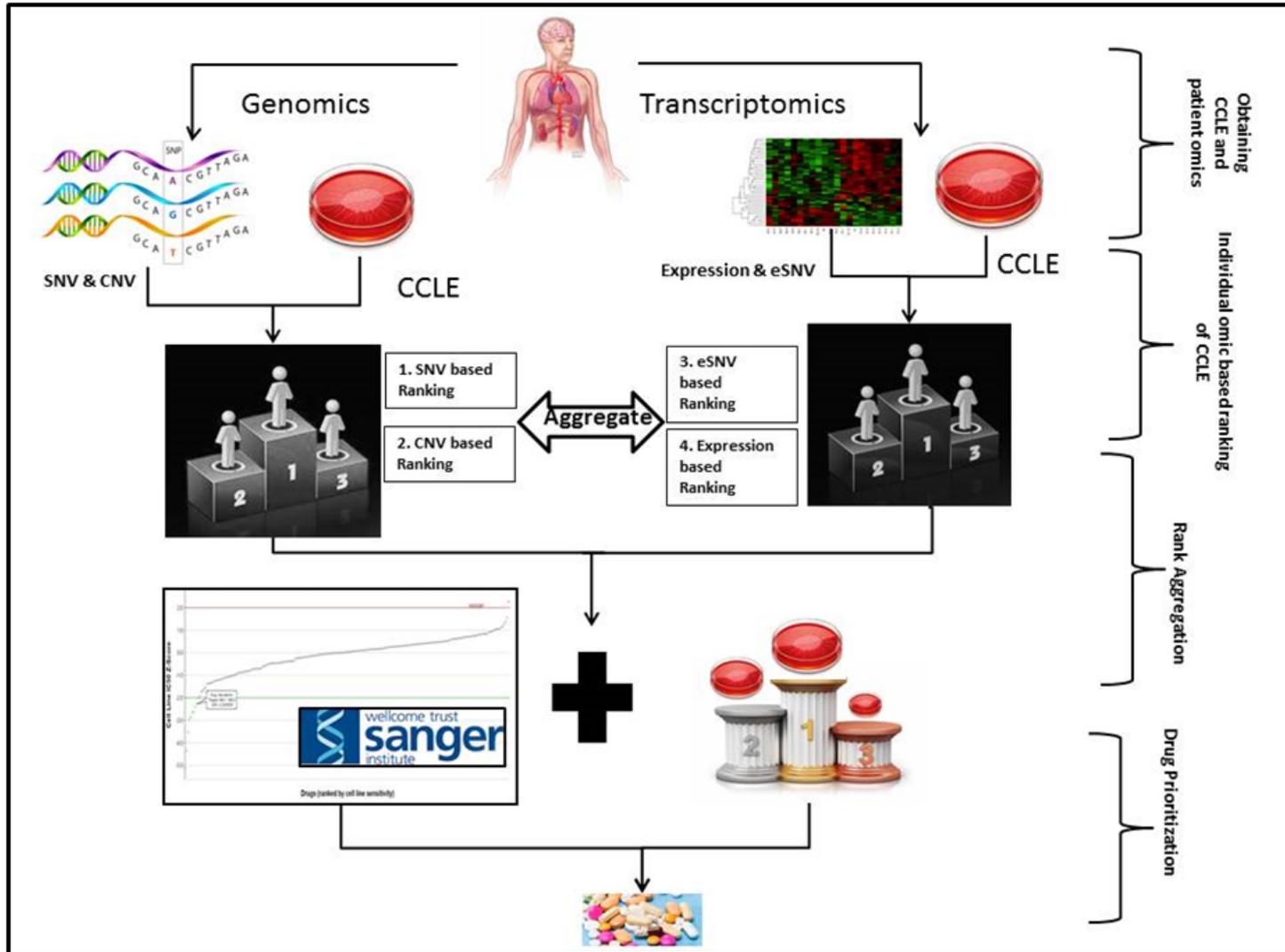


Kalari KR, Sinnwell JP, Thompson KJ, Tang X, Carlson EE, Yu J, Vedell PT, Ingle JN, Weinshilboum RM, Boughey JC, Wang L, Goetz MP, Suman V. JCO Clin Cancer Inform. 2018 Dec;(2):1-11. doi: 10.1200/CCI.18.00012.



