

# 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 Provisional Dataset.

# What happens in a regular query?

QUERY

DOWNLOAD DATA

Select Studies:

1 studies selected (532 samples) Deselect all View summary

Search...

|                       |    |  |
|-----------------------|----|--|
| PanCancer Studies     | 3  | <div>Select all listed studies matching filter (15)</div> <div>CNS/Brain</div> <div>Diffuse Glioma</div> <div><input type="checkbox"/> Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 514 samples</div> <div><input checked="" type="checkbox"/> Brain Lower Grade Glioma (TCGA, Provisional) 532 samples</div> <div><input type="checkbox"/> Low-Grade Gliomas (UCSF, Science 2014). 61 samples</div> <div><input type="checkbox"/> Merged Cohort of LGG and GBM (TCGA, Cell 2016) 1122 samples</div> <div>→ GLIOBLASTOMA</div> <div><input type="checkbox"/> Glioblastoma (TCGA, Cell 2013) 585 samples</div> <div><input type="checkbox"/> Glioblastoma (TCGA, Nature 2008) 206 samples</div> <div><input type="checkbox"/> Glioblastoma Multiforme (TCGA, PanCancer Atlas) 592 samples</div> <div><input type="checkbox"/> Glioblastoma Multiforme (TCGA, Provisional) 604 samples</div> <div>→ OLIGODENDROGLIOMA</div> <div><input type="checkbox"/> NGS in Anaplastic Oligodendroglioma and Anaplastic Oligoastrocytom... 22 samples</div> <div>Embryonal Tumor</div> <div>→ MEDULLOBLASTOMA</div> |
| Cell lines            | 2  |  |
| Adrenal Gland         | 2  |  |
| Ampulla of Vater      | 1  |  |
| Biliary Tract         | 6  |  |
| Bladder/Urinary Tract | 12 |  |
| Bone                  | 2  |  |
| Bowel                 | 7  |  |
| Breast                | 14 |  |
| CNS/Brain             | 15 |  |

Select Genomic Profiles:

☒ Mutations

☒ Putative copy-number alterations from GISTIC

☐ mRNA Expression. Select one of the profiles below:

☐ mRNA Expression z-Scores (microarray)

☐ mRNA Expression z-Scores (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.

Tumor Samples with sequencing and CNA data (283)

Enter Genes:

Advanced: Onco Query Language (OQL)

User-defined List

Select from Recurrently Mutated Genes (MutSig)

Select Genes from Recurrent CNAs (Gistic)

IDH1 IDH2 EGFR TP53

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 three different alterations (Amplification, Deep Deletion or Mutation) define a sample as having an alteration in a query gene.



# What happens in a regular query?

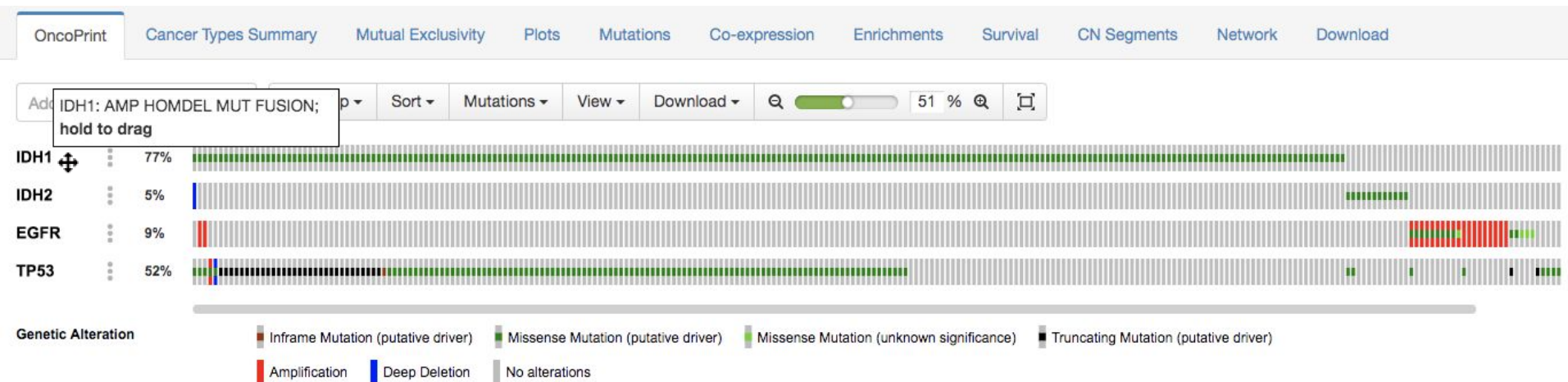
But how were those three types of alterations selected? How do we know if an alteration isn't present in the data or just isn't being examined in this query?

For example, are there samples with fusions or 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, amplification or homozygous/deep deletion in the query gene.



# What happens in a regular query?

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

Because "FUSION" is included in the OQL but is not shown in OncoPrint, we know that there are no reported fusions involving these genes in the data uploaded to cBioPortal for this study. An important caveat here is that many studies do not standardly report fusions, so the absence of data does not necessarily reflect that fusions are not truly present in the samples.

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

\* 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  |
|-----------------|---|--|
| <b>DRIVER</b>   | Mutations<br>Fusions<br>Copy Number Alterations | Include only mutations, fusions and copy number alterations which are driver events, as defined in OncoPrint (default: OncoKB and CancerHotspots). |
| <b>GERMLINE</b> | Mutations                                       | Include only mutations that are defined as germline events by the study.   |
| <b>SOMATIC</b>  | Mutations                                       | Include all mutations that are not defined as germline.  |

# 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

DOWNLOAD DATA

Select Studies:

1 studies selected (532 samples) Deselect all [View summary](#)

Search...

PerCancer Studies 3

Cell lines 2

Adrenal Gland 2

Ampulla of Vater 1

Biliary Tract 6

Bladder/Urinary Tract 12

Bone 2

Bowel 7

Breast 14

CNS/Brain 15

☐ Select all listed studies matching filter (15)

**CNS/Brain**

Diffuse Glioma

☐ Brain Lower Grade Glioma (TCGA, PanCancer Atlas)

☒ Brain Lower Grade Glioma (TCGA, Provisional)

☐ Low-Grade Gliomas (UCSF, Science 2014)

☐ Merged Cohort of LGG and GBM (TCGA, Cell 2016)

514 samples

532 samples

61 samples

1122 samples

→ **GLIOBLASTOMA**

☐ Glioblastoma (TCGA, Cell 2013)

☐ Glioblastoma (TCGA, Nature 2008)

☐ Glioblastoma Multiforme (TCGA, PanCancer Atlas)

☐ Glioblastoma Multiforme (TCGA, Provisional)

585 samples

206 samples

592 samples

604 samples

→ **OLIGODENDROGLIOMA**

☐ NGS in Anaplastic Oligodendroglioma and Anaplastic Oligastrocytom...

22 samples

Embryonal Tumor

→ **MEDULLOBLASTOMA**

Select Genomic Profiles:

☒ Mutations

☒ Putative copy-number alterations from GISTIC

☐ mRNA Expression. Select one of the profiles below:

☐ mRNA Expression z-Scores (microarray)

☐ mRNA Expression z-Scores (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.

Tumor Samples with sequencing and CNA data (283)

Enter Genes:

Advanced: Onco Query Language (OQL)

User-defined List

Select from Recurrently Mutated Genes (MutSig)

Select Genes from Recurrent CNAs (Gistic)

IDH1 IDH2 EGFR TP53

All gene symbols are valid.

Submit Query

## Enter Genes:

Advanced: Onco Query Language (OQL)

### User-defined List

Select from Recurrently Mutated Genes (MutSig)

Select Genes from Recurrent CNAs (Gistic)

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.

## Enter Genes:

Advanced: Onco Query Language (OQL)

### User-defined List

Select from Recurrently Mutated Genes (MutSig)

Select Genes from Recurrent CNAs (Gistic)

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:

Advanced: Onco Query Language (OQL)

User-defined List ✕ ▼

Select from Recurrently Mutated Genes (MutSig) Select Genes from Recurrent CNAs (Gistic)

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:

Advanced: Onco Query Language (OQL)

User-defined List ✕ ▼

Select from Recurrently Mutated Genes (MutSig) Select Genes from Recurrent CNAs (Gistic)

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:**  
Advanced: Onco Query Language (OQL)

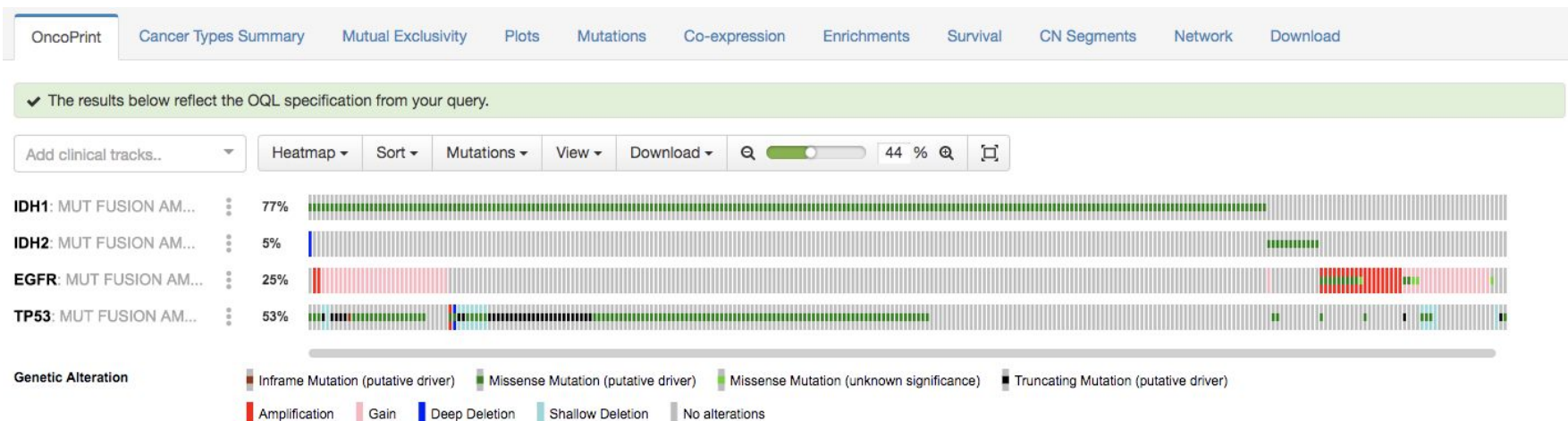
User-defined List

Select from Recurrently Mutated Genes (MutSig)

Select Genes from Recurrent CNAs (Gistic)

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 the mutation in OQL:

**Enter Genes:**  
Advanced: Onco Query Language (OQL)

User-defined List

Select from Recurrently Mutated Genes (MutSig)

Select Genes from Recurrent CNAs (Gistic)

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 the deep deletion in IDH2:

**Enter Genes:**  
Advanced: Onco Query Language (OQL)

User-defined List

Select from Recurrently Mutated Genes (MutSig)

Select Genes from Recurrent CNAs (Gistic)

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

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Enrichments Survival CN Segments Network Download

✓ The results below reflect the OQL specification from your query.

Add clinical tracks..

Heatmap

Sort

Mutations

View

Download

Q

61 %

Q

IDH1: MUT=R132C

⋮

4%



IDH2: MUT

⋮

4%



EGFR: MUT FUSION AM...

⋮

25%



TP53: MUT FUSION HO...

⋮

53%



Genetic Alteration

■

Inframe Mutation (putative driver)

■

Missense Mutation (putative driver)

■

Missense Mutation (unknown significance)

■

Truncating Mutation (putative driver)

■

Amplification

■

Gain

■

Deep Deletion

■

Shallow Deletion

■

No alterations

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

DOWNLOAD DATA

Select Studies:

1 studies selected (563 samples) Deselect all View summary

ovarian X

Ovary/Fallopian Tube 4

☐ Select all listed studies matching filter (4)

Ovary/Fallopian Tube

Ovarian Epithelial Tumor

→ SEROUS OVARIAN CANCER

☒ Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011) 563 samples

☐ Ovarian Serous Cystadenocarcinoma (TCGA, PanCancer Atlas) 585 samples

☐ Ovarian Serous Cystadenocarcinoma (TCGA, Provisional) 606 samples

→ SMALL CELL CARCINOMA OF THE OVARY

☐ Small Cell Carcinoma of the Ovary (MSKCC, Nat Genet 2014) 12 samples

Select Genomic Profiles:

☒ Mutations

☒ Putative copy-number alterations (GISTIC)

☐ mRNA Expression. Select one of the profiles below:

- ☐ mRNA expression Z-scores (all genes)
- ☐ 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.

Tumors with sequencing and CNA data (316) X

Enter Genes:

Advanced: Onco Query Language (OQL)

User-defined List X

BRCA1 BRCA2

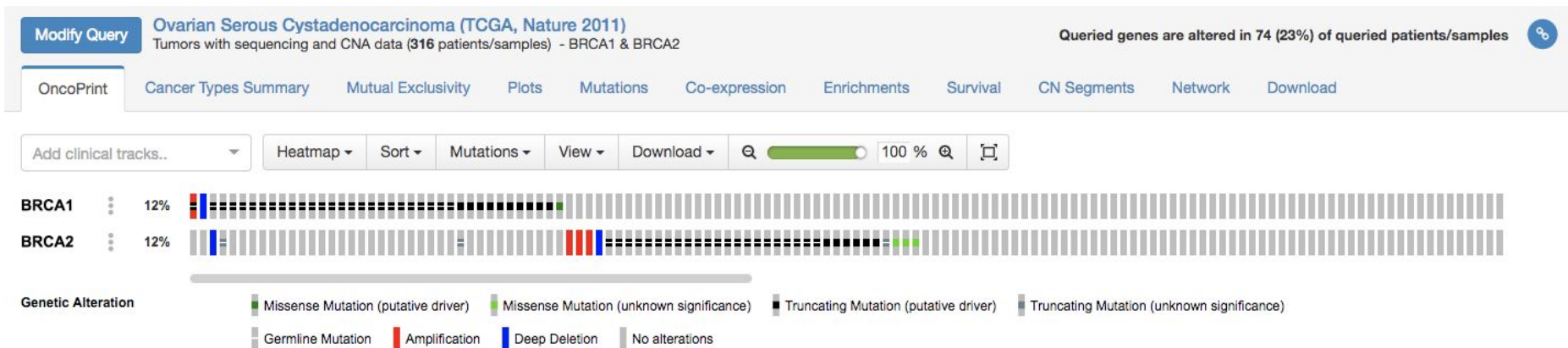
All gene symbols are valid.

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 homozygous deletions:

Enter Genes:

