cBioPortal Tutorial #2: Single Study Query

Query one or multiple genes in a single dataset

Last update: October 11, 2019

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
  - Enrichments
  - Survival
  - Network
  - CN Segments
  - Download
  - Bookmark
- 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.

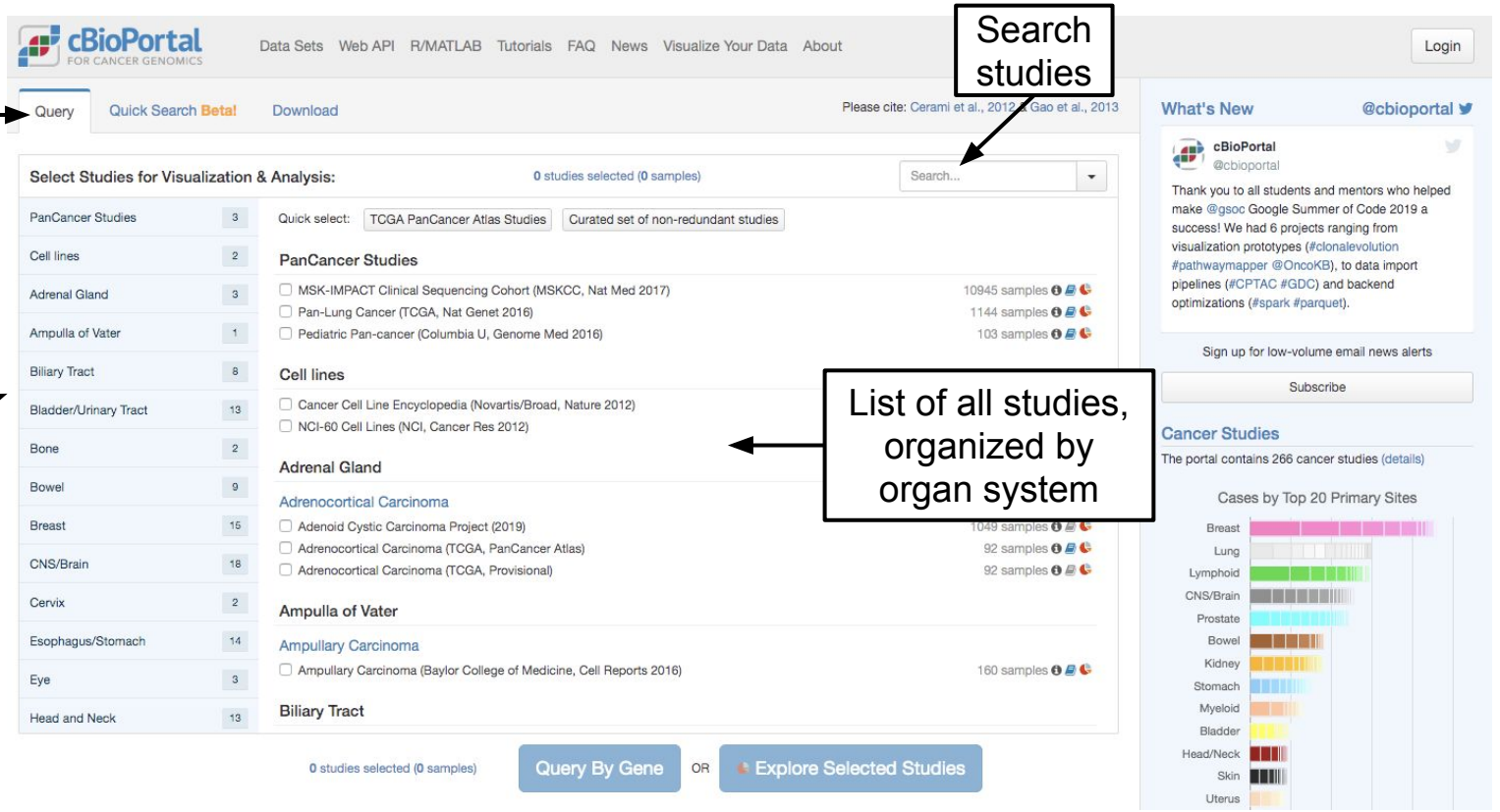
Overview of Tabs in a Single Study Query

Note that depending on the data available for a particular study, not all of these will be present (e.g. a study without outcome data will not have a Survival 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
- **Enrichments:** Explore which genes are altered in the set of samples with query gene alterations or in the set of samples without query gene alterations
- **Survival:** Compare survival of patients with alterations in query genes to the rest of the cohort
- **CN Segments:** Explore copy number changes with the Integrated Genomics Viewer (IGV)
- **Network:** Explore gene networks centered on the query genes
- **Download:** Download data or copy sample lists
- **Bookmark:** Link to save the query

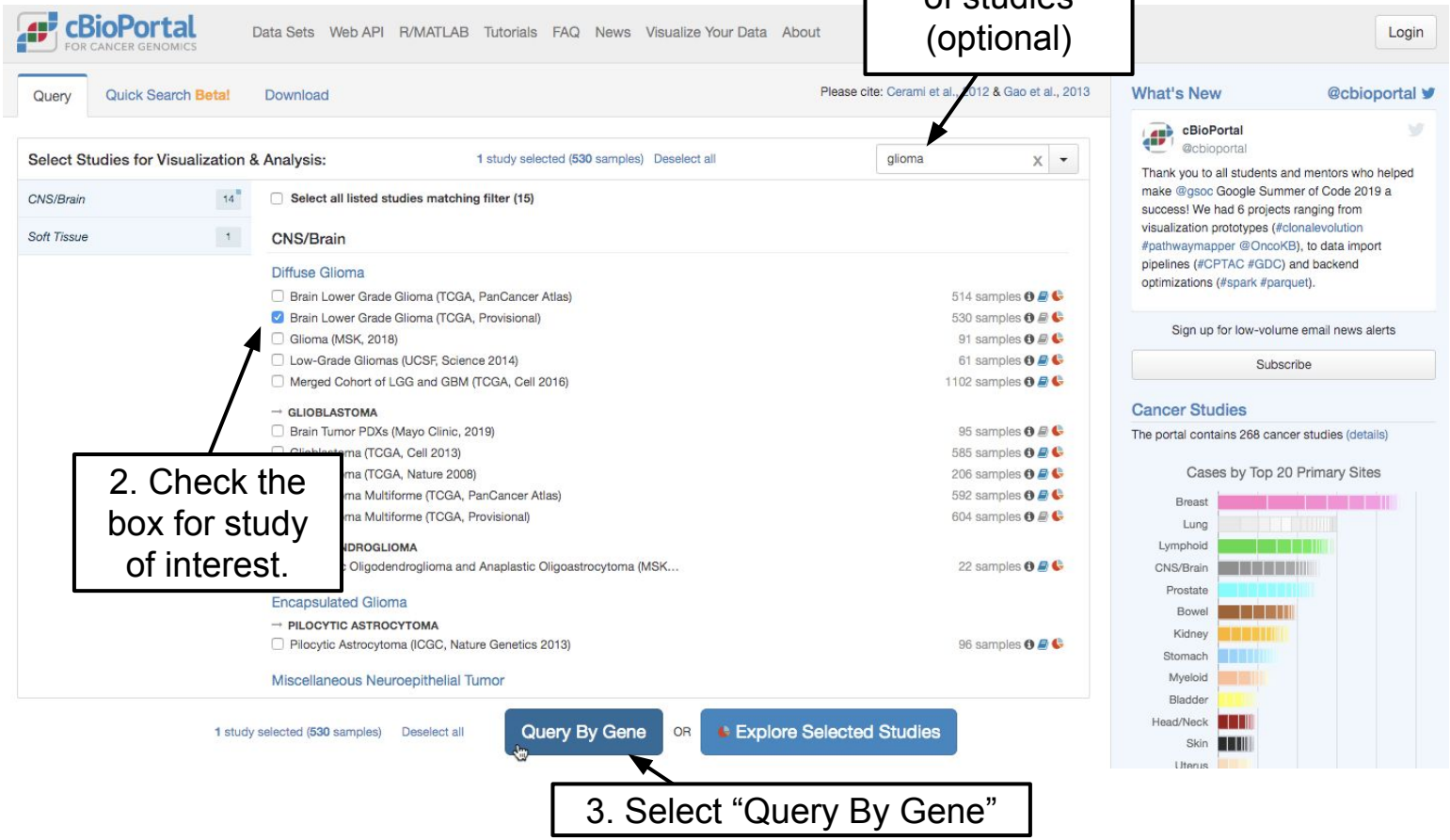
We’re going to run a query in the TCGA Lower-Grade Glioma study. The next few slides will show how to run this query from the Query page. You can also run the same query from a Single Study Exploration, as we did in Tutorial #1.