Oncomatch – matching best cancer cell line to a patient

# OncoMatch





# 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<sup>†</sup>, Xiaojia Tang<sup>†</sup>, 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\*

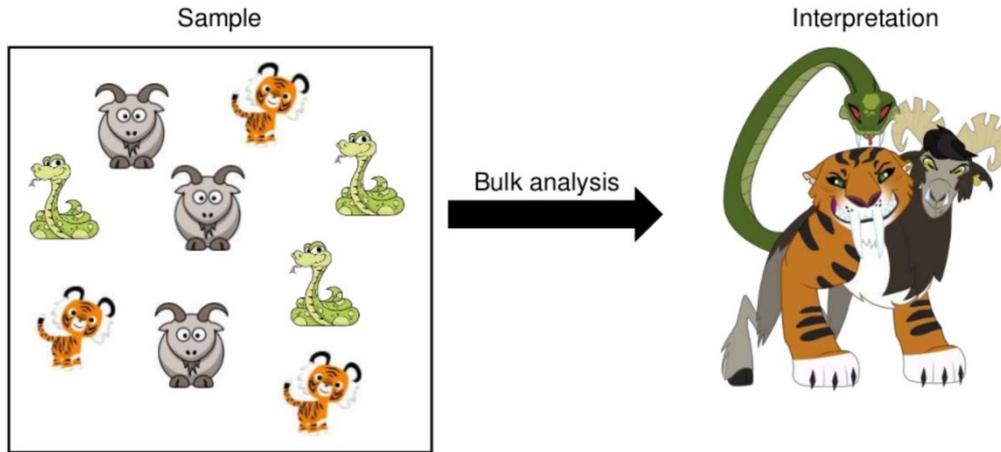
Under review



# Single-cell RNA-Sequencing

# Single Cell vs. Bulk Samples

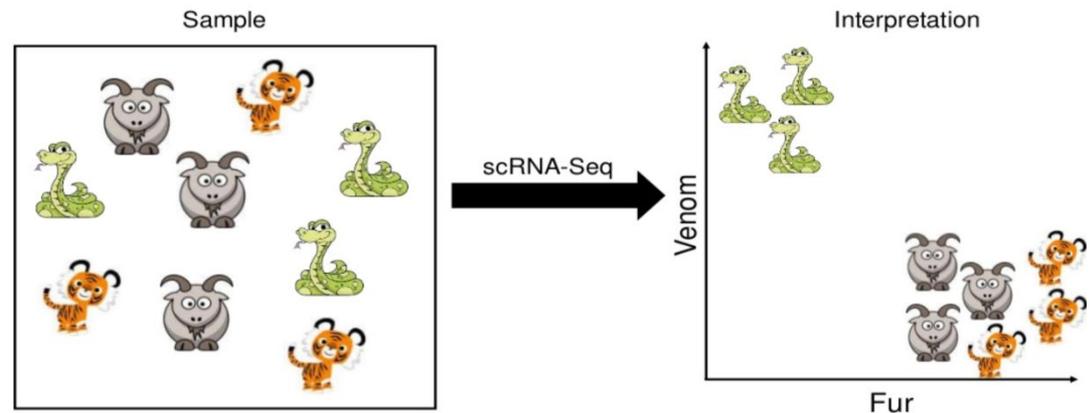
Conclusions from bulk analysis can be representative of nothing



