cBioPortal Tutorial #2: Single Study Query

Query one or multiple genes in a single dataset

Last update: December 21, 2021

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.

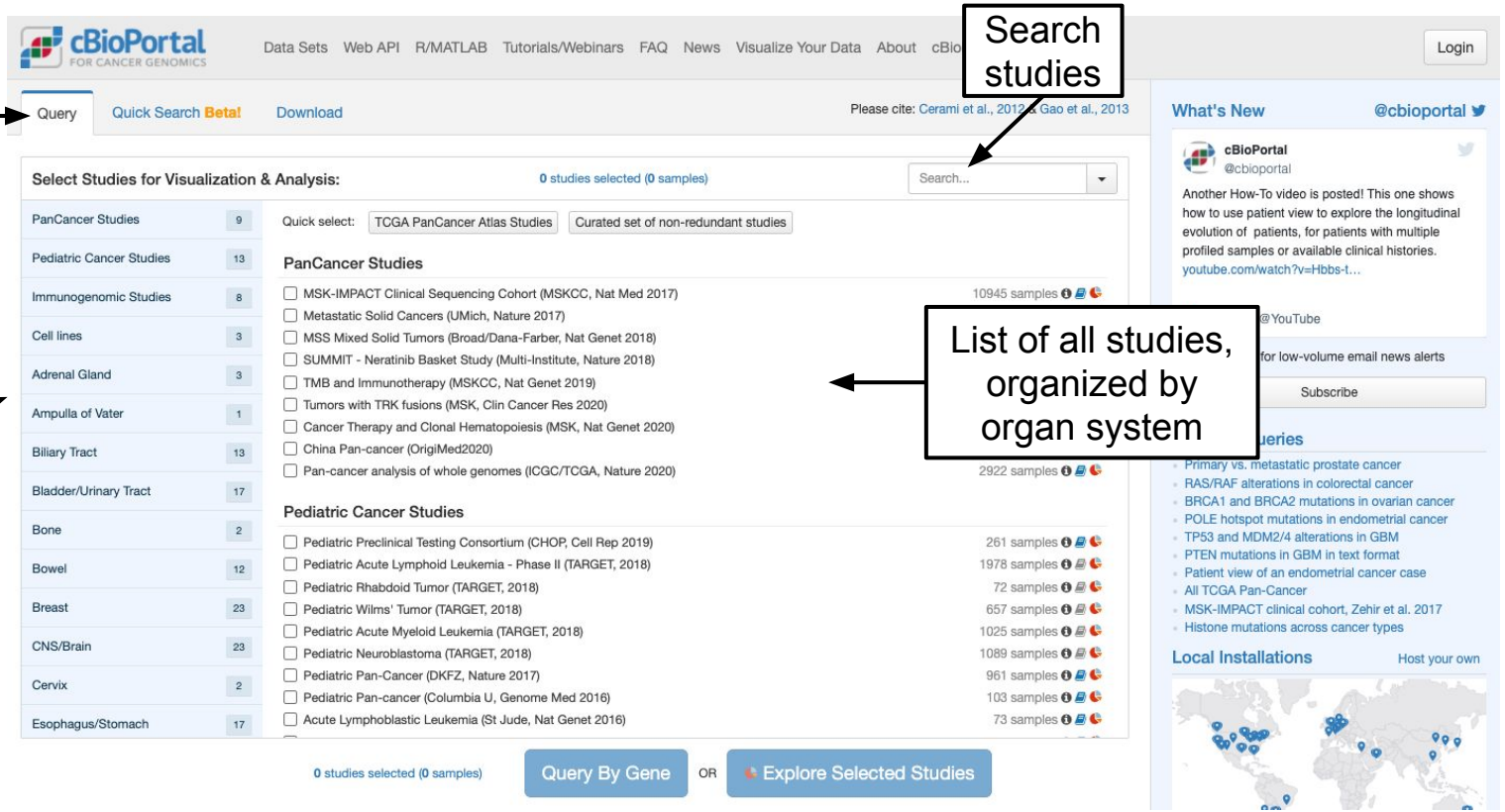
Overview of Tabs in a Single Study Query

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

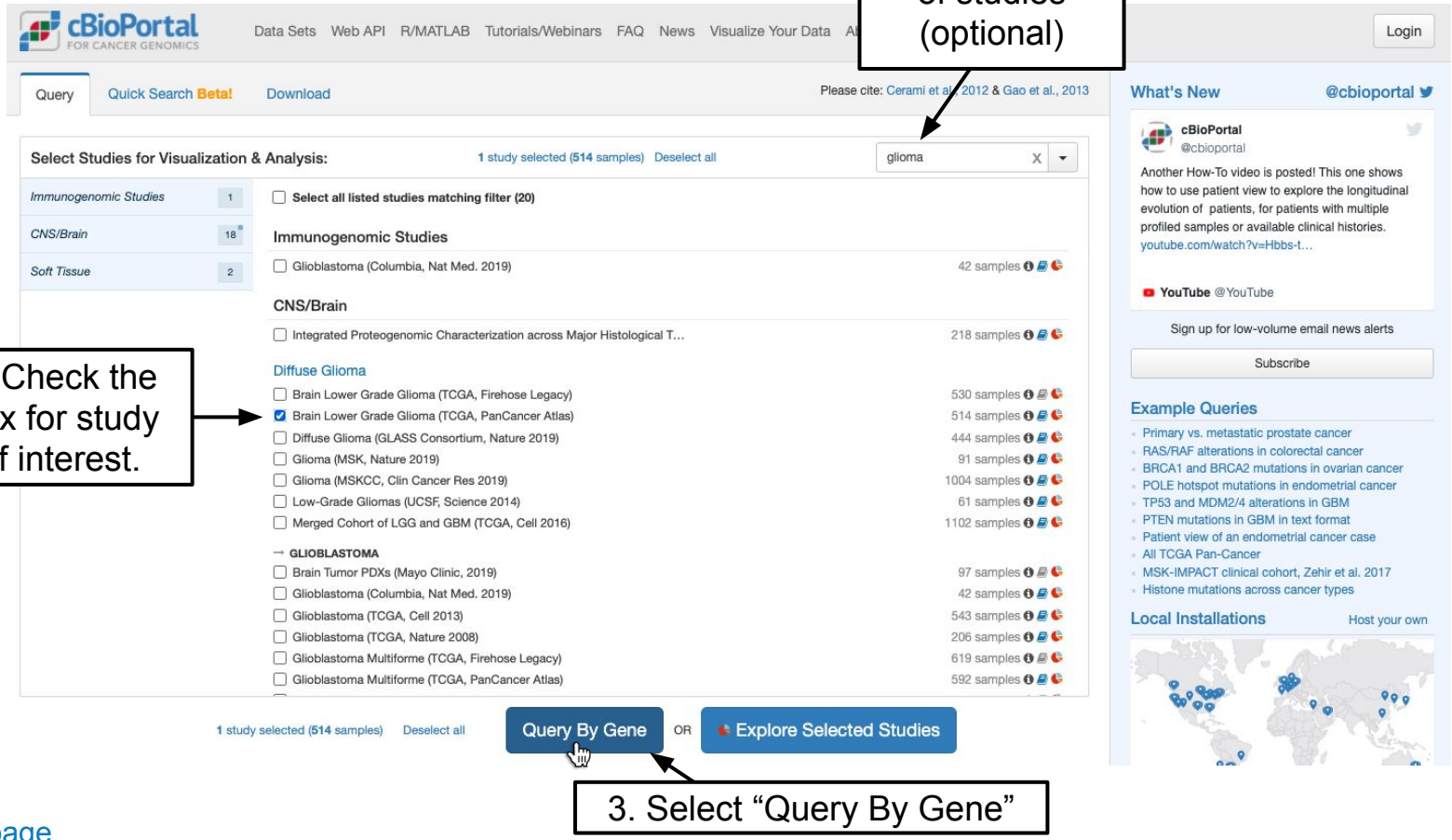
We’re going to run a query in a TCGA Lower-Grade Glioma study. The next few slides will show how to run this query from the Query page. You can also run the same query from a Single Study Exploration, as we did in Tutorial #1.

Query overview



Link to this page

Single study query

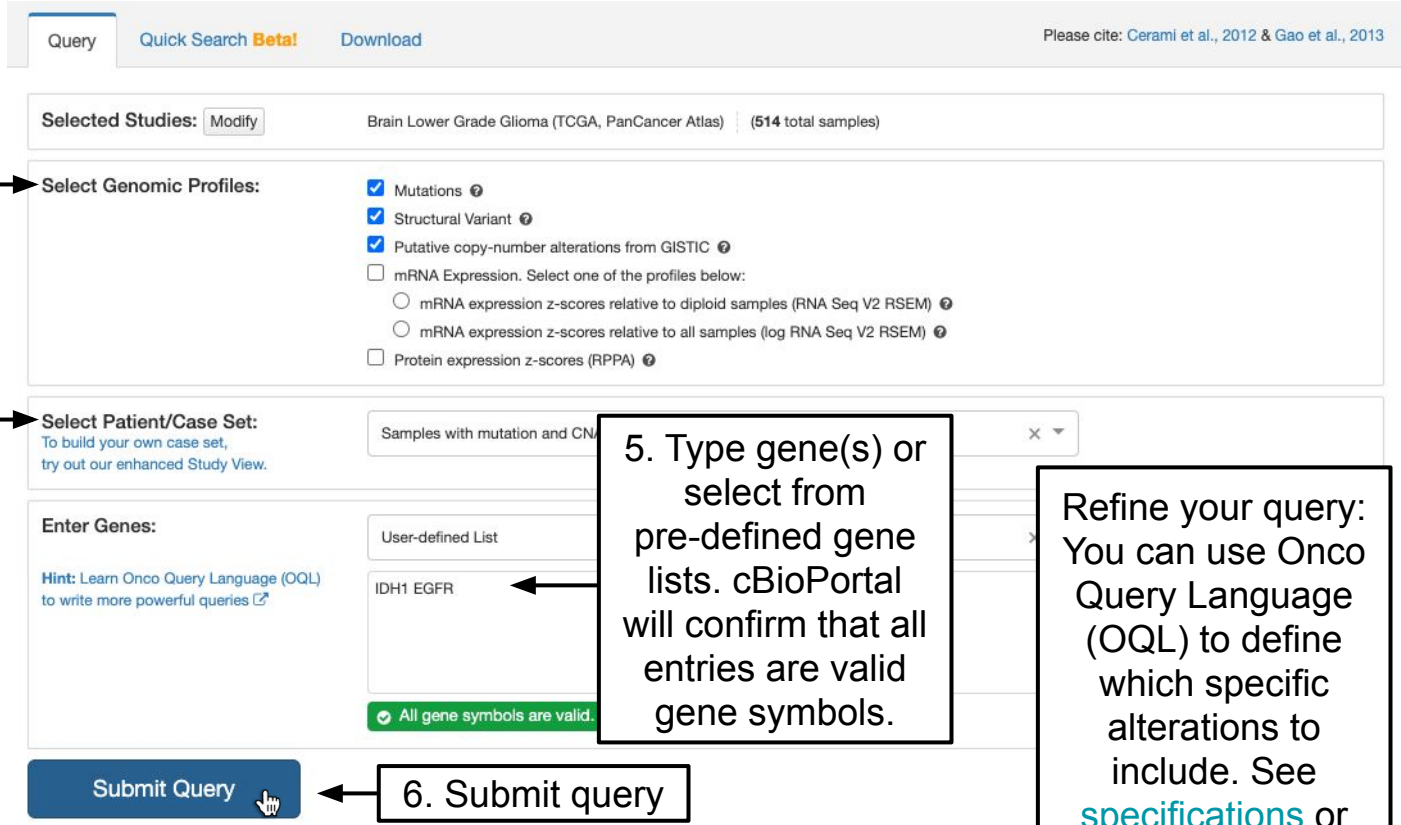


Link to this page

Single study query

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.

Link to this page

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 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.