Query overview



Link to this page

Single study query

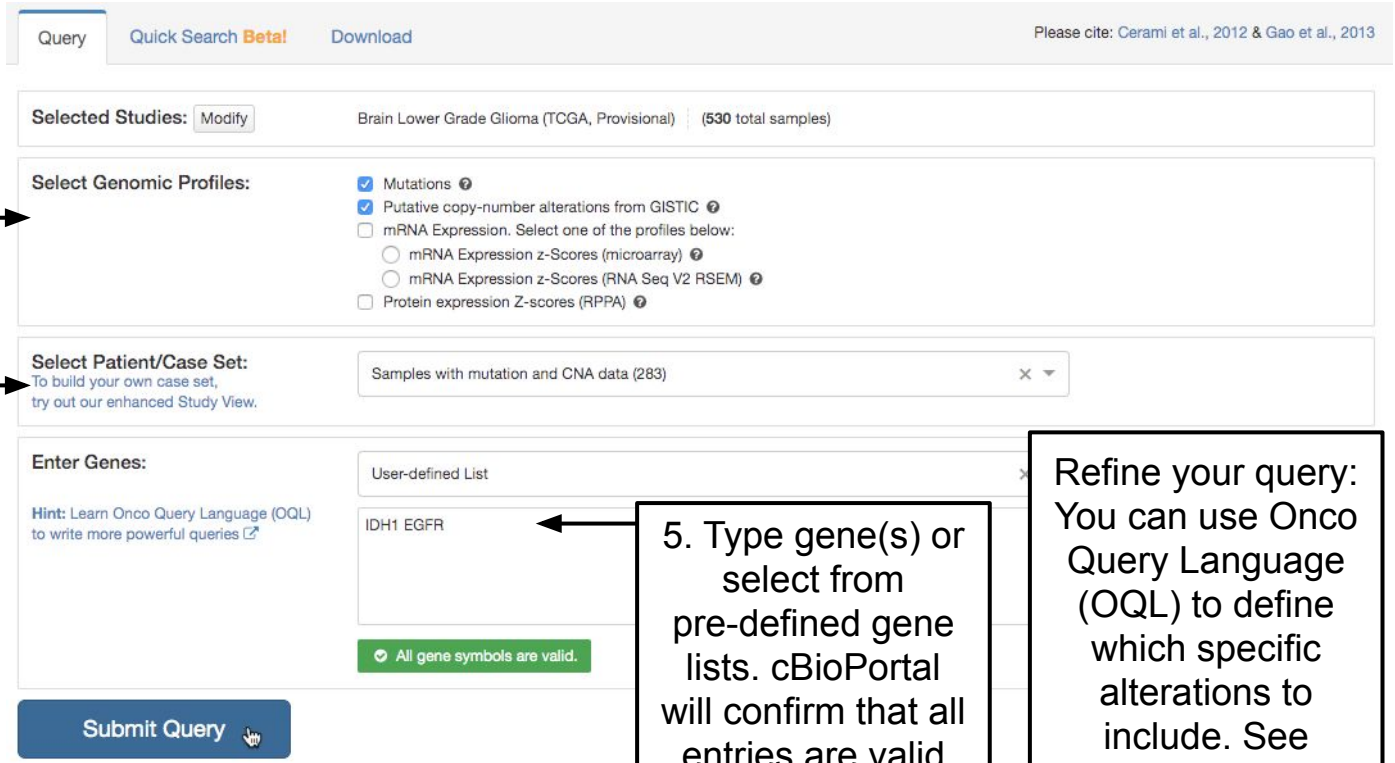


Link to this page

Single study query

4. This section lists include all data types available for the selected study. Select data types to query.

5. Select sample set. For most studies, an appropriate sample set will be automatically selected given the data types selected in Step 3.



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

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

6. Submit query

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

The name of the study. Click to view the queried samples 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.

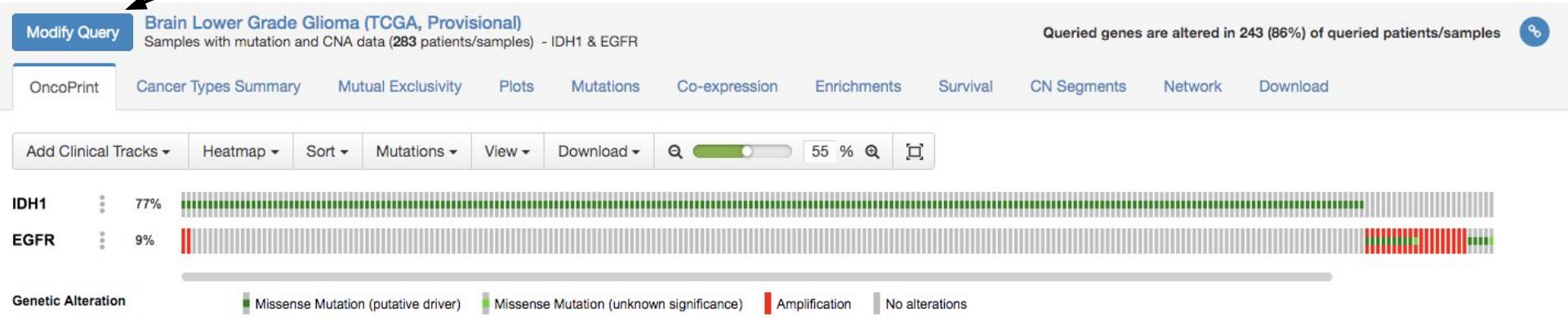
Save a link to the current session. Useful for sharing with others or returning to a query at a later date.

Link to this page

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

Modify Query

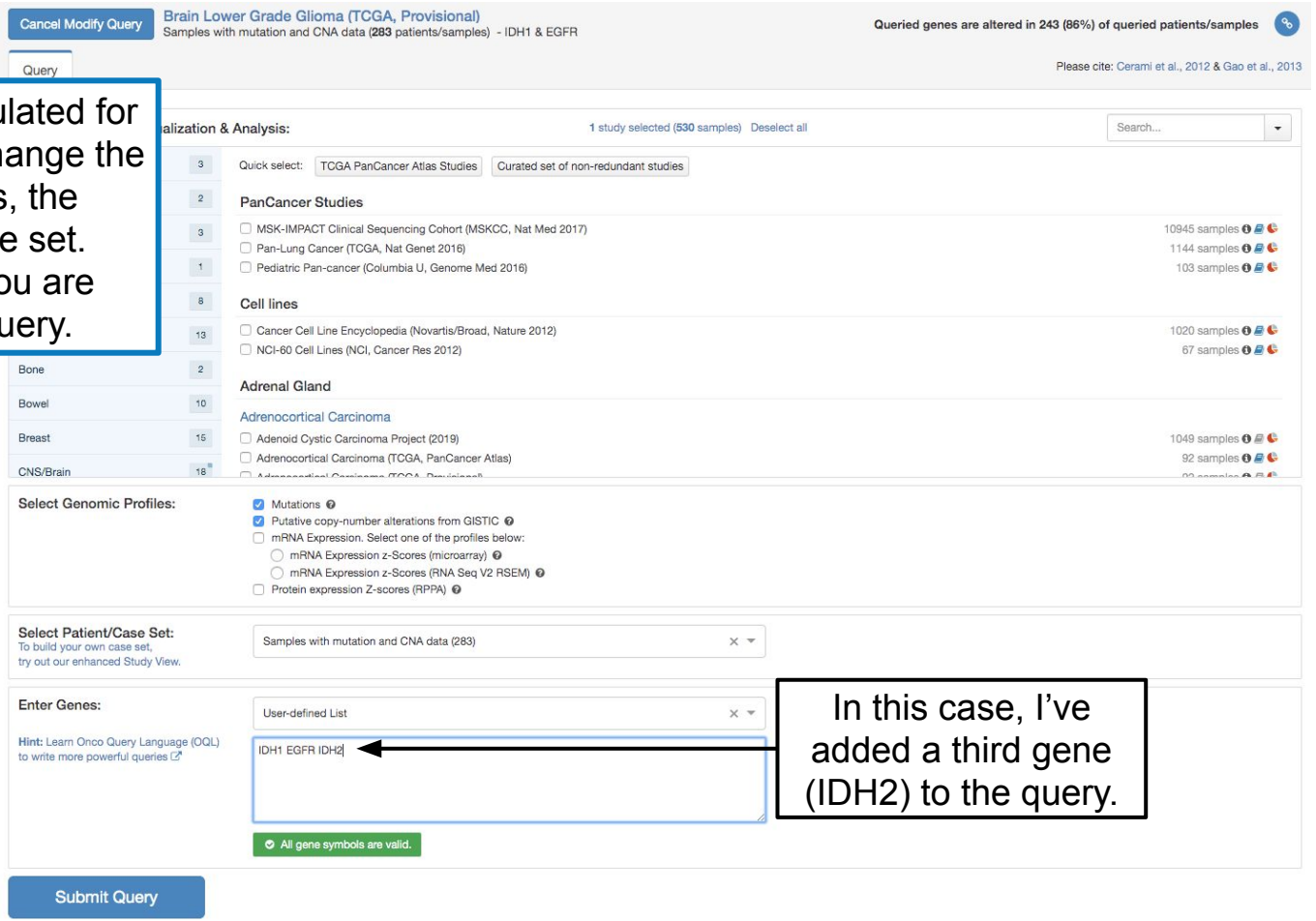
Click on “Modify Query”. This button is available on all tabs and can be used at any time. This will bring up the query interface from the homepage (see next slide for a screenshot).



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.



In this case, I’ve added a third gene (IDH2) to the 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

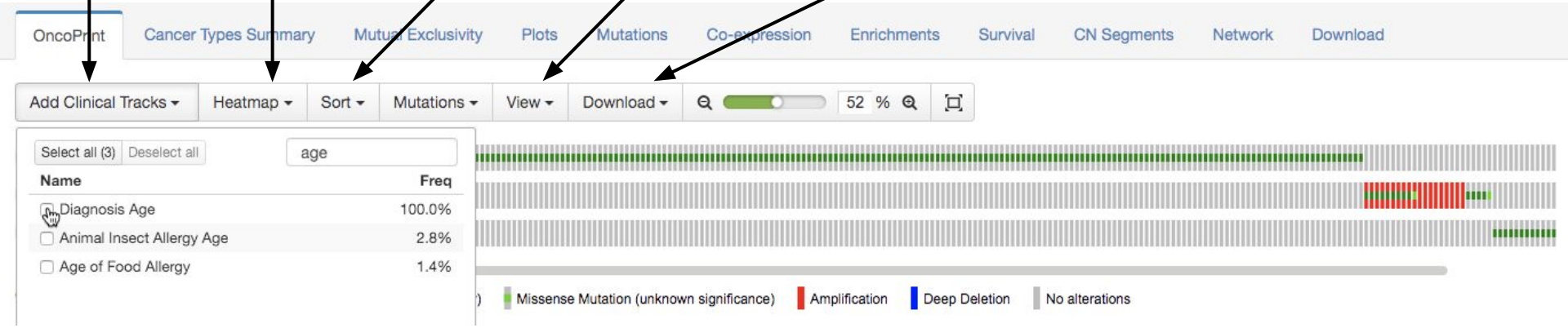
Add clinical tracks (options will vary depending on the data available for each study)