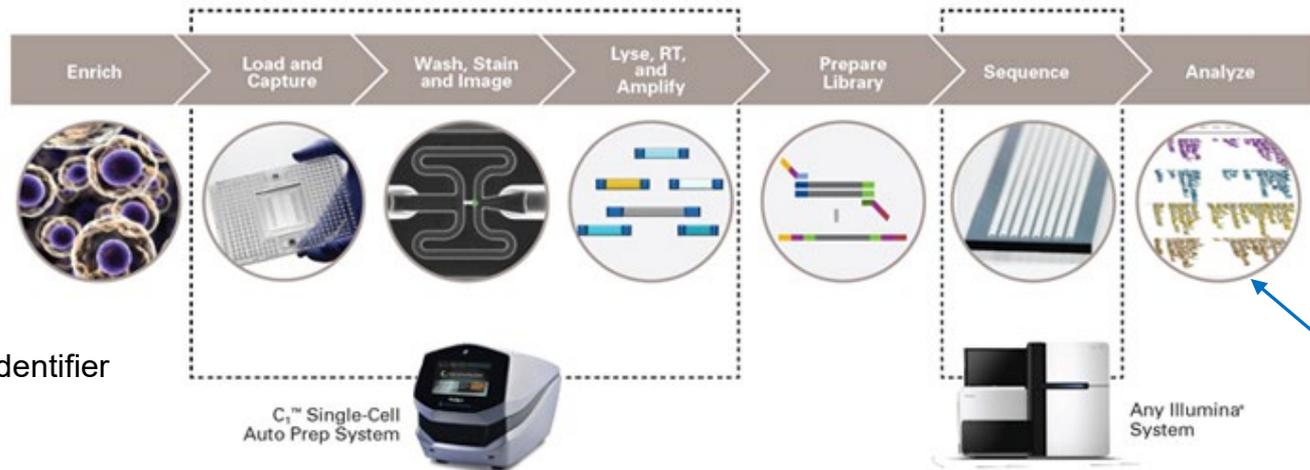
David Cook, SlideShare, 2017

Single-cell exposes this heterogeneity

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

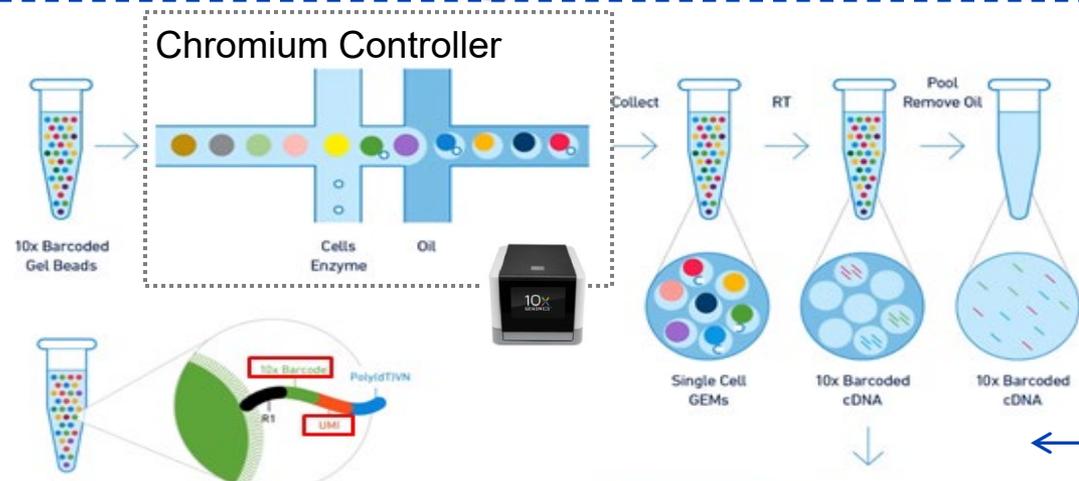


# Two scRNA-Seq Platforms at MGF



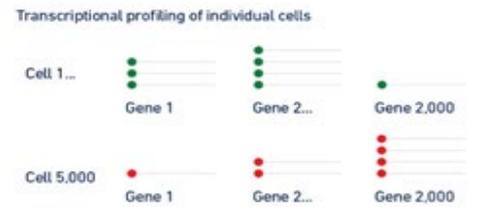
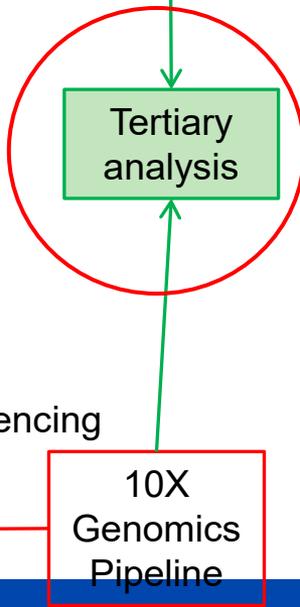
Fluidigm C1  
(fluidic circuits)  
(non-UMI)

UMI=Unique Molecular Identifier



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

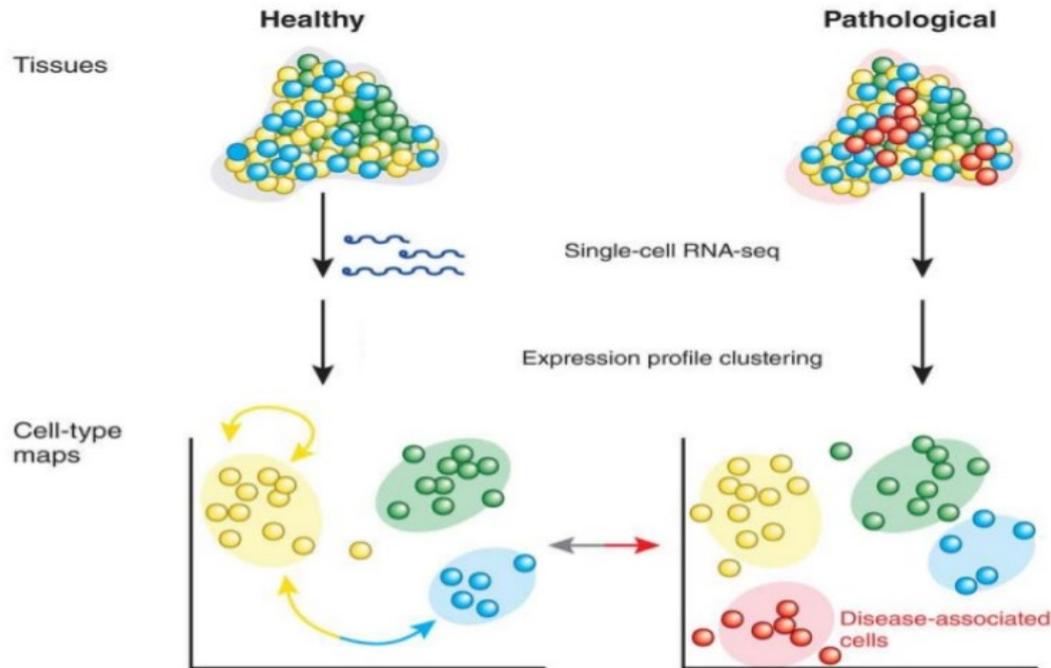
- Input: Single cells in suspension + 10x Gel Beads and Reagents
- Output: Digital gene expression profiles from every partitioned cell



