

cBioPortal Tutorial #2: Single Study Query

Query one or multiple genes in a single dataset

Tutorial Objectives

- 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/Survival
 - CN Segments
 - Pathways
 - Download
- Show how to modify and re-run a query

In this tutorial, blue boxes provide an overview of each tab on cBioPortal ...

... while green boxes ask a biological question that we can answer using cBioPortal.

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/Survival:** 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

We're going to run a query in a TCGA Lower-Grade Glioma study. The next few slides will show how to run this query from the Query page. You can also run the same query from a Single Study Exploration, as we did in Tutorial #1.

Query overview

Search studies

Browse available datasets and select studies to explore or query

Number of studies for each tissue of origin (click to filter)

List of all studies, organized by organ system

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...

Category	Count
PanCancer Studies	9
Pediatric Cancer Studies	13
Immunogenomic Studies	8
Cell lines	3
Adrenal Gland	3
Ampulla of Vater	1
Biliary Tract	13
Bladder/Urinary Tract	17
Bone	2
Bowel	12
Breast	23
CNS/Brain	23
Cervix	2
Esophagus/Stomach	17

PanCancer Studies

- MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017) 10945 samples
- Metastatic Solid Cancers (UMich, Nature 2017)
- MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018)
- SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018)
- TMB and Immunotherapy (MSKCC, Nat Genet 2019)
- Tumors with TRK fusions (MSK, Clin Cancer Res 2020)
- Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020)
- China Pan-cancer (Origimed2020)
- Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020) 2922 samples

Pediatric Cancer Studies

- Pediatric Preclinical Testing Consortium (CHOP, Cell Rep 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
- Pediatric Pan-Cancer (DKFZ, Nature 2017) 961 samples
- Pediatric Pan-cancer (Columbia U, Genome Med 2016) 103 samples
- Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016) 73 samples

0 studies selected (0 samples) Query By Gene OR Explore Selected Studies

What's New @cbioportal

Local Installations Host your own

Single study query

1. Filter the list of studies (optional)

Select Studies for Visualization & Analysis: 1 study selected (514 samples) Deselect all

glioma X

Immunogenomic Studies 1 Select all listed studies matching filter (20)

CNS/Brain 18 Glioblastoma (Columbia, Nat Med. 2019) 42 samples

Soft Tissue 2 Integrated Proteogenomic Characterization across Major Histological T... 218 samples

Diffuse Glioma

- Brain Lower Grade Glioma (TCGA, Firehose Legacy) 530 samples
- Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 514 samples
- Diffuse Glioma (GLASS Consortium, Nature 2019) 444 samples
- Glioma (MSK, Nature 2019) 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) 97 samples
- Glioblastoma (Columbia, Nat Med. 2019) 42 samples
- Glioblastoma (TCGA, Cell 2013) 543 samples
- Glioblastoma (TCGA, Nature 2008) 206 samples
- Glioblastoma Multiforme (TCGA, Firehose Legacy) 619 samples
- Glioblastoma Multiforme (TCGA, PanCancer Atlas) 592 samples

1 study selected (514 samples) Deselect all

Query By Gene OR **Explore Selected Studies**

2. Check the box for study of interest.

3. Select "Query By Gene"

Single study query

The screenshot shows the 'Query' section of the cBioPortal interface. At the top, there are tabs for 'Query', 'Quick Search Beta!', and 'Download'. A citation 'Please cite: Cerami et al., 2012 & Gao et al., 2013' is visible in the top right. The main content area is divided into several sections:

- Selected Studies:** A dropdown menu shows 'Brain Lower Grade Glioma (TCGA, PanCancer Atlas)' with '(514 total samples)'. A 'Modify' button is next to it.
- Select Genomic Profiles:** A list of checkboxes for data types: 'Mutations' (checked), 'Structural Variant' (checked), 'Putative copy-number alterations from GISTIC' (checked), 'mRNA Expression' (unchecked), and 'Protein expression z-scores (RPPA)' (unchecked). Under 'mRNA Expression', there are two radio button options for z-score profiles.
- Select Patient/Case Set:** A dropdown menu shows 'Samples with mutation and CNV'. Below it is a link: 'To build your own case set, try out our enhanced Study View.'
- Enter Genes:** A text input field contains 'IDH1 EGFR'. Below it is a green status bar that says 'All gene symbols are valid.' A 'Submit Query' button is at the bottom left.

Numbered callouts are present:

- 4.** Points to the 'Select Genomic Profiles' section.
- 5.** Points to the 'Enter Genes' section.
- 6.** Points to the 'Submit Query' button.

4. This section lists all data types available for the selected study. Select data types to query. By default, Mutations and CNA will be selected (if available).

5. Select sample set. For most studies, an appropriate sample set will be automatically selected given the data types selected in Step 4.

5. Type gene(s) or select from pre-defined gene lists. cBioPortal will confirm that all entries are valid gene symbols.

6. Submit query

Refine your query: You can use Onco Query Language (OQL) to define which specific alterations to include. See [specifications](#) or [OQL tutorial](#).

Performing a query as shown in the previous slides or as shown in Tutorial #1 will both bring you to Results View, shown on the next slide.

Results View is made up of multiple tabs, each with specific functionality, which all share a header.

Results View Header: General Information

The name of the study.
Click to view the full
study in Study View.

The number (percentage) of
samples/patients with an alteration
in any of the query genes

The screenshot shows the header of the OncoPrint interface. On the left, there is a 'Modify Query' button and a hamburger menu icon. The main header text reads 'Brain Lower Grade Glioma (TCGA, PanCancer Atlas)' followed by 'Samples with mutation and CNA data (511 patients/samples) - IDH1 & EGFR'. Below this is a navigation bar with links for 'OncoPrint', 'Cancer Types Summary', 'Mutual Exclusivity', 'Plots', 'Mutations', 'Co-expression', 'Comparison/Survival', 'CN Segments', 'Pathways', and 'Download'. On the right side of the header, it states 'Queried genes are altered in 444 (87%) of queried patients/samples' next to a circular icon containing the number '86'.

The number of samples and patients included
in the query. Note that these numbers can
differ from each other if some patients have
more than one tumor sample profiled.

Click on the number of patients/samples to go
to Study View for just the queried samples.

Save a link to the current session.
Useful for sharing with others or
returning to a query at a later date.

Results View Header: Variant Settings

Use this menu to control how alterations are visualized. Changes made here are immediately reflected across Results View. However over the **i** to confirm how individual tabs reflect these selections.

Brain Lower Grade Glioma (TCGA, PanCancer Atlas)
Samples with mutation and CNA data (511 patients/samples) - IDH1 & EGFR

Queried genes are altered in 444 (87%) of queried patients/samples

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

Annotate Data

- Putative drivers vs VUS:
 - OncoKB driver annotation
 - Hotspots
 - cBioPortal ≥ 0
 - COSMIC ≥ 0

Filter Data

- Exclude alterations (mutations, structural variants and copy number) of unknown significance
- Exclude germline mutations
- Exclude unprofiled samples
- Exclude samples that are unprofiled in any queried gene or profile
- Exclude samples that are unprofiled in every queried gene and profile.

Genetic Alteration

Missense

Set the definition of a putative driver vs variant of unknown significance (VUS).

Check boxes to exclude VUS (as defined above) or germline alterations. When checked, VUS or germline alterations are considered not present, so a sample with only VUS or germline alterations will be treated as an unaltered sample.

Check box to exclude samples where queried genes are not profiled or genomic profiles are not available.

But wait! What if I changed my mind?
Can we modify a query?

Modify Query

Click on “Modify Query”. This button is available on all tabs and can be used at any time. This will bring up the query interface from the homepage (see next slide for a screenshot).

Brain Lower Grade Glioma (TCGA, PanCancer Atlas)
Samples with mutation and CNA data (511 patients/samples) - IDH1 & EGFR

Queried genes are altered in 444 (87%) of queried patients/samples

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

Add Tracks | Sort | Mutations | View | Download | 30 %

Gene	Alteration Type	Percentage
IDH1	Missense Mutation (putative driver)	77%
EGFR	Amplification	11%

Genetic Alteration

- Missense Mutation (putative driver)
- Missense Mutation (unknown significance)
- Structural Variant (unknown significance)
- Amplification
- No alterations

You can also click the  for a quick edit of the queried genes, including OQL edits. To change other query settings, use the “Modify Query” button.

Modify Query

The existing query is pre-populated for your convenience. You can change the study, the genomic profiles, the patient/case set or the gene set. Simply hit “Submit” when you are happy with the modified query.

Cancel Modify Query **Brain Lower Grade Glioma (TCGA, PanCancer Atlas)** Queried genes are altered in 444 (87%) of queried patients/samples

Samples with mutation and CNA data (511 patients/samples) - IDH1 & EGFR

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

Select Studies for Visualization & Analysis: 1 study selected (514 samples) Deselect all

PanCancer Studies	9
Pediatric Cancer Studies	13
Genomic Studies	8
Lines	3
Salivary Gland	3
Uterus of Venter	1
Rectum	13
Bladder/Urinary Tract	17
Bone	2
Bowel	12

Quick select:

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, Clin Cancer Res 2020) 106 samples
- Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020) 24146 samples
- China Pan-cancer (Origimed2020) 10194 samples
- Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020) 2922 samples

Pediatric Cancer Studies

- Pediatric Preclinical Testing Consortium (CHOP, Cell Rep 2019) 261 samples
- Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018) 1978 samples

Select Genomic Profiles:

- Mutations
- Structural Variant
- 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)
- Protein expression z-scores (RPPA)

Select Patient/Case Set:

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

Enter Genes:

IDH1 EGFR IDH2 ←

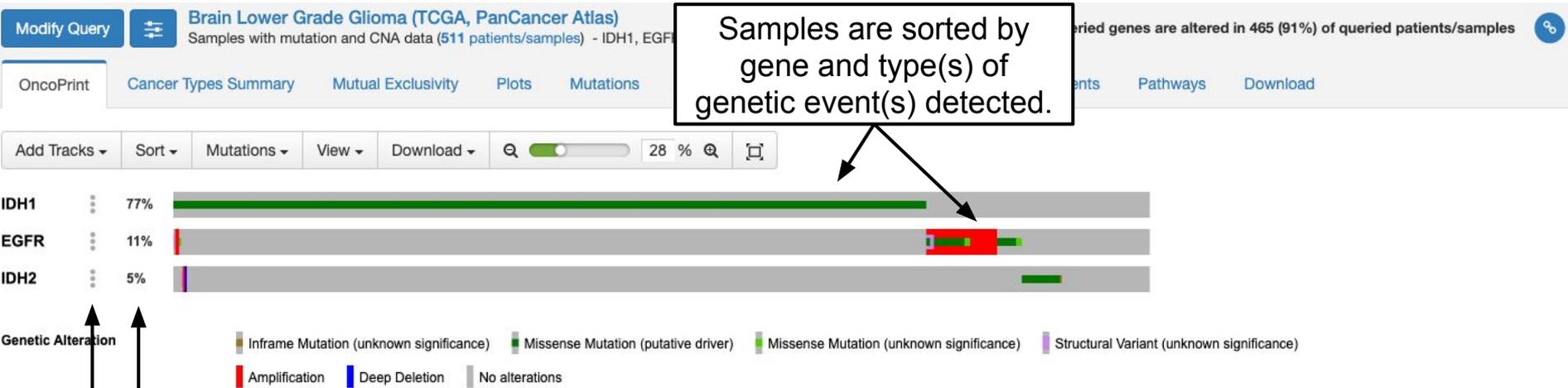
Hint: Learn Onco Query Language (OQL) to write more powerful queries

All gene symbols are valid.