Add a heatmap with RNA or protein levels

Change the sample sorting order

Customize visualization

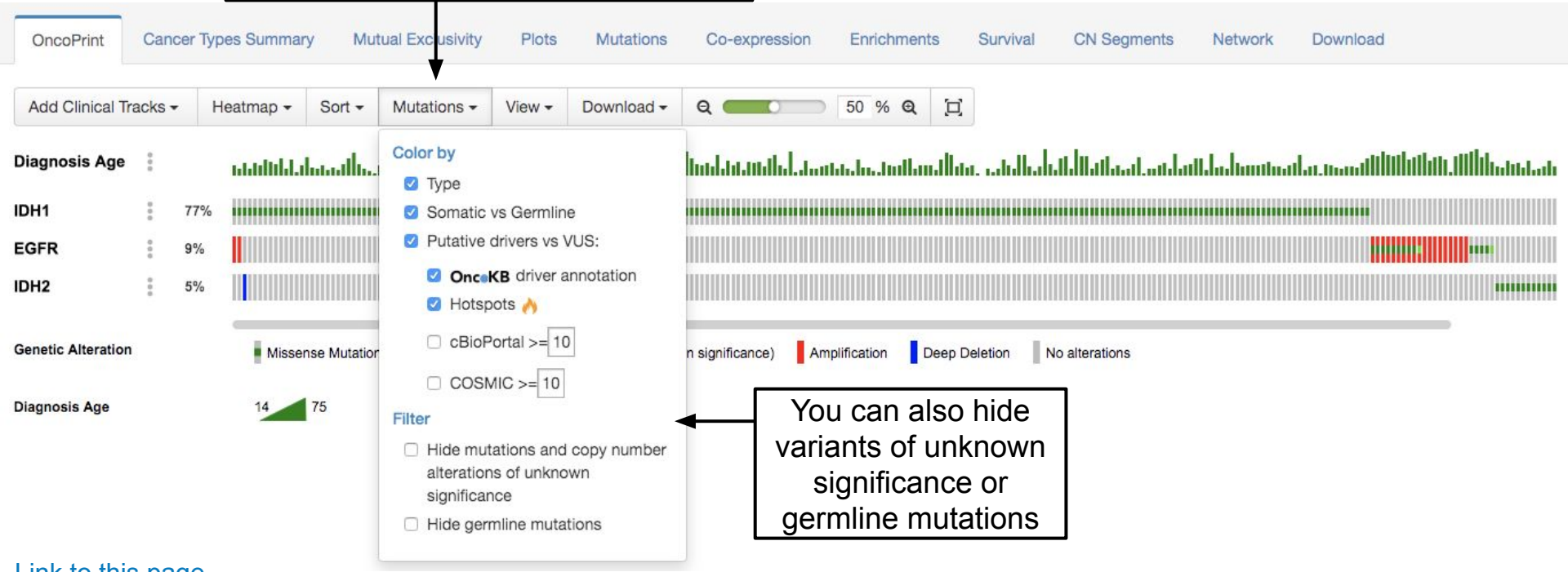
Download figure as PNG, PDF or SVG. Download patient/sample IDs in same order as OncoPrint.



Link to this page

OncoPrint: Mutation Classification Rules

Change the rules by which mutations are colored. This includes the rules to classify a mutation as a putative driver or passenger.



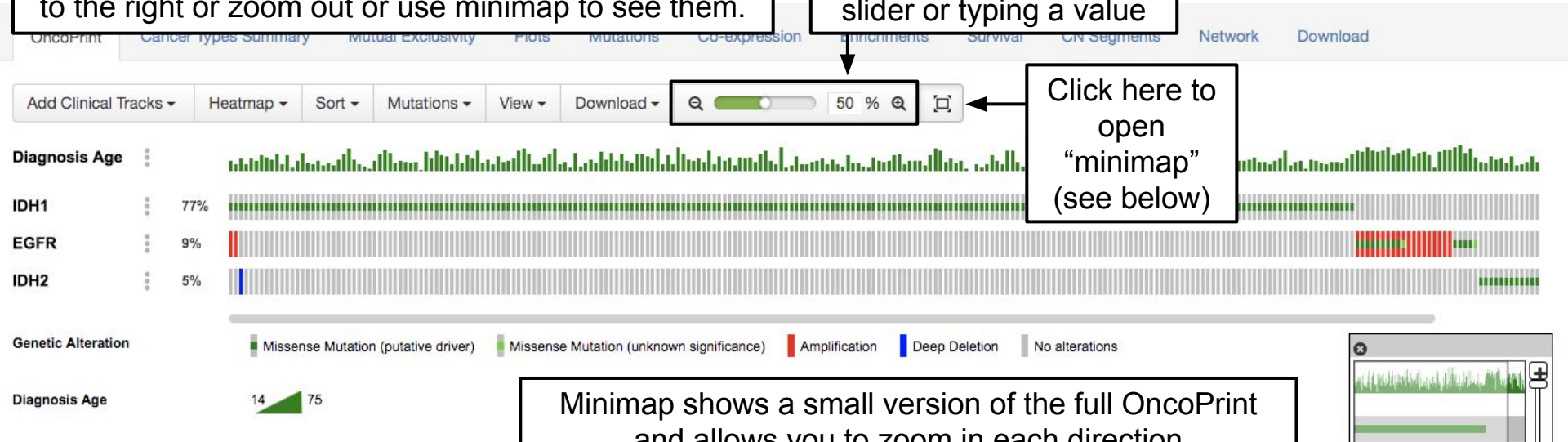
You can also hide variants of unknown significance or germline mutations

Link to this page

OncoPrint: Zoom

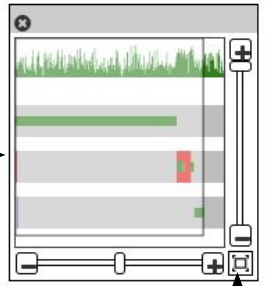
There may be more samples hiding off-screen. Scroll to the right or zoom out or use minimap to see them.

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)

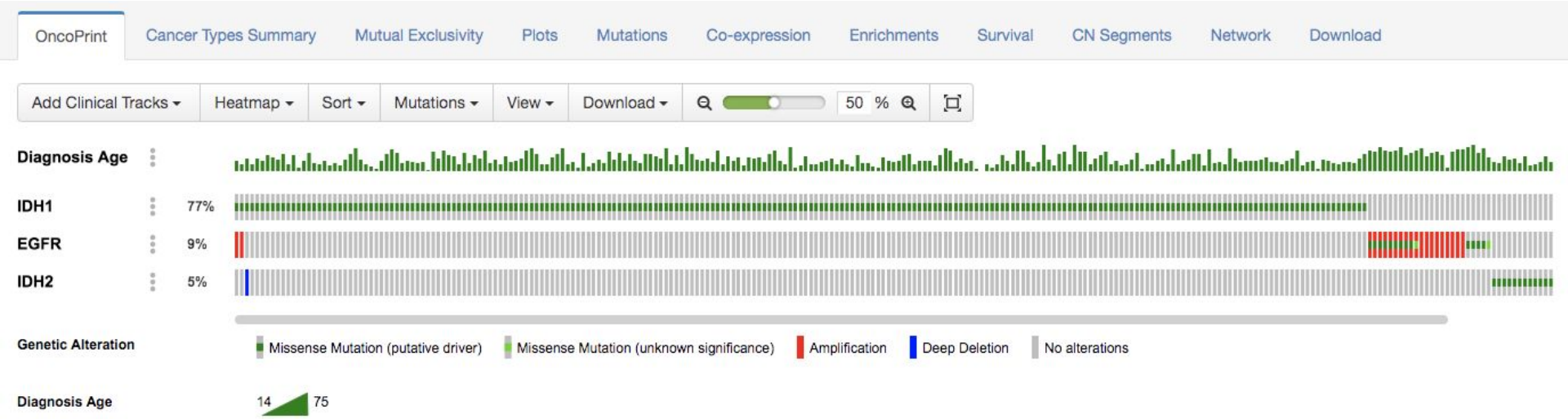
Minimap shows a small version of the full OncoPrint and allows you to zoom in each direction independently. The rectangle can be dragged to move around OncoPrint or resized to change the zoom.



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 are 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

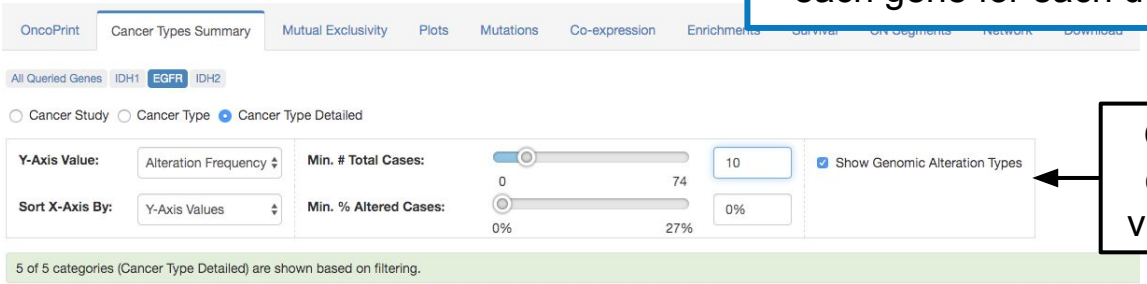
Now we're going to go through all the other tabs and ask some questions about alterations in *IDH1*, *IDH2* and *EGFR* in the TCGA Lower-Grade Glioma study.