# 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	<b>High cell capture efficiency, cell size &lt;50µm, nuclei suspensions can be studied, lower system cost</b>
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



## Types of analyses

### Within cell type

- Stochasticity, variability of transcription
- Regulatory network inference
- Allelic expression patterns
- Scaling laws of transcription

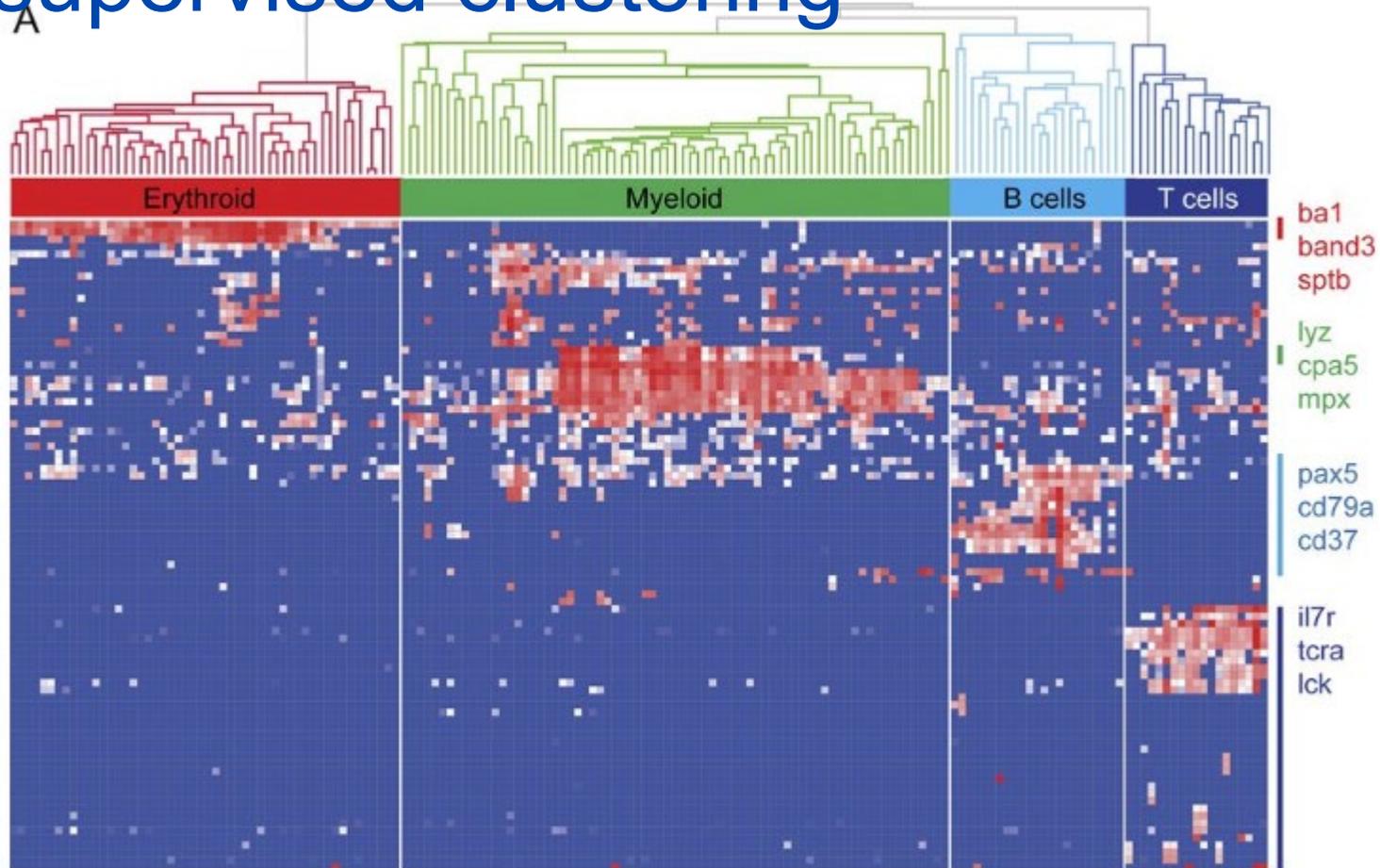
### Between cell types

- Identify biomarkers
- (Post)-transcriptional differences

### Between tissues

- Cell-type compositions
