

# cBioPortal Tutorial #5: Onco Query Language (OQL)

Use OQL to refine your queries

# Tutorial Objectives

- Introduce Onco Query Language (OQL)
- Explain the basic rules of OQL
- Show the utility of OQL with several examples

# Onco Query Language (OQL) Overview

## **What is OQL?**

OQL defines the specific types of alterations to be considered when running a query.

## **Why is OQL necessary or useful?**

When you run a query on one or more genes, OQL defines which genomic alterations count towards a sample being altered.

## **What does that actually mean?**

Let's look at an example. On the next slide is a query for IDH1, IDH2, EGFR and TP53 in the TCGA LGG PanCan Atlas Dataset.

# What happens in a regular query?

[Query](#) [Quick Search \*\*Beta!\*\*](#) [Download](#) Please cite: [Cerami et al., 2012](#) & [Gao et al., 2013](#)

**Selected Studies:** [Modify](#) Brain Lower Grade Glioma (TCGA, PanCancer Atlas) (514 total 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.](#)

Samples with mutation and CNA data (511) [×](#) [▼](#)

**Enter Genes:**

User-defined List [×](#) [▼](#)

IDH1 IDH2 EGFR TP53

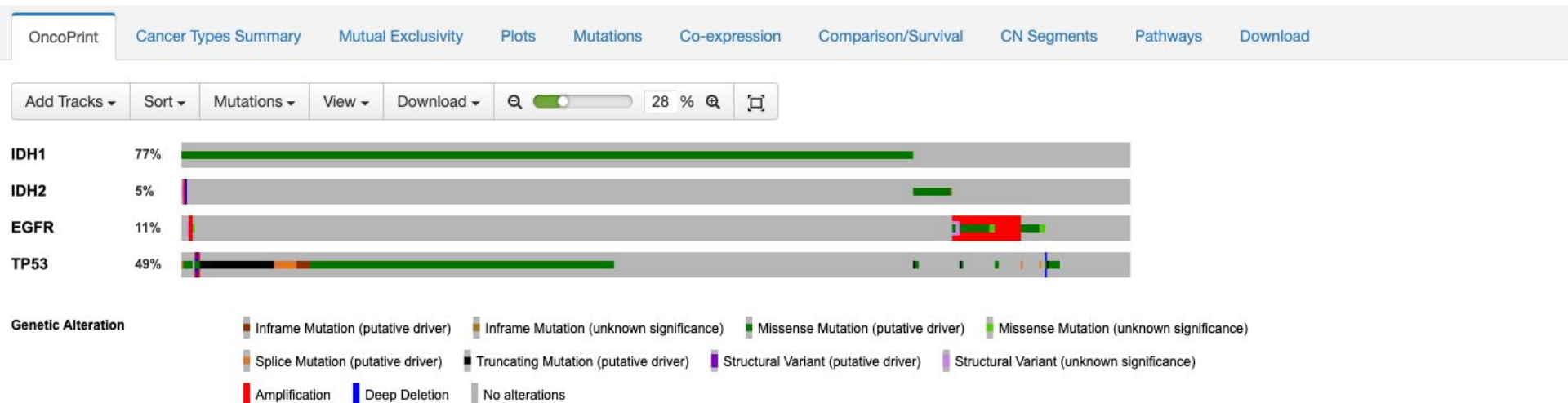
[Hint:](#) Learn Onco Query Language (OQL) to write more powerful queries [↗](#)

[✔](#) All gene symbols are valid.

Submit Query

# What happens in a regular query?

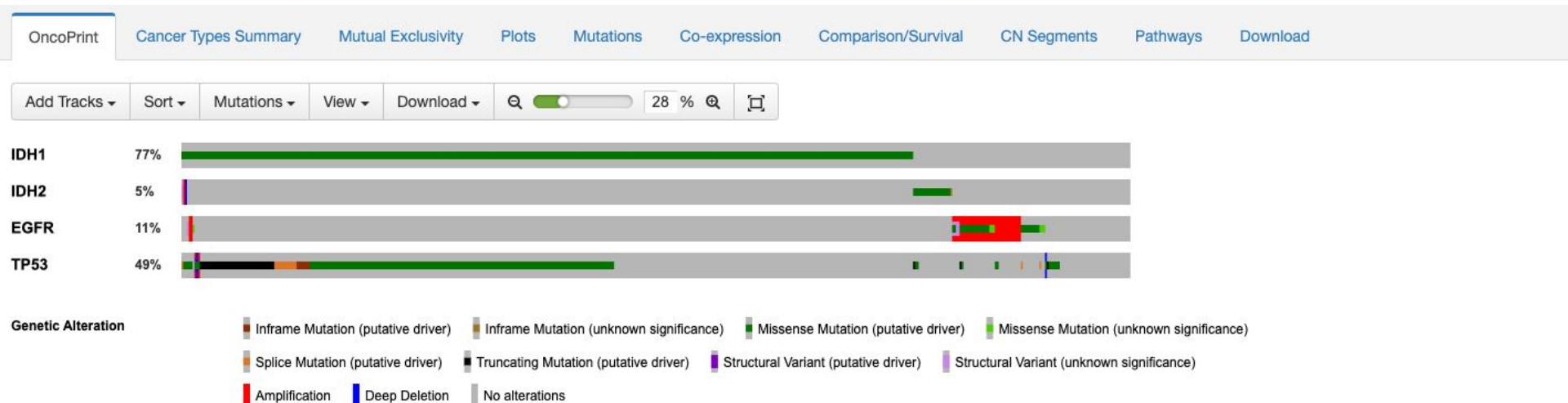
This query looks for samples with alterations in IDH1, IDH2, EGFR and TP53. We can see that the presence of any of four different alterations (Amplification, Deep Deletion, Mutation or Structural Variant) define a sample as having an alteration in a query gene.