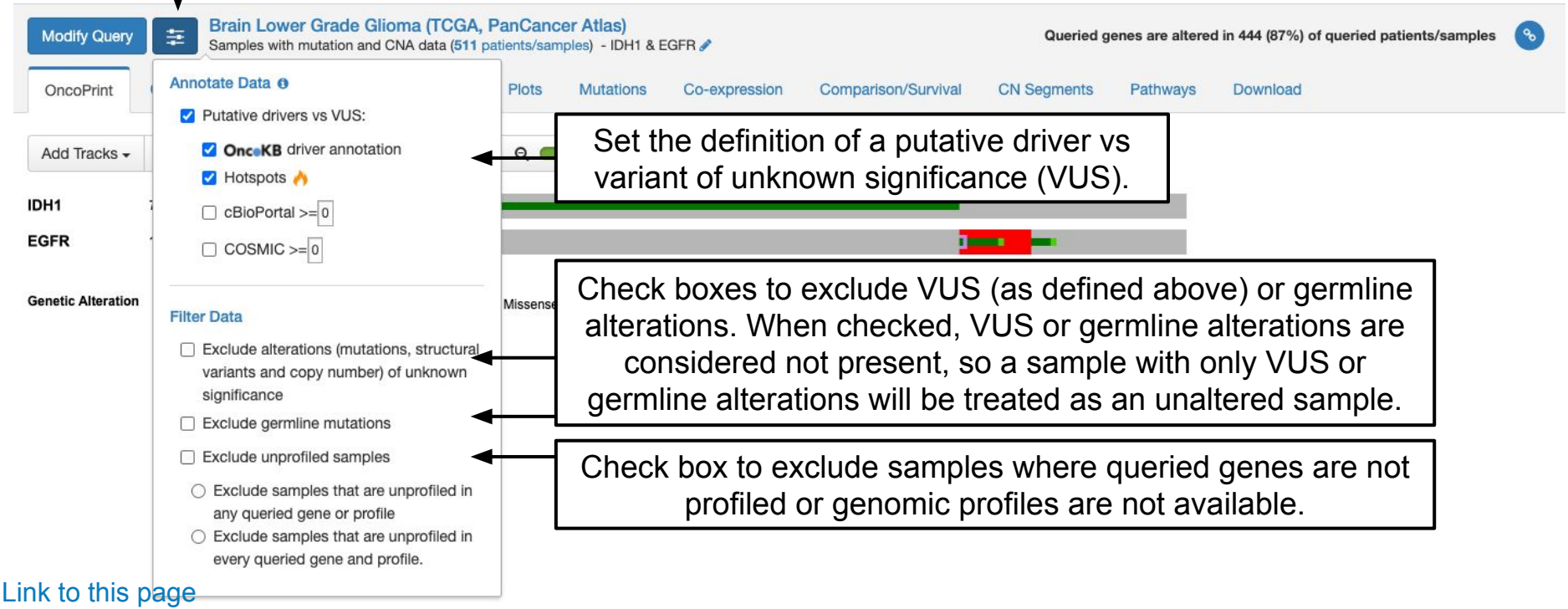
Click on the number of patients/samples to go to Study View for just the queried samples.

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

Link to this page

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.



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.

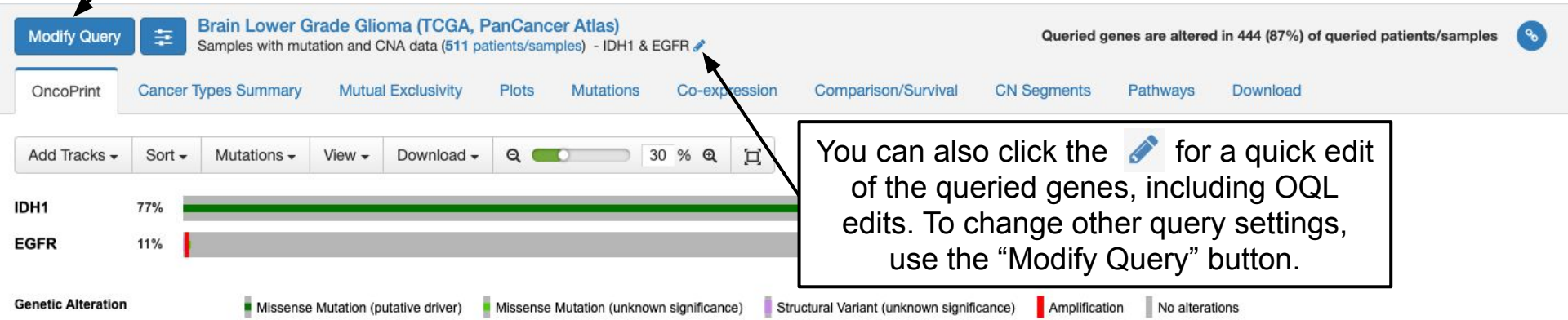
Check box to exclude samples where queried genes are not profiled or genomic profiles are not available.

Link to this page

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).

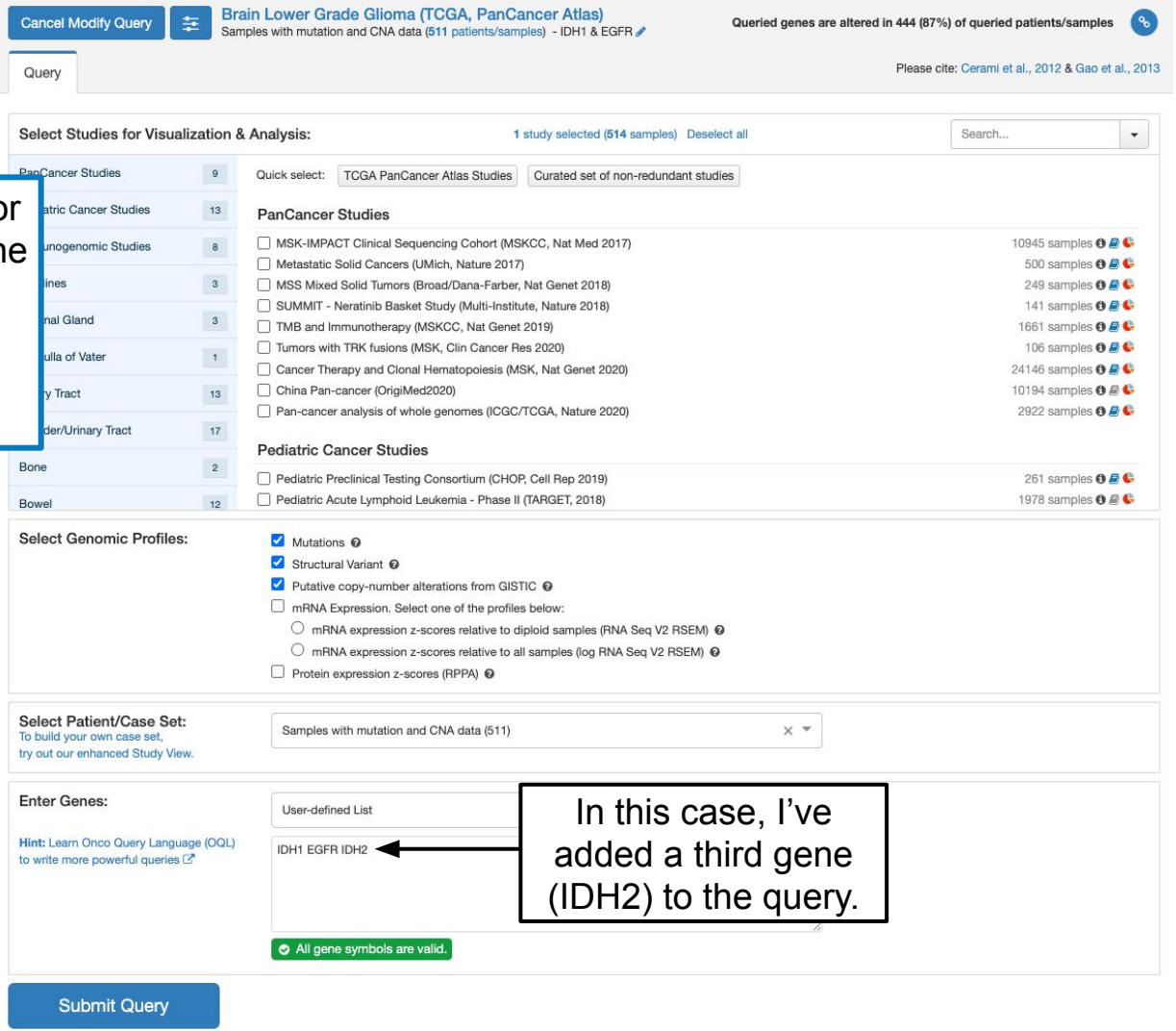


You can also click the for a quick edit of the queried genes, including OQL edits. To change other query settings, use the “Modify Query” button.

Link to this page

Modify Query

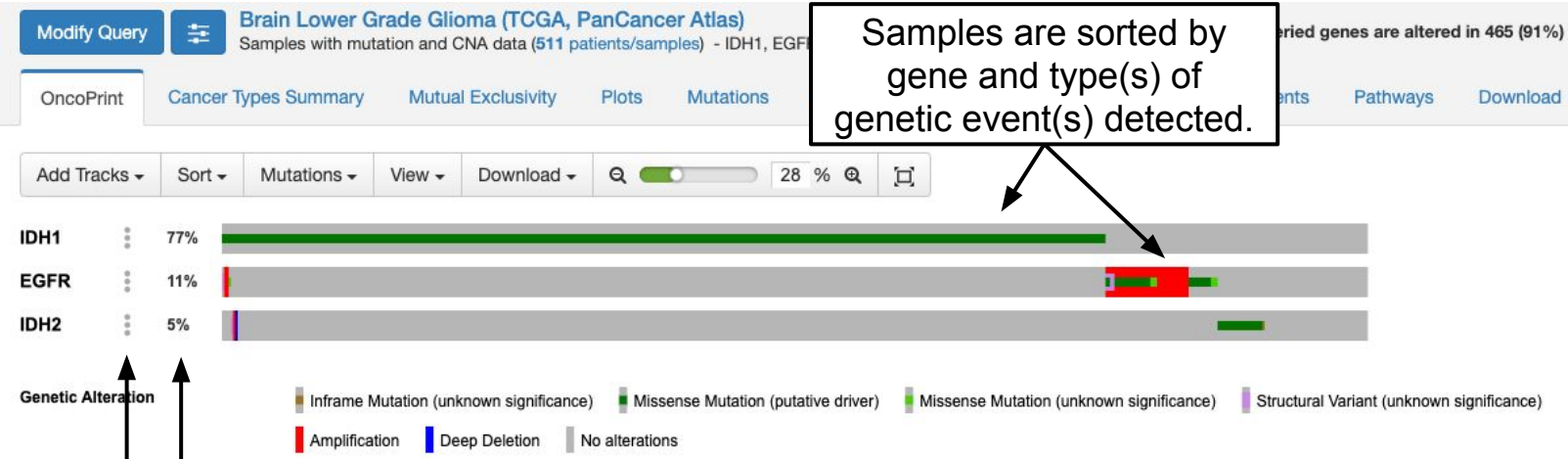
The existing query is pre-populated for your convenience. You can change the study, the genomic profiles, the patient/case set or the gene set. Simply hit “Submit” when you are happy with the modified query.



Link to this page

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.

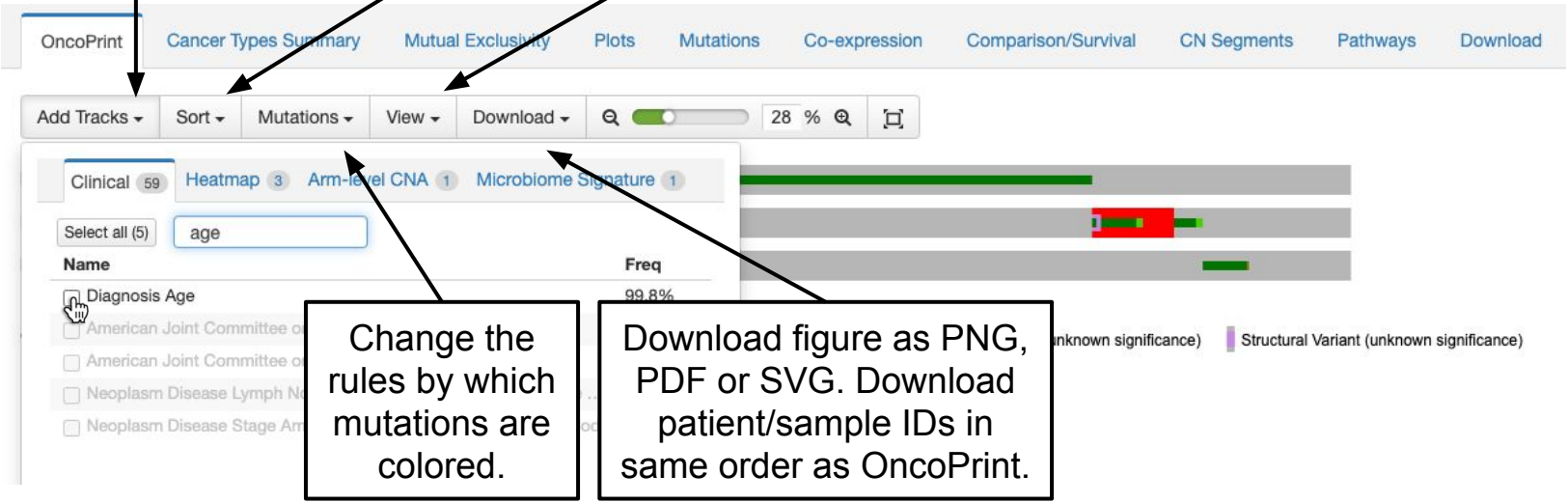
Link to this page

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.