- Altered transcription in matched cell types

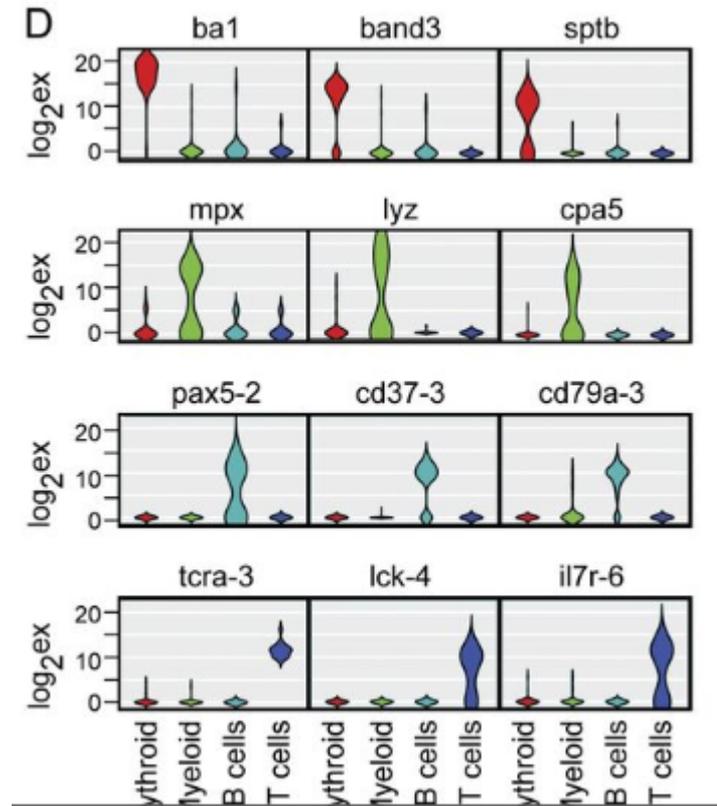
# Unsupervised clustering



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

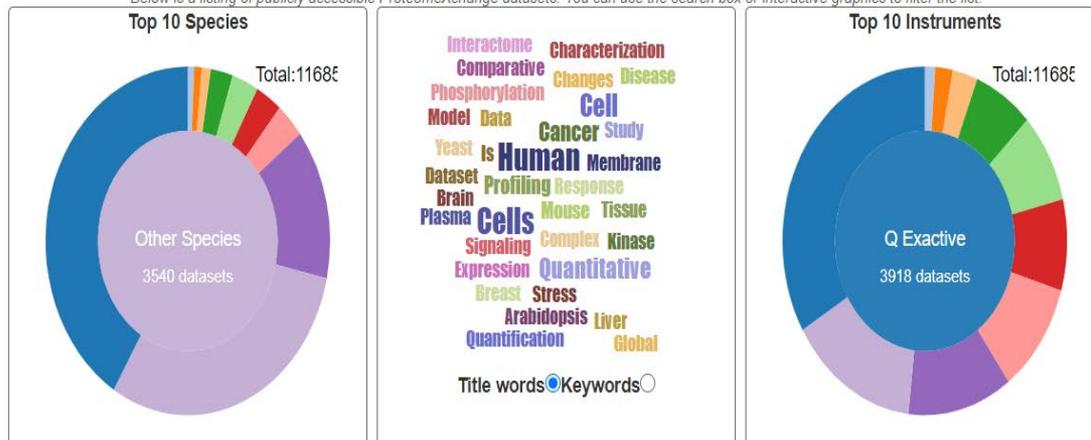
<http://jem.rupress.org/content/213/6/979>

# Proteomics datasets

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



Below is a listing of publicly accessible ProteomeXchange datasets. You can use the search box or interactive graphics to filter the list.



# 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