# What happens in a regular query?

But how were those four types of alterations selected? How do we know if an alteration isn't present in the data or just isn't being examined in this query?

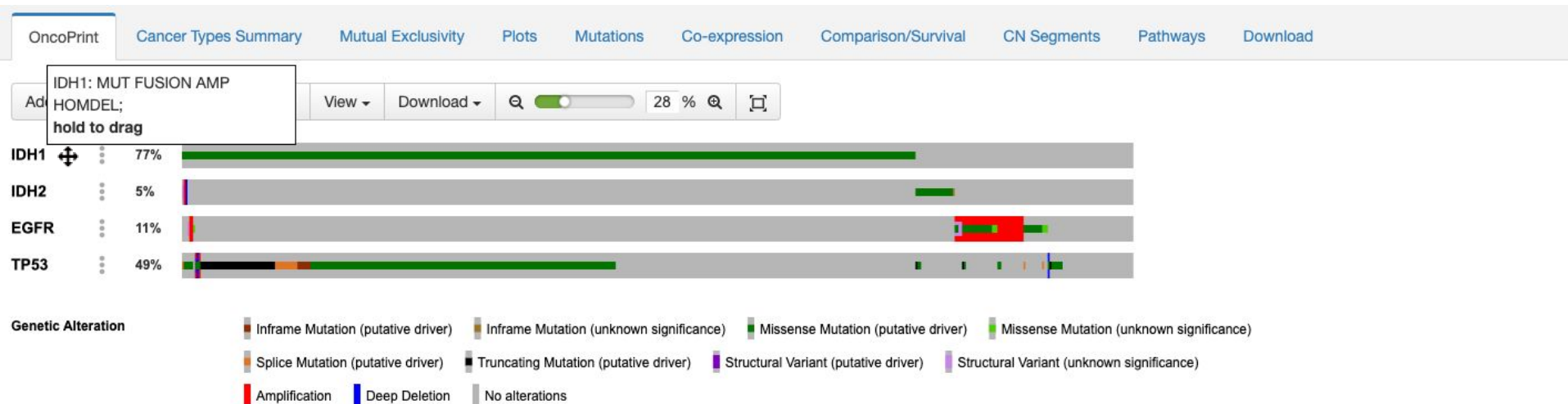
For example, are there samples with shallow deletions in any of the query genes?



# What happens in a regular query?

If you hover over a gene name, you can see the specific alterations which were included in the query: “MUT”, “FUSION”, “AMP”, “HOMDEL”. These are the default OQL options and will highlight any mutation, fusion/structural variant, amplification or homozygous/deep deletion in the query gene.

Note: Not all studies have all datatypes, for example many studies do not have fusion/structural variant calls.



# What happens in a regular query?

So let's come back to this question: are there samples with shallow deletions in any of the query genes?

Shallow deletions were not included in the OQL for this query, so there may be shallow deletions affecting these genes, but we won't see them because the query didn't look for them.

What if we want to include shallow deletions? How do we do that? Let's learn how to use OQL!



# The Rules of OQL

OQL uses keywords to define the alterations to include in a query. To the right is a table defining the general keywords (top) and the modifiers which can be applied to certain keywords (bottom). The complete specifications can be found [here](#).

## OQL Keywords

Users can define specific subsets of genetic alterations for five data types:

Data Type	Keywords and Syntax	Default*
Mutations	<code>MUT</code> All non-synonymous mutations <code>MUT = &lt;protein change&gt;</code> Specific amino acid changes (e.g. <code>V600E</code> or <code>V600</code> ) <code>MUT = &lt;mutation type&gt;</code> Acceptable values are: <code>MISSENSE</code> , <code>NONSENSE</code> , <code>NONSTART</code> , <code>NONSTOP</code> , <code>FRAMESHIFT</code> , <code>INFRAME</code> , <code>SPLICE</code> , <code>TRUNC</code>	<code>MUT</code>
Fusions	<code>FUSION</code> All fusions (note that many studies lack fusion data)	<code>FUSION</code>
Copy Number Alterations	<code>AMP</code> Amplifications <code>HOMDEL</code> Deep Deletions <code>GAIN</code> Gains <code>HETLOSS</code> Shallow Deletions Comparison operators can also be used with <code>CNA</code> (e.g. <code>CNA &gt;= GAIN</code> is the same as <code>AMP GAIN</code> )	<code>AMP</code> <code>HOMDEL</code>
mRNA Expression	<code>EXP &lt; -x</code> mRNA expression is less than <code>x</code> standard deviations (SD) below the mean <code>EXP &gt; x</code> mRNA expression is greater than <code>x</code> SD above the mean The comparison operators <code>&lt;=</code> and <code>&gt;=</code> also work	<code>EXP &gt;= 2</code> <code>EXP &lt;= -2</code>
Protein/phosphoprotein level	<code>PROT &lt; -x</code> Protein expression is less than <code>x</code> standard deviations (SD) below the mean <code>PROT &gt; x</code> Protein expression is greater than <code>x</code> SD above the mean The comparison operators <code>&lt;=</code> and <code>&gt;=</code> also work	<code>PROT &gt;= 2</code> <code>PROT &lt;= -2</code>

\* These are the default OQL keywords used for each data type when a gene is queried without any explicit OQL.