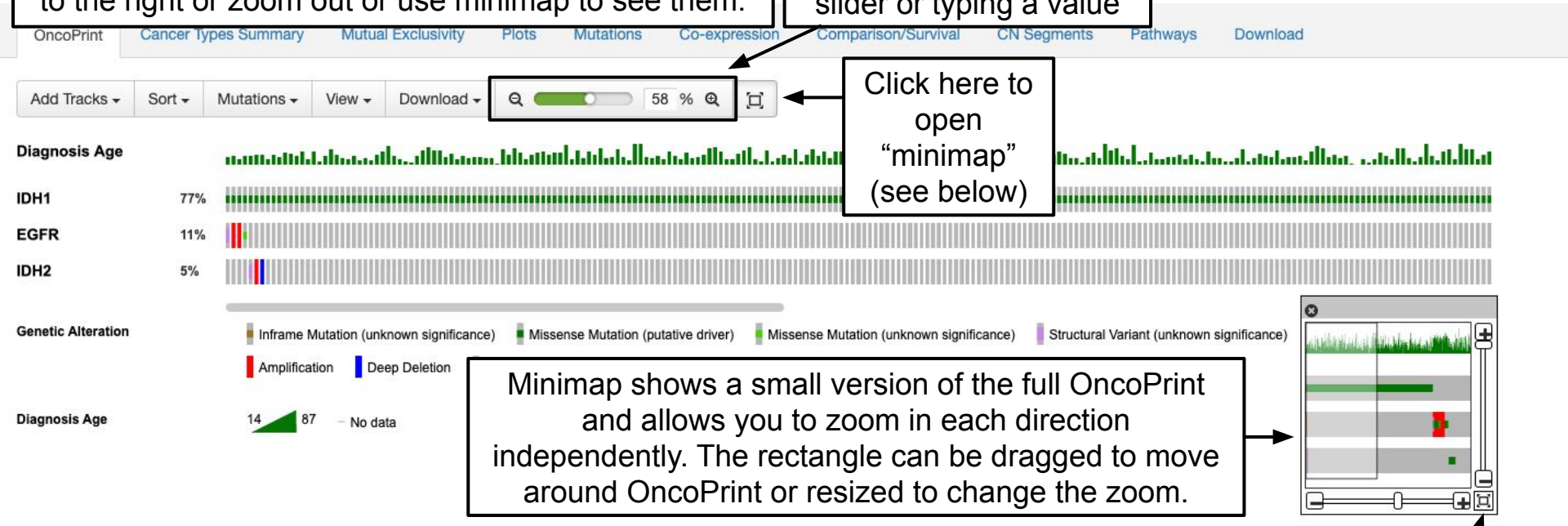
Download figure as PNG, PDF or SVG. Download patient/sample IDs in same order as OncoPrint.

Link to this page

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



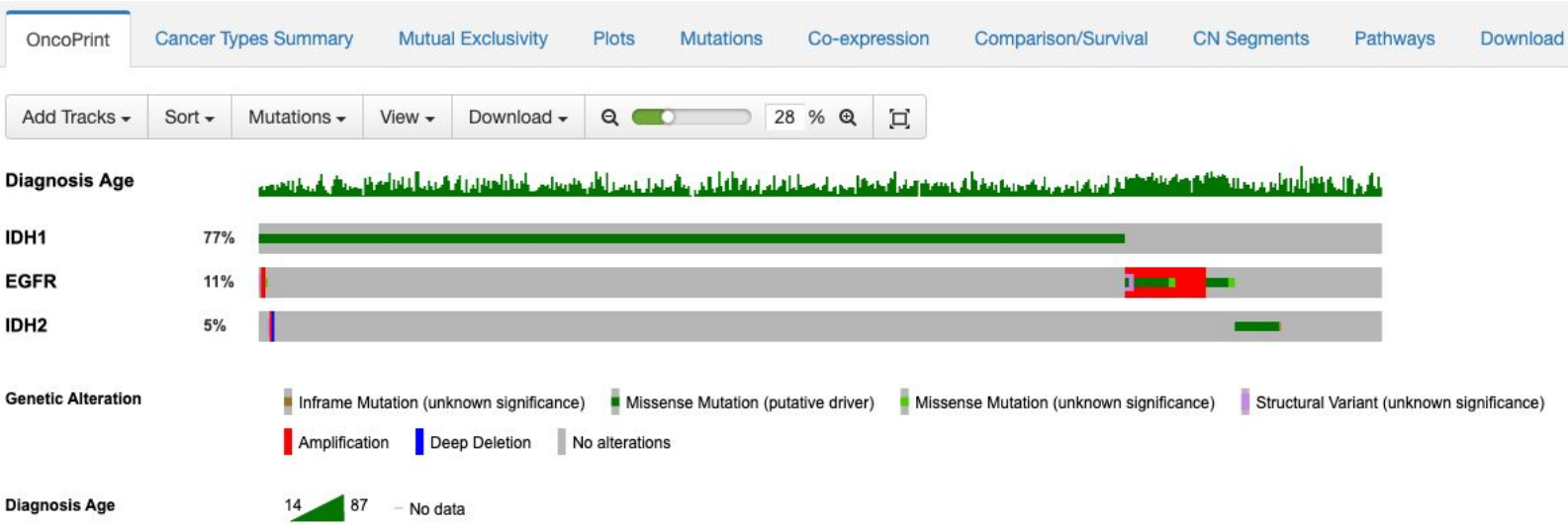
Click here to open "minimap" (see below)

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

Link to this page

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

Link to this page

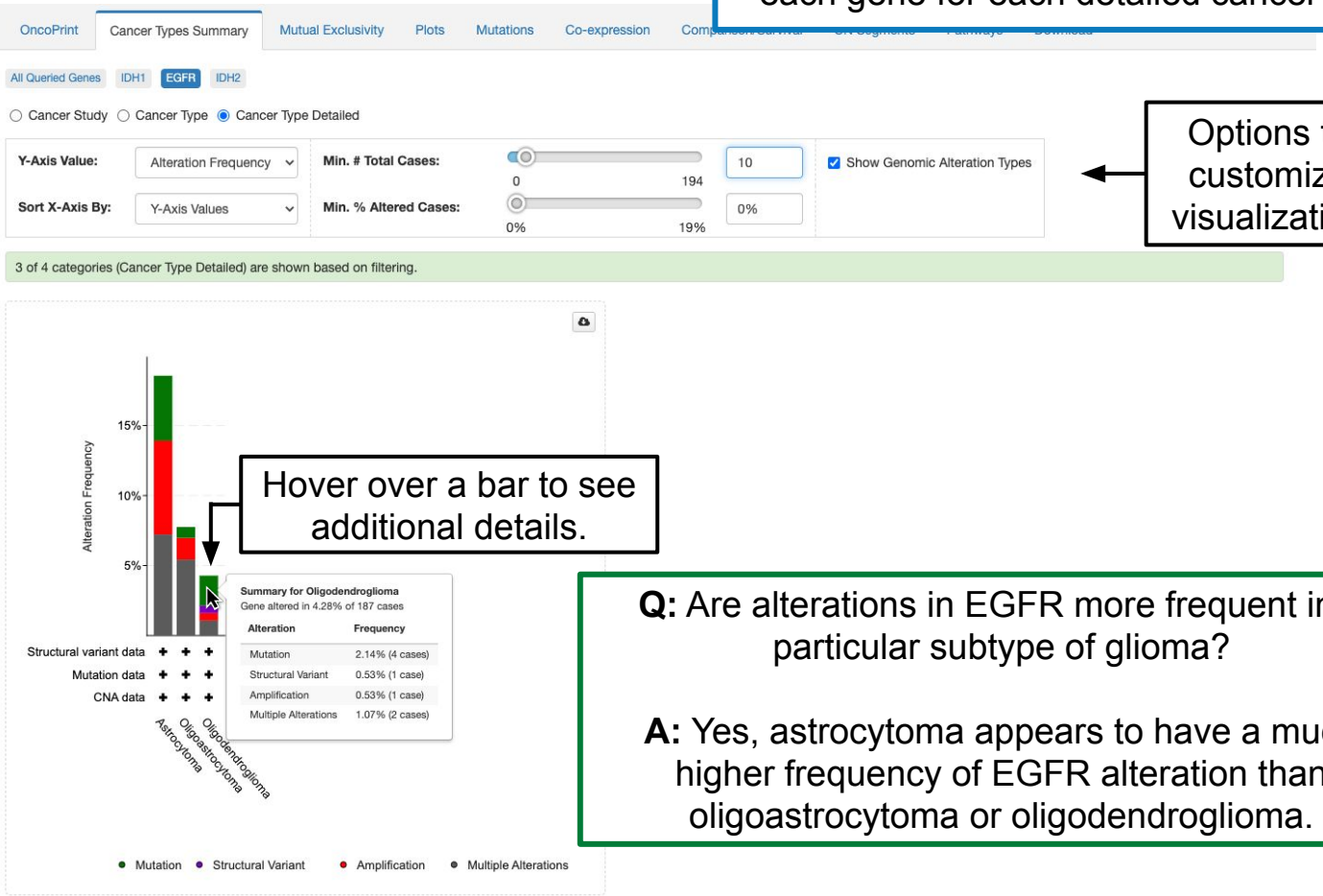
Now we're going to go through all the other tabs and ask some questions about alterations in *IDH1*, *IDH2* and *EGFR* in the TCGA Lower-Grade Glioma study.

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.



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?

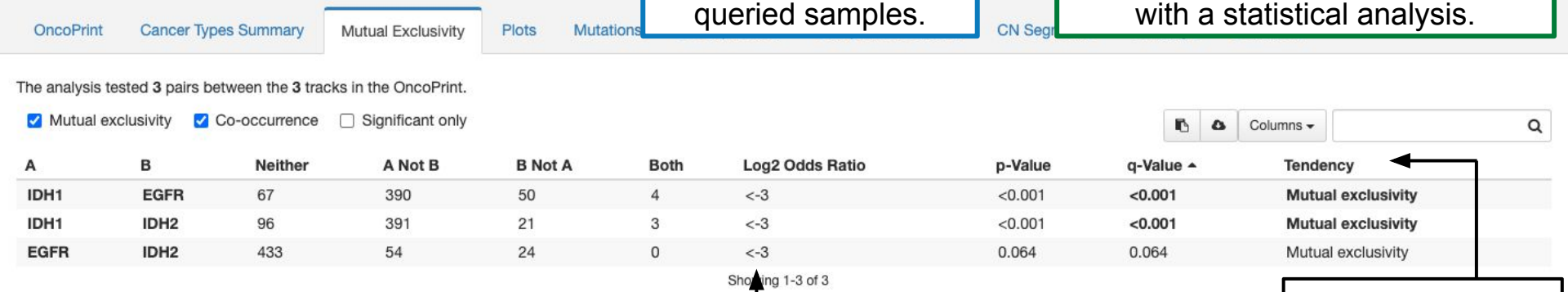
A: Yes, astrocytoma appears to have a much higher frequency of EGFR alteration than oligoastrocytoma or oligodendroglioma.

Link to this page

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.



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.