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

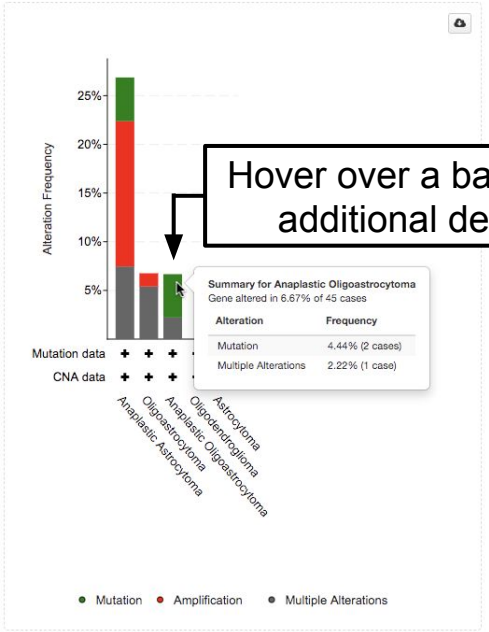
Cancer Types Summary

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?

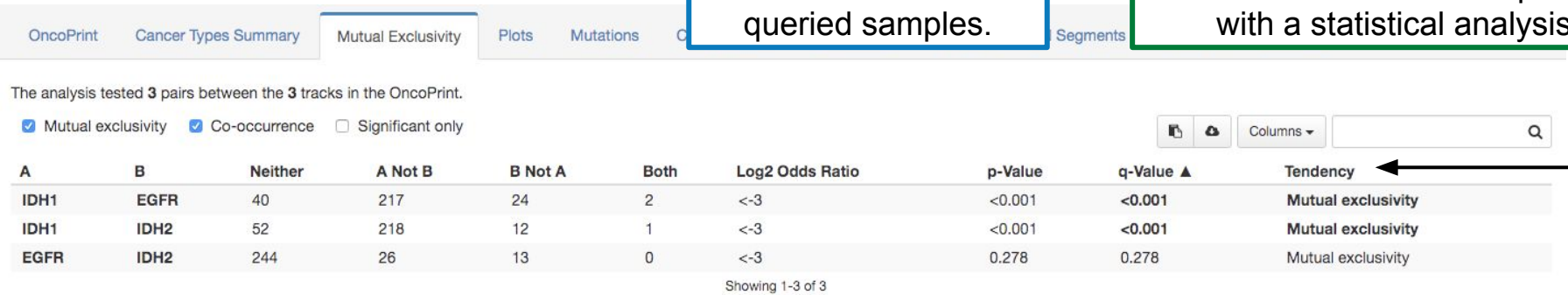
A: Yes, anaplastic astrocytoma appears to have a much higher frequency of EGFR alteration than oligoastrocytoma or anaplastic oligoastrocytoma.

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.

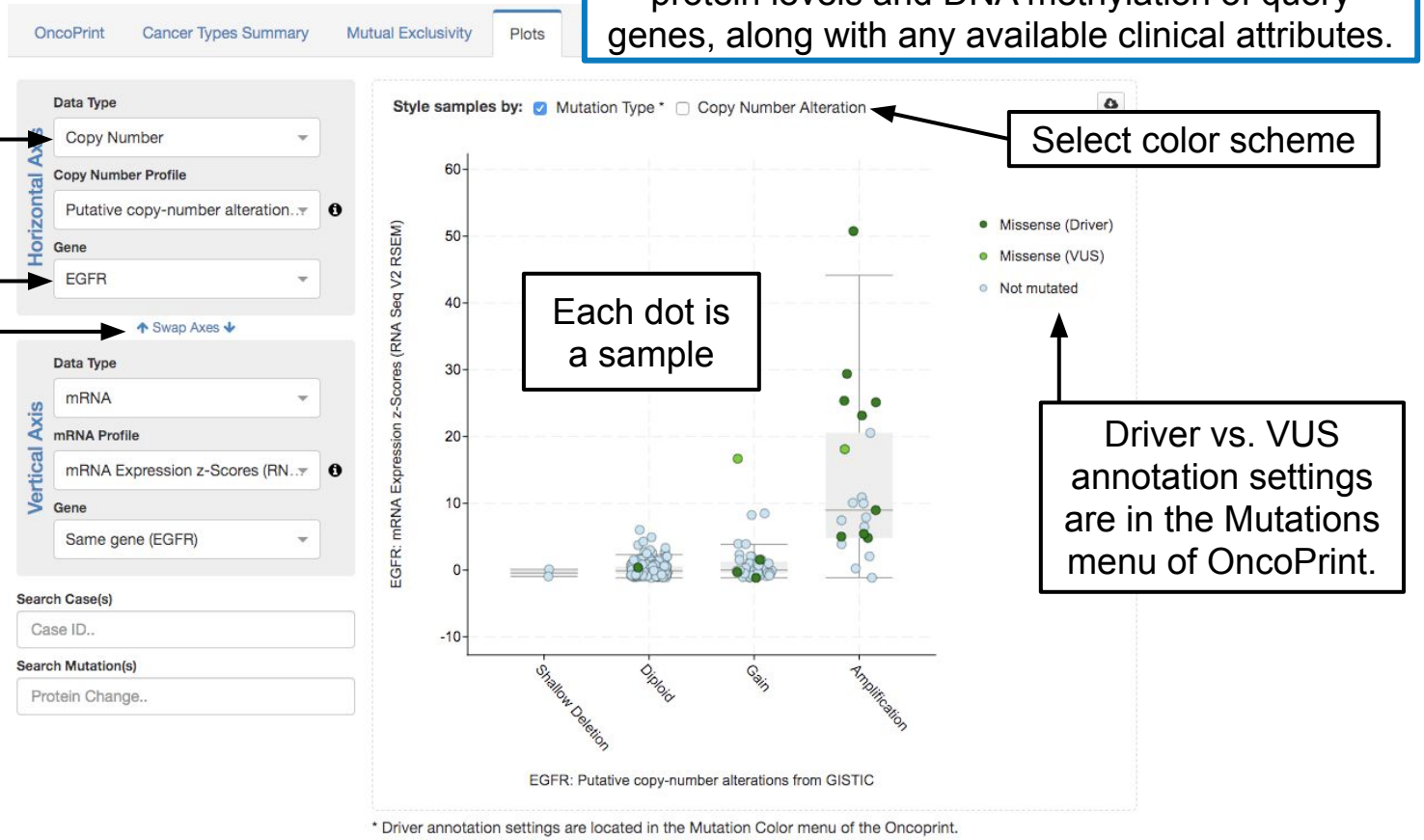
Click on any column header to sort. Hover over the column names for more details about how values are calculated.

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

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.



Choose type of data

Select a query gene

Swap horizontal & vertical axis

Select color scheme

Each dot is a sample

Driver vs. VUS annotation settings are in the Mutations menu of OncoPrint.

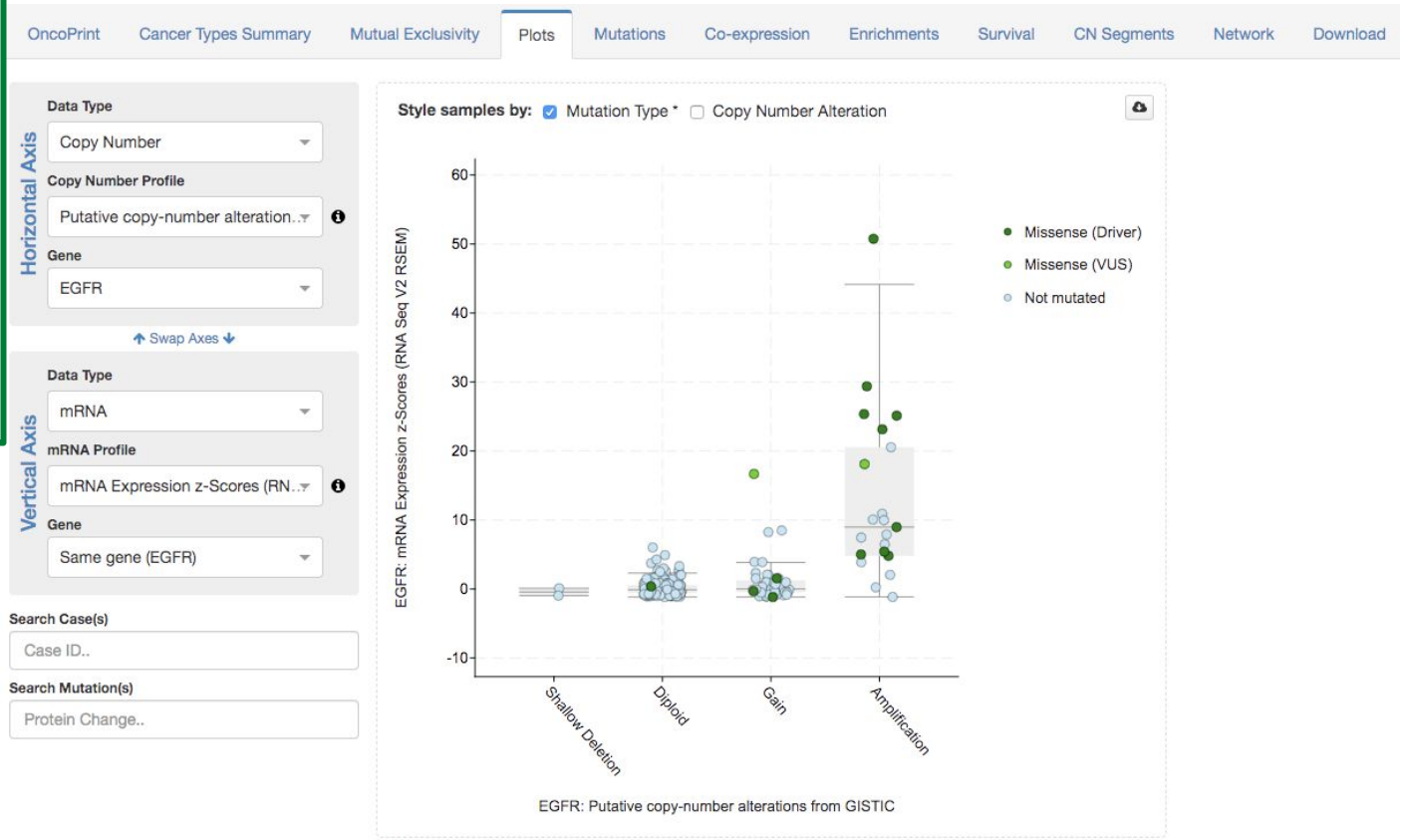
Link to this page

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

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

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

Mutations

This tab shows details about all mutations called in a particular gene.