## OQL modifiers

Mutations and copy number alterations can be further refined using modifiers:

Keyword	Applicable Data Type	Explanation
<code>DRIVER</code>	Mutations Fusions Copy Number Alterations	Include only mutations, fusions and copy number alterations which are driver events, as defined in OncoPrint (default: OncoKB and CancerHotspots).
<code>GERMLINE</code>	Mutations	Include only mutations that are defined as germline events by the study.
<code>SOMATIC</code>	Mutations	Include all mutations that are not defined as germline.
<code>(a-b)</code> (protein position range)	Mutations	Include all mutations that overlap with the protein position range <code>a-b</code> , where <code>a</code> and <code>b</code> are integers. If you add a <code>*</code> (i.e. <code>(a-b*)</code> ) then it will only include those mutations that are fully contained inside <code>a-b</code> . The open-ended ranges <code>(a-)</code> and <code>(-b)</code> are also allowed.

# Using OQL

Let's re-create our initial query. On the left is the query as we ran it before. On the right are two different ways to write the exact same query using OQL.

Query

Quick Search **Beta!**

Download

Please c

Selected Studies:

Modify

Brain Lower Grade Glioma (TCGA, PanCancer Atlas)

(514 total 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.

Samples with mutation and CNA data (511)

×

▼

Enter Genes:

User-defined List

×

▼

IDH1 IDH2 EGFR TP53

✔ All gene symbols are valid.

Submit Query

## Enter Genes:

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

## Enter Genes:

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

User-defined List

×

▼

IDH1: MUT FUSION AMP HOMDEL  
IDH2: MUT FUSION AMP HOMDEL  
EGFR: MUT FUSION AMP HOMDEL  
TP53: MUT FUSION AMP HOMDEL

✔ All gene symbols are valid.

User-defined List

×

▼

DATATYPES: MUT FUSION AMP HOMDEL  
IDH1 IDH2 EGFR TP53

✔ All gene symbols are valid.

# Using OQL

The general format for OQL is “GENE: ALTERATION1 ALTERATION2 ...”. But as shown in the bottom example, the “DATATYPES” command allows a user to select the same set of alterations for multiple genes all at once.

Here each gene is listed on its own line, followed by a colon and then the list of alterations.

Enter Genes:

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

User-defined List ✕ ▼

IDH1: MUT FUSION AMP HOMDEL  
IDH2: MUT FUSION AMP HOMDEL  
EGFR: MUT FUSION AMP HOMDEL  
TP53: MUT FUSION AMP HOMDEL

✔ All gene symbols are valid.

Rather than writing the same alterations after each gene, the “DATATYPES” command can be used to set the alterations for multiple genes at once.

Enter Genes:

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

User-defined List ✕ ▼

DATATYPES: MUT FUSION AMP HOMDEL  
IDH1 IDH2 EGFR TP53

✔ All gene symbols are valid.

# Using OQL

Now let's adjust the default query. Let's look for gains in EGFR and shallow deletions in TP53. Add "GAIN" and "HETLOSS" to the query:

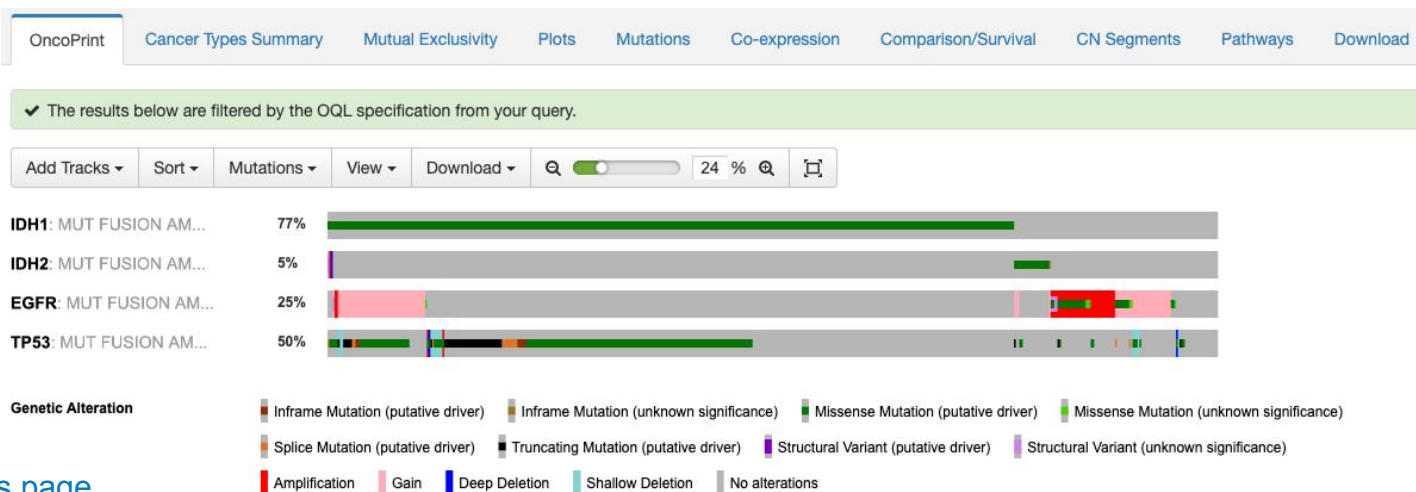
Enter Genes:

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

User-defined List

IDH1: MUT FUSION AMP HOMDEL  
IDH2: MUT FUSION AMP HOMDEL  
EGFR: MUT FUSION AMP HOMDEL GAIN  
TP53: MUT FUSION AMP HOMDEL HETLOSS

OncoPrint now includes gains in EGFR and shallow deletions in TP53:



# Using OQL

What if we want to look at IDH1 R132C mutations, but no other IDH1 alteration?