In this case, I've added a third gene (IDH2) to the query.

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.



Samples are sorted by gene and type(s) of genetic event(s) detected.

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

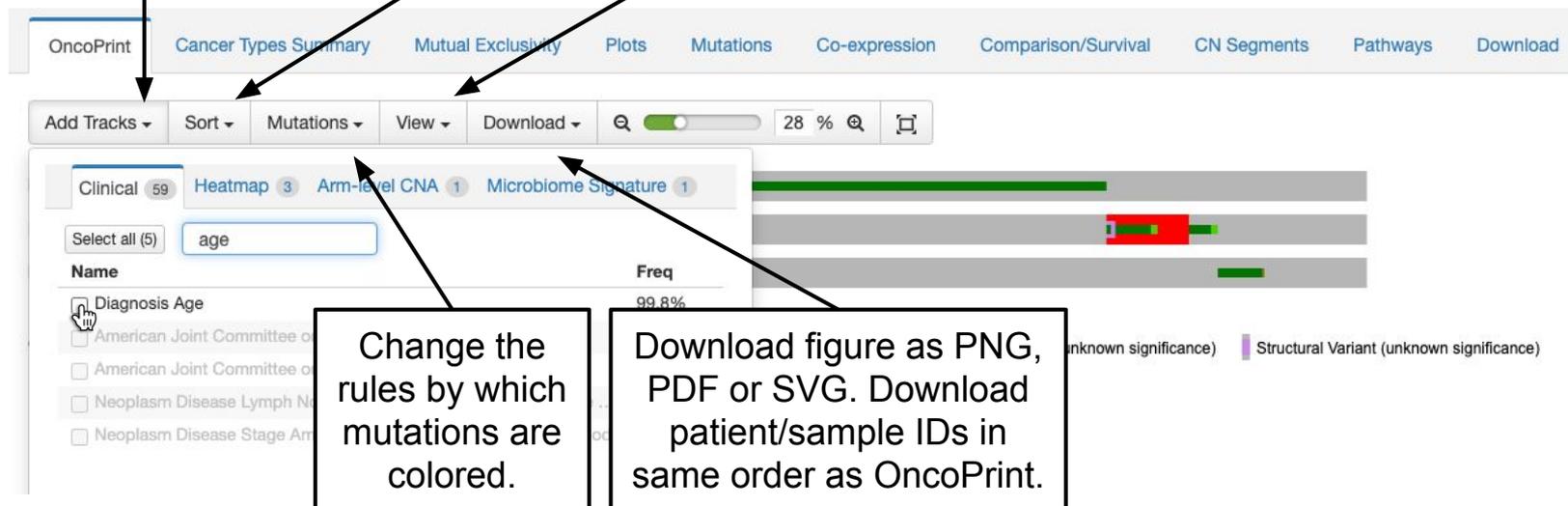
To change the order, click on a gene name and drag, or click on the . Samples will re-sort based on this new order.

OncoPrint: Features

Add clinical tracks, heatmaps (eg RNA levels) or other data (eg Arm-level CNA). Available data varies by study.

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.

OncoPrint: Zoom

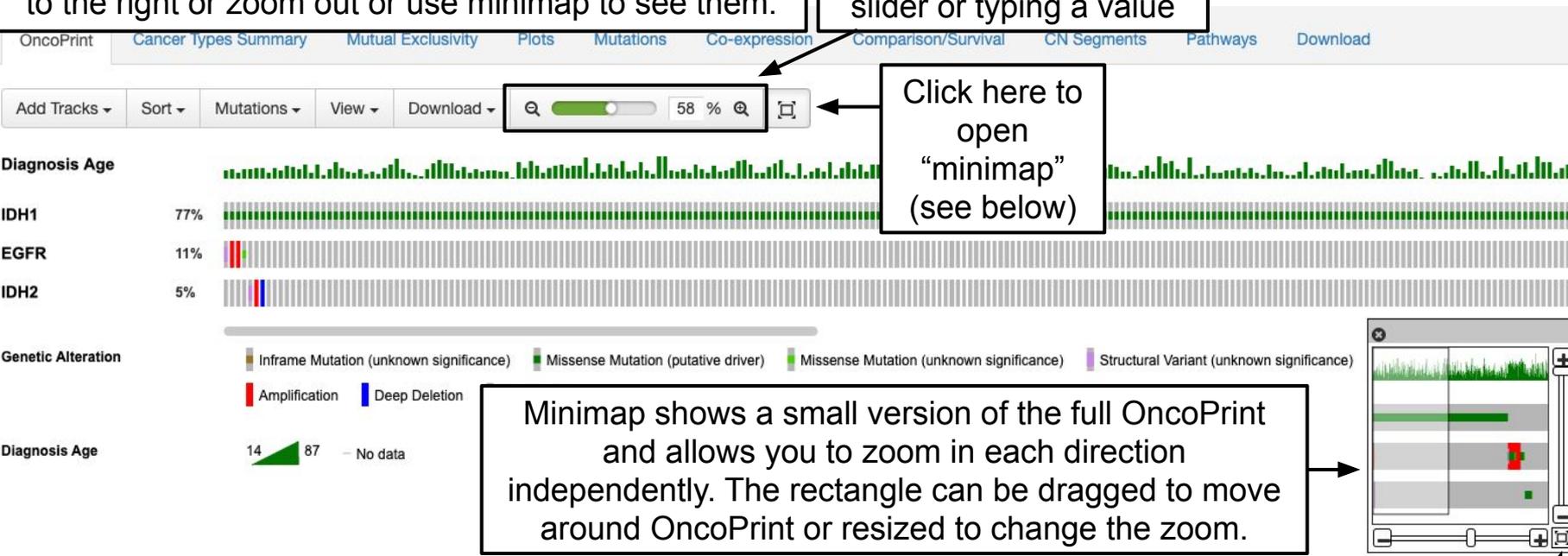
There may be more samples hiding off-screen. Scroll to the right or zoom out or use minimap to see them.

Change the zoom by clicking the zoom in/out icons or moving the slider or typing a value

Click here to open "minimap" (see below)

Minimap shows a small version of the full OncoPrint and allows you to zoom in each direction independently. The rectangle can be dragged to move around OncoPrint or resized to change the zoom.

This button zooms OncoPrint to show all samples with alterations



OncoPrint: What can we learn?



Q: Are genetic alterations in these genes mutually exclusive?

A: We can see that samples with alterations in one gene tend to not have alterations in the other genes.

Q: Is there an association between alterations in a particular query gene and age?

A: We can see that patients with mutations or amplifications in EGFR tend to be older than those with mutations in IDH1/IDH2

Now we're going to go through all the other tabs and ask some questions about alterations in *IDH1*, *IDH2* and *EGFR* in the TCGA Lower-Grade Glioma study.

Note: Depending on the data available for a particular study, not all of the following tabs will be present (e.g. a study without expression data will not have a Co-expression tab)

Cancer Types Summary

Histogram of the frequency of alterations in each gene for each detailed cancer type.

Plots for all queried genes together and each individual gene are available as separate tabs.

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression

All Queried Genes IDH1 EGFR IDH2

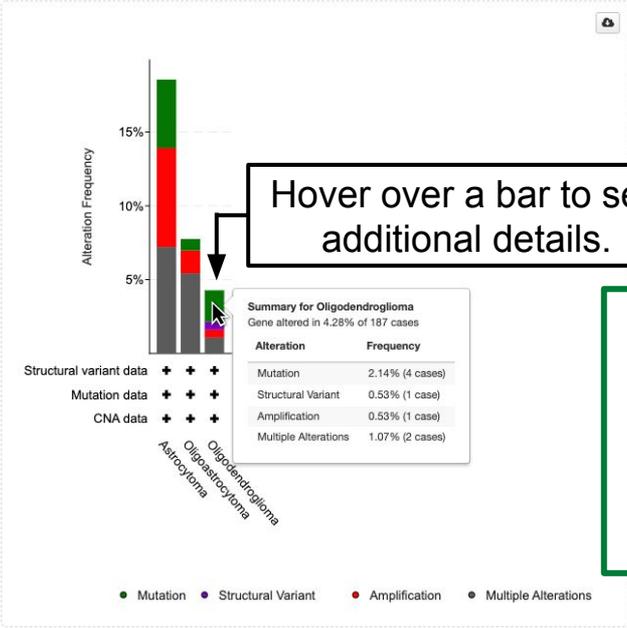
Cancer Study Cancer Type Cancer Type Detailed

Y-Axis Value: Alteration Frequency Min. # Total Cases: 0 194 10 Show Genomic Alteration Types

Sort X-Axis By: Y-Axis Values Min. % Altered Cases: 0% 19% 0%

3 of 4 categories (Cancer Type Detailed) are shown based on filtering.

Options to customize visualization



Hover over a bar to see additional details.

Q: Are alterations in EGFR more frequent in a particular subtype of glioma?

A: Yes, astrocytoma appears to have a much higher frequency of EGFR alteration than oligoastrocytoma or oligodendroglioma.

Mutual Exclusivity

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 Cancer Types Summary Mutual Exclusivity Plots Mutations CN Seg

The analysis tested 3 pairs between the 3 tracks in the OncoPrint.

Mutual exclusivity Co-occurrence Significant only

Columns

A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value ^	Tendency
IDH1	EGFR	67	390	50	4	<-3	<0.001	<0.001	Mutual exclusivity
IDH1	IDH2	96	391	21	3	<-3	<0.001	<0.001	Mutual exclusivity
EGFR	IDH2	433	54	24	0	<-3	0.064	0.064	Mutual exclusivity

Showing 1-3 of 3

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.

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

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

Plots

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

Example plot settings

Choose type of data

Select a query gene

Swap horizontal & vertical axis

OncoPrint Cancer Types Summary Mutual Exclusivity **Plots** Mutations

Examples: Mut# vs Dx FGA vs Dx Mut# vs FGA mRNA vs Dx mRNA vs mut type mRNA

Showing 511 samples with data in both profiles (axes)

Color samples by: EGFR Mutation Type * Structural Variant* Copy Number

Horizontal Axis

Data Type: Copy Number

Copy Number Profile: Putative copy-number alteration...

Gene: EGFR

Filter categories: Select...

Sort Categories by Median

Vertical Axis

Data Type: mRNA

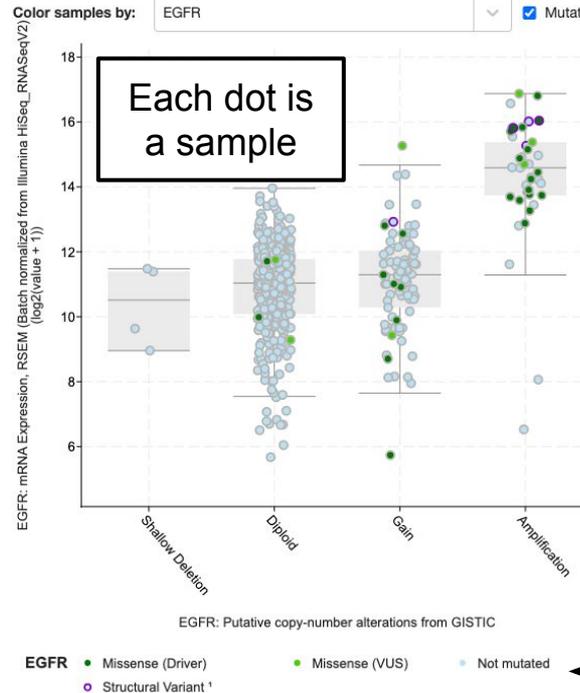
mRNA Profile: mRNA Expression, RSEM (Batc...)