Each gene appears on a separate tab

Statistics about mutation frequency

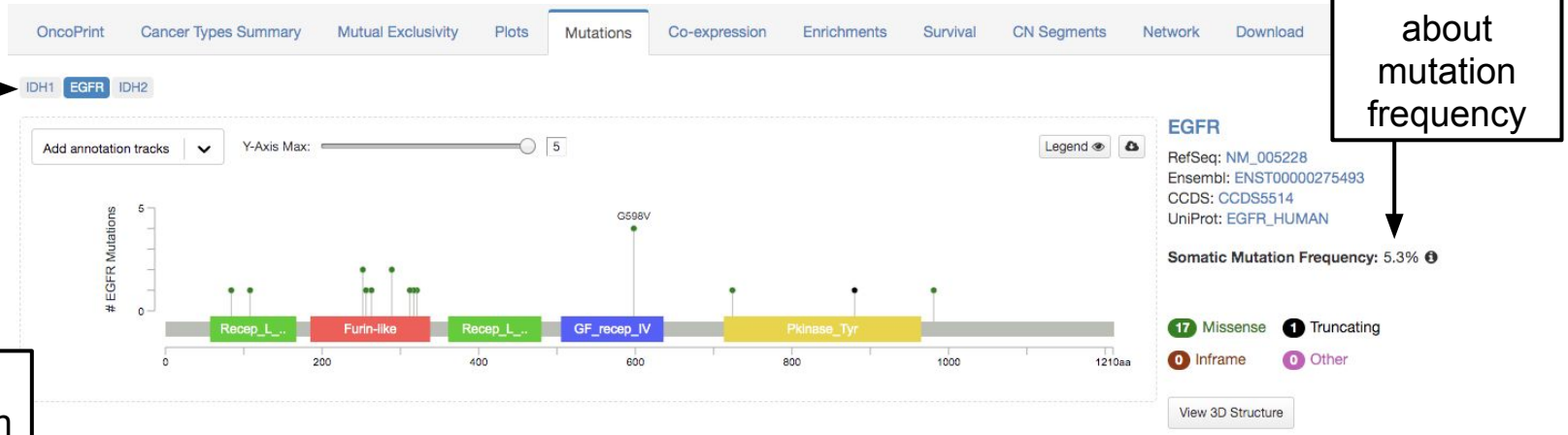
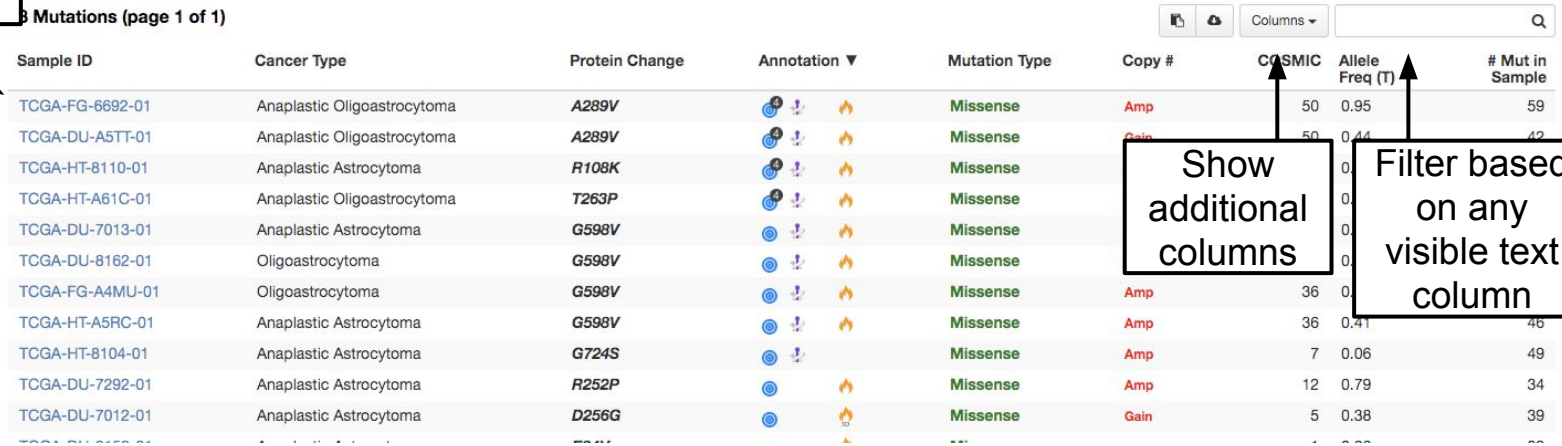