We can specify a specific mutation in OQL:

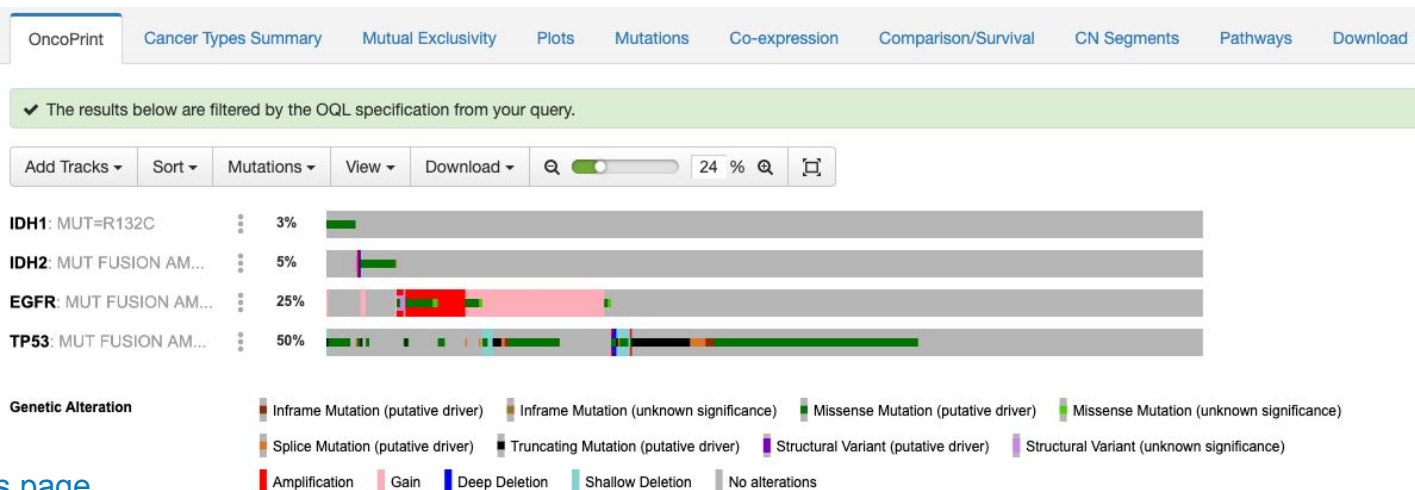
Enter Genes:

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

User-defined List

IDH1: MUT=R132C  
IDH2: MUT FUSION AMP HOMDEL  
EGFR: MUT FUSION AMP HOMDEL GAIN  
TP53: MUT FUSION AMP HOMDEL HETLOSS

We then see that there are many fewer samples with mutations in IDH1 since we have limited the query to the relatively rare R132C.



# Using OQL

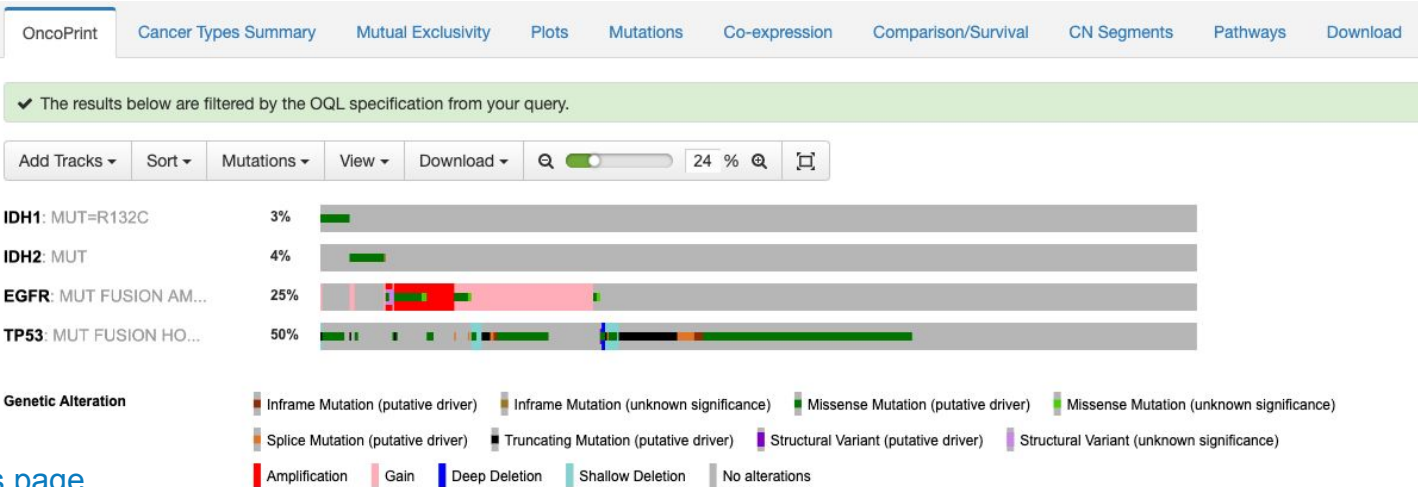
We can further refine the query by removing alteration types that are not biologically relevant, like deep deletions in IDH2 & EGFR:

Enter Genes:

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

User-defined List

IDH1: MUT=R132C  
IDH2: MUT  
EGFR: MUT FUSION AMP GAIN  
TP53: MUT FUSION HOMDEL HETLOSS



OQL Example:

BRCA1/2 inactivation in ovarian cancer

# OQL Example: BRCA1/2 inactivation

Loss of BRCA1 is a common event in ovarian cancer. What percentage of samples lose BRCA1? Let's run a query to find out:

Query Quick Search **Beta!** Download Please cite: Cerami et al., 2012 & Gao et al., 2013

**Selected Studies:** Modify

Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011) | (489 total samples)

**Select Genomic Profiles:**

☒ Mutations ⓘ  
☒ Putative copy-number alterations from GISTIC ⓘ  
☐ mRNA Expression. Select one of the profiles below:  
☐ mRNA expression z-scores relative to diploid samples ⓘ  
☐ mRNA expression z-scores relative to all samples (log microarray) ⓘ  
☐ microRNA expression Z-scores ⓘ  
☐ mRNA/miRNA expression Z-scores (all genes) ⓘ

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

Samples with mutation and CNA data (316) × ▼

**Enter Genes:**

User-defined List × ▼

BRCA1 BRCA2

✓ All gene symbols are valid.