Log Scale

Gene: Same gene (EGFR)

Search Case(s): Case ID..

Search Mutation(s): Protein Change..



Each dot is a sample

Select color scheme

Driver vs. VUS annotation settings are in the  menu in the header of the page.

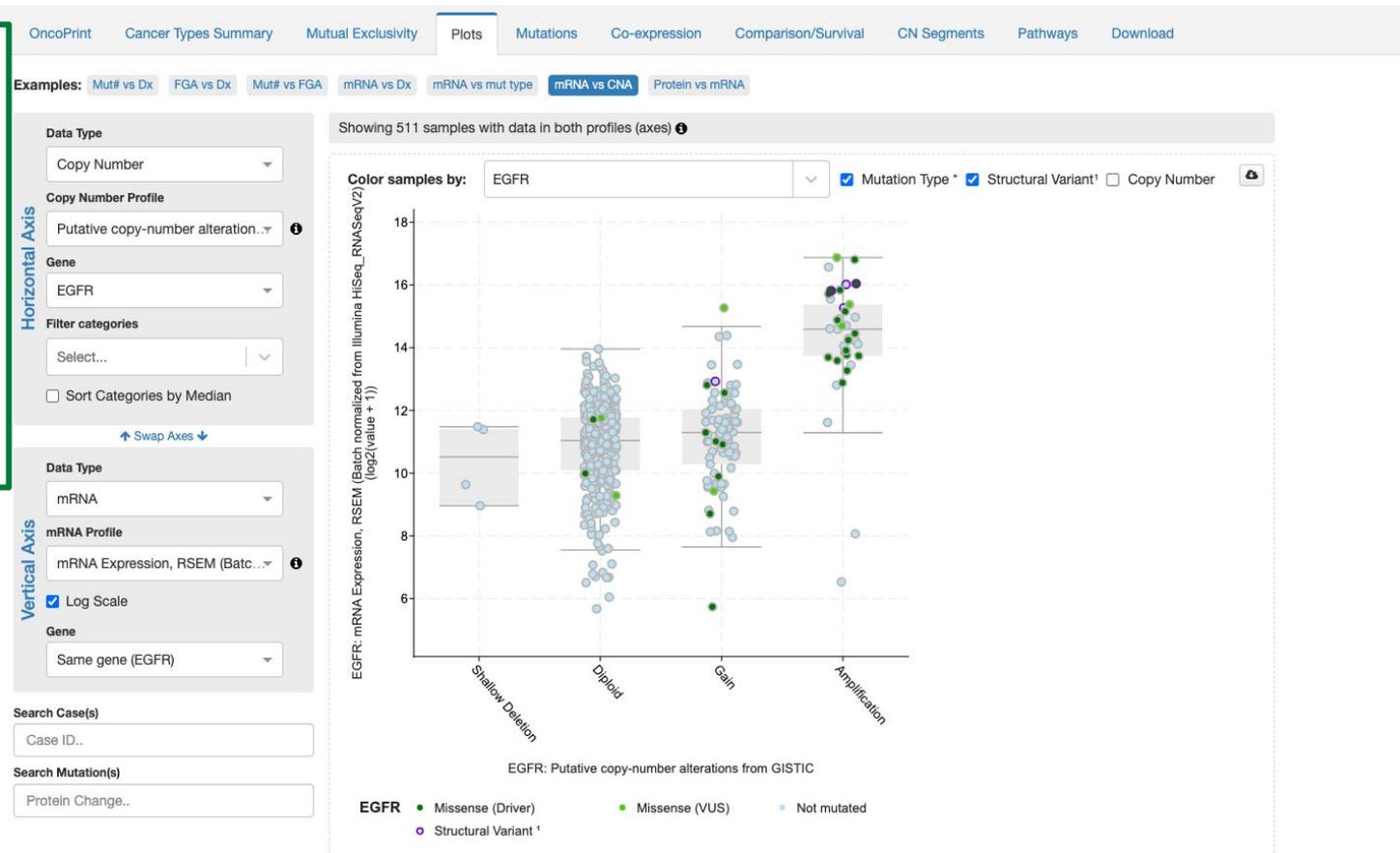
* Driver annotation settings are located in the settings menu  at the top of the page.

¹ Structural variants are shown instead of copy number alterations when a sample has both.

Plots

Q: Does amplification of EGFR alter gene expression?

A: Yes, we can see that higher copy number of EGFR (x-axis) is associated with increased expression (y-axis).



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

† Structural variants are shown instead of copy number alterations when a sample has both.

Mutations

This tab shows details about all mutations called in each query gene.

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

Each gene appears on a separate tab

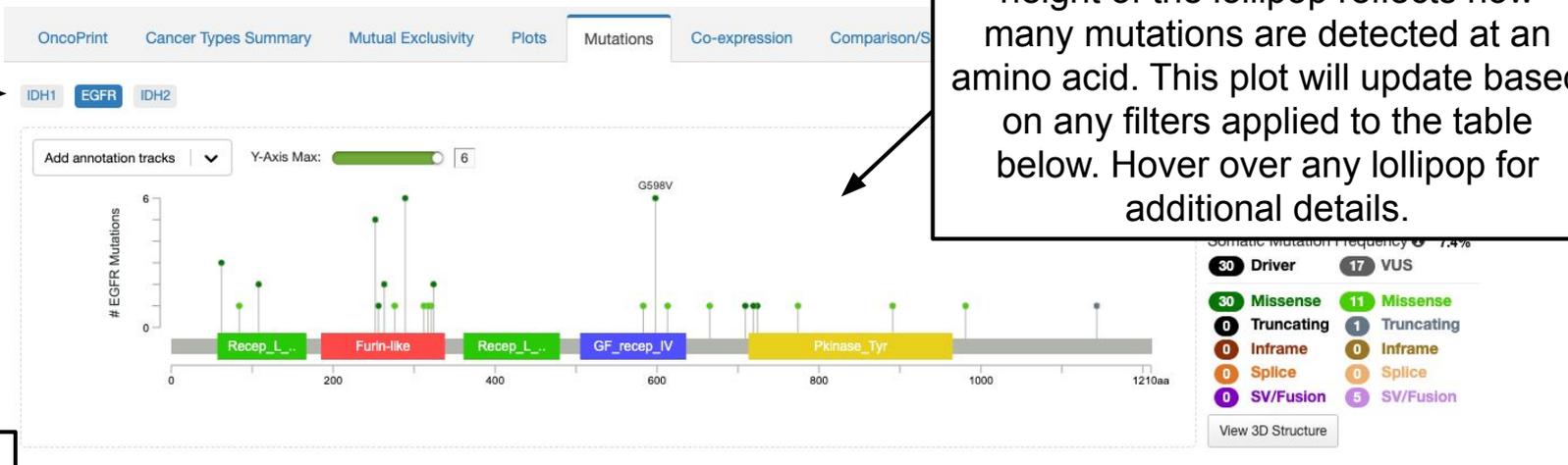


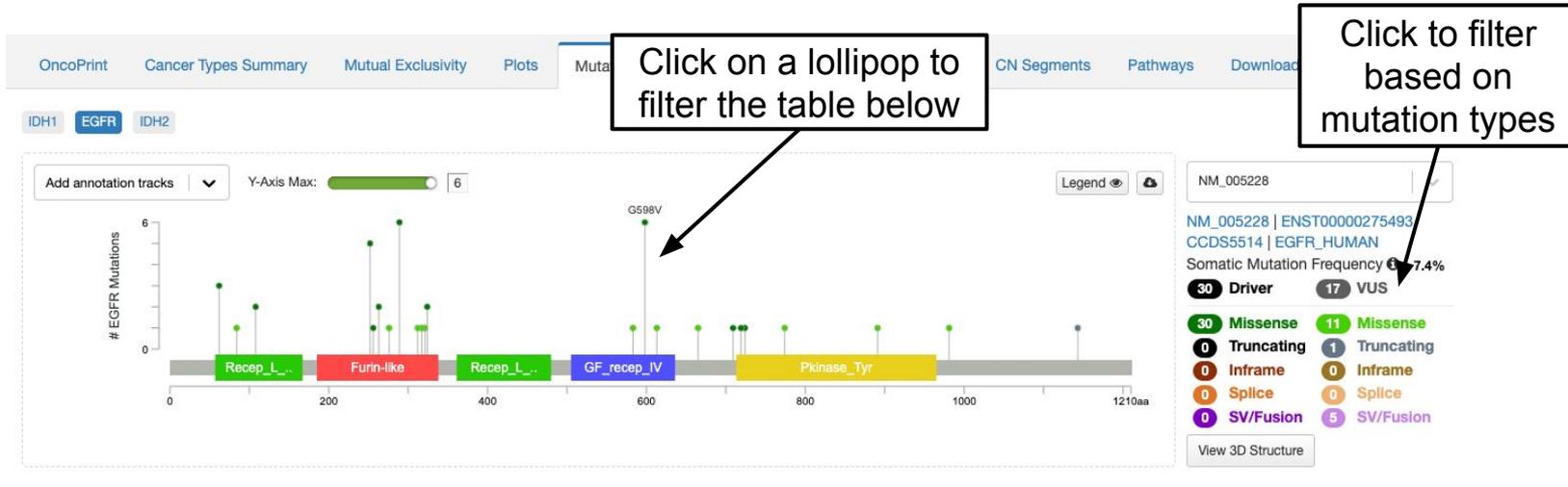
Table of all mutations with annotations

47 Mutations (page 1 of 2)

Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	CO-MIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	G719D	⊙ 3b ↓ 🔥	Missense	Amp		125 0.18	39
TCGA-DU-A5TT...	Oligodendroglioma	A289V	⊙ 4 ↓ 🔥	Missense	Gain			45
TCGA-E1-A7YM...	Astrocytoma	A289V	⊙ 4 ↓ 🔥	Missense	Gain			35
TCGA-FG-6692...	Oligodendroglioma	A289V	⊙ 4 ↓ 🔥	Missense	Amp			63
TCGA-FG-A70Z...	Oligoastrocytoma	A289V	⊙ 4 ↓ 🔥	Missense	Amp			47
TCGA-KT-A7W1...	Astrocytoma	A289V	⊙ 4 ↓ 🔥	Missense	Amp			39
TCGA-TM-A7C3...	Astrocytoma	A289V	⊙ 4 ↓ 🔥	Missense	Amp	50	0.71	56
TCGA-HT-8107-01	Oligodendroglioma	R108K	⊙ 4 ↓ 🔥	Missense	Diploid	17	0.15	1
TCGA-HT-8110-01	Astrocytoma	R108K	⊙ 4 ↓ 🔥	Missense	Amp	17	0.94	30
TCGA-HT-A61C...	Oligodendroglioma	T263P	⊙ 4 ↓ 🔥	Missense	Gain	7	0.24	38
TCGA-DU-7013...	Astrocytoma	G598V	⊙ 4 ↓ 🔥	Missense	Amp	36	0.96	36
TCGA-DU-8162...	Oligoastrocytoma	G598V	⊙ 4 ↓ 🔥	Missense	Amp	36	0.57	22