Table of all mutations with annotations

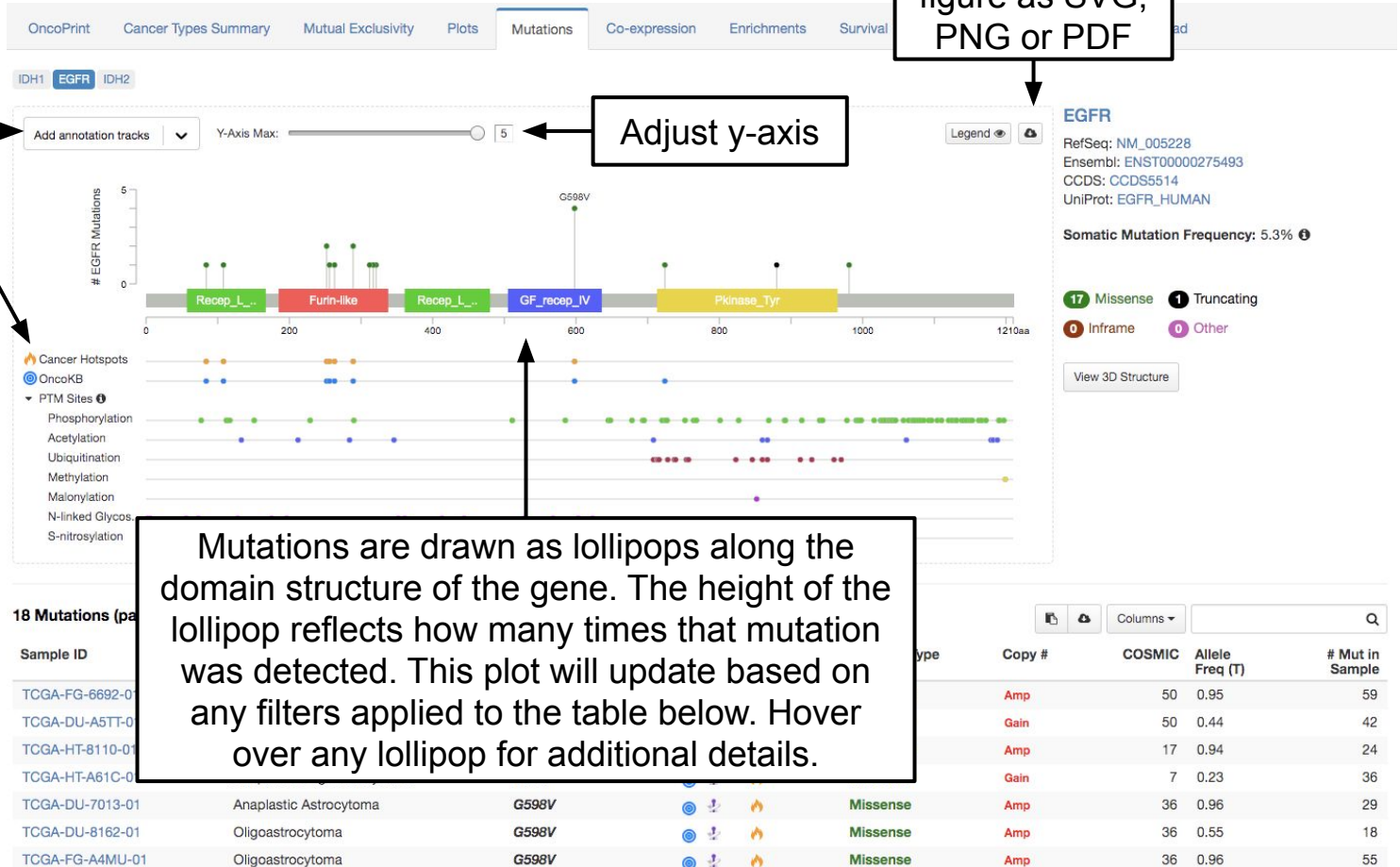


Show additional columns

Filter based on any visible text column

Link to this page

Mutations

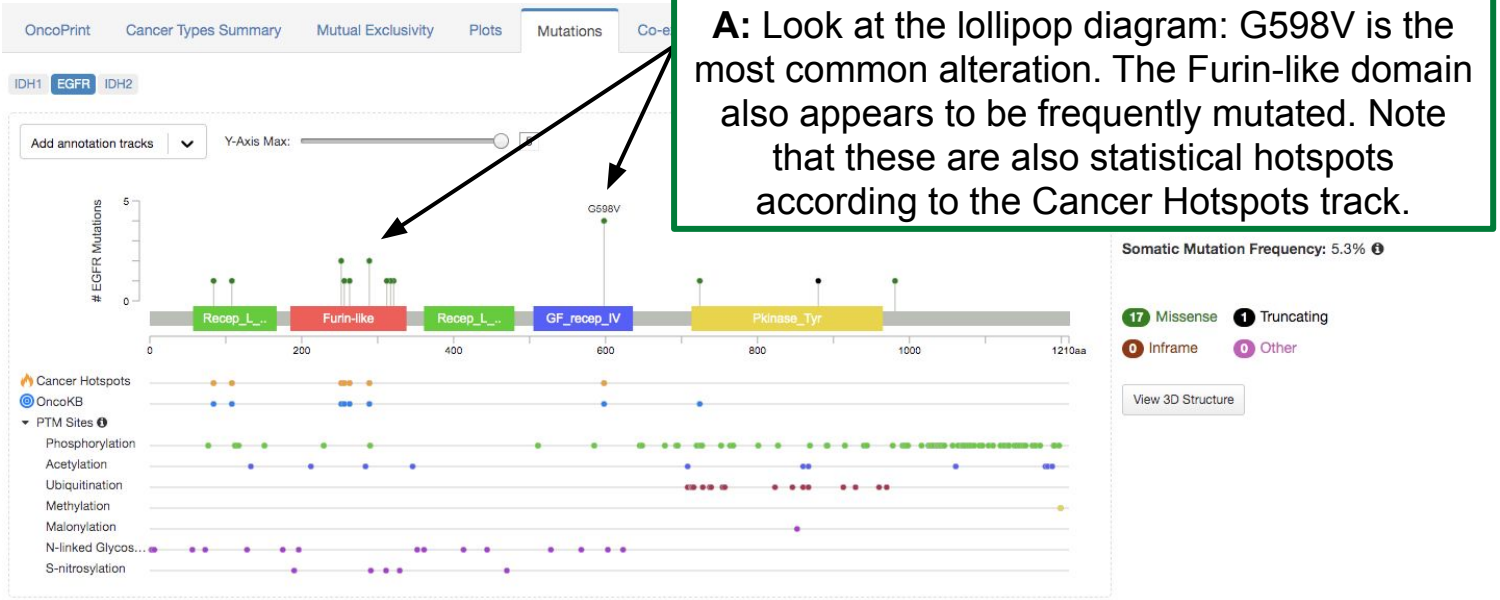


Link to this page

Mutations

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

A: Look at the lollipop diagram: G598V is the most common alteration. The Furin-like domain also appears to be frequently mutated. Note that these are also statistical hotspots according to the Cancer Hotspots track.



18 Mutations (page 1 of 1)

Columns

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

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.

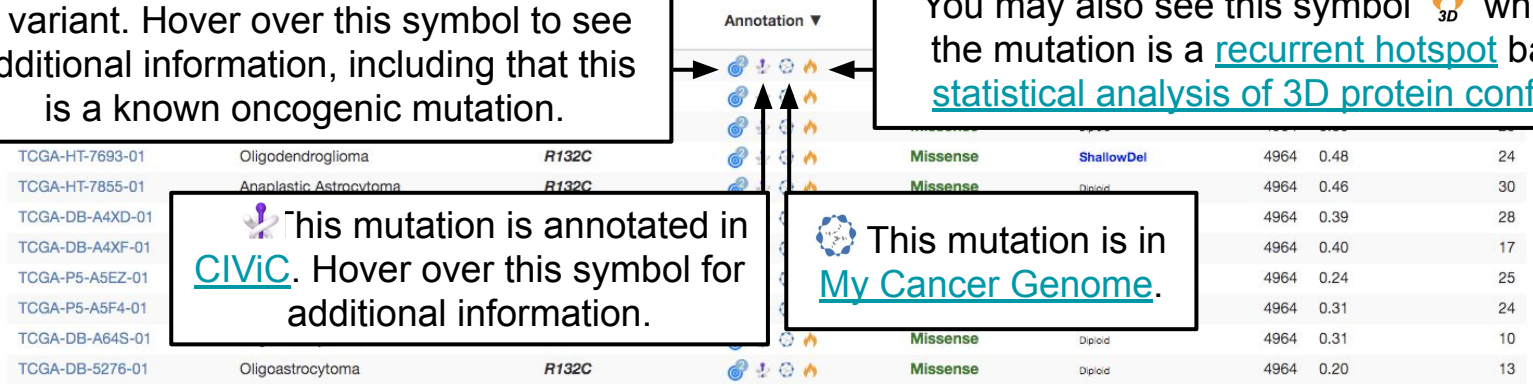
This mutation is a recurrent hotspot based on a statistical analysis of mutation frequency.

You may also see this symbol 3D 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 2 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.



Link to this page

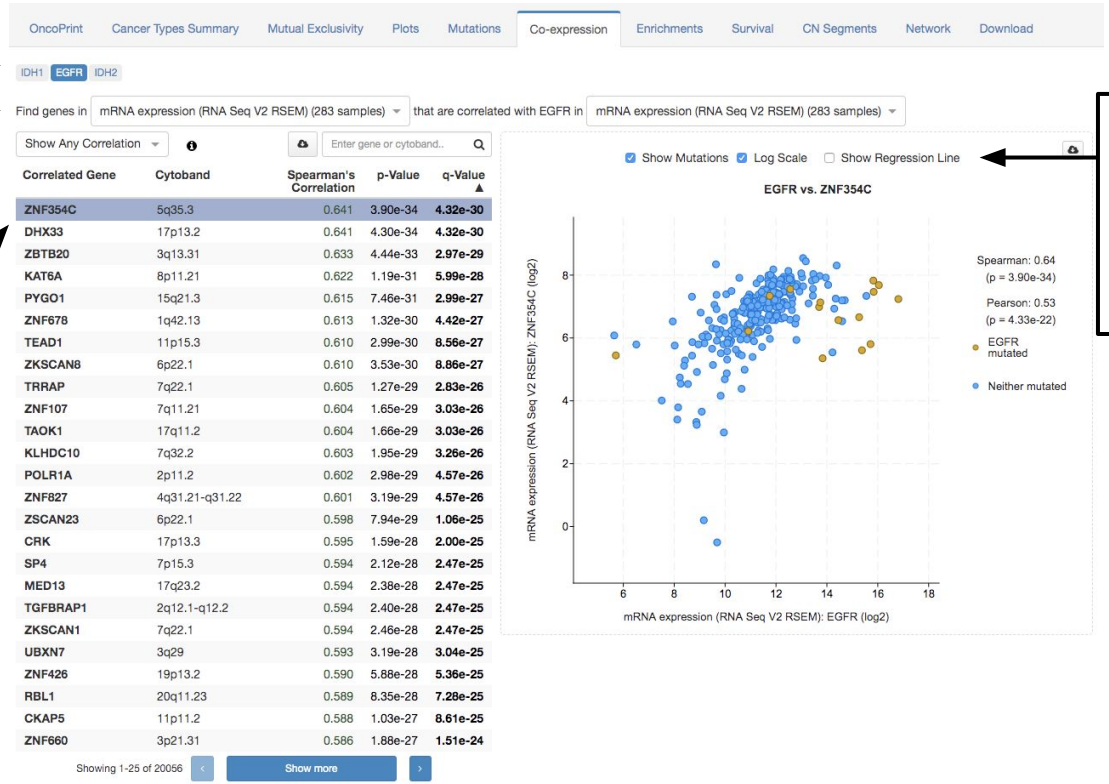
Co-Expression

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 or change x- or y-axis to log scale

Link to this page

Co-Expression

Q: Which genes have expression that correlates with EGFR expression across the cohort?

A: EGFR is on chr7 and many other genes located on chr7 have expression levels correlated with EGFR expression (see table on the left). Chr7 is frequently gained in some subtypes of glioma which could explain these correlated increases in expression.

OncoPrint | Cancer Types Summary | Mutual Exclusiv

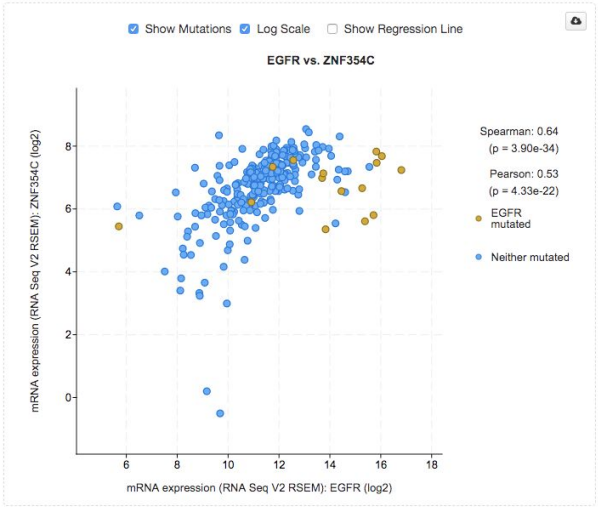
IDH1 EGFR IDH2

Find genes in mRNA expression (RNA Seq V2 RSEM) (283 sa

Show Any Correlation | Enter gene or cytoband..