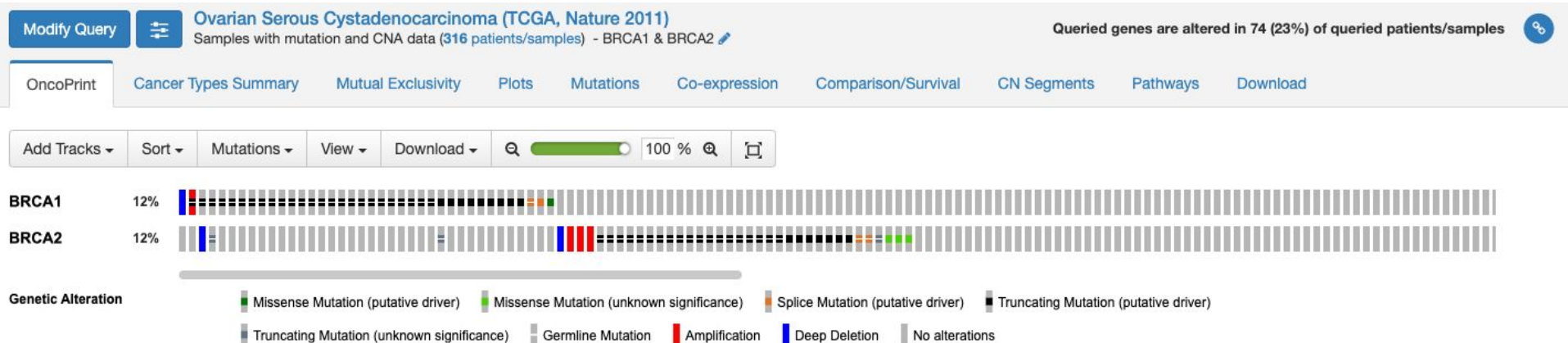
[Hint: Learn Onco Query Language \(OQL\) to write more powerful queries ↗](#)

Submit Query



# OQL Example: BRCA1/2 inactivation

Looking at OncoPrint, we can see that 12% of cases have an alteration in each of BRCA1 and BRCA2. However, this includes amplifications, which will not result in a loss of function. We can use OQL to make the query more specific.



# OQL Example: BRCA1/2 inactivation

Modify the query to include only mutations and deep deletions:

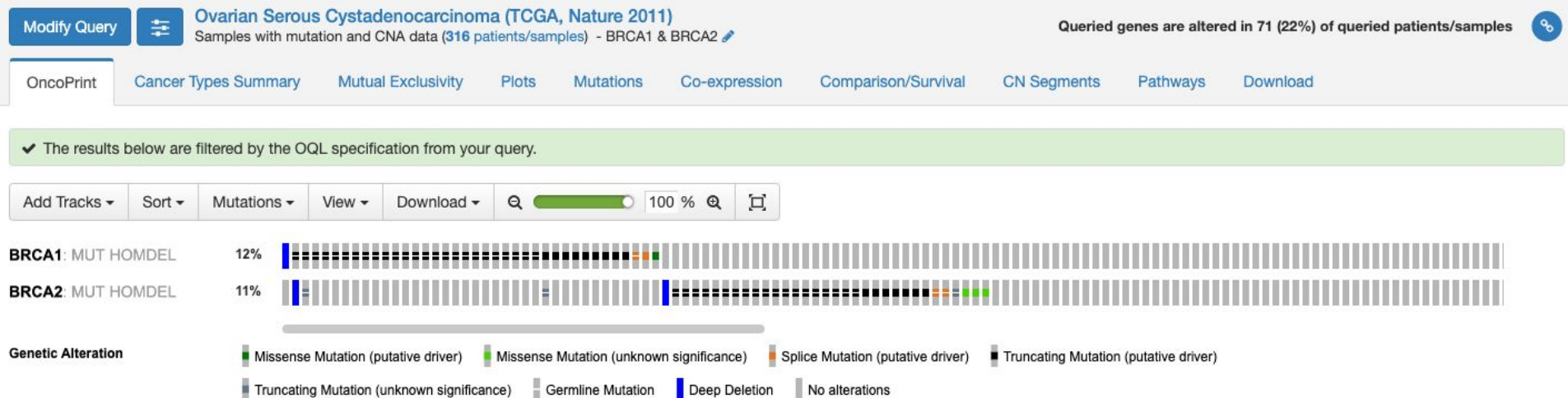
Enter Genes:

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

User-defined List

BRCA1: MUT HOMDEL  
BRCA2: MUT HOMDEL

OncoPrint now shows a more accurate estimate of the frequency of BRCA1/2 loss:



# OQL Example: BRCA1/2 inactivation

However, mutations and deletions are not the only way to decrease the levels of functional protein in a cell. DNA methylation can lead to decreased mRNA expression. We can use the “Plots” tab to examine the relationship between DNA methylation and gene expression. Note that the lower right quadrant contains samples with low expression and high DNA methylation. We can also use OQL to identify these samples.