Show additional columns

Mutations



47 Mutations (page 1 of 2)

Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	G719D	Missense	Missense	Amp	125	0.18	39
TCGA-DU-A5TT...	Oligodendroglioma		Missense	Missense	Gain	50	0.45	
TCGA-E1-A7YM...	Astrocytoma		Missense	Missense	Gain	50	0.18	
TCGA-FG-6692...	Oligodendroglioma		Missense	Missense	Amp	50	0.95	
TCGA-FG-A70Z...	Oligoastrocytoma		Missense	Missense	Amp	50	0.88	
TCGA-KT-A7W1...	Astrocytoma		Missense	Missense	Amp	50	0.02	
TCGA-TM-A7C3...	Astrocytoma		Missense	Missense	Amp	50	0.71	
TCGA-HT-8107-01	Oligodendroglioma	R108K	Missense	Missense	Diploid	17	0.15	1
TCGA-HT-8110-01	Astrocytoma	R108K	Missense	Missense	Amp	17	0.94	30
TCGA-HT-A61C...	Oligodendroglioma	T263P	Missense	Missense	Gain	7	0.24	38
TCGA-DU-7013...	Astrocytoma	G598V	Missense	Missense	Amp	36	0.96	36
TCGA-DU-8162...	Oligoastrocytoma	G598V	Missense	Missense	Amp	36	0.57	22

Click here (visible when you hover over a column) to filter on a specific column

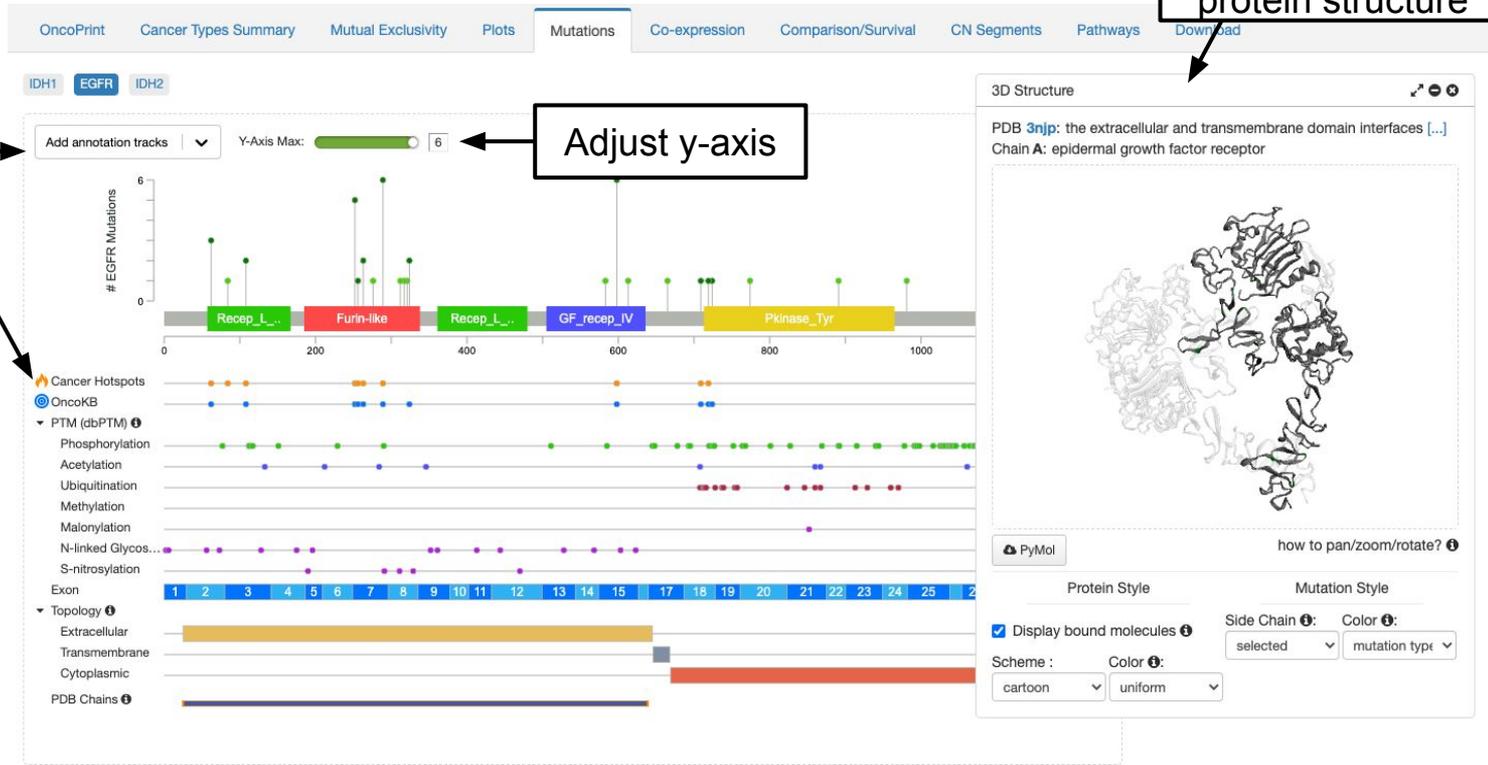
Filter based on any visible text column

Mutations

View mutations in context of 3D protein structure

Add annotation tracks to the plot

Adjust y-axis



47 Mutations (page 1 of 2)

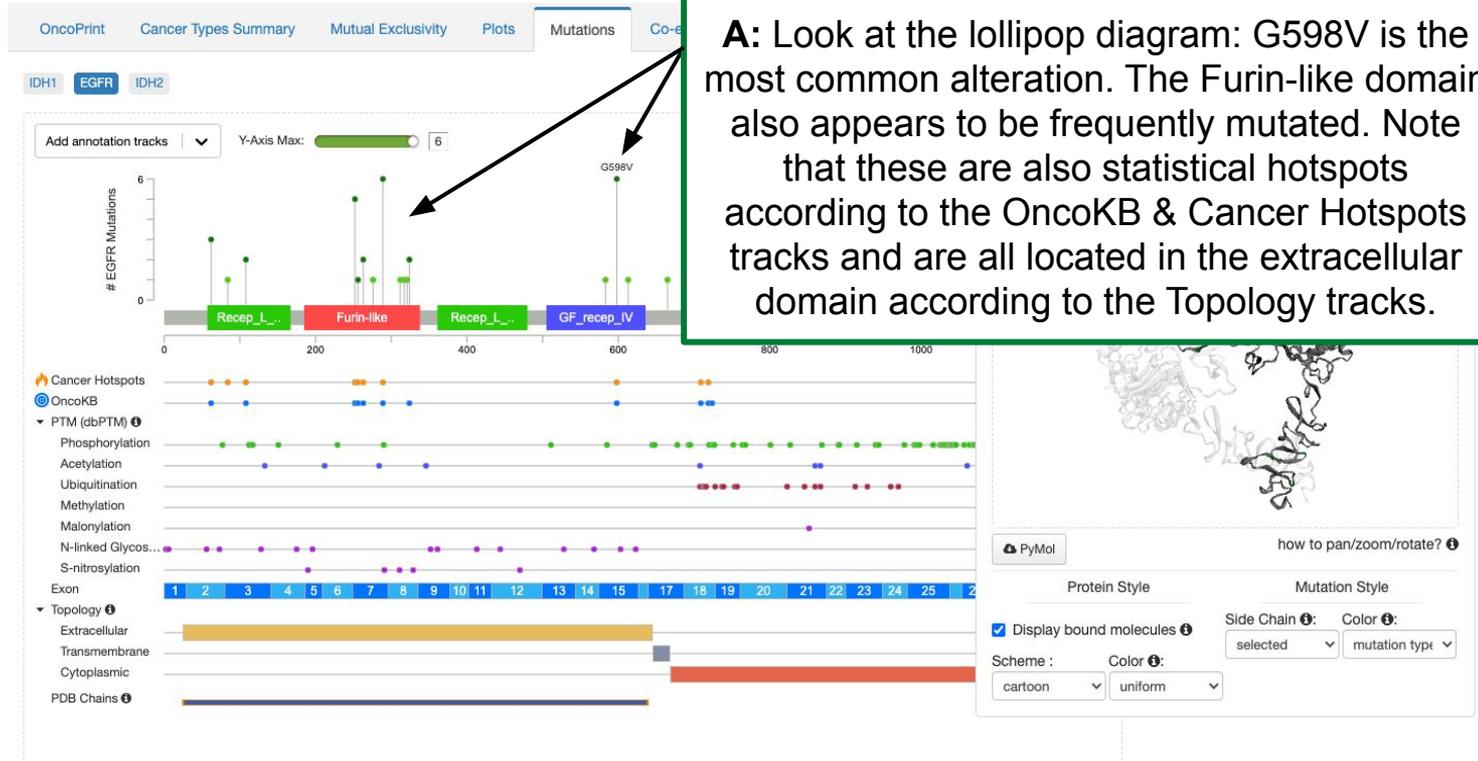
Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	G719D		Missense	Amp	125	0.18	39
TCGA-DU-A5TT...	Oligodendroglioma	A289V		Missense	Gain	50	0.45	45

[Link to this page](#)

Mutations

Q: Where are the hotspots for EGFR mutation in glioma?

A: Look at the lollipop diagram: G598V is the most common alteration. The Furin-like domain also appears to be frequently mutated. Note that these are also statistical hotspots according to the OncoKB & Cancer Hotspots tracks and are all located in the extracellular domain according to the Topology tracks.



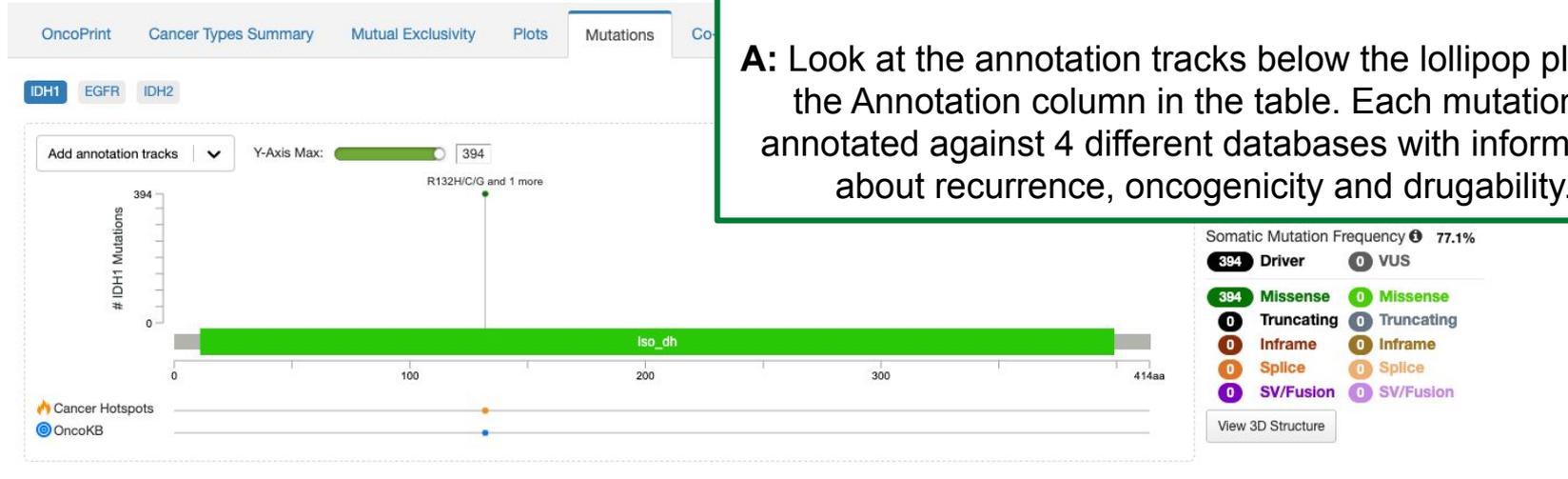
47 Mutations (page 1 of 2)

Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	G719D		Missense	Amp	125	0.18	39
TCGA-DU-A5TT...	Oligodendroglioma	A289V		Missense	Gain	50	0.45	45

Mutations

Q: The mutations in IDH1 appear to be highly recurrent. Are these mutations known hotspots? Known oncogenic drivers? Biomarkers for any drugs?

