cBioPortal Tutorial #5: Onco Query Language (OQL)

Use OQL to refine your queries

Last update: December 23, 2021

Tutorial Objectives

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

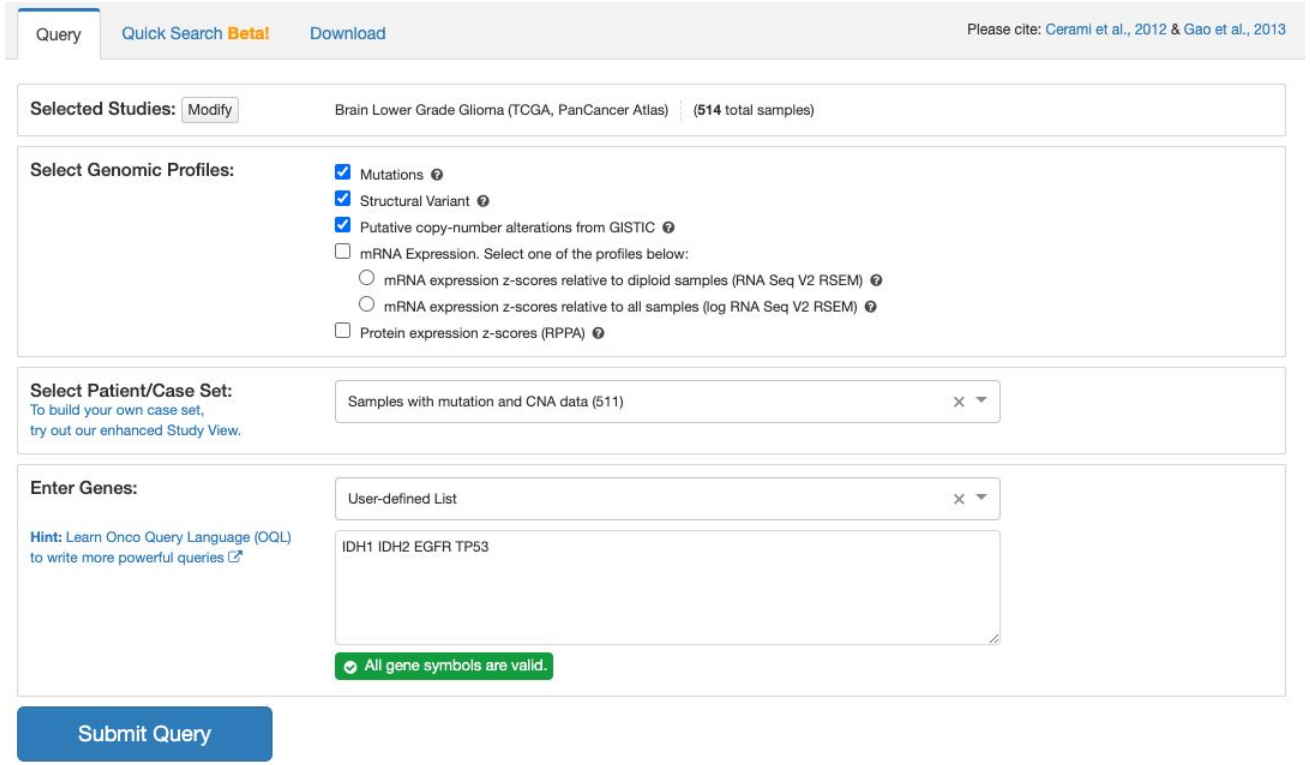
Onco Query Language (OQL) Overview

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.

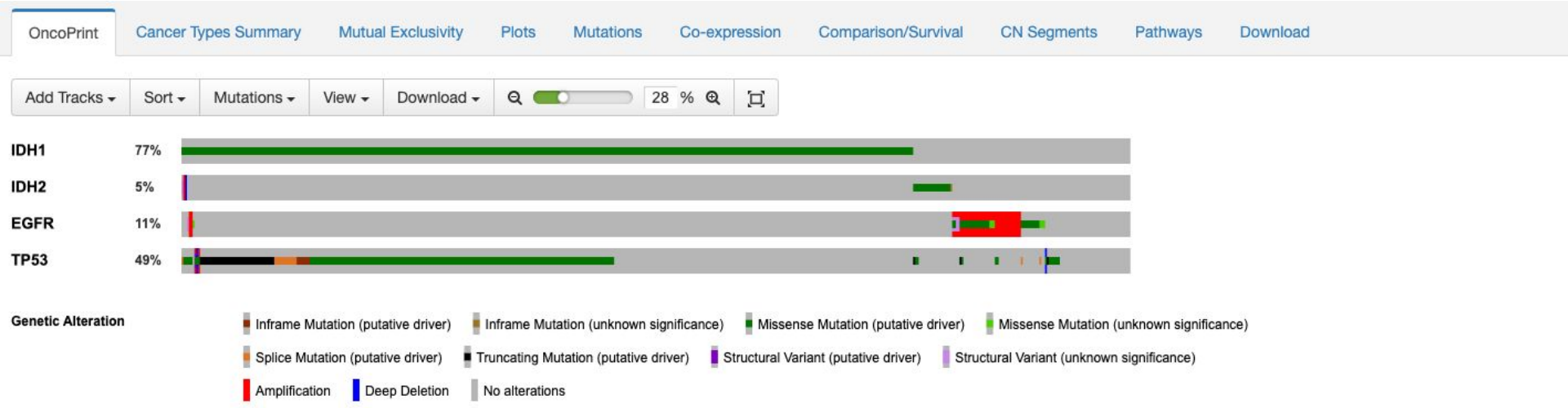
What does that actually mean?
Let’s look at an example. On the next slide is a query for IDH1, IDH2, EGFR and TP53 in the TCGA LGG PanCan Atlas Dataset.

What happens in a regular query?



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.