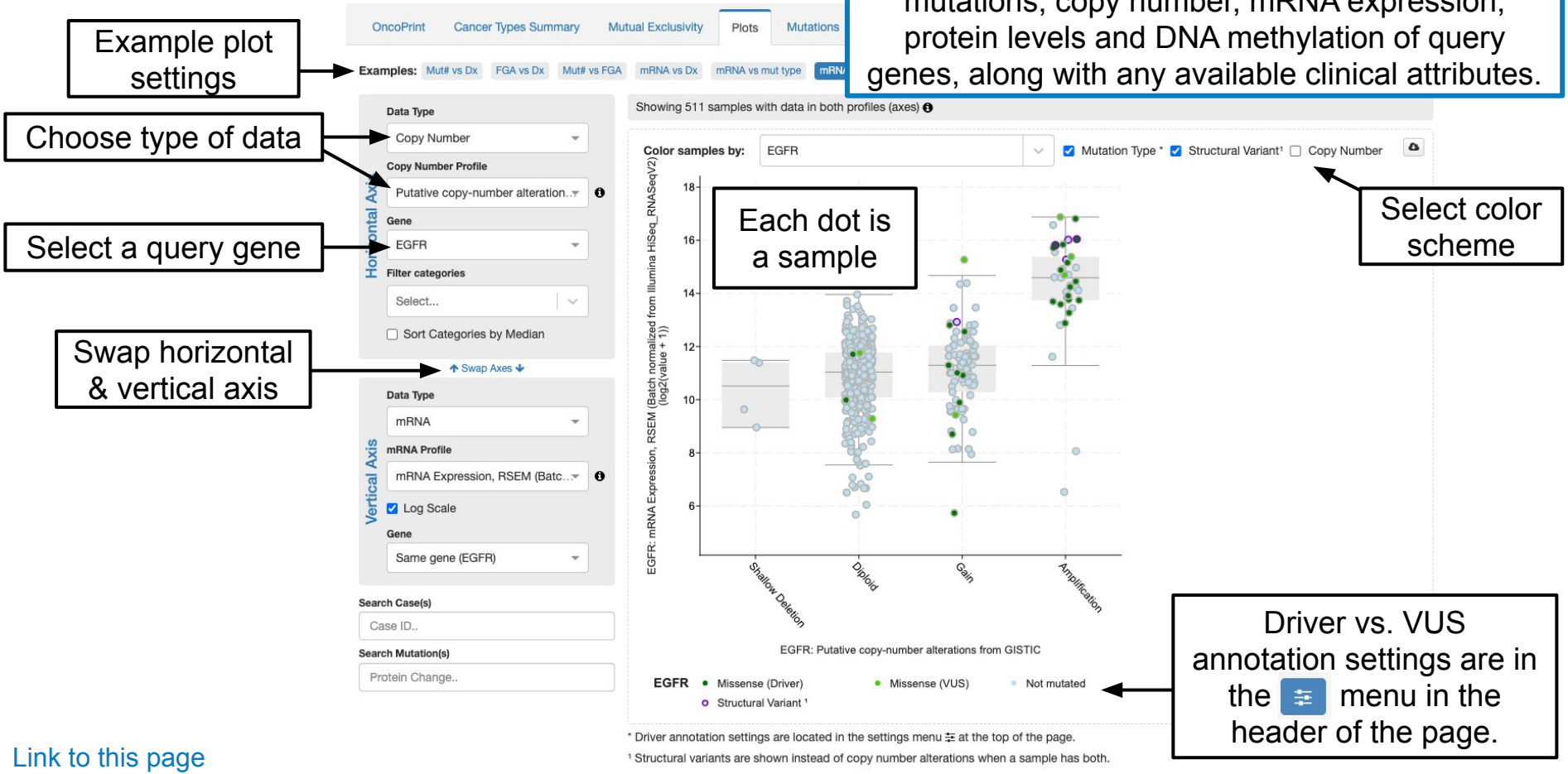
Click on any column header to sort. Hover over the column names for more details about how values are calculated.

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

Link to this page

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.



Link to this page

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).

Link to this page

Mutations

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

Each gene appears on a separate tab

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.

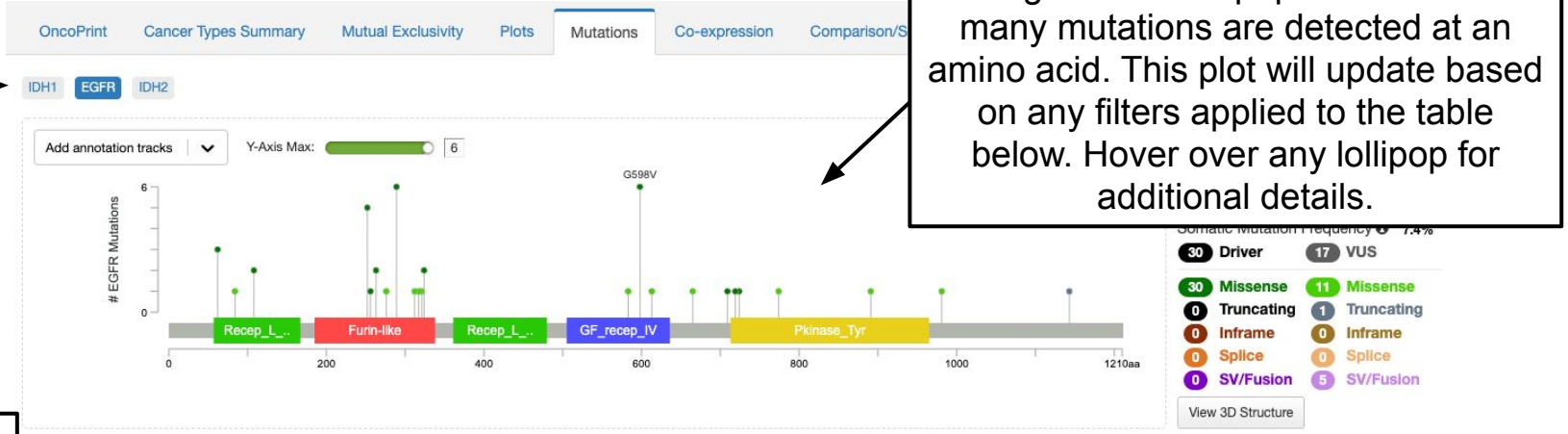


Table of all mutations with annotations

47 Mutations (page 1 of 2)

Columns (9 / 87)

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			45
TCGA-E1-A7YM...	Astrocytoma	*A289V*		Missense	Gain			35
TCGA-FG-6692-...	Oligodendroglioma	*A289V*		Missense	Amp			63
TCGA-FG-A70Z...	Oligoastrocytoma	*A289V*		Missense	Amp			47
TCGA-KT-A7W1...	Astrocytoma	*A289V*		Missense	Amp			39
TCGA-TM-A7C3...	Astrocytoma	*A289V*		Missense	Amp	50	0.71	56
TCGA-HT-8107-01	Oligodendroglioma	*R108K*		Missense	Diploid	17	0.15	1
TCGA-HT-8110-01	Astrocytoma	*R108K*		Missense	Amp	17	0.94	30
TCGA-HT-A61C...	Oligodendroglioma	*T263P*		Missense	Gain	7	0.24	38
TCGA-DU-7013...	Astrocytoma	*G598V*		Missense	Amp	36	0.96	36
TCGA-DU-8162...	Oligoastrocytoma	*G598V*		Missense	Amp	36	0.57	22

Show additional columns

Link to this page

Mutations

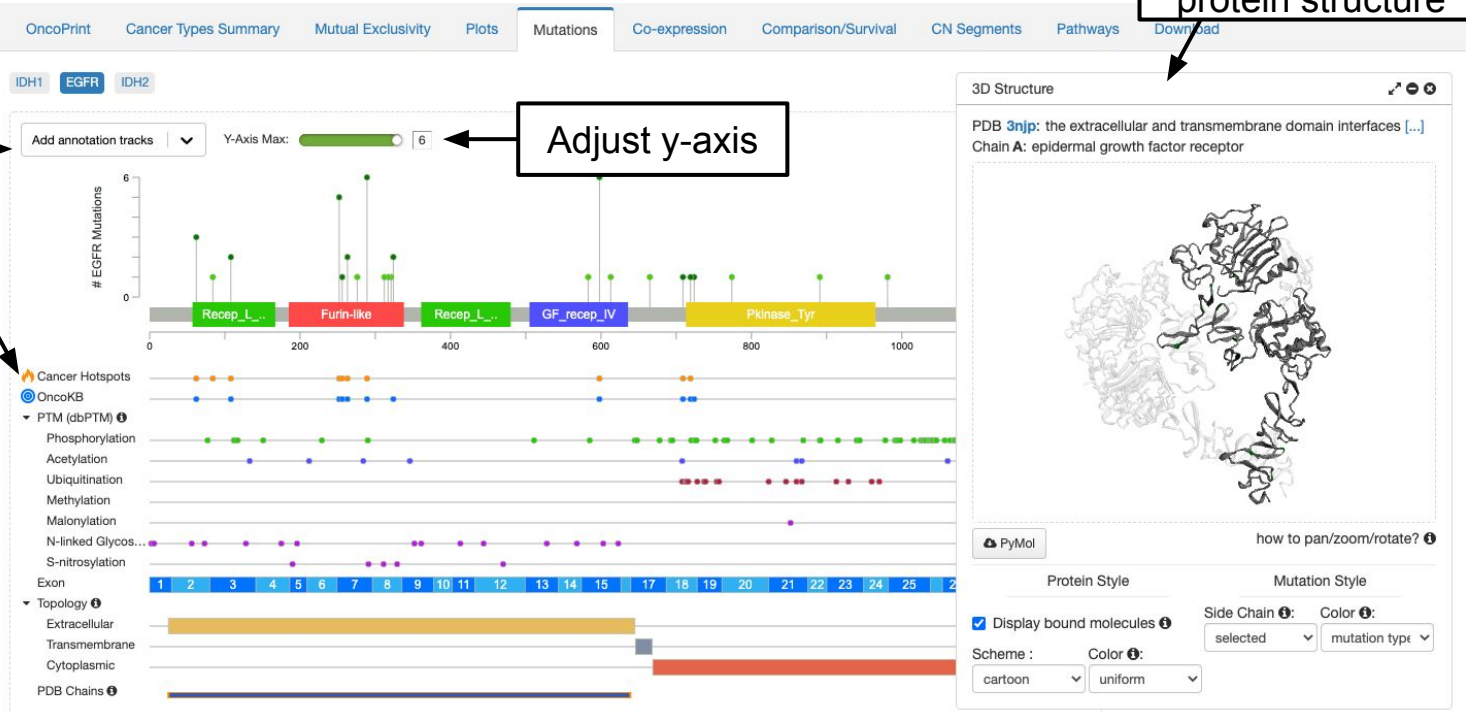
Link to this page

Mutations

View mutations in context of 3D protein structure

Add annotation tracks to the plot

Adjust y-axis



47 Mutations (page 1 of 2)

Columns (9 / 87)

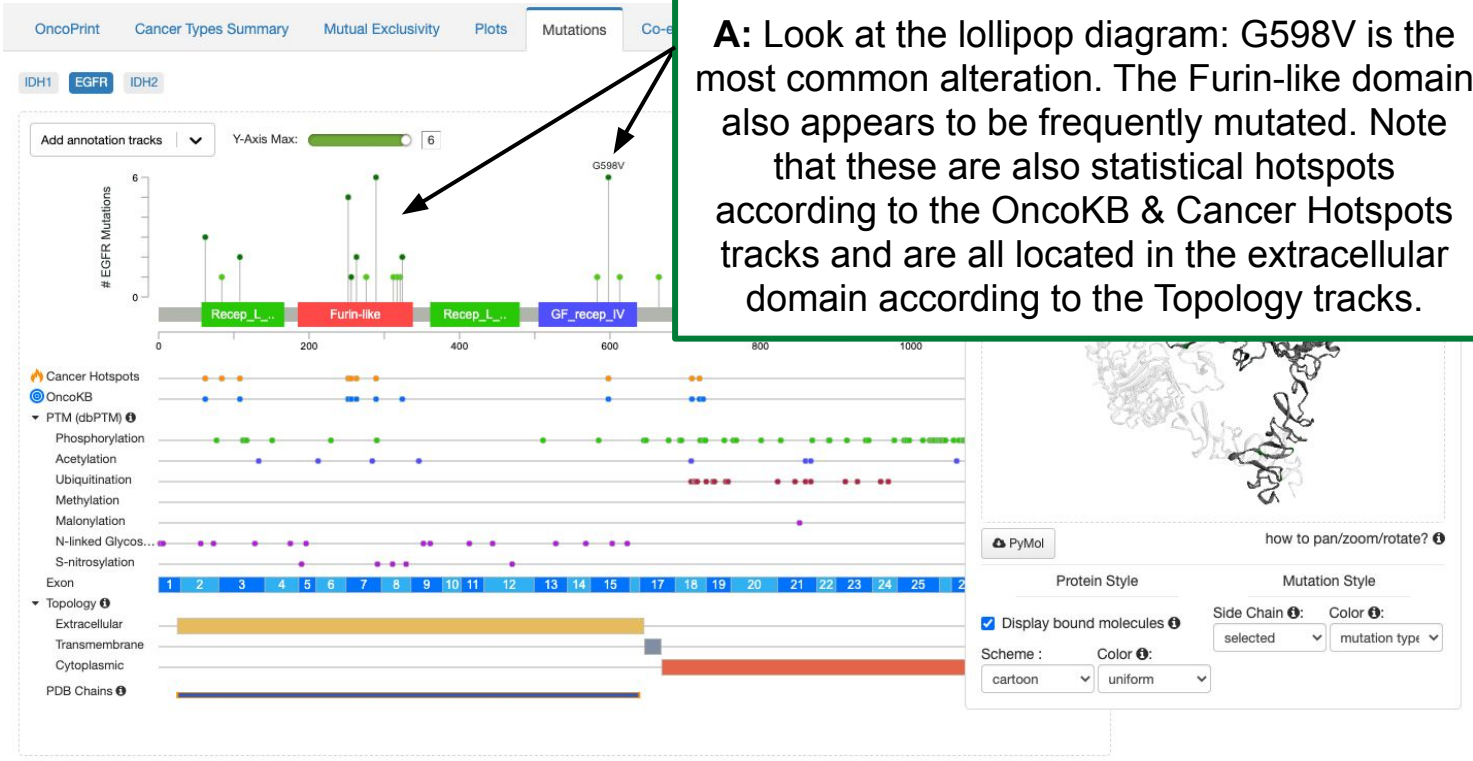
Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	*G719D*		Missense	Amp	125	0.18	39
TCGA-DU-A5TT...	Oligodendroglioma	*A289V*		Missense	Gain	50	0.45	45

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.