A: Look at the annotation tracks below the lollipop plot and the Annotation column in the table. Each mutation is annotated against 4 different databases with information about recurrence, oncogenicity and drugability.



 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 a [recurrent hotspot](#) based on a [statistical analysis of mutation frequency](#).

 This mutation is annotated in [CIViC](#). Hover over this symbol for additional information.

 This mutation is in [My Cancer Genome](#).

TCGA-ID	Cancer Type	Gene	Annotation	Mutation Type	Copy #	COSMIC	Allele	# Mut in
TCGA-DB-A4XF...	Astrocytoma	R132C	  	Missense	Diploid	4964	0.40	29
TCGA-DB-A64S...	Oligoastrocytoma	R132C	  	Missense	Diploid	4964	0.32	12
TCGA-DB-A75...			  			4964	0.20	25
TCGA-FG-8185...			 			4964	0.38	36
TCGA-HT-7479-D						4964	0.25	22
TCGA-HT-7693-D						4964	0.47	34
TCGA-HT-7855-D				Missense	Diploid	4964	0.47	36

Co-Expression

Compares mRNA/protein level expression of your query genes against all other genes.

Each gene appears on a separate tab

Select from available data types

Click on a gene name to see correlation plot

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

IDH1 **EGFR** IDH2

Find genes in mRNA Expression, RSEM (Batch normalized from Illumina HiSeq RNASeqV2) that are correlated with EGFR in mRNA Expression, RSEM (Batch normalized from Illumina HiSeq RNASeqV2)

Show Any Correlation Enter gene or cytoband...

Correlated Gene	Cytoband	Spearman's Correlation	p-Value	q-Value
ZNF354C	5q35.3	0.600	2.93e-51	5.87e-47
ZKSCAN8	6p22.1	0.585	3.40e-48	2.34e-44
ZNF107	7q11.21	0.585	3.51e-48	2.34e-44
CHD9	16q12.2	0.582	9.65e-48	4.83e-44
ZNF426	19p13.2	0.582	1.33e-47	5.31e-44
SP4	7p15.3	0.580	2.43e-47	8.12e-44
TEAD1	11p15.3	0.580	3.00e-47	8.58e-44
KAT6A	8p11.21	0.574	4.60e-46	1.04e-42
SON	21q22.11	0.574	4.70e-46	1.04e-42
PYGO1	15q21.3	0.572	8.55e-46	1.71e-42
TRRAP	7q22.1	0.572	1.20e-45	2.19e-42
ZBTB20	3q13.31	0.570	2.17e-45	3.62e-42
RNASEK	17p13.1	-0.569	3.07e-45	4.66e-42
BAZ1B	7q11.23	0.569	3.26e-45	4.66e-42
ZNF699	19p13.2	0.569	3.83e-45	5.11e-42
MED13	17q23.2	0.568	6.64e-45	8.30e-42
KMT2C	7q36.1	0.566	1.08e-44	1.27e-41
DHX33	17p13.2	0.566	1.35e-44	1.50e-41
ZNF791	19p13.13	0.563	4.31e-44	4.54e-41
SMAD5	5q31.1	0.561	1.07e-43	1.07e-40
ECHS1	10q26.3	-0.560	1.33e-43	1.27e-40
RBL1	20q11.23	0.560	1.70e-43	1.55e-40
ZSCAN23	6p22.1	0.558	3.95e-43	3.35e-40
N4BP2	4p14	0.558	4.02e-43	3.35e-40
ZNF800	7q31.33	0.557	6.65e-43	5.32e-40

Showing 1-25 of 20008

EGFR vs. ZNF354C

Show Mutations Log Scale Show Regression Line

mRNA Expression, RSEM (Batch normalized from Illumina HiSeq RNASeqV2): ZNF354C (log2)

mRNA Expression, RSEM (Batch normalized from Illumina HiSeq RNASeqV2): EGFR (7p11.2) (log2)

Spearman: 0.60 (p = 2.93e-51)
Pearson: 0.49 (p = 1.03e-32)

● EGFR mutated
● Neither mutated

Check boxes to color-code sample dots by mutation status, change x- or y-axis to log scale, or add a regression line.

Co-Expression

Q: Several genes on chr7 show high expression correlation with EGFR within this cohort (see table on the left). Why might that be?

A: EGFR is also located on chr7 and is frequently gained in some subtypes of glioma which could explain these correlated increases in expression. This can be further explored in the “CN Segments” tab.

OncoPrint Cancer Types Summary Mutual Ex

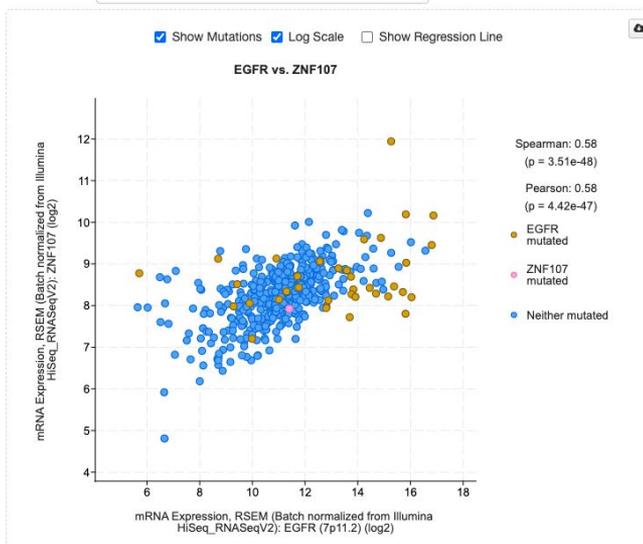
IDH1 EGFR IDH2

Find genes in mRNA Expression, RSEM (Batch normalized)

Show Any Correlation

Correlated Gene	Cytoband	Spearman's Correlation	p-Value	q-Value
ZNF354C	5q35.3	0.600	2.93e-51	5.87e-47
ZKSCAN8	6p22.1	0.585	3.40e-48	2.34e-44
ZNF107	7q11.21	0.585	3.51e-48	2.34e-44
CHD9	16q12.2	0.582	9.65e-48	4.83e-44
ZNF426	19p13.2	0.582	1.33e-47	5.31e-44
SP4	7p15.3	0.580	2.43e-47	8.12e-44
TEAD1	11p15.3	0.580	3.00e-47	8.58e-44
KAT6A	8p11.21	0.574	4.60e-46	1.04e-42
SON	21q22.11	0.574	4.70e-46	1.04e-42
PYGO1	15q21.3	0.572	8.55e-46	1.71e-42
TRRAP	7q22.1	0.572	1.20e-45	2.19e-42
ZBTB20	3q13.31	0.570	2.17e-45	3.62e-42
RNASEK	17p13.1	-0.569	3.07e-45	4.66e-42
BAZ1B	7q11.23	0.569	3.26e-45	4.66e-42
ZNF699	19p13.2	0.569	3.83e-45	5.11e-42
MED13	17q23.2	0.568	6.64e-45	8.30e-42
KMT2C	7q36.1	0.566	1.08e-44	1.27e-41
DHX33	17p13.2	0.566	1.35e-44	1.50e-41
ZNF791	19p13.13	0.563	4.31e-44	4.54e-41
SMAD5	5q31.1	0.561	1.07e-43	1.07e-40
ECHS1	10q26.3	-0.560	1.33e-43	1.27e-40
RBL1	20q11.23	0.560	1.70e-43	1.55e-40
ZSCAN23	6p22.1	0.558	3.95e-43	3.35e-40
N4BP2	4p14	0.558	4.02e-43	3.35e-40
ZNF800	7q31.33	0.557	6.65e-43	5.32e-40

Showing 1-25 of 20008



Comparison

This tab enables the comparison of all available data types between samples with or without alterations in the query genes. This tab replaces and enhances the old “Enrichments” tab.

The Comparison tab is the same as the Group Comparison functionality that is accessible from Study View. See the [Group Comparison Tutorial](#) for more details about the functionality of this tab.

The screenshot shows the OncoPrint interface with the 'Comparison/Survival' tab selected. The top navigation bar includes 'OncoPrint', 'Cancer Types Summary', 'Mutual Exclusivity', 'Plots', 'Mutations', 'Co-expression', 'Comparison/Survival', 'CN Segments', 'Pathways', and 'Download'. Below this, the 'Groups' section displays several buttons: 'Altered group (465)' (highlighted in red), 'Unaltered group (46)' (highlighted in blue), 'IDH1 (394)', 'EGFR (54)', 'IDH2 (24)', 'Select all', and 'Deselect all'. Below the groups, there are tabs for 'Overlap', 'Survival', 'Clinical', 'Genomic Alterations', 'mRNA', 'Protein', and 'Microbiome Signature'. Two callout boxes with arrows point to the 'Altered group' and 'IDH1' buttons respectively.