# OQL Example: BRCA1/2 inactivation

Modify the query to also include samples with decreased expression (don't forget to select "mRNA Expression" in the "Genomics Profiles" section):

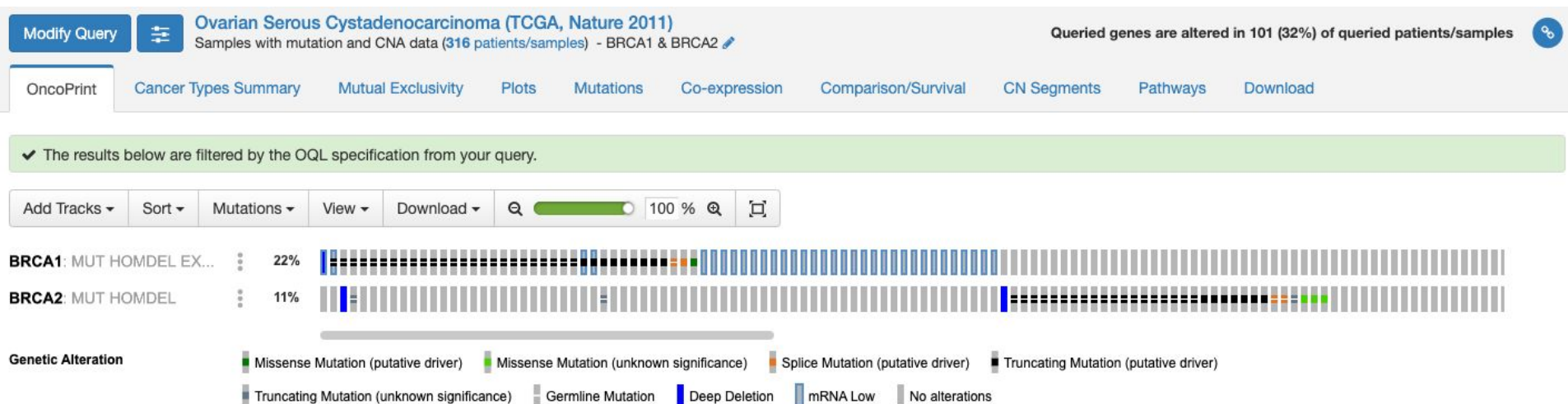
Enter Genes:

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

User-defined List

BRCA1: MUT HOMDEL EXP<-1.5  
BRCA2: MUT HOMDEL

OncoPrint now shows a more accurate estimate of the frequency of BRCA1 loss:



[Link to this page](#)

# OQL Example: BRCA1/2 inactivation

Some mutations in the OncoPrint are variants of unknown significance. Recall from the [Single Study Query Tutorial](#) that mutations are annotated as “putative drivers” or “unknown significance” based on this settings menu in the header:

The screenshot displays the OncoPrint web interface for a query titled "Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011)". The header indicates that 316 patients/samples have mutation and CNA data for BRCA1 & BRCA2, and that 101 (32%) of queried genes are altered. The "Annotate Data" settings menu is open, showing the following options:

- ☒ Putative drivers vs VUS:
- ☒ OncoKB driver annotation
- ☒ Hotspots
- ☐ cBioPortal  $\geq 0$
- ☐ COSMIC  $\geq 0$

Below the settings menu, the "Filter Data" section includes:

- ☐ 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

The main visualization area shows a genomic track with a search bar set to 100% zoom. A legend at the bottom identifies various mutation types: Missense Mutation (unknown significance), Splice Mutation (putative driver), Truncating Mutation (putative driver), Germline Mutation, Deep Deletion, mRNA Low, and No alterations.

# OQL Example: BRCA1/2 inactivation

We can further refine the query by only including those mutations which are putative drivers, as defined by the settings menu. We can do this by:

- Adding `_DRIVER` to the `MUT` term in OQL. This will include only mutations that are putative drivers (see BRCA1 below)
- Or, replace the entire OQL string with `DRIVER`. This will include mutations, fusions/structural variant and copy number changes that are putative drivers (see BRCA2 below)

Enter Genes:

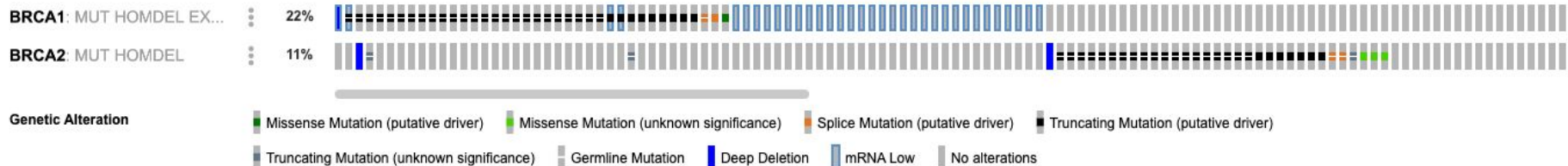
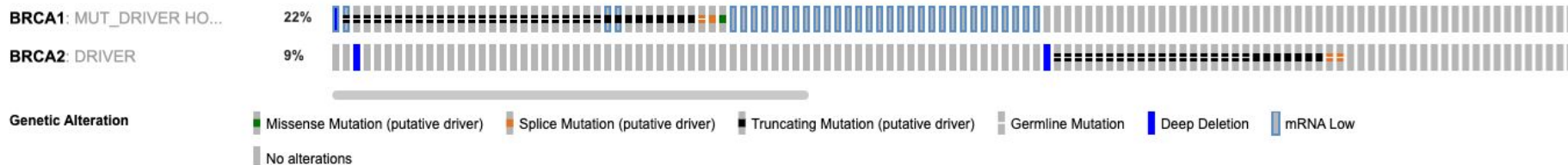
**Hint:** [Learn Onco Query Language \(OQL\)](#)  
to write more powerful queries [↗](#)

User-defined List ✕ ▼

BRCA1: `MUT_DRIVER HOMDEL EXP<-1.5`  
BRCA2: `DRIVER`

# OQL Example: BRCA1/2 inactivation

Compare the result of this latest query (top) with the previous query (bottom) and see that the mutations of unknown significance are no longer present.