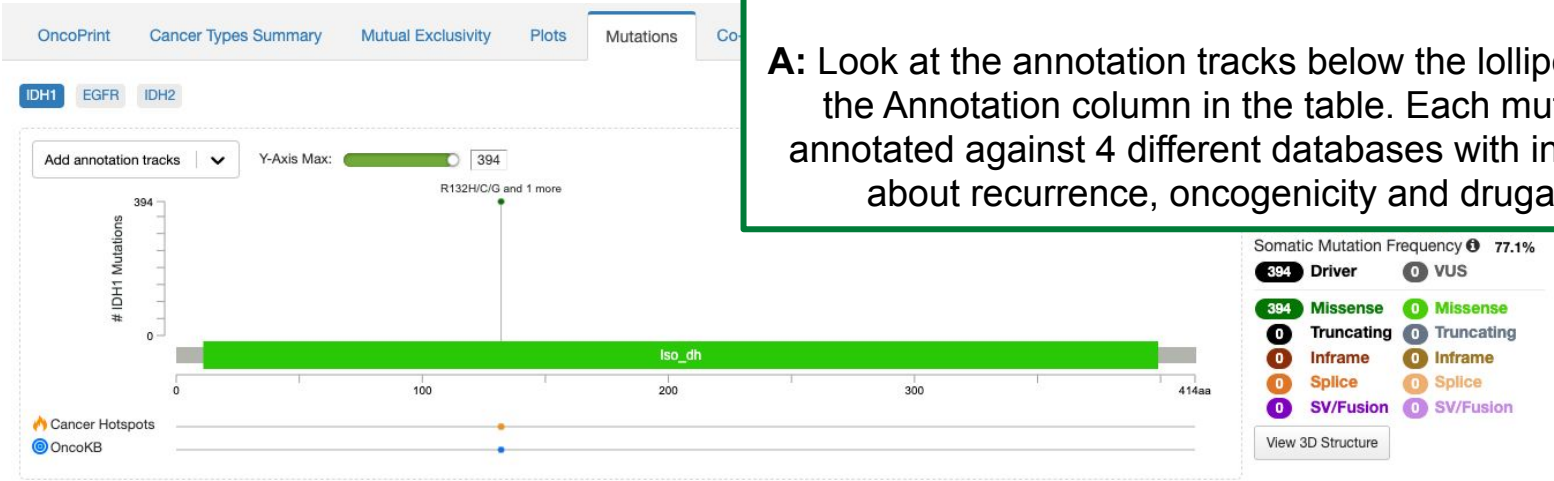


Link to this page

Mutations

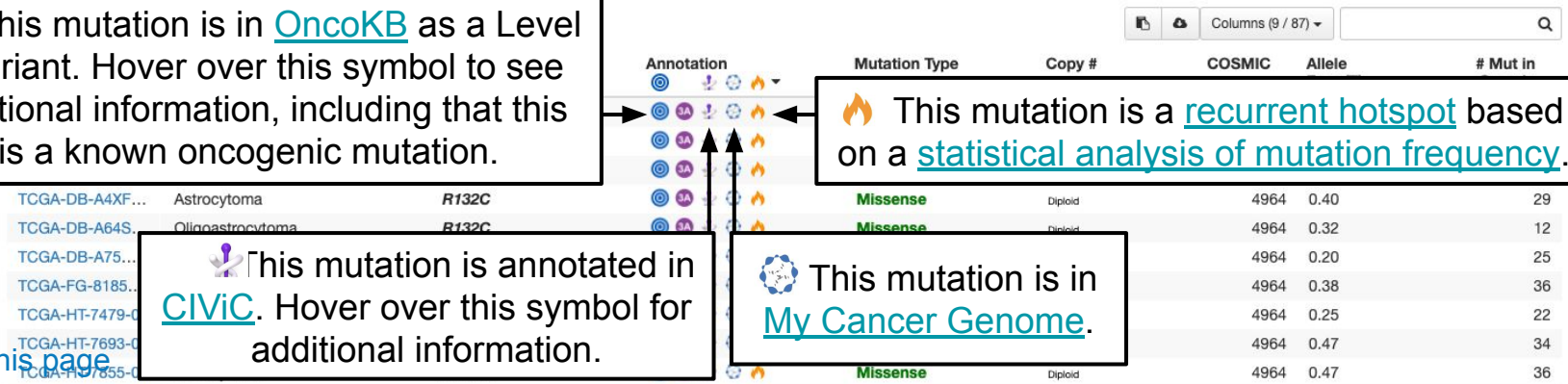
Q: The mutations in IDH1 appear to be highly recurrent. Are these mutations known hotspots? Known oncogenic drivers? Biomarkers for any drugs?

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.

Link to this page

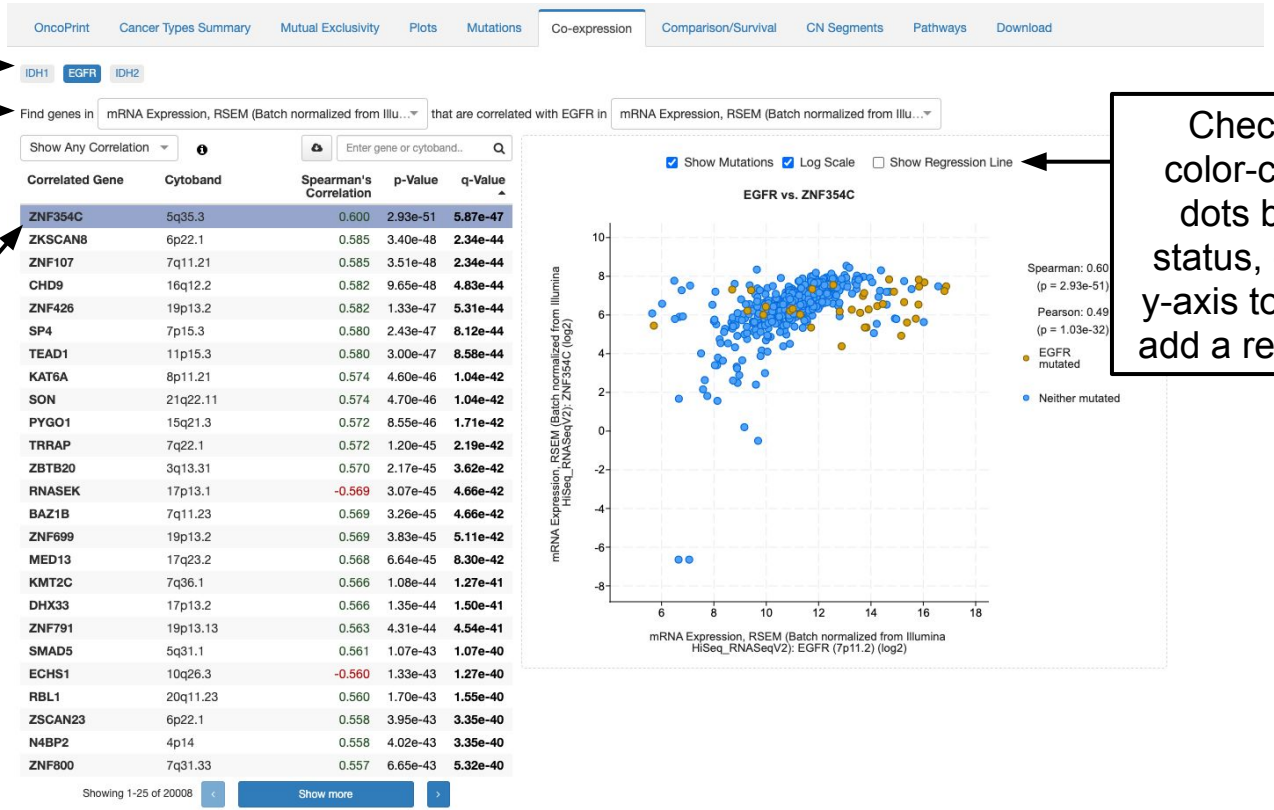
Co-Expression

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



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

Link to this page

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 Exc

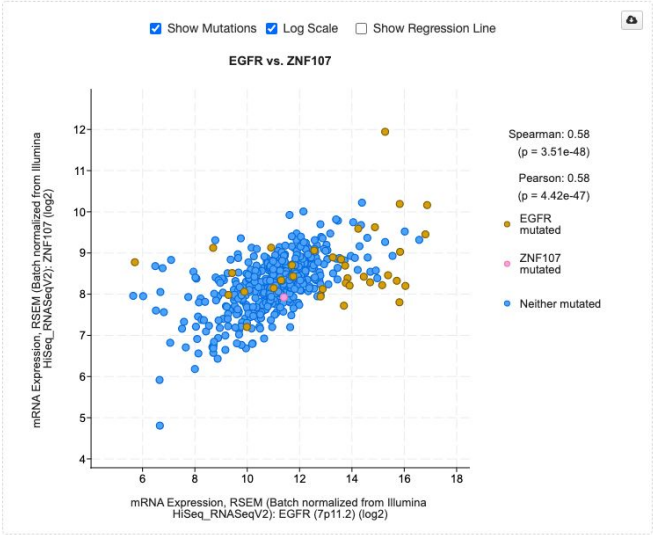
IDH1 | EGFR | IDH2

Find genes in mRNA Expression, RSEM (Batch normalize

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 | Show more

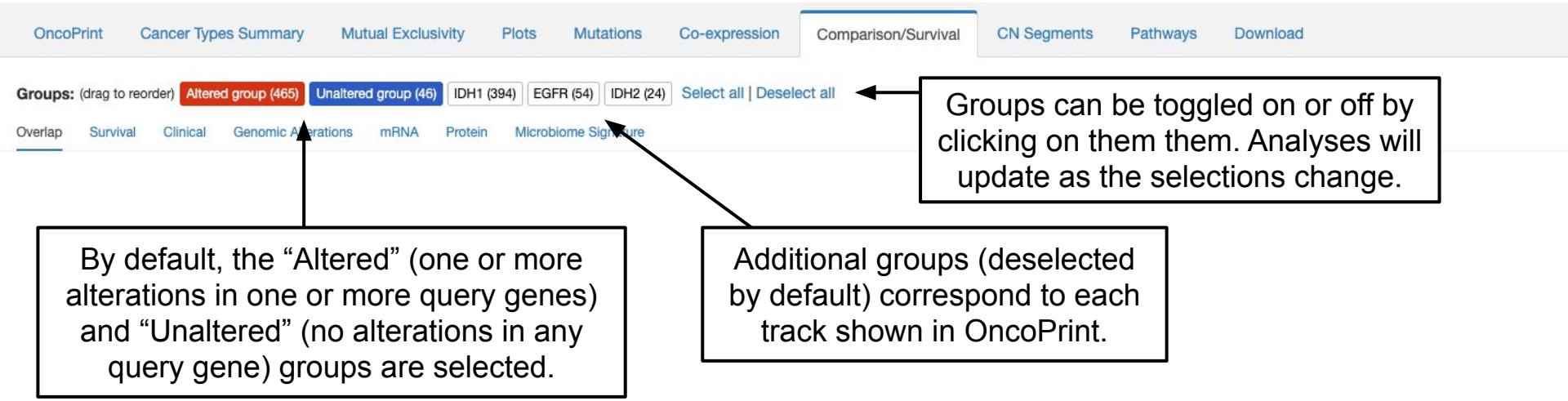


Link to this page

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.