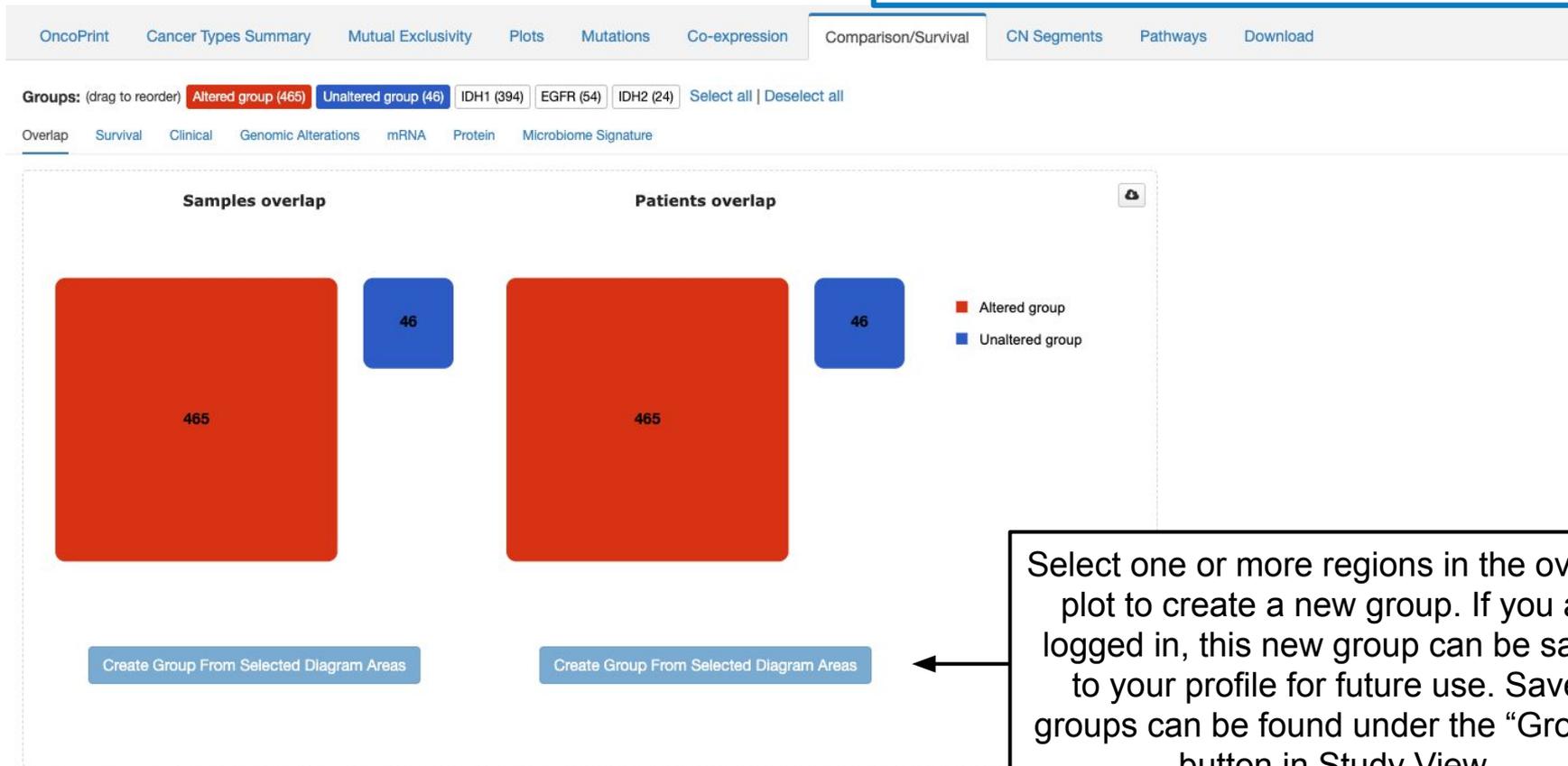
By default, the “Altered” (one or more alterations in one or more query genes) and “Unaltered” (no alterations in any query gene) groups are selected.

Additional groups (deselected by default) correspond to each track shown in OncoPrint.

Groups can be toggled on or off by clicking on them. Analyses will update as the selections change.

Comparison: Overlap

The Overlap subtab shows samples or patients that may overlap among the selected groups.

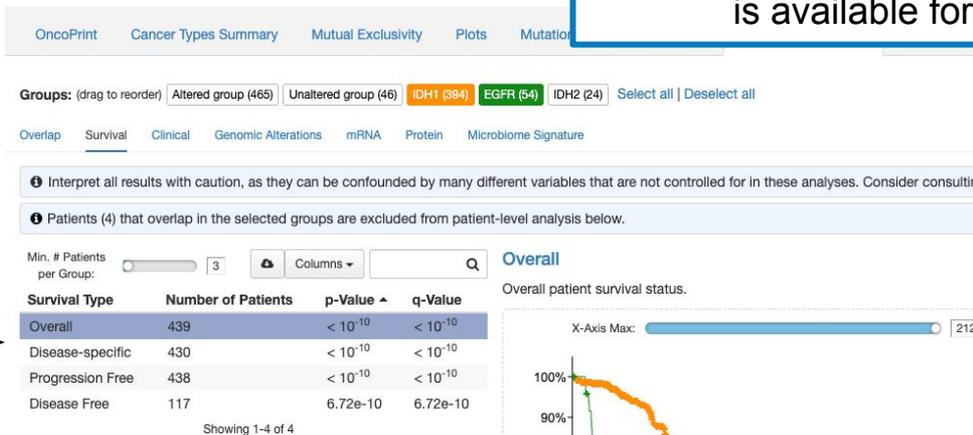


Select one or more regions in the overlap plot to create a new group. If you are logged in, this new group can be saved to your profile for future use. Saved groups can be found under the “Groups” button in Study View.

Comparison: Survival

The Survival subtab replaces the old “Survival” tab. This subtab will only be visible if outcome data is available for the selected study.

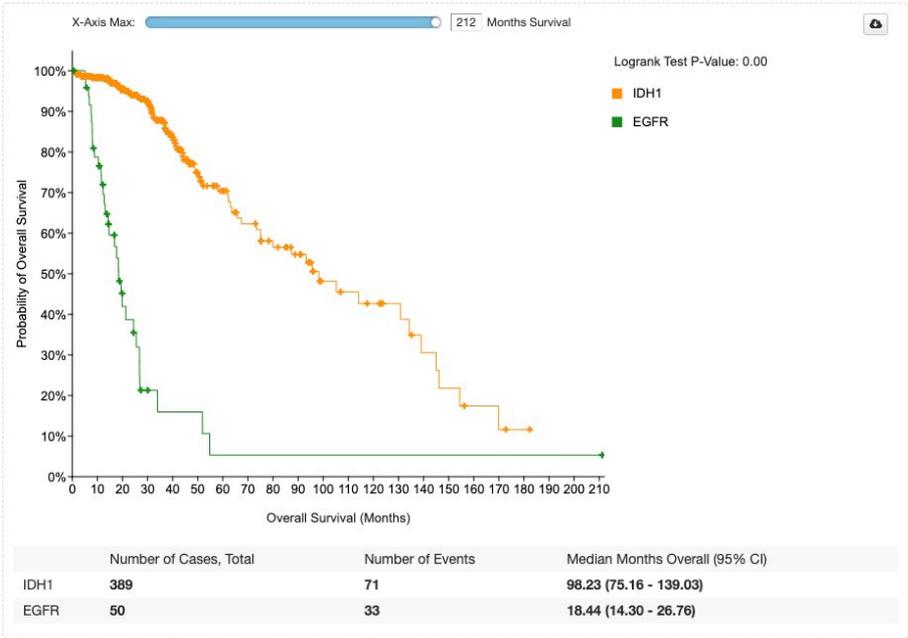
Select among different outcome measures. Options here depend on data availability for the study.



Note: These plots reflect data as provided by the study. We do not perform any additional processing.

Q: Do patients with alterations in IDH1 have different outcomes compared to patients with alterations EGFR?

A: Patients with alterations in IDH1 have significantly better OS than patients with alterations in EGFR.



Comparison: Clinical

The Clinical subtab compares all available clinical data among the selected groups.

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

Groups: (drag to reorder) Altered group (465) Unaltered group (46) IDH1 (394) EGFR (54) IDH2 (24) Select all | Deselect all

Exclude overlapping samples and patients

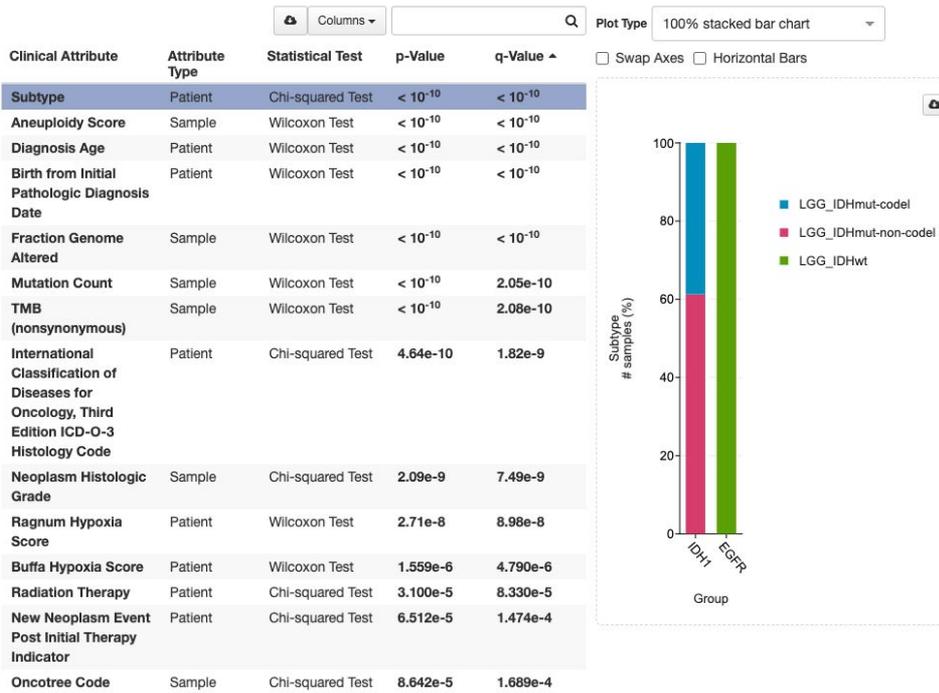
Overlap Survival Clinical Genomic Alterations mRNA Protein Microbiome Signature

Interpret all results with caution, as they can be confounded by many different variables that are not controlled for in these analyses. Consider consulting a statistician.

Samples (4) that overlap in the selected groups are excluded from sample-level analysis below.

Patients (4) that overlap in the selected groups are excluded from patient-level analysis below.