# OQL Example: BRCA1/2 inactivation

This study is one of the few in cBioPortal that includes germline mutations. We can make one final adjustment to our query to ask a slightly different question: what percentage of samples have putative driver germline mutations in BRCA1/BRCA2? Note that the OQL for BRCA1 and BRCA2 are equivalent as the GERMLINE term only applies to mutations.

Enter Genes:

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

User-defined List

BRCA1: MUT\_GERMLINE\_DRIVER  
BRCA2: GERMLINE\_DRIVER

OncoPrint

Cancer Types Summary

Mutual Exclusivity

Plots

Mutations

Co-expression

Comparison/Survival

CN Segments

Pathways

Download

✓ The results below are filtered by the OQL specification from your query.

Add Tracks ▾

Sort ▾

Mutations ▾

View ▾

Download ▾

Q

100 %

Q

Q

BRCA1: MUT\_GERMLINE\_...

8%



BRCA2: MUT\_GERMLINE\_...

7%



Genetic Alteration

Splice Mutation (putative driver)

Truncating Mutation (putative driver)

Germline Mutation

No alterations

[Link to this page](#)



OQL Example:

RTK pathway alterations

# OQL Example: RTK pathway alterations

Alterations in RTK signaling pathway members are common in colorectal adenocarcinoma. What is the pattern of alterations across the different levels of the signaling pathway?

Recall that RTKs (e.g. EGFR, ERBB2) activate RAS (KRAS, NRAS, HRAS) which in turn activate RAF (BRAF, ARAF, RAF1) which in turn activate MEK (MAP2K1, MAP2K2). Let's query all of these genes:

[Query](#) [Quick Search](#) [Beta!](#) [Download](#) Please cite: [Cerami et al., 2012](#) & [Gao et al., 2013](#)

**Selected Studies:** [Modify](#)

Colorectal Adenocarcinoma (TCGA, PanCancer Atlas) (594 total 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) [?](#)

☐ mRNA expression z-scores relative to normal samples (log RNA Seq V2 RSEM) [?](#)

☐ Protein/phosphoprotein level. Select one of the profiles below:

☐ Protein expression z-scores (RPPA) [?](#)

☐ Protein level z-scores (mass spectrometry by CPTAC) [?](#)

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

Samples with mutation and CNA data (526) [x](#) [v](#)

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

User-defined List [x](#) [v](#)

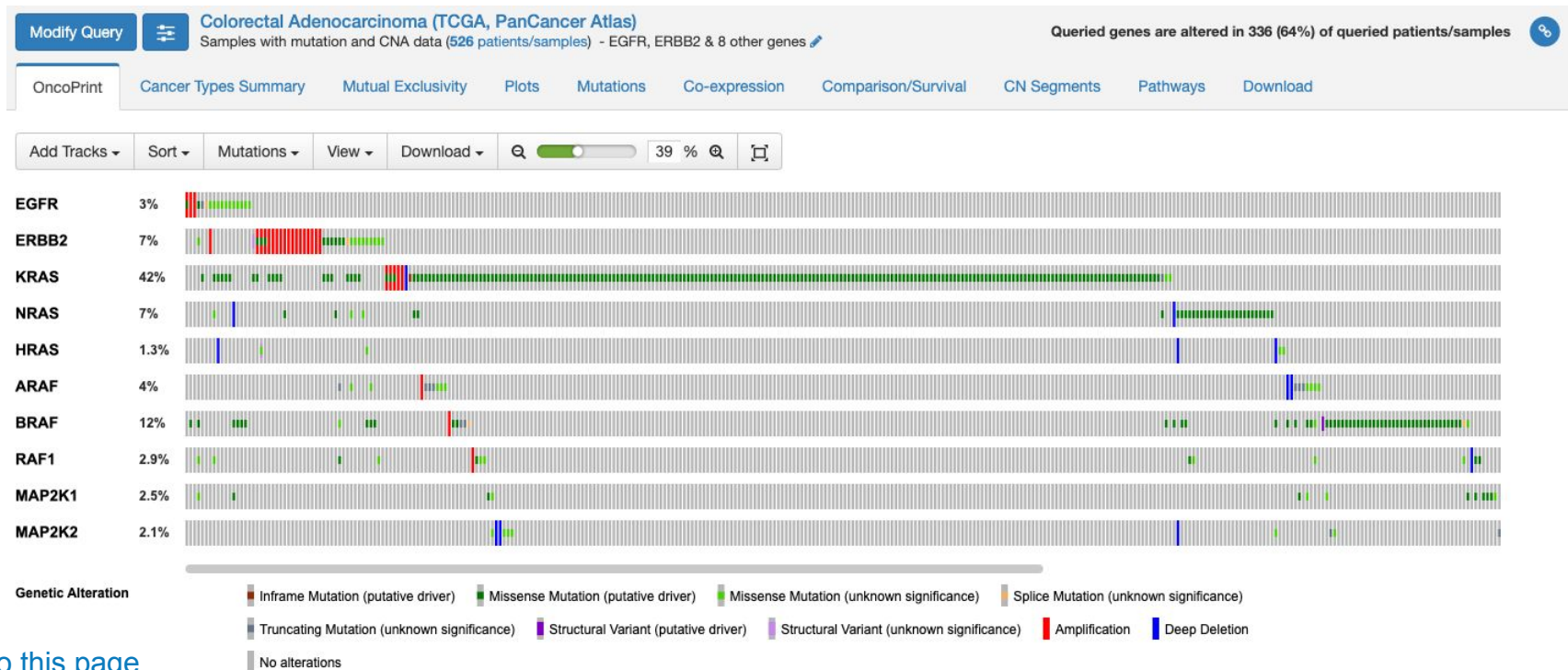
EGFR ERBB2  
KRAS NRAS HRAS  
ARAF BRAF RAF1  
MAP2K1 MAP2K2