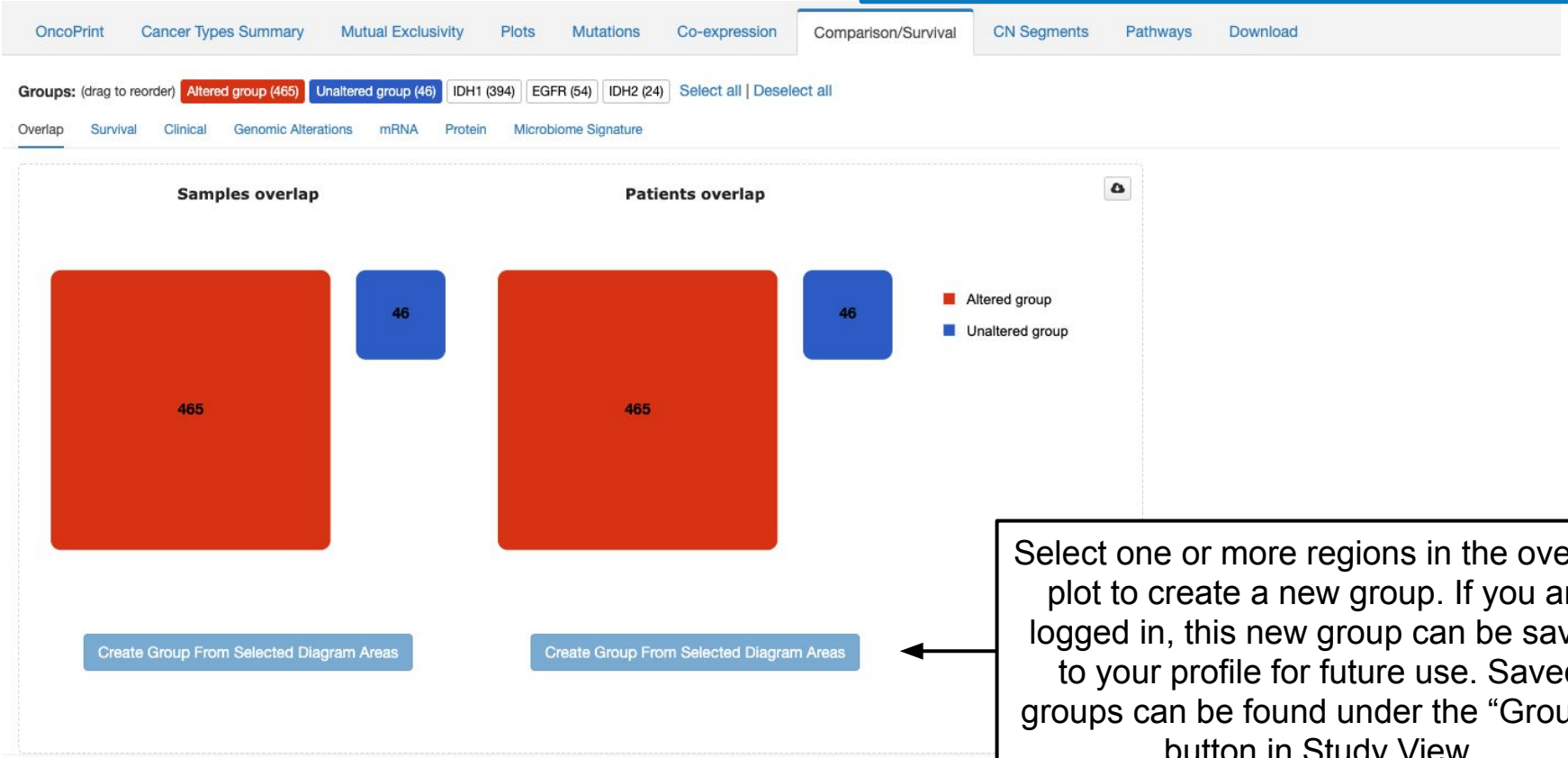
The Comparison tab is the same as the Group Comparison functionality that is accessible from Study View. See the Group Comparison Tutorial for more details about the functionality of this tab.



Link to this page

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.

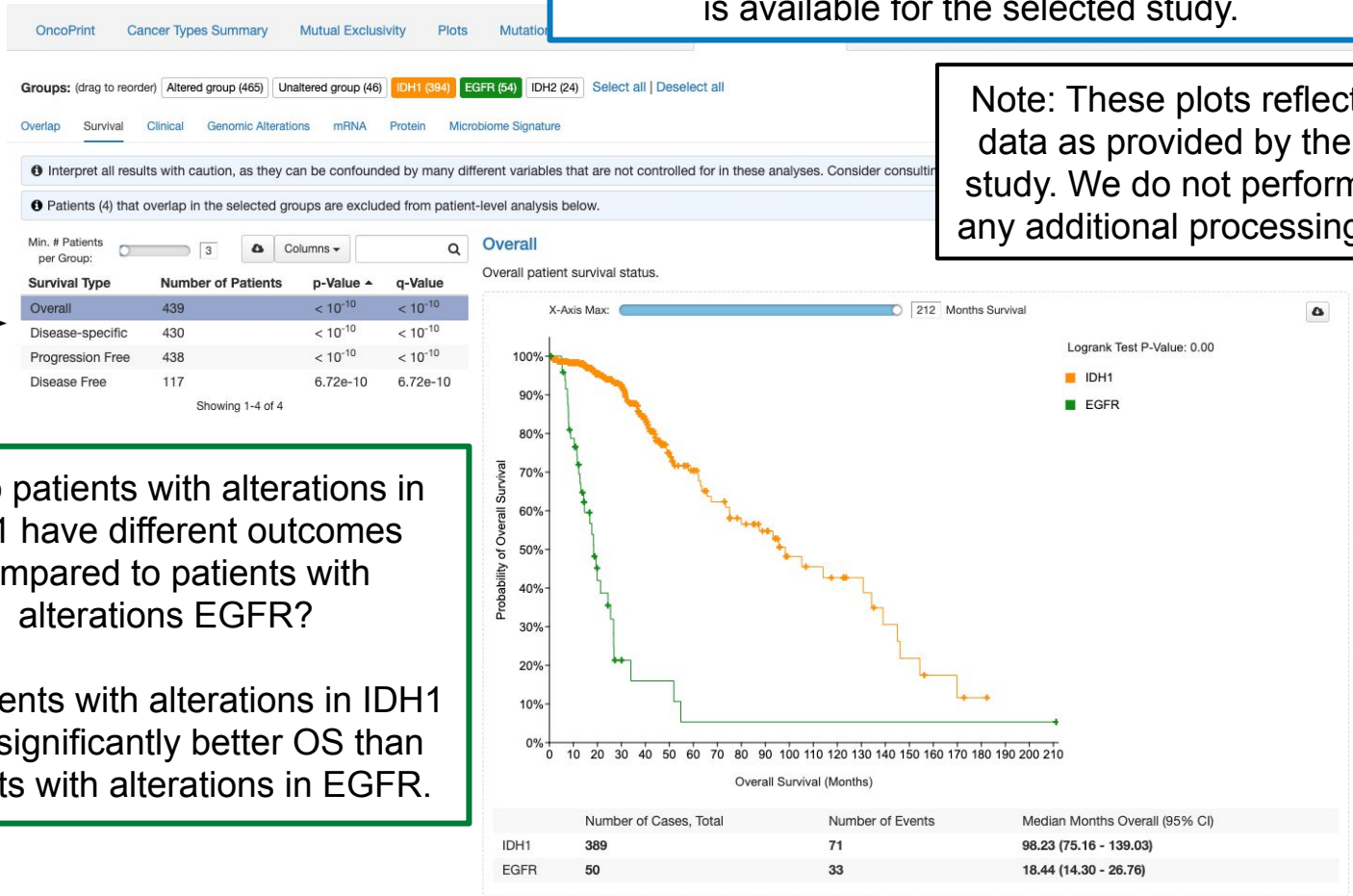
Link to this page

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?

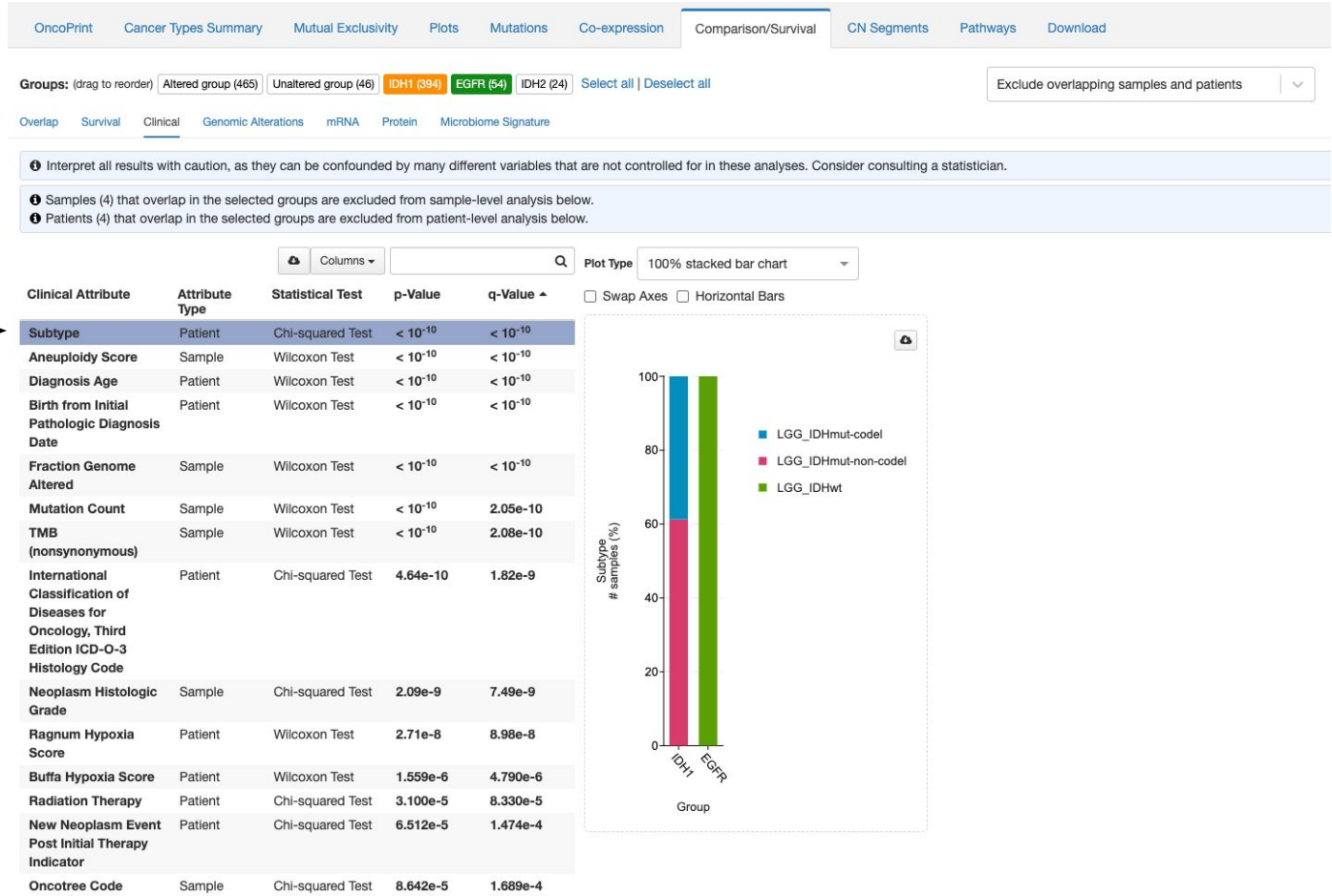
A: Patients with alterations in IDH1 have significantly better OS than patients with alterations in EGFR.

Link to this page

Comparison: Clinical

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

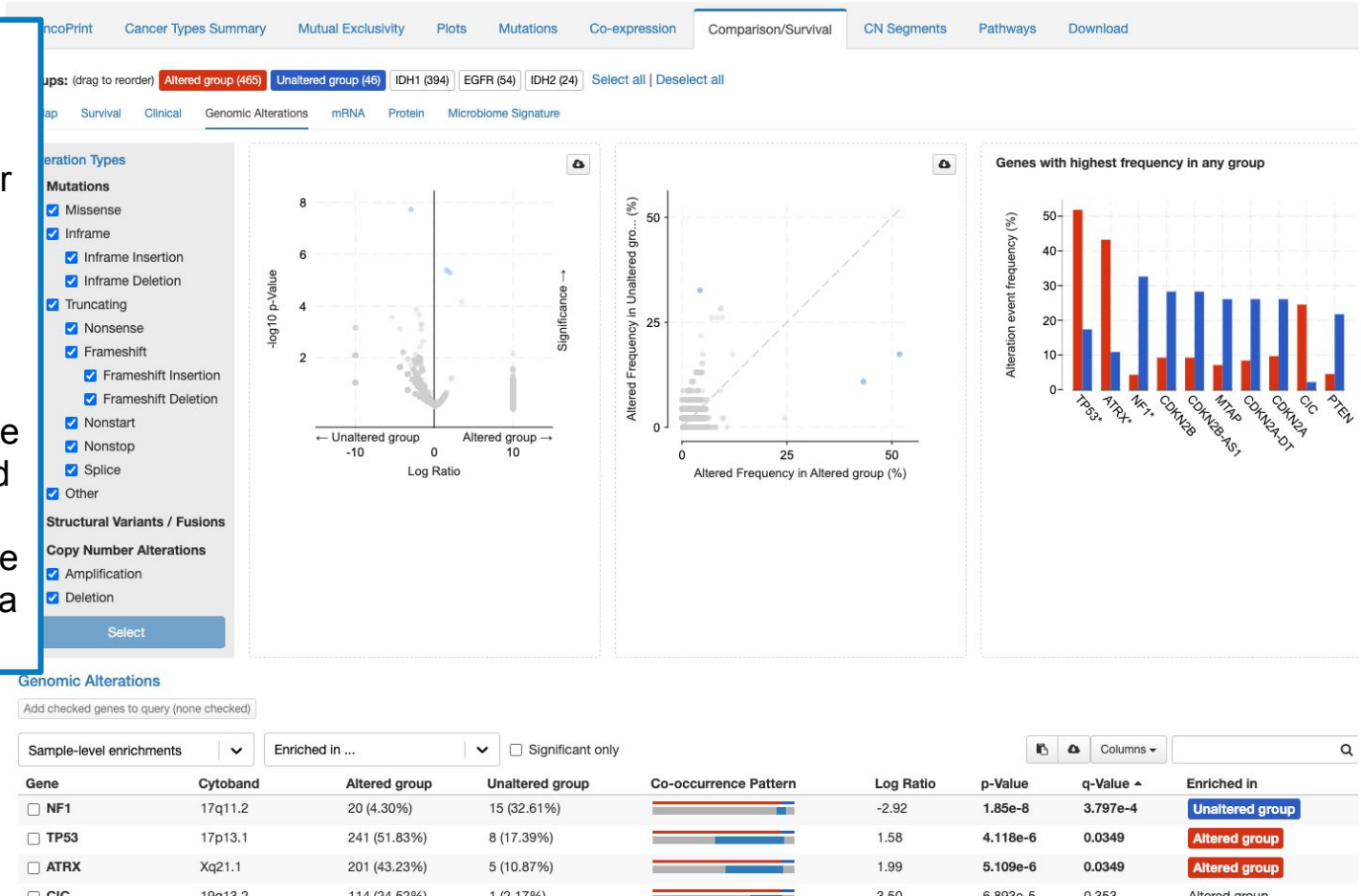
Click on a clinical attribute to visualize the data in the plot on the right.



Link to this page

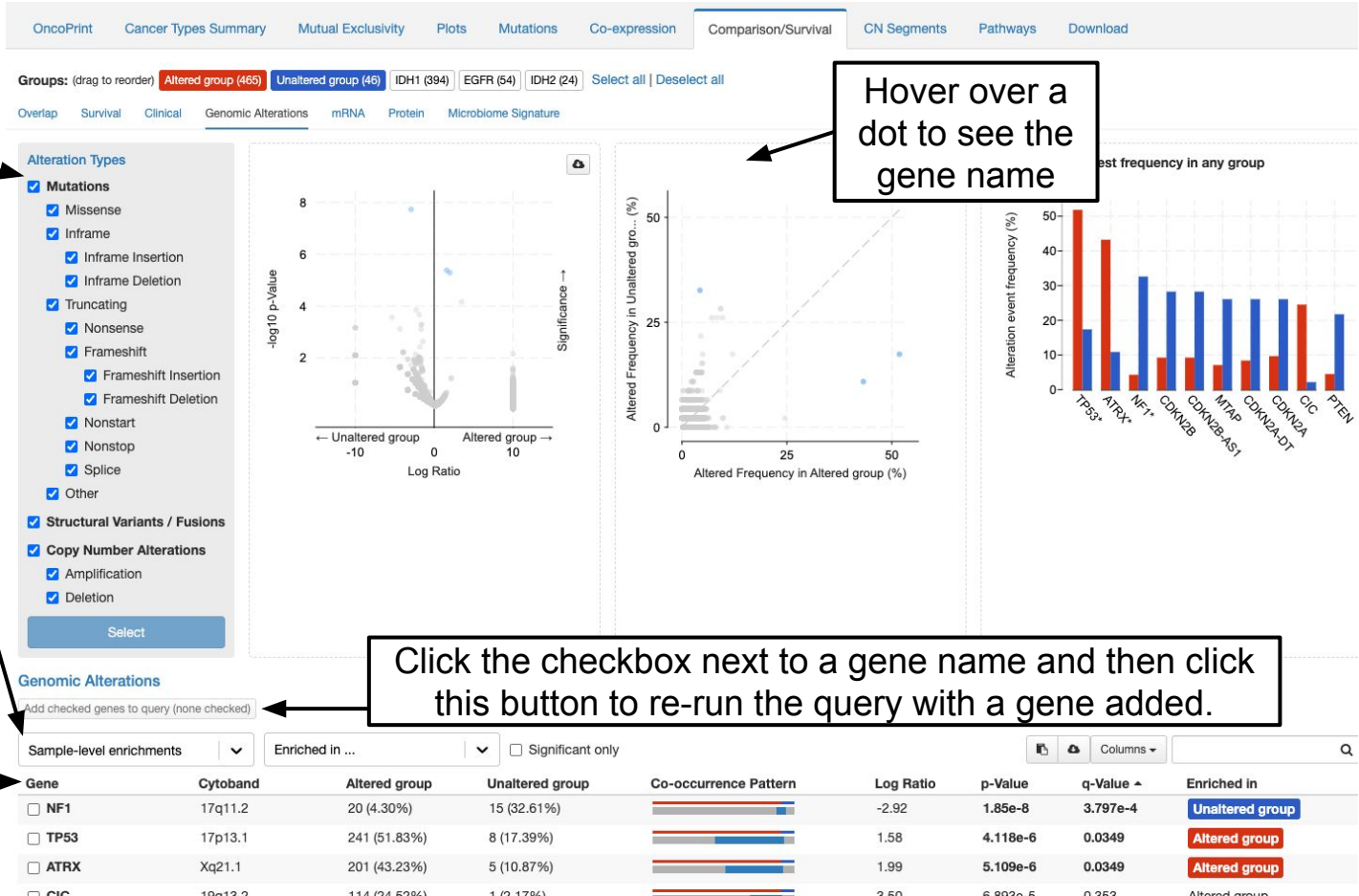
Comparison: Molecular Profiles

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 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.



Link to this page

Comparison: Molecular Profiles



Select which types of alterations to include in the analysis

Hover over a dot to see the gene name

Select sample-level or patient-level analysis

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

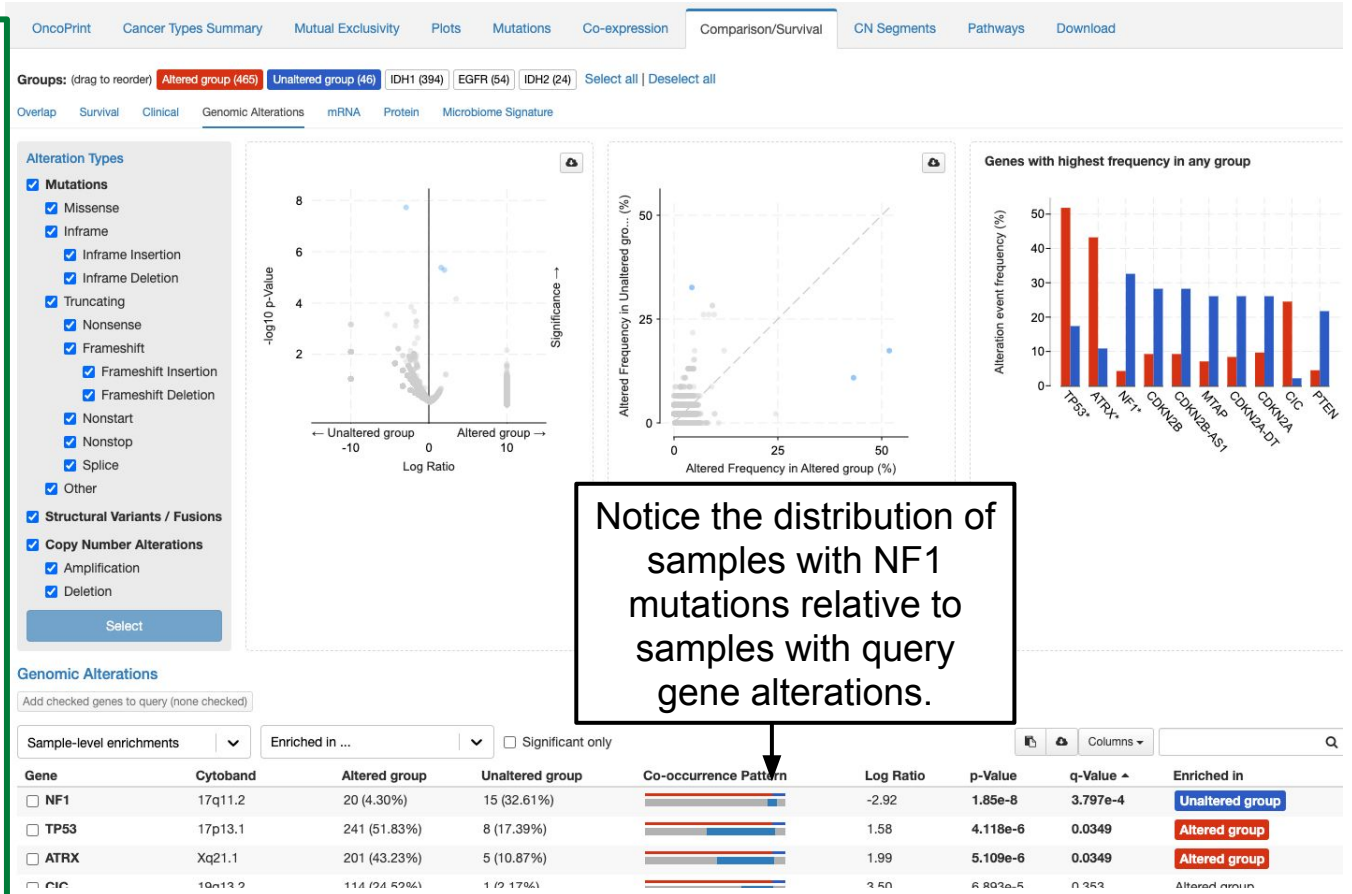
Click the checkbox next to a gene name and then click this button to re-run the query with a gene added.

Link to this page

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).

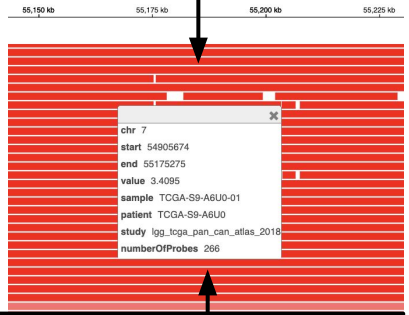
Plots for each gene appear on a separate tab.