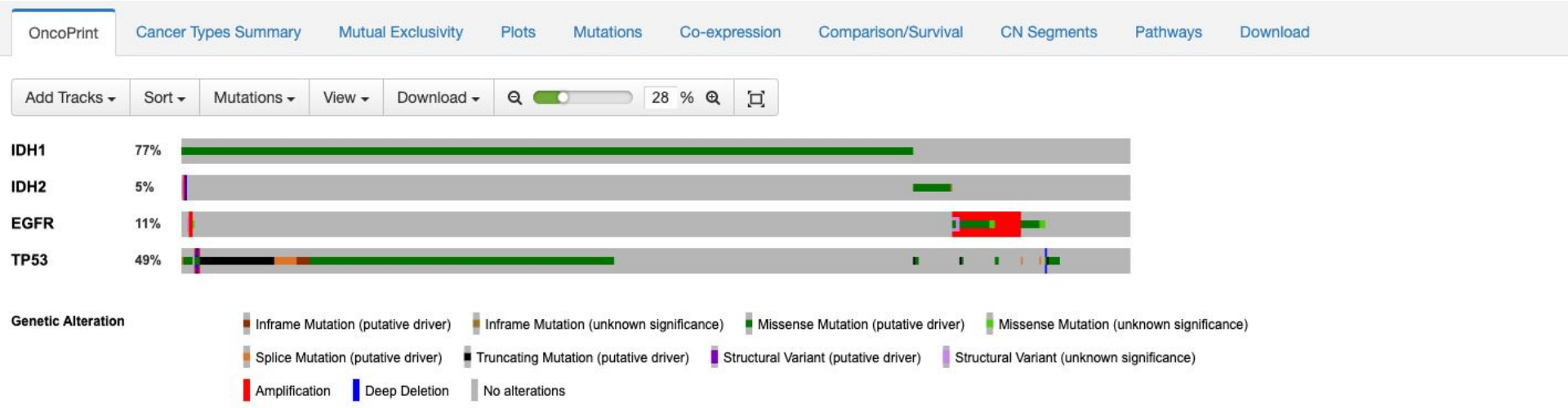


Link to this page

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?

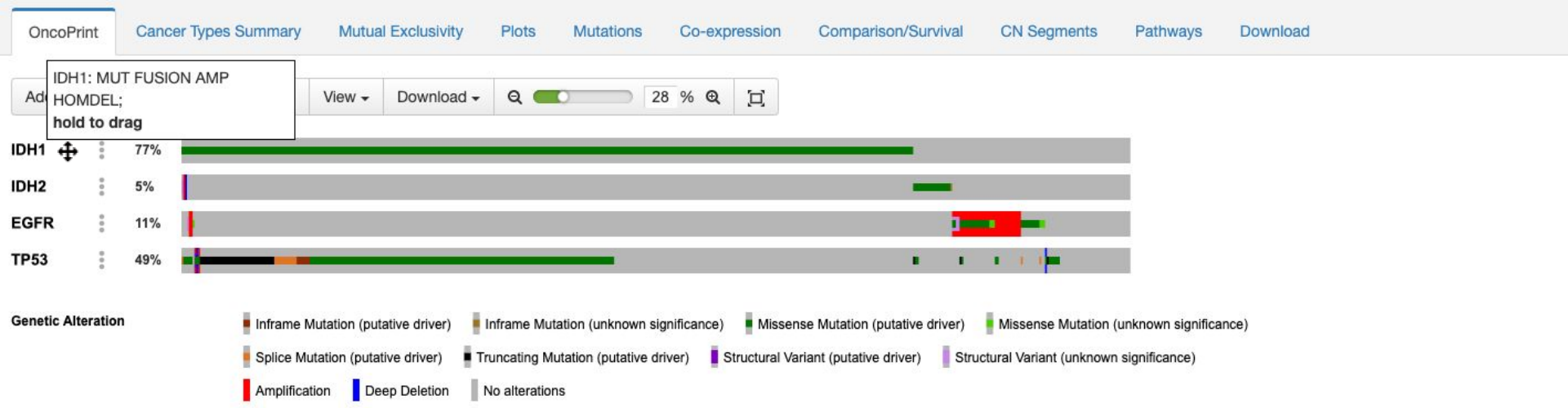


Link to this page

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.



Link to this page

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.

Link to this page

OQL Keywords

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

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

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

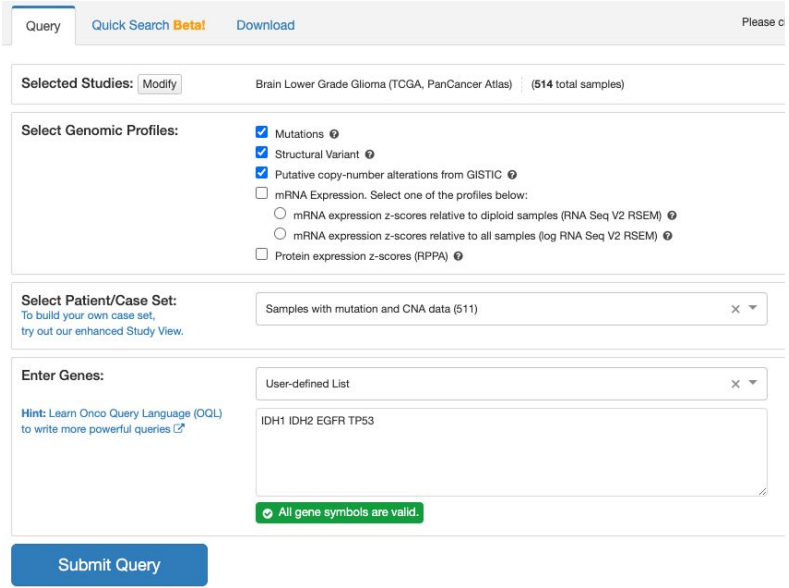
OQL modifiers

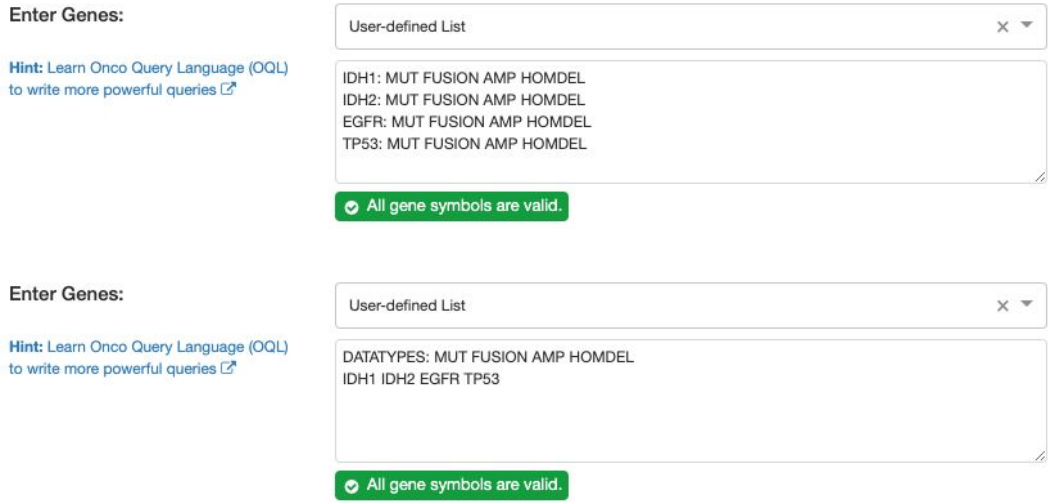
Mutations and copy number alterations can be further refined using modifiers:

Keyword	Applicable Data Type	Explanation
`DRIVER`	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).
`GERMLINE`	Mutations	Include only mutations that are defined as germline events by the study.
`SOMATIC`	Mutations	Include all mutations that are not defined as germline.
`(a-b)` (protein position range)	Mutations	Include all mutations that overlap with the protein position range `a-b`, where `a` and `b` are integers. If you add a `*` (i.e. `(a-b*)`) then it will only include those mutations that are fully contained inside `a-b`. The open-ended ranges `(a-)` and `(-b)` 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.