Correlated Gene	Cytoband	Spearman's Correlation	p-Value	q-Value
ZNF354C	5q35.3	0.641	3.90e-34	4.32e-30
DHX33	17p13.2	0.641	4.30e-34	4.32e-30
ZBTB20	3q13.31	0.633	4.44e-33	2.97e-29
KAT6A	8p11.21	0.622	1.19e-31	5.99e-28
PYGO1	15q21.3	0.615	7.46e-31	2.99e-27
ZNF678	1q42.13	0.613	1.32e-30	4.42e-27
TEAD1	11p15.3	0.610	2.99e-30	8.56e-27
ZKSCAN8	6p22.1	0.610	3.53e-30	8.86e-27
TRRAP	7q22.1	0.605	1.27e-29	2.83e-26
ZNF107	7q11.21	0.604	1.65e-29	3.03e-26
TAOK1	17q11.2	0.604	1.66e-29	3.03e-26
KLHDC10	7q32.2	0.603	1.95e-29	3.26e-26
POLR1A	2p11.2	0.602	2.98e-29	4.57e-26
ZNF827	4q31.21-q31.22	0.601	3.19e-29	4.57e-26
ZSCAN23	6p22.1	0.598	7.94e-29	1.06e-25
CRK	17p13.3	0.595	1.59e-28	2.00e-25
SP4	7p15.3	0.594	2.12e-28	2.47e-25
MED13	17q23.2	0.594	2.38e-28	2.47e-25
TGFBRAP1	2q12.1-q12.2	0.594	2.40e-28	2.47e-25
ZKSCAN1	7q22.1	0.594	2.46e-28	2.47e-25
UBXN7	3q29	0.593	3.19e-28	3.04e-25
ZNF426	19p13.2	0.590	5.88e-28	5.36e-25
RBL1	20q11.23	0.589	8.35e-28	7.28e-25
CKAP5	11p11.2	0.588	1.03e-27	8.61e-25
ZNF660	3p21.31	0.586	1.88e-27	1.51e-24

Showing 1-25 of 20056 | Show more



Link to this page

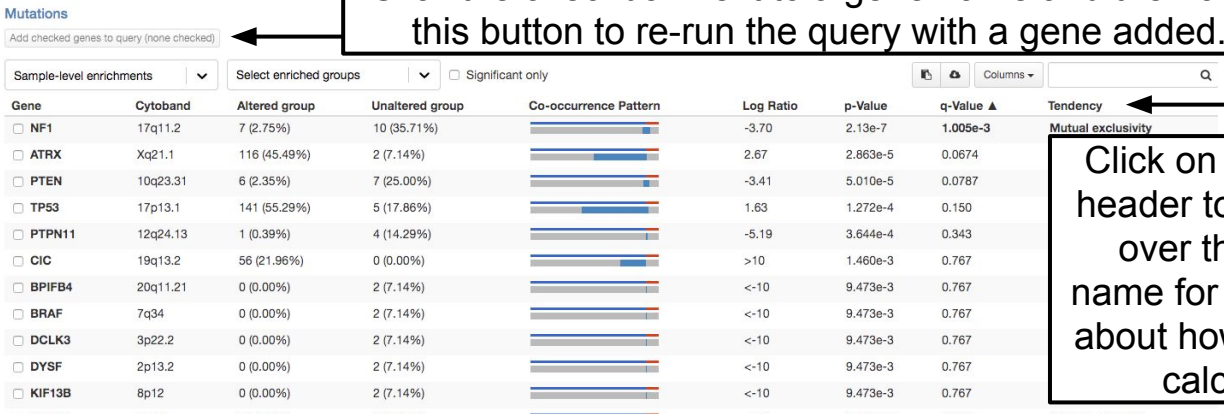
Enrichments

This tab takes samples with alterations in any query gene as a set and looks to see whether other genes are frequently altered in the same set of samples (co-occurring) or in the set of samples without query gene alterations (mutually exclusive).

Select type of data to examine

Hover over a dot to see the gene name

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



Gene	Cytoband	Altered group	Unaltered group	Co-occurrence Pattern	Log Ratio	p-Value	q-Value ▲	Tendency
NF1	17q11.2	7 (2.75%)	10 (35.71%)		-3.70	2.13e-7	1.005e-3	Mutual exclusivity
ATRX	Xq21.1	116 (45.49%)	2 (7.14%)		2.67	2.863e-5	0.0674	
PTEN	10q23.31	6 (2.35%)	7 (25.00%)		-3.41	5.010e-5	0.0787	
TP53	17p13.1	141 (55.29%)	5 (17.86%)		1.63	1.272e-4	0.150	
PTPN11	12q24.13	1 (0.39%)	4 (14.29%)		-5.19	3.644e-4	0.343	
CIC	19q13.2	56 (21.96%)	0 (0.00%)		>10	1.460e-3	0.767	
BPIFB4	20q11.21	0 (0.00%)	2 (7.14%)		<-10	9.473e-3	0.767	
BRAF	7q34	0 (0.00%)	2 (7.14%)		<-10	9.473e-3	0.767	
DCLK3	3p22.2	0 (0.00%)	2 (7.14%)		<-10	9.473e-3	0.767	
DYSF	2p13.2	0 (0.00%)	2 (7.14%)		<-10	9.473e-3	0.767	
KIF13B	8p12	0 (0.00%)	2 (7.14%)		<-10	9.473e-3	0.767	

Select sample-level or patient-level analysis

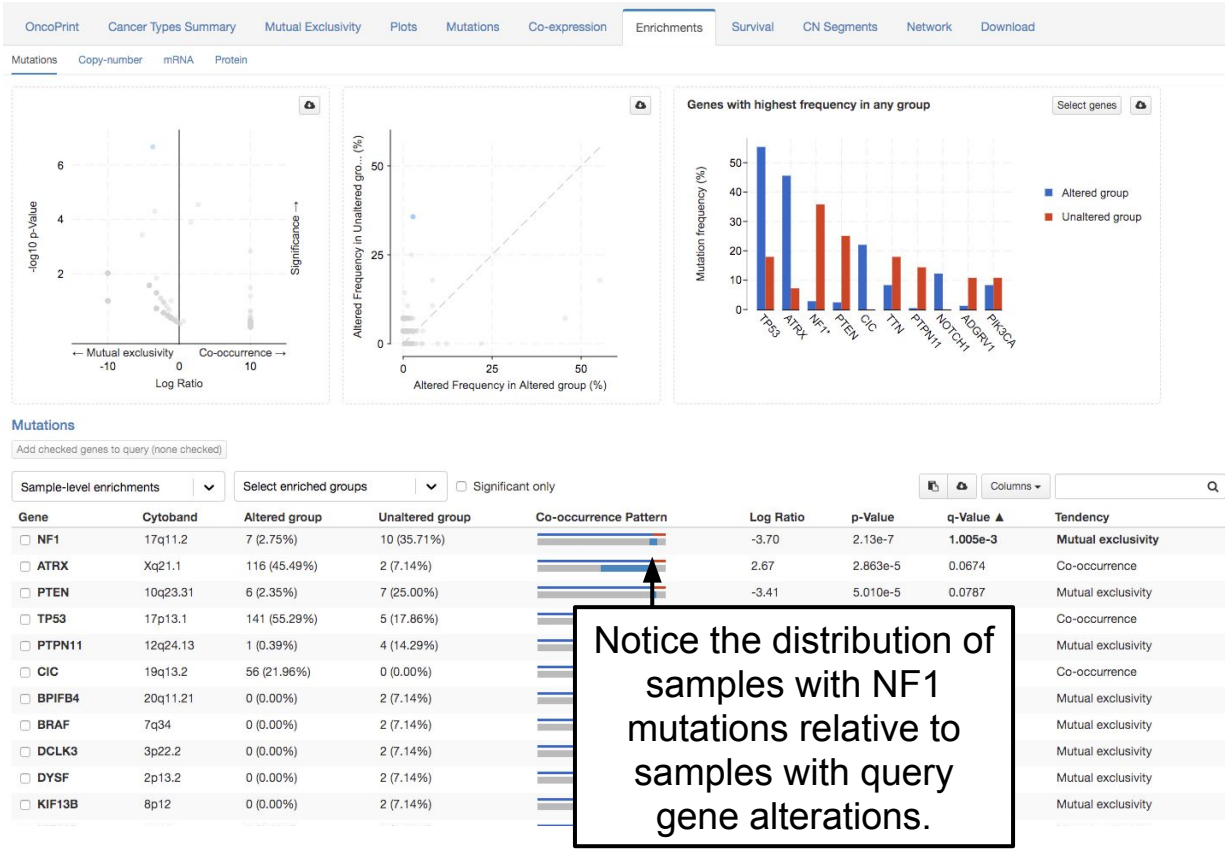
Click on any column header to sort. Hover over the column name for more details about how values are calculated.

Link to this page

Enrichments

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 mutations in one or more other genes?

A: Mutations 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.

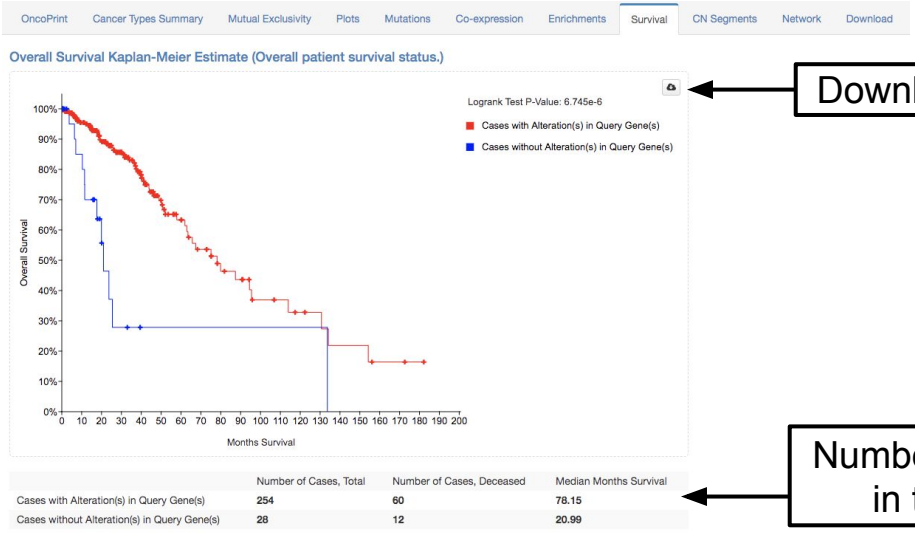


Survival

For studies with outcome data, this tab has Overall Survival and Disease Free Survival Kaplan-Meier plots. In red are cases with one or more alterations in the query gene(s). In blue are all other cases in the study.

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

Link to this page



Download plots or data

Number of cases in the plot

Q: Do patients with alterations in IDH1, IDH2 or EGFR have different outcomes compared to patients without alterations in any of those genes?