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

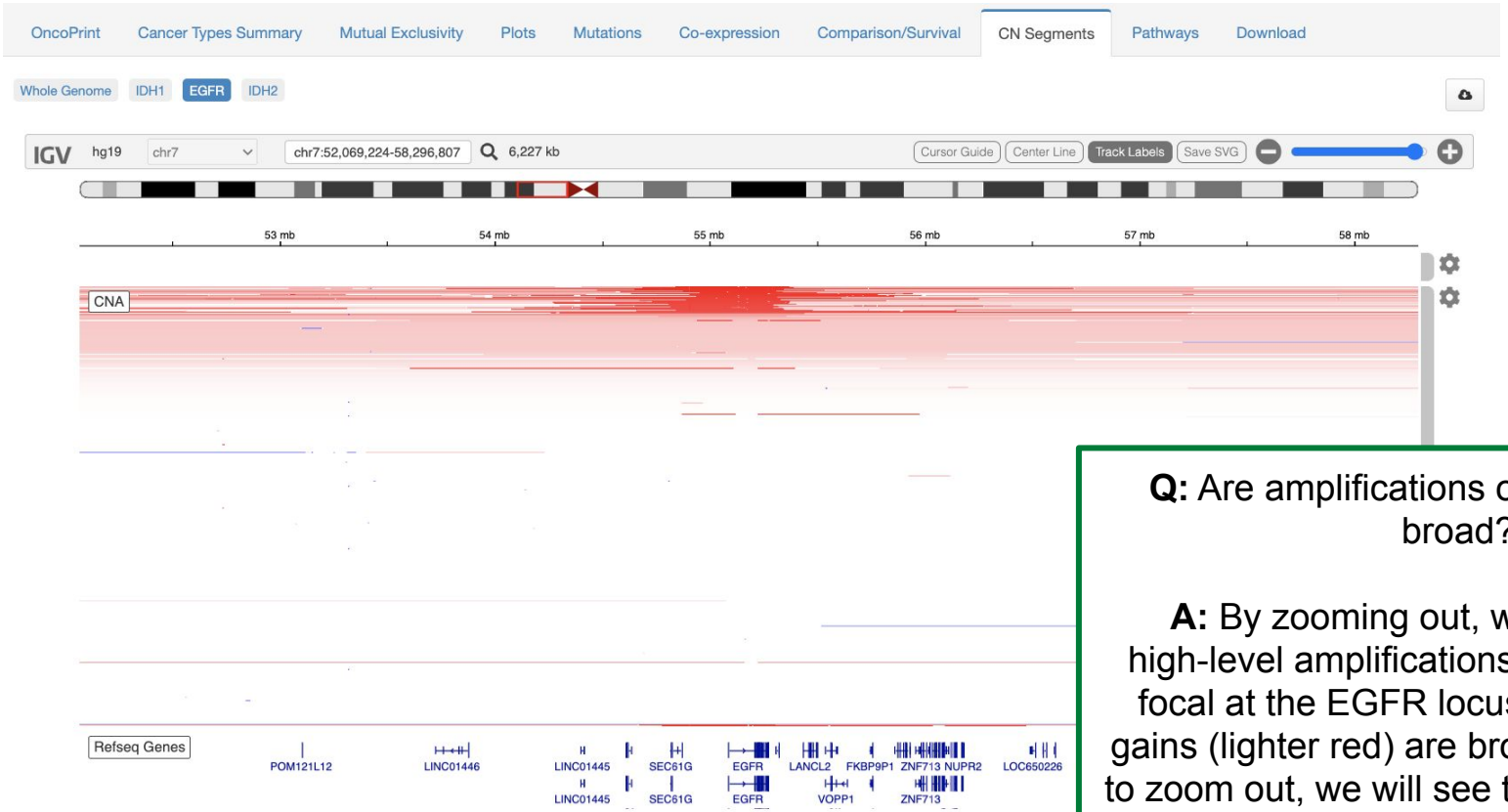
Gene structures



Click on a read for details

Link to this page

CN Segments



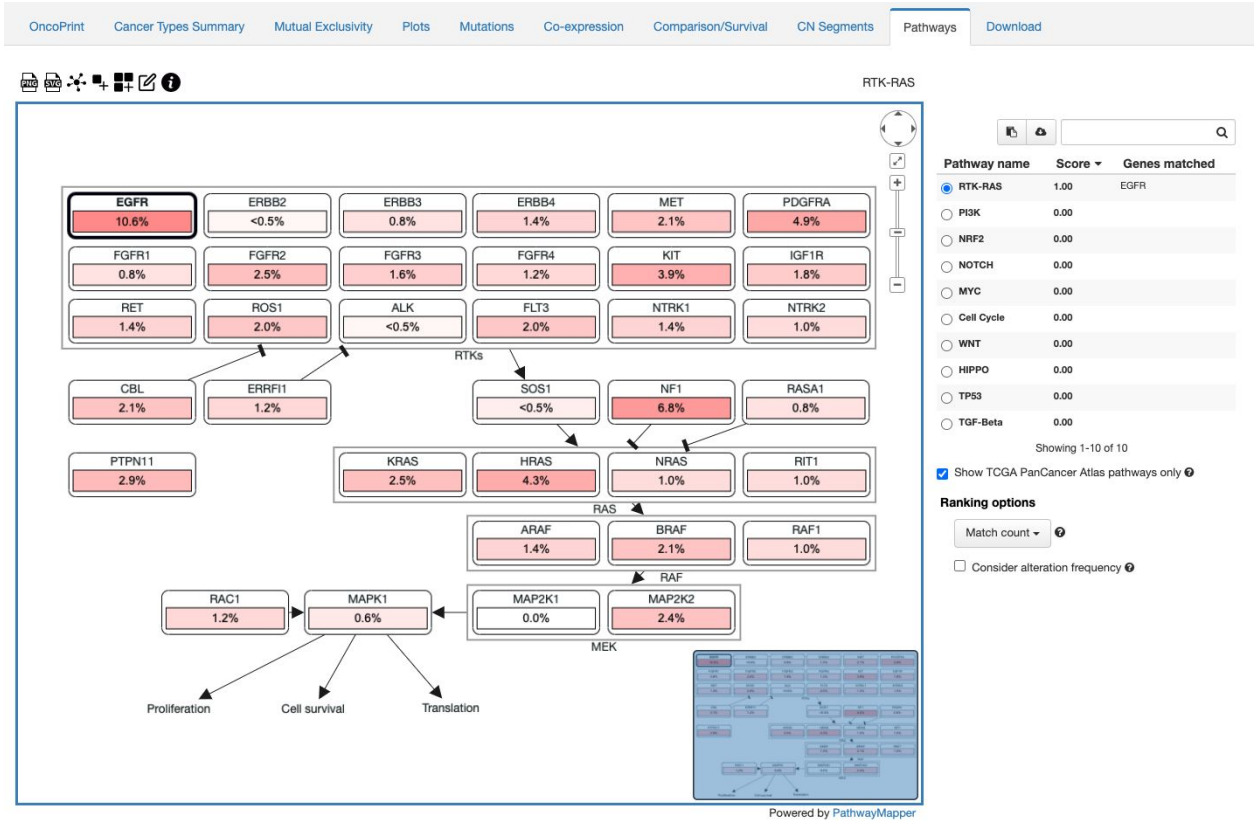
Q: Are amplifications of EGFR focal or broad?

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.

Link to this page

Pathways

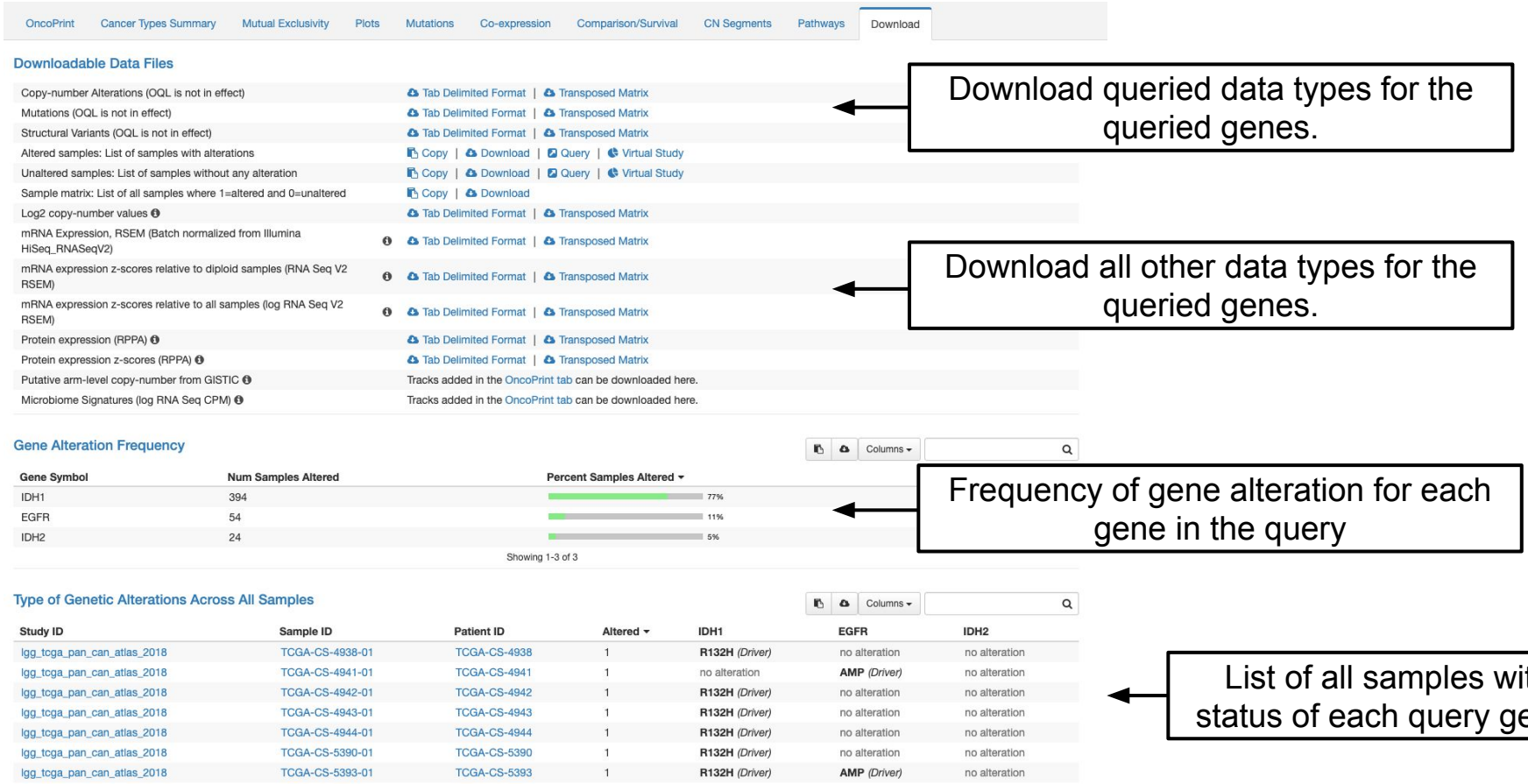
The Pathways tab replaces the now retired “Network” tab. This tab in an integration with PathwayMapper. The tab enables exploration of the queried genes in the context of Pathways defined by TCGA. For more detail on this tab, refer to the Pathways Tutorial.



Link to this page

Download

Download data or copy lists of samples.



Download queried data types for the queried genes.

Download all other data types for the queried genes.

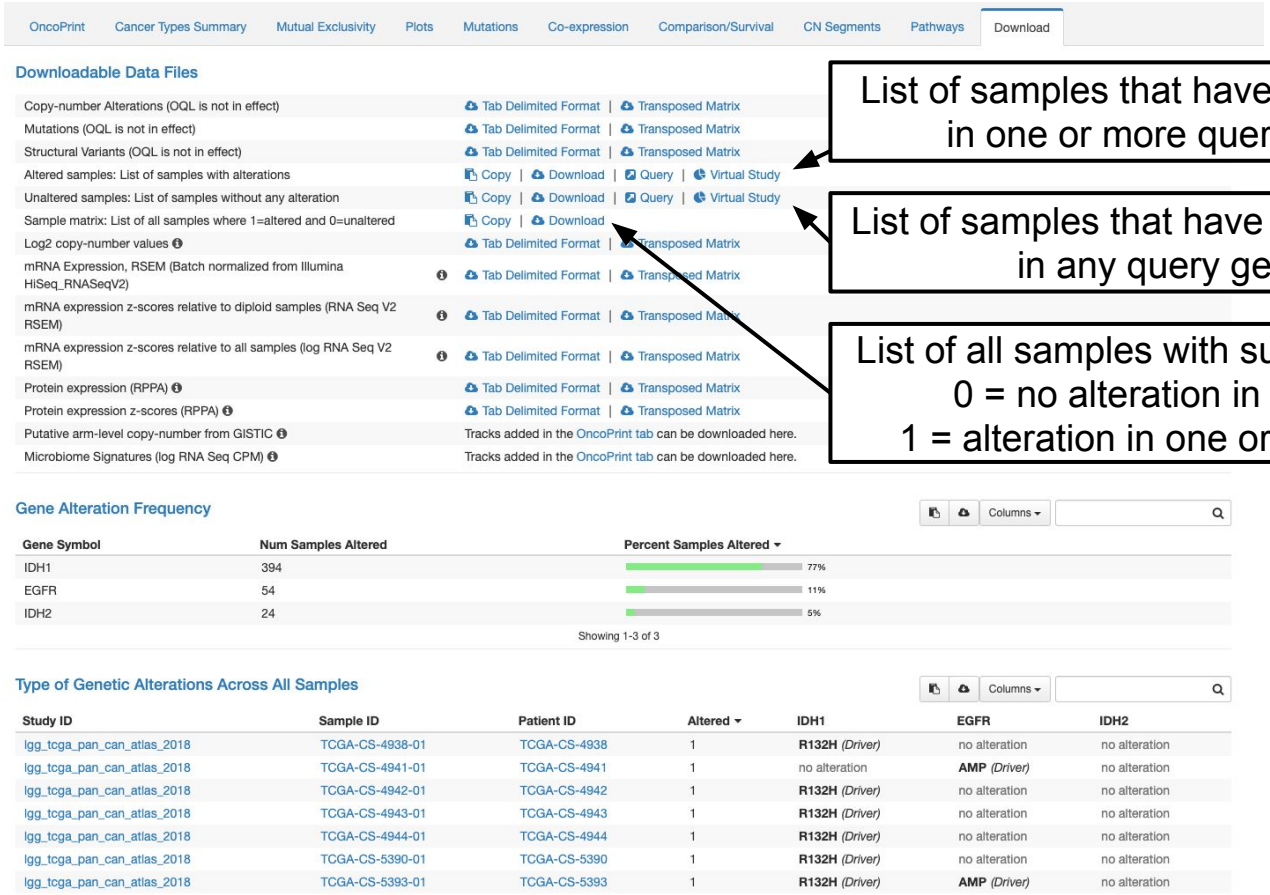
Frequency of gene alteration for each gene in the query

List of all samples with status of each query gene.

Link to this page

Download

Download data or copy lists of samples.



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

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.

Link to this page

Questions?
Check out our other tutorials
or email us at:
cbioportal@googlegroups.com