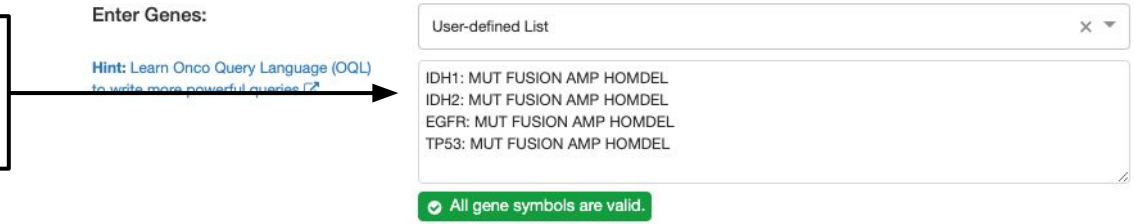




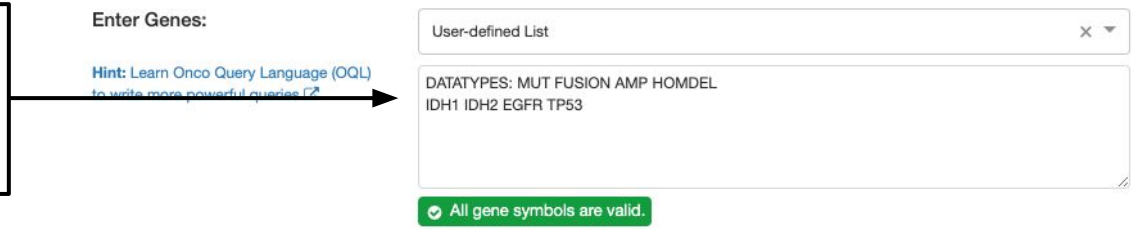
Using OQL

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.

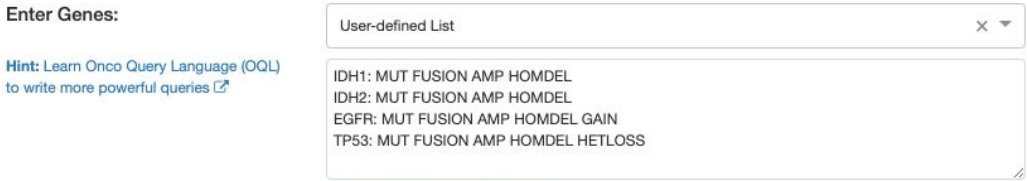


Rather than writing the same alterations after each gene, the “DATATYPES” command can be used to set the alterations for multiple genes at once.

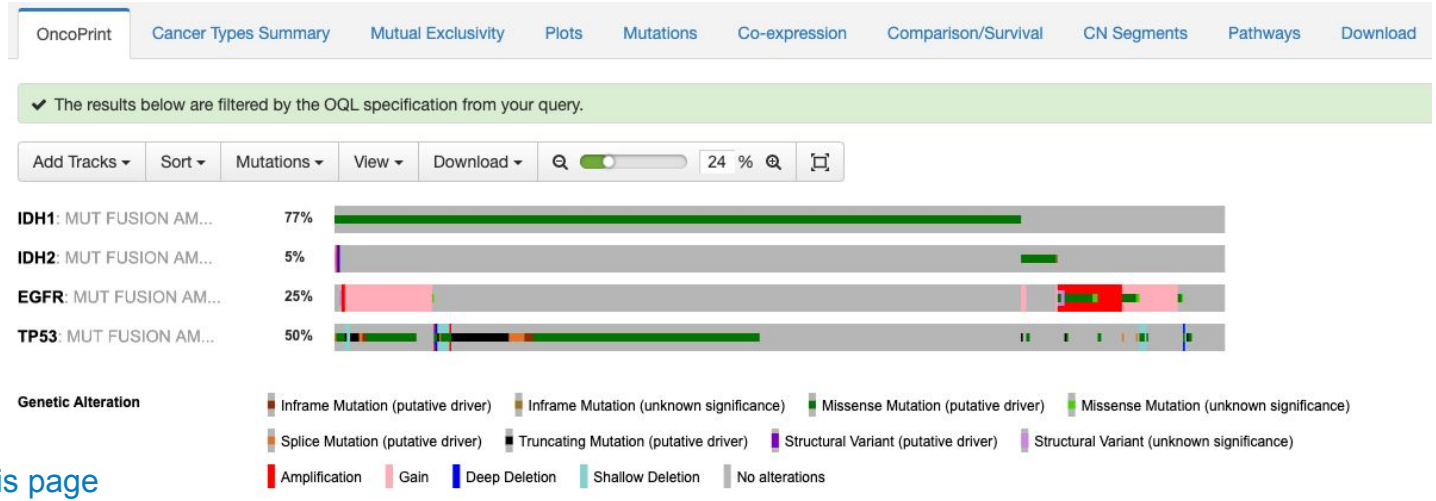


Using OQL

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:



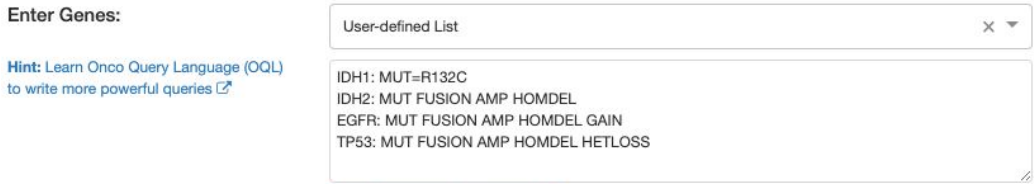
OncoPrint now includes gains in EGFR and shallow deletions in TP53:



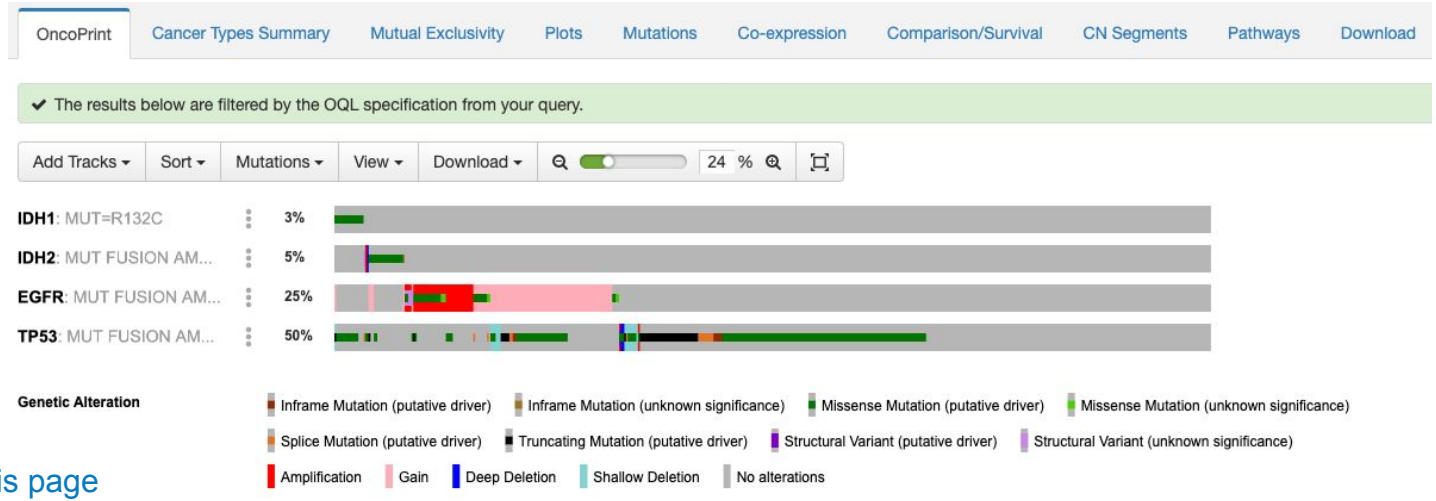
Link to this page

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:



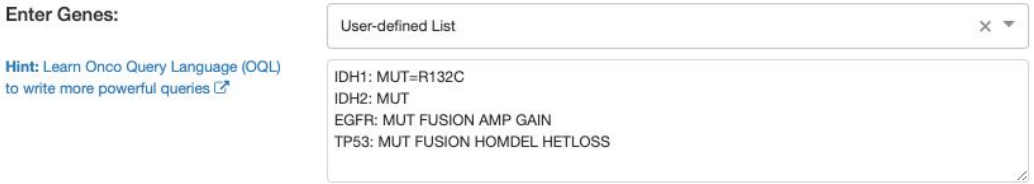
We then see that there are many fewer samples with mutations in IDH1 since we have limited the query to the relatively rare R132C.

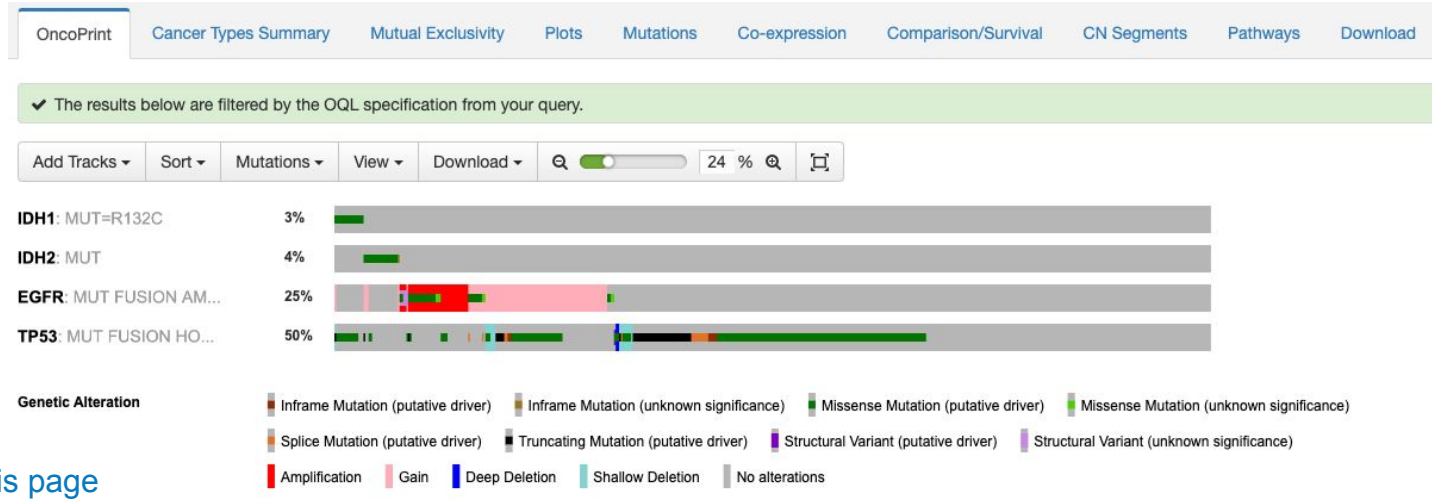


Link to this page

Using OQL

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





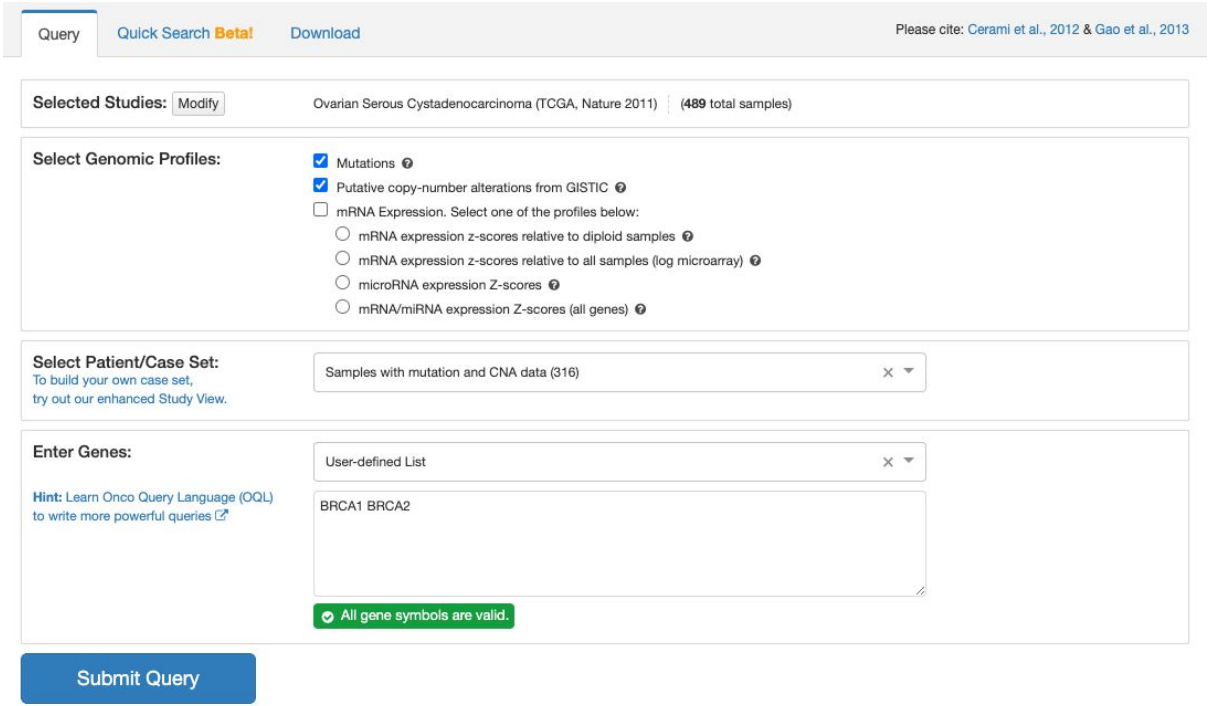
Link to this page

OQL Example:

BRCA1/2 inactivation in ovarian cancer

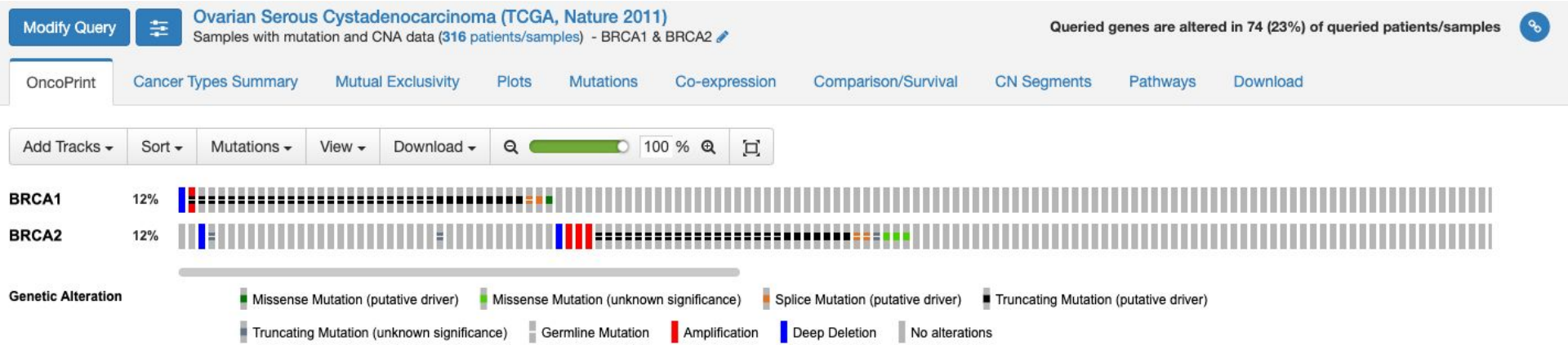
OQL Example: BRCA1/2 inactivation

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



OQL Example: BRCA1/2 inactivation

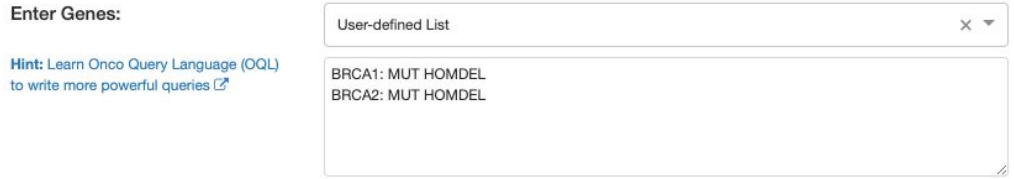
Looking at OncoPrint, we can see that 12% of cases have an alteration in each of BRCA1 and BRCA2. However, this includes amplifications, which will not result in a loss of function. We can use OQL to make the query more specific.



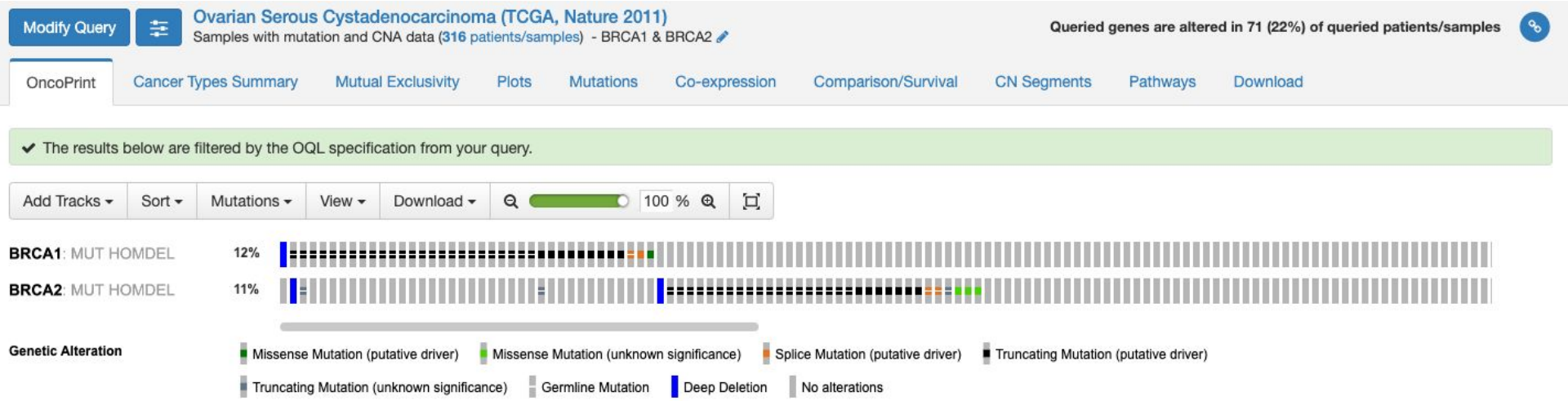
Link to this page

OQL Example: BRCA1/2 inactivation

Modify the query to include only mutations and deep deletions:



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



Link to this page

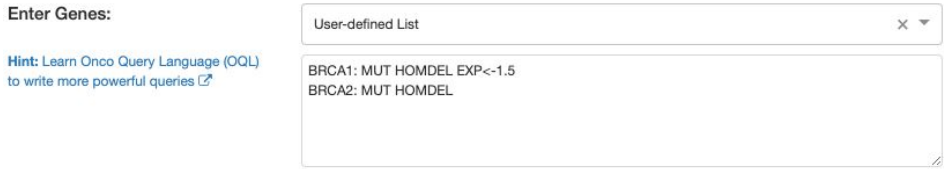
OQL Example: BRCA1/2 inactivation

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.