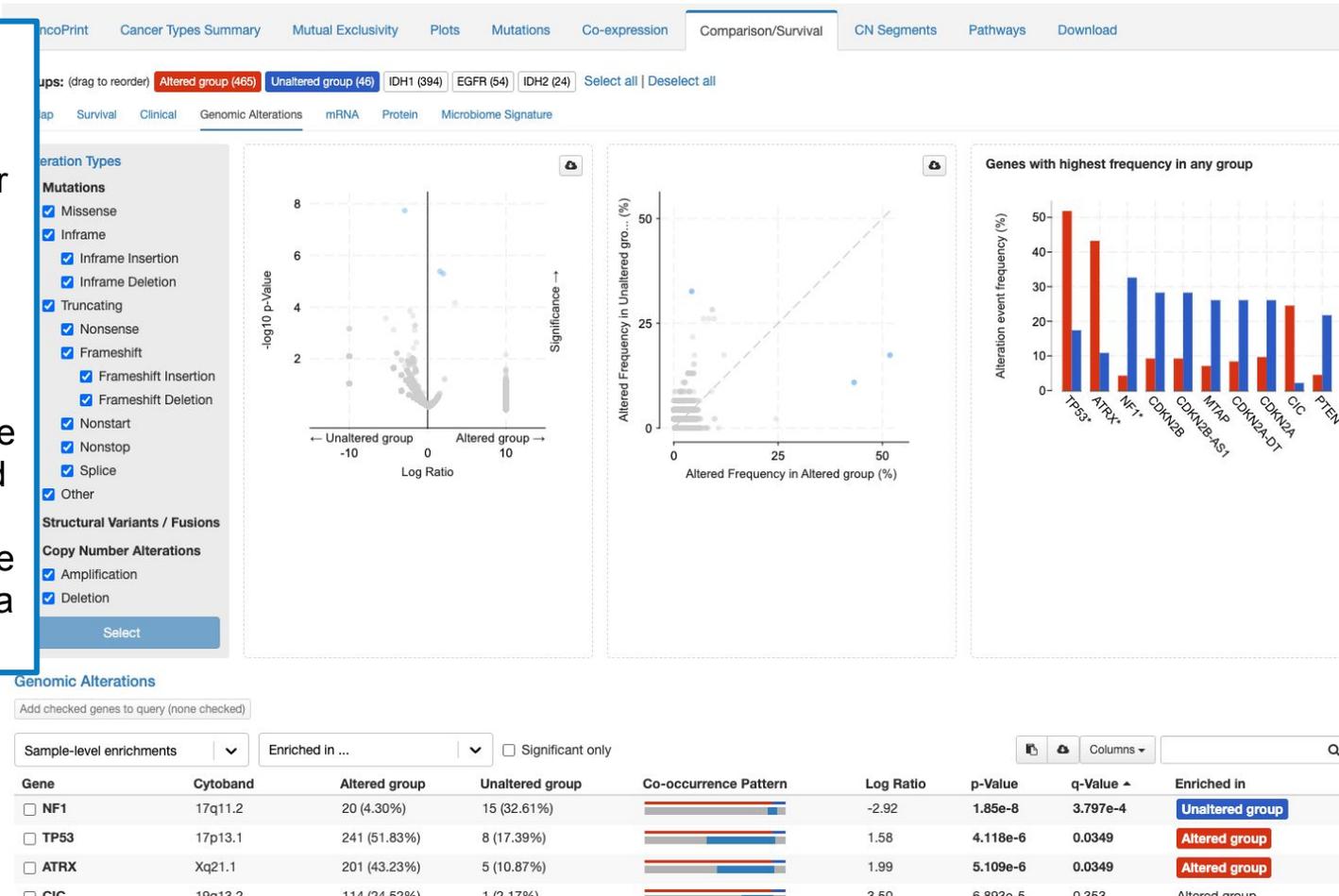
Click on a clinical attribute to visualize the data in the plot on the right.



Comparison: Molecular Profiles

The molecular profiles subtabs replace the old “Enrichments” tab.

These analyses ask whether Genomic Alterations (mutations/copy-number alterations) or mRNA expression or protein expression or protein expression in a particular gene is enriched in one of the selected groups. These, and additional subtabs like Microbiome Signature, will be visible depending on the data available for each study.



Comparison: Molecular Profiles

Groups: (drag to reorder) **Altered group (465)** Unaltered group (46) IDH1 (394) EGFR (54) IDH2 (24) Select all | Deselect all

Overlap Survival Clinical Genomic Alterations mRNA Protein Microbiome Signature

Select which types of alterations to include in the analysis

Alteration Types

- Mutations
 - Missense
 - Inframe
 - Inframe Insertion
 - Inframe Deletion
 - Truncating
 - Nonsense
 - Frameshift
 - Frameshift Insertion
 - Frameshift Deletion
 - Nonstart
 - Nonstop
 - Splice
- Other
- Structural Variants / Fusions
- Copy Number Alterations
 - Amplification
 - Deletion

Select

Select sample-level or patient-level analysis

Genomic Alterations
Add checked genes to query (none checked)

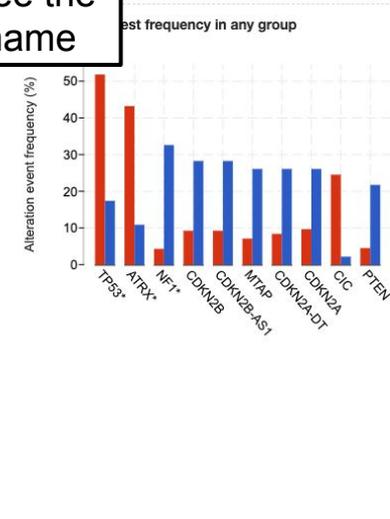
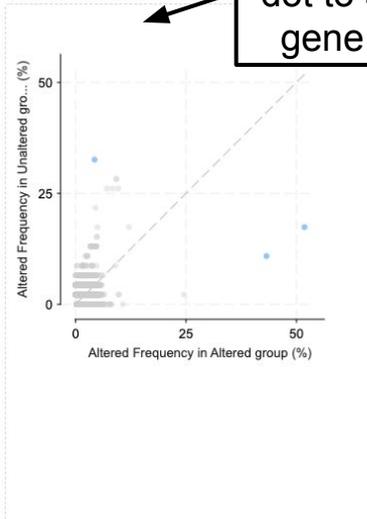
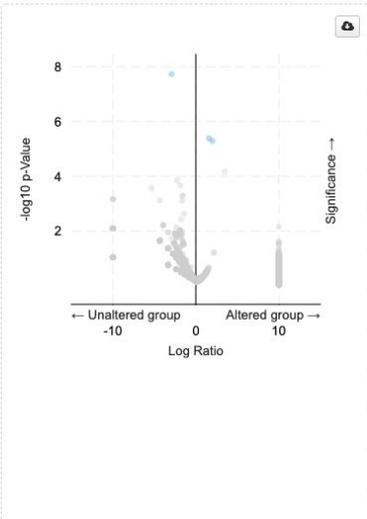
Sample-level enrichments Enriched in ... Significant only

Gene	Cytoband	Altered group	Unaltered group	Co-occurrence Pattern	Log Ratio	p-Value	q-Value	Enriched in
<input type="checkbox"/> NF1	17q11.2	20 (4.30%)	15 (32.61%)		-2.92	1.85e-8	3.797e-4	Unaltered group
<input type="checkbox"/> TP53	17p13.1	241 (51.83%)	8 (17.39%)		1.58	4.118e-6	0.0349	Altered group
<input type="checkbox"/> ATRX	Xq21.1	201 (43.23%)	5 (10.87%)		1.99	5.109e-6	0.0349	Altered group
<input type="checkbox"/> CIC	19p13.2	114 (24.52%)	1 (2.17%)		2.50	6.892e-5	0.253	Altered group

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

Click the checkbox next to a gene name and then click this button to re-run the query with a gene added.

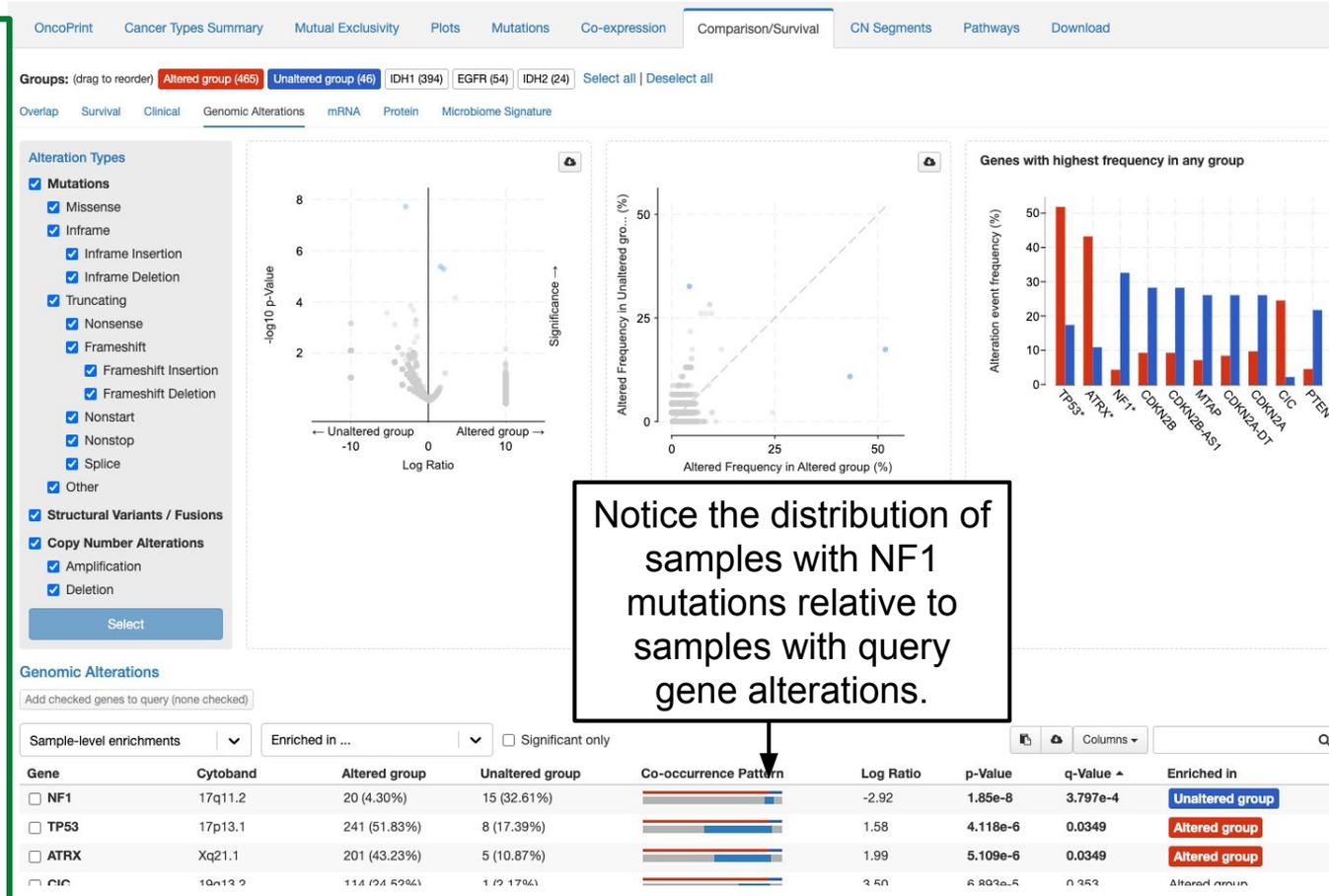
Hover over a dot to see the gene name



Comparison

Q: Alterations in IDH1, IDH2 and EGFR are mutually exclusive but some samples have alterations in none of these genes. Do samples without IDH1, IDH2 or EGFR alterations commonly have genomic alterations in one or more other genes?

A: Alterations in NF1 are significantly mutually exclusive with alterations in IDH1, IDH2 and EGFR (see table). Try adding NF1 to the query (check the box next to NF1 and then click “Add checked genes to query”) and examine the OncoPrint and the Mutual Exclusivity tabs.



CN Segments

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

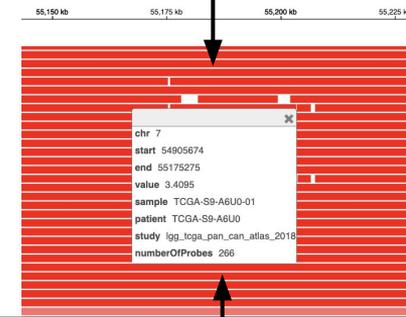


Toggle track labels, a vertical line marking the center of the viewing screen, and a vertical line that moves with your cursor. Use to zoom in or out.