[All gene symbols are valid.](#)

[Submit Query](#)

# OQL Example: RTK pathway alterations

We see here an overview of each individual gene in the pathway. However, it can be informative to instead see each level of the pathway grouped together.



# OQL Example: RTK pathway alterations

We can use gene tracks to group genes together in the OncoPrint. The format is [“optional track name” GENE1 GENE2 ... ]:

Enter Genes:

User-defined List

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

[EGFR ERBB2]  
["RAS" KRAS NRAS HRAS]  
["RAF" ARAF BRAF RAF1]  
["MEK" MAP2K1 MAP2K2]

Modify Query



Colorectal Adenocarcinoma (TCGA, PanCancer Atlas)

Samples with mutation and CNA data (526 patients/samples) - EGFR, ERBB2 & 8 other genes

Queried genes are altered in 336 (64%) 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 ▾



38 %



EGFR / ERBB2

10%



RAS

48%



RAF

17%



MEK

4%



Genetic Alteration



[Link to this page](#)

# OQL Example: RTK pathway alterations

Gene tracks can be combined with other OQL terms, either using the DATATYPES command as shown here, or attaching OQL to genes within the square brackets.

Now we can clearly visualize the pattern of mutual exclusivity of driver alterations at each level of the pathway.

Enter Genes:

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

User-defined List

DATATYPES: DRIVER  
[EGFR ERBB2]  
["RAS" KRAS NRAS HRAS]  
["RAF" ARAF BRAF RAF1]  
["MEK" MAP2K1 MAP2K2]

OncoPrint

Cancer Types Summary

Mutual Exclusivity

Plots

Mutations

Co-expression

Comparison/Survival

CN Segments

Pathways

Download

✓ The results below are filtered by the OQL specification from your query.

Add Tracks ▾

Sort ▾

Mutations ▾

View ▾

Download ▾

Q

40 %

Q

Q

EGFR / ERBB2

5%



RAS

46%



RAF

12%



MEK

1.5%



Genetic Alteration

Inframe Mutation (putative driver)

Missense Mutation (putative driver)


Structural Variant (putative driver)

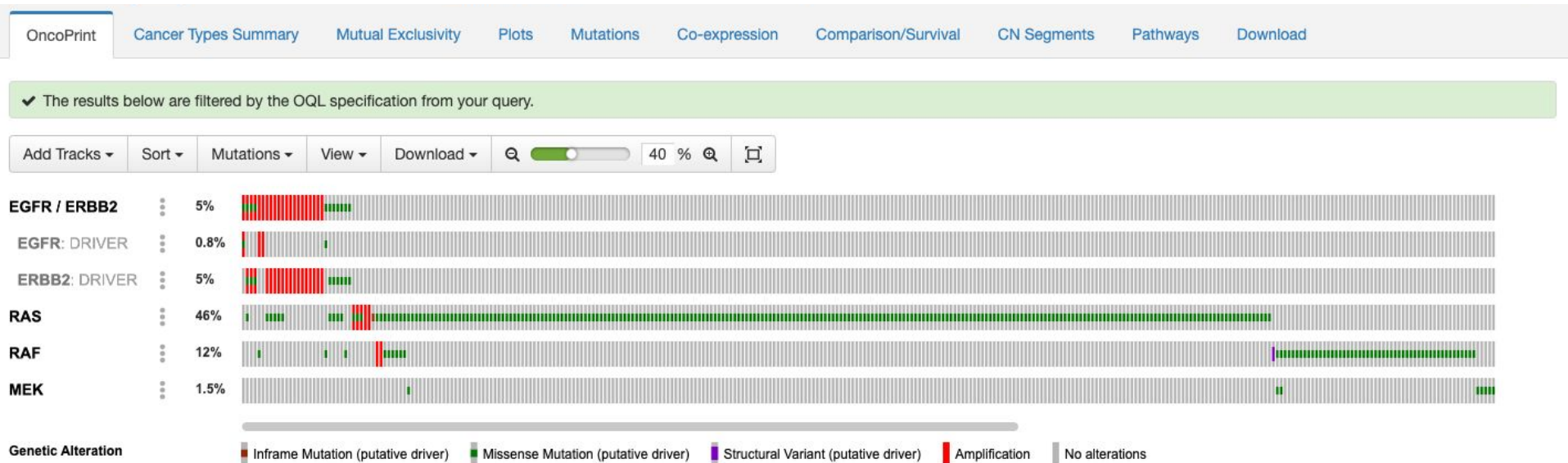
Amplification

No alterations

[Link to this page](#)

# OQL Example: RTK pathway alterations

Gene tracks can also be expanded to see tracks for individual genes. To expand, click the  symbol next to the track.



Note that OncoPrint, Mutual Exclusivity and Group Comparison are the only tabs that currently support gene tracks. All other tabs show individual genes rather than gene tracks.

Questions?

Check out the [OQL specification](#),

or our other tutorials,

or email us at:

`cbioportal@googlegroups.com`