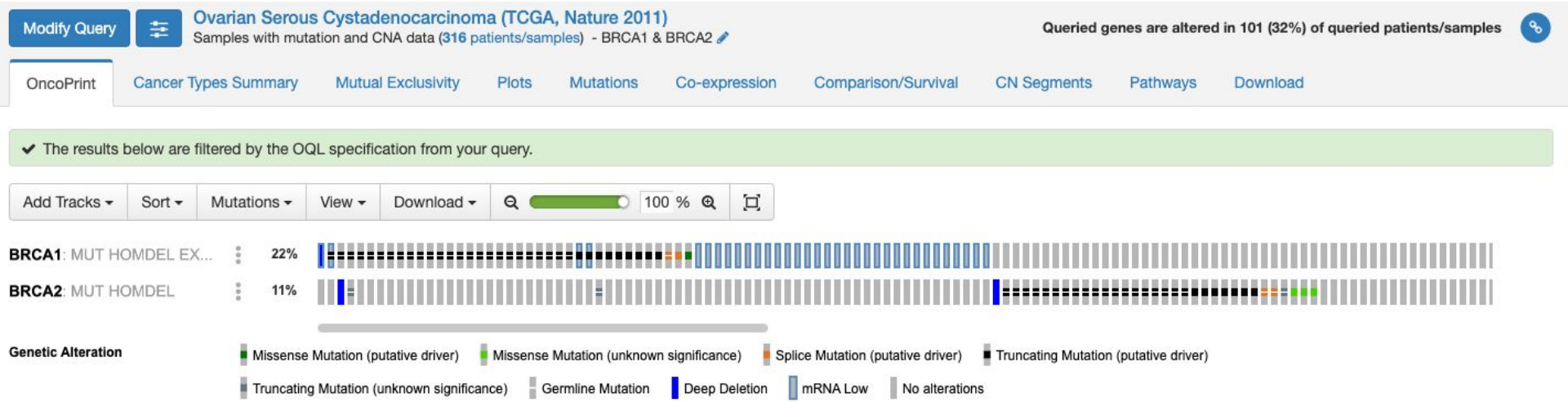
Link to this page

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):



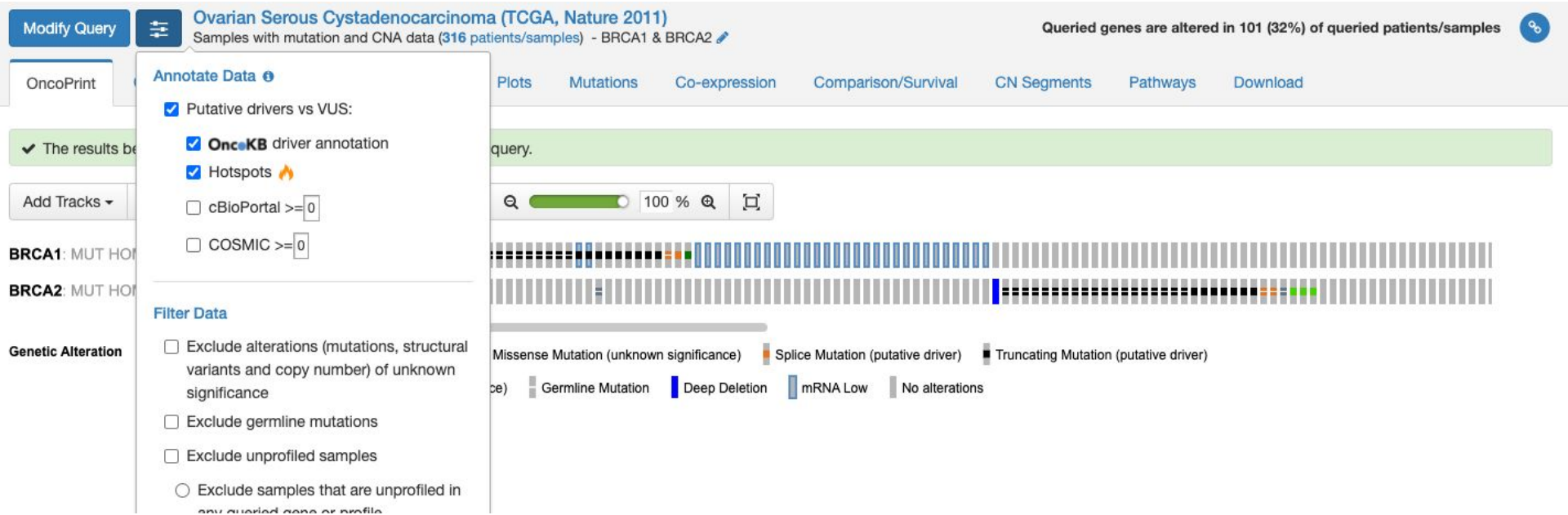
OncoPrint now shows a more accurate estimate of the frequency of BRCA1 loss:



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:

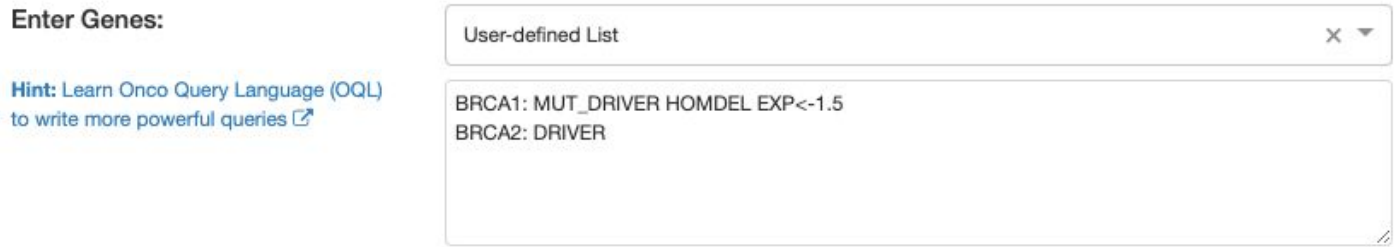


Link to this page

OQL Example: BRCA1/2 inactivation

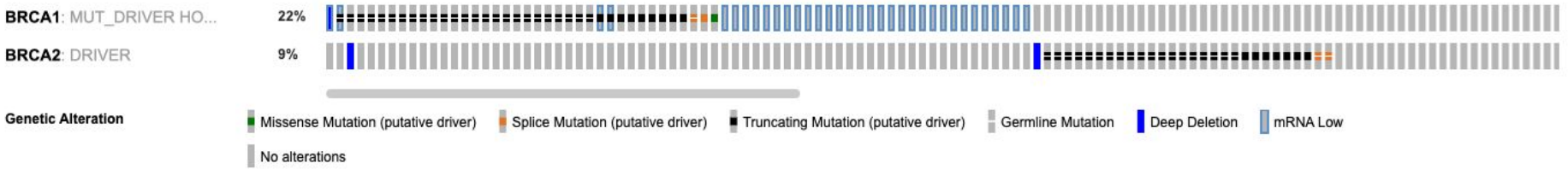
We can further refine the query by only including those mutations which are putative drivers, as defined by the settings menu. We can do this by:

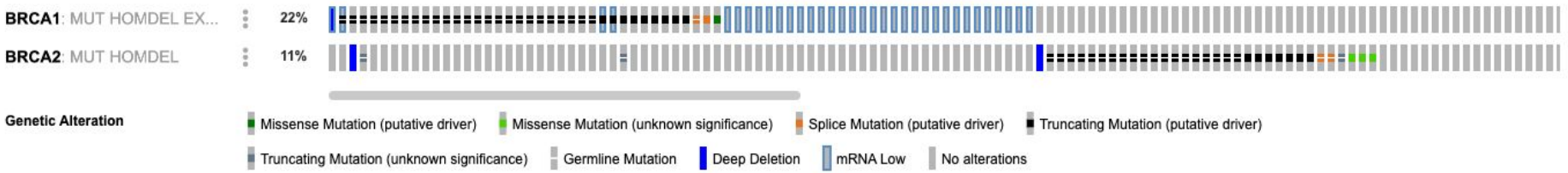
- Adding _DRIVER to the MUT term in OQL. This will include only mutations that are putative drivers (see BRCA1 below)
- Or, replace the entire OQL string with DRIVER. This will include mutations, fusions/structural variant and copy number changes that are putative drivers (see BRCA2 below)



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.