A: Patients with alterations in IDH1, IDH2 or EGFR have significantly better OS and DFS than patients without those alterations.

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)

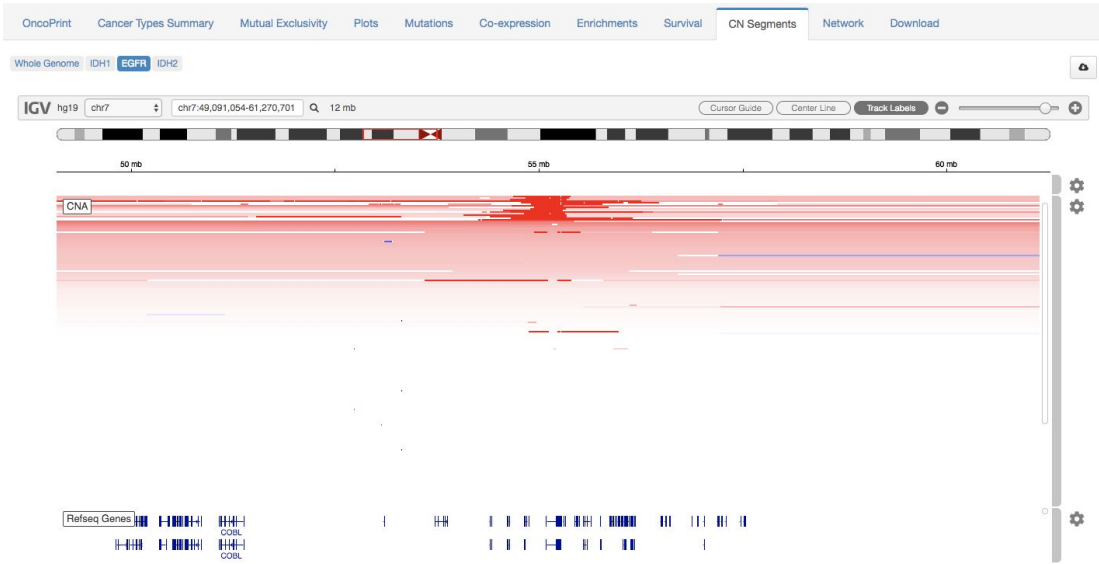
Each row is a single sample

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 the low-level gain encompasses the entire chromosome.

Link to this page

Network

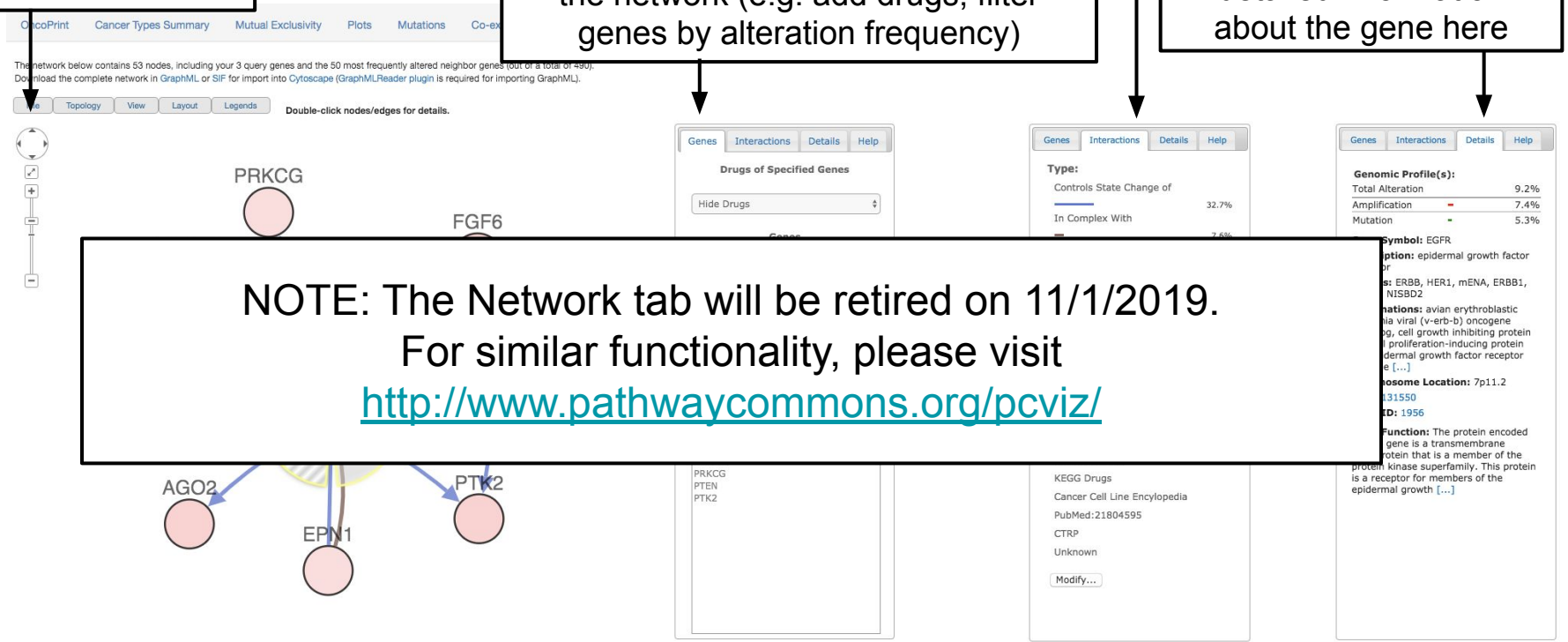
Visualize biological interaction networks centered on your query genes, with color-coding and filter options based on the frequency of genomic alterations in each gene. Click on the “Help” tab for a more detailed explanation.

Change zoom and move around network

View or modify the nodes included in the network (e.g. add drugs, filter genes by alteration frequency)

View or modify the types of interactions (edges) utilized in the plot

Click on any node to see detailed information about the gene here



NOTE: The Network tab will be retired on 11/1/2019.
For similar functionality, please visit
http://www.pathwaycommons.org/pcviz/

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.

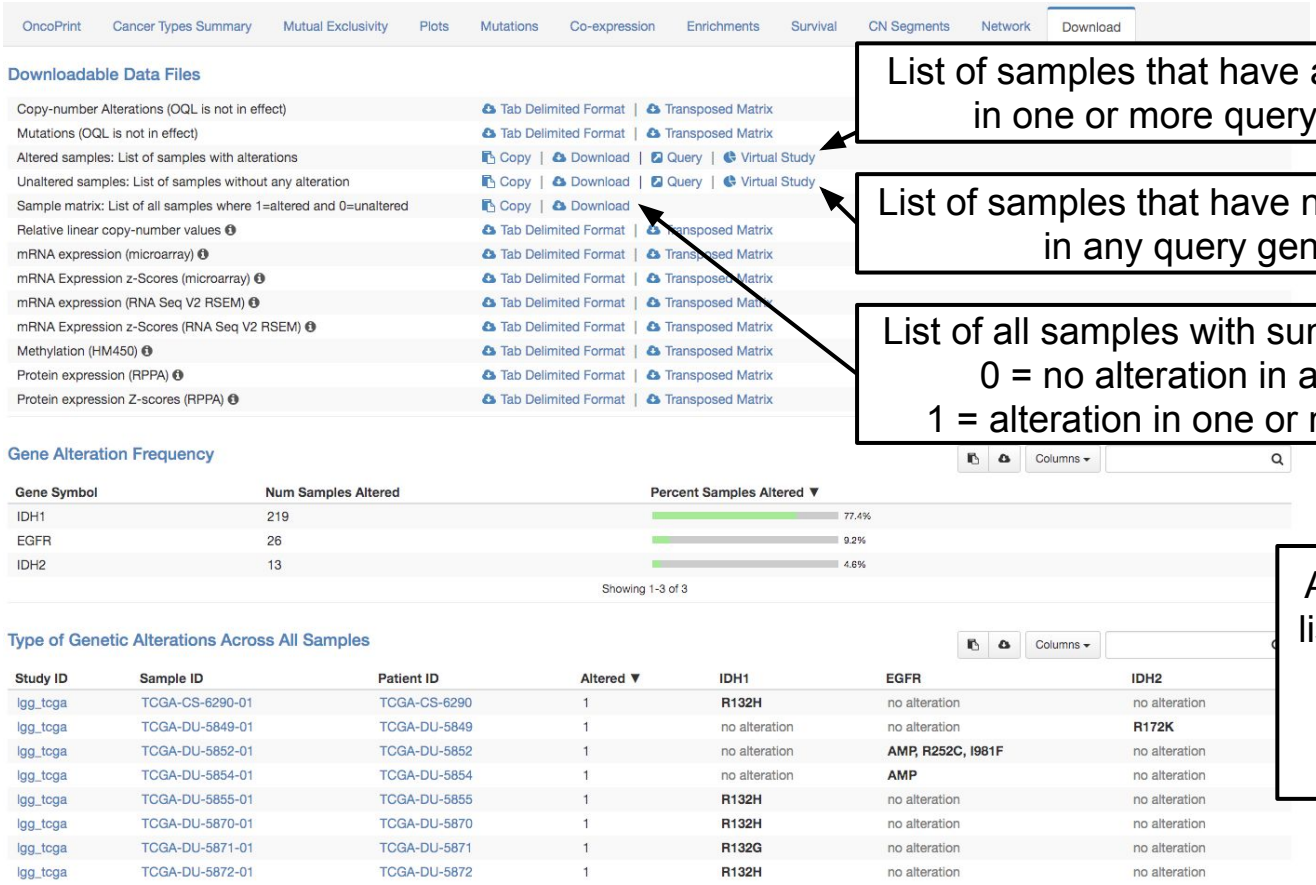
Frequency of gene alteration for each gene in the query

List of all samples with status of each query gene for each type of alteration queried (see OQL tutorial for explanation of alteration types included in a query)

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