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

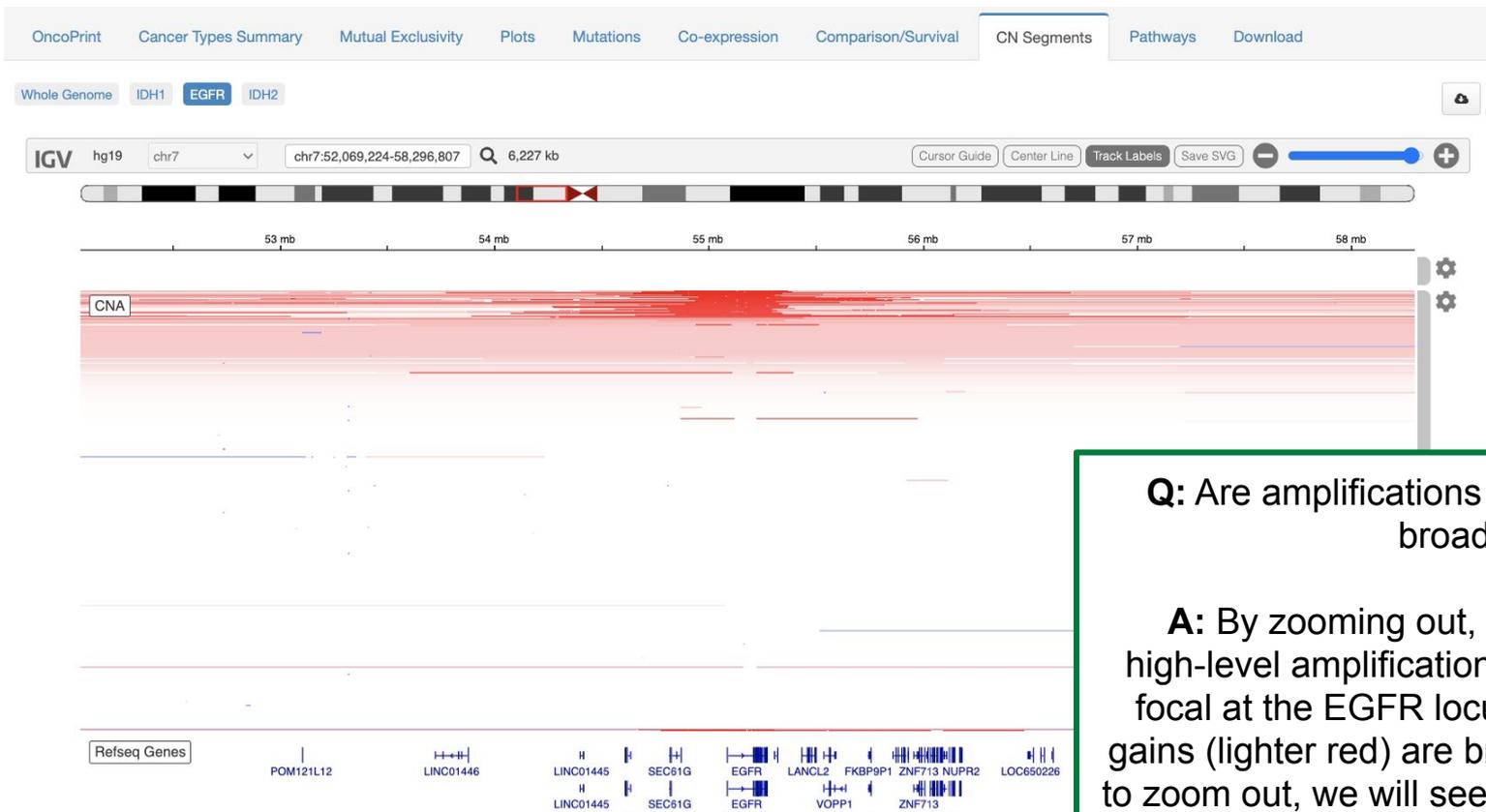
Each row is a single sample

Gene structures



Click on a read for details

CN Segments

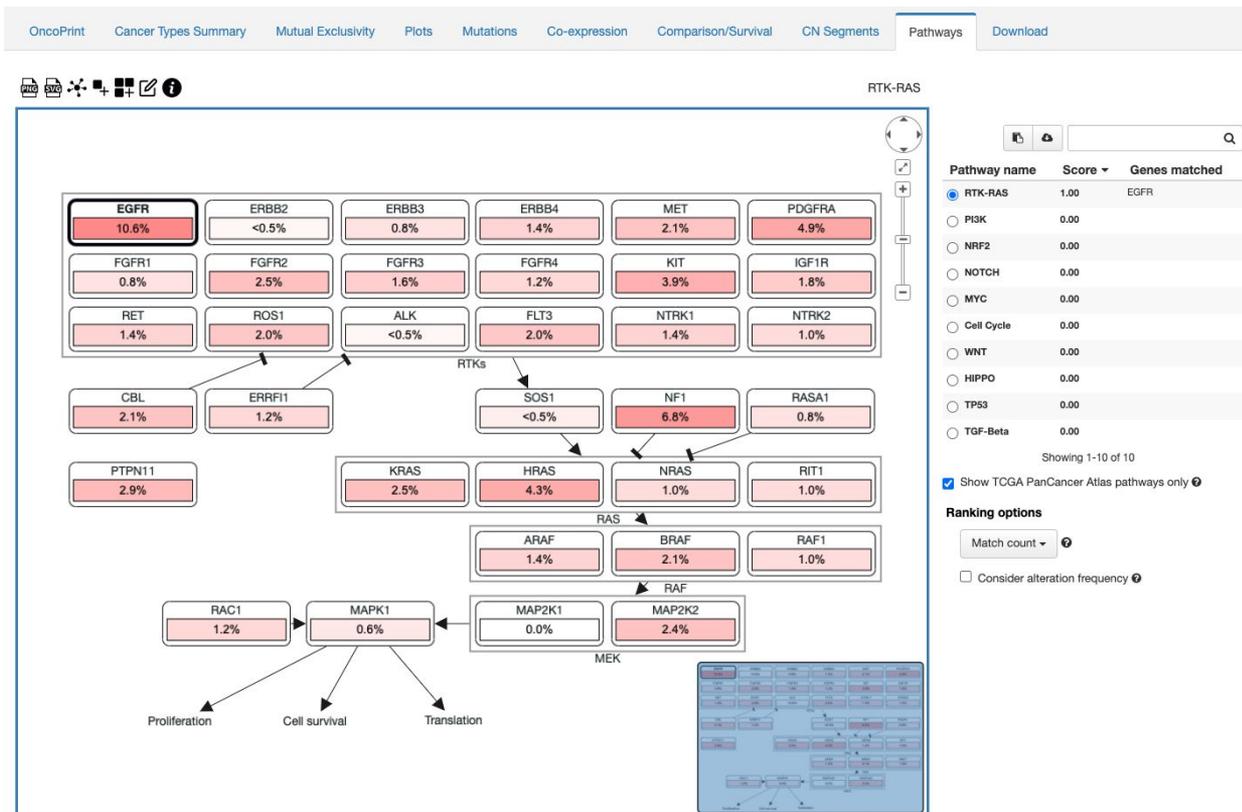


Q: Are amplifications of EGFR focal or broad?

A: By zooming out, we can see that high-level amplifications (deeper red) are focal at the EGFR locus, while low-level gains (lighter red) are broad. If we continue to zoom out, we will see that low-level gains often encompass the entire chromosome.

Pathways

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



Download

Download data or copy lists of samples.

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

Downloadable Data Files

Copy-number Alterations (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Mutations (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Structural Variants (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Altered samples: List of samples with alterations	Copy Download Query Virtual Study
Unaltered samples: List of samples without any alteration	Copy Download Query Virtual Study
Sample matrix: List of all samples where 1=altered and 0=unaltered	Copy Download
Log2 copy-number values	Tab Delimited Format Transposed Matrix
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2)	Tab Delimited Format Transposed Matrix
mRNA expression z-scores relative to diploid samples (RNA Seq V2 RSEM)	Tab Delimited Format Transposed Matrix
mRNA expression z-scores relative to all samples (log RNA Seq V2 RSEM)	Tab Delimited Format Transposed Matrix
Protein expression (RPPA)	Tab Delimited Format Transposed Matrix
Protein expression z-scores (RPPA)	Tab Delimited Format Transposed Matrix
Putative arm-level copy-number from GISTIC	Tracks added in the OncoPrint tab can be downloaded here.
Microbiome Signatures (log RNA Seq CPM)	Tracks added in the OncoPrint tab can be downloaded here.

Download queried data types for the queried genes.

Download all other data types for the queried genes.

Gene Alteration Frequency

Gene Symbol	Num Samples Altered	Percent Samples Altered
IDH1	394	77%
EGFR	54	11%
IDH2	24	5%

Showing 1-3 of 3

Frequency of gene alteration for each gene in the query

Type of Genetic Alterations Across All Samples

Study ID	Sample ID	Patient ID	Altered	IDH1	EGFR	IDH2
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4938-01	TCGA-CS-4938	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4941-01	TCGA-CS-4941	1	no alteration	AMP (Driver)	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4942-01	TCGA-CS-4942	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4943-01	TCGA-CS-4943	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4944-01	TCGA-CS-4944	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-5390-01	TCGA-CS-5390	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-5393-01	TCGA-CS-5393	1	R132H (Driver)	AMP (Driver)	no alteration

List of all samples with status of each query gene.

Download

Download data or copy lists of samples.

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

Downloadable Data Files

Copy-number Alterations (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Mutations (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Structural Variants (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Altered samples: List of samples with alterations	Copy Download Query Virtual Study
Unaltered samples: List of samples without any alteration	Copy Download Query Virtual Study
Sample matrix: List of all samples where 1=altered and 0=unaltered	Copy Download
Log2 copy-number values	Tab Delimited Format Transposed Matrix
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2)	Tab Delimited Format Transposed Matrix
mRNA expression z-scores relative to diploid samples (RNA Seq V2 RSEM)	Tab Delimited Format Transposed Matrix
mRNA expression z-scores relative to all samples (log RNA Seq V2 RSEM)	Tab Delimited Format Transposed Matrix
Protein expression (RPPA)	Tab Delimited Format Transposed Matrix
Protein expression z-scores (RPPA)	Tab Delimited Format Transposed Matrix
Putative arm-level copy-number from GISTIC	Tracks added in the OncoPrint tab can be downloaded here.
Microbiome Signatures (log RNA Seq CPM)	Tracks added in the OncoPrint tab can be downloaded here.

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

List of samples that have no alterations in any query genes

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

Gene Alteration Frequency

Gene Symbol	Num Samples Altered	Percent Samples Altered
IDH1	394	77%
EGFR	54	11%
IDH2	24	5%

Showing 1-3 of 3

Type of Genetic Alterations Across All Samples

Study ID	Sample ID	Patient ID	Altered	IDH1	EGFR	IDH2
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4938-01	TCGA-CS-4938	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4941-01	TCGA-CS-4941	1	no alteration	AMP (Driver)	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4942-01	TCGA-CS-4942	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4943-01	TCGA-CS-4943	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4944-01	TCGA-CS-4944	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-5390-01	TCGA-CS-5390	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-5393-01	TCGA-CS-5393	1	R132H (Driver)	AMP (Driver)	no alteration

Advanced feature: use these lists to build a custom sample list to run a new query, to create [virtual studies](#) or to build [custom groups](#).

Questions?

Check out our other tutorials
or email us at:

cbioportal@googlegroups.com