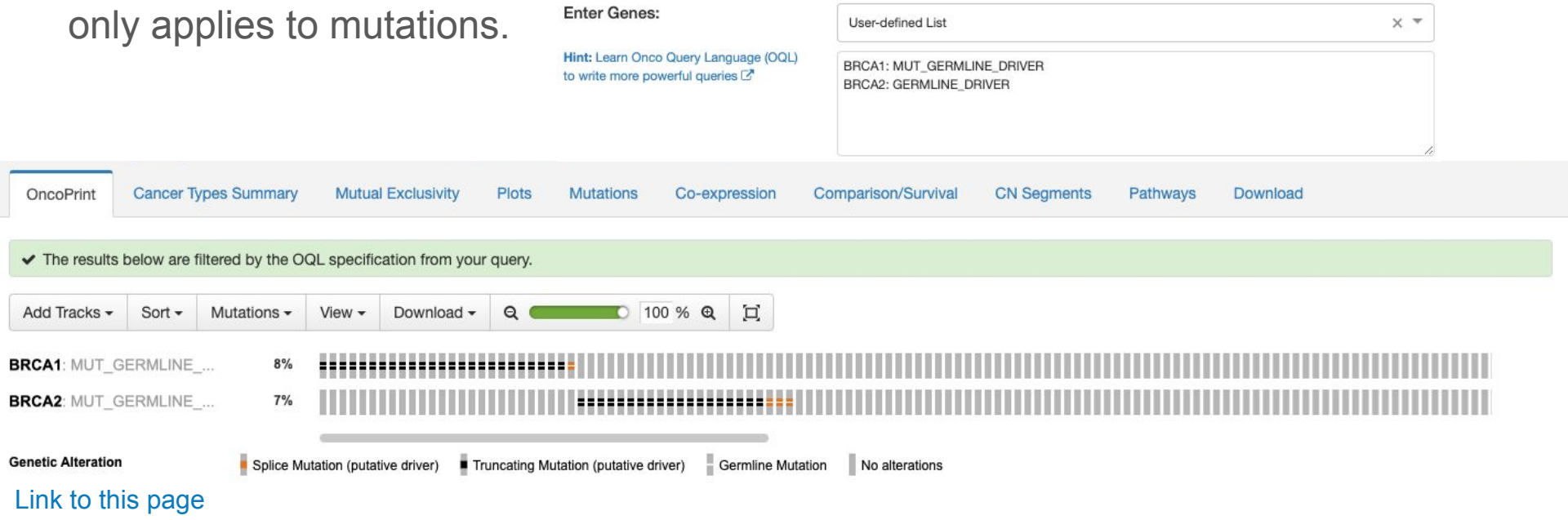




Link to this page

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.



Link to this page

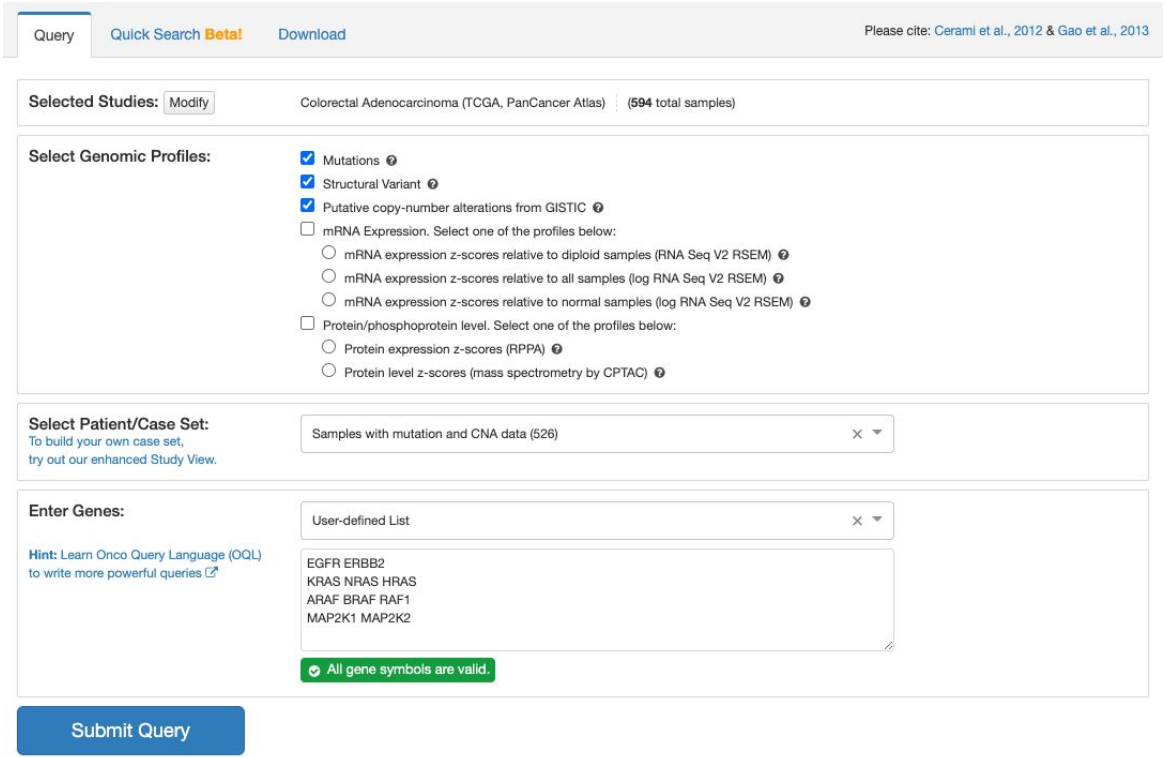
OQL Example:

RTK pathway alterations

OQL Example: RTK pathway alterations

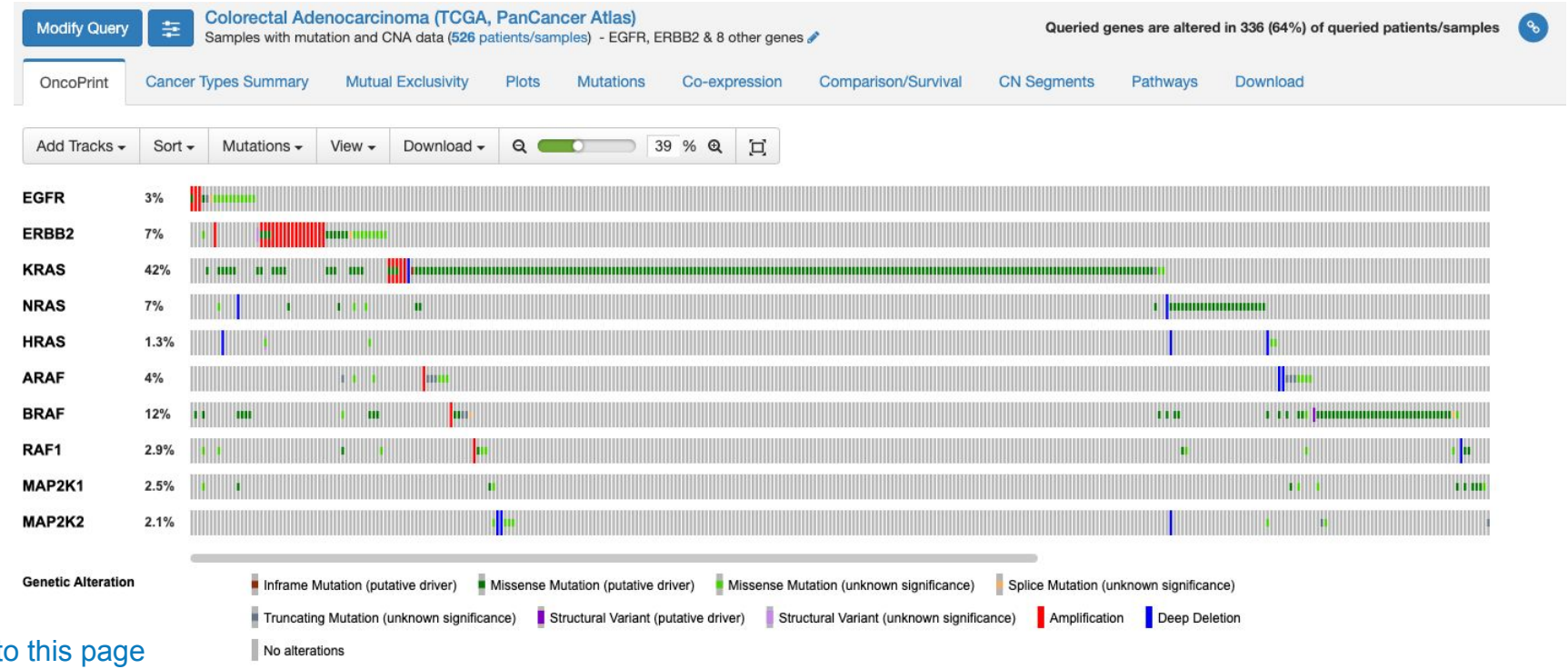
Alterations in RTK signaling pathway members are common in colorectal adenocarcinoma. What is the pattern of alterations across the different levels of the signaling pathway?

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:



OQL Example: RTK pathway alterations

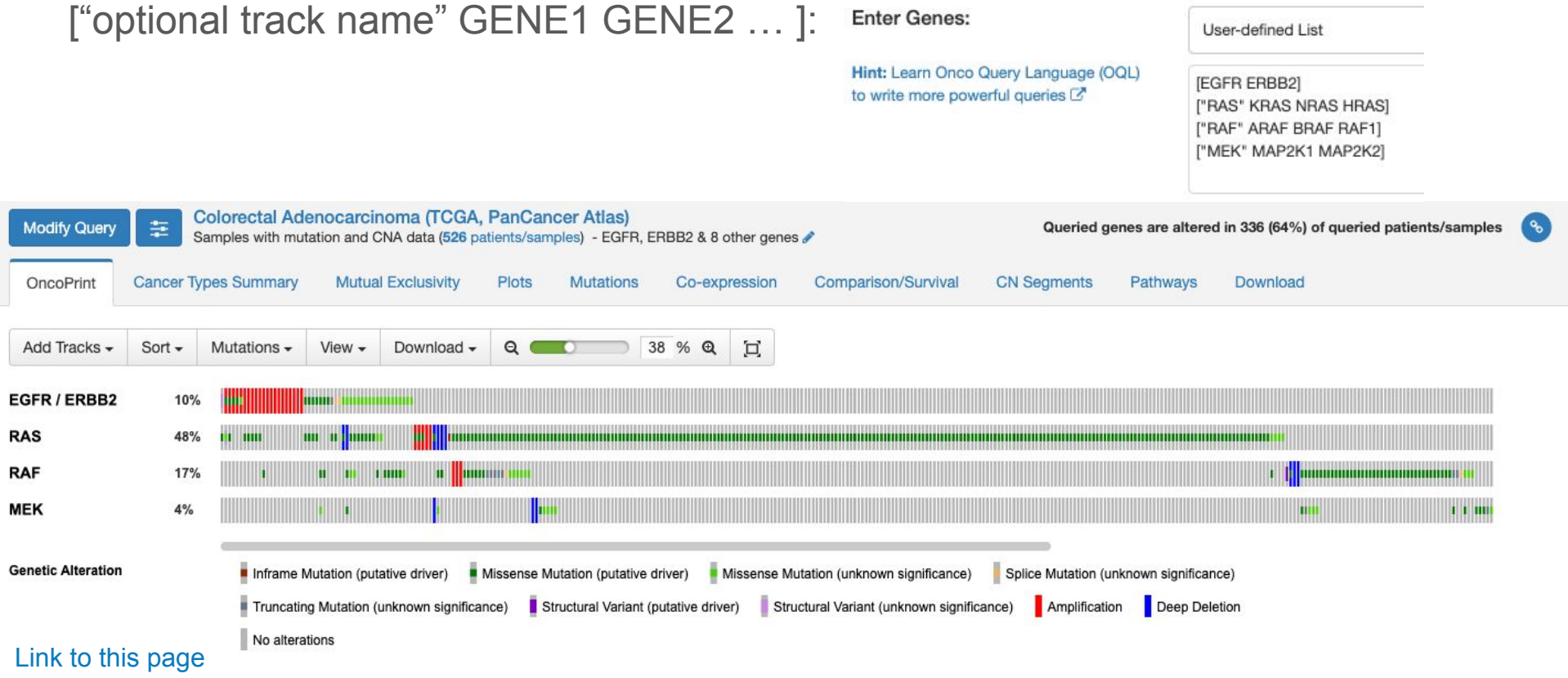
We see here an overview of each individual gene in the pathway. However, it can be informative to instead see each level of the pathway grouped together.



Link to this page

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 …]:



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.

Enter Genes:

User-defined List

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

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

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

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.

Link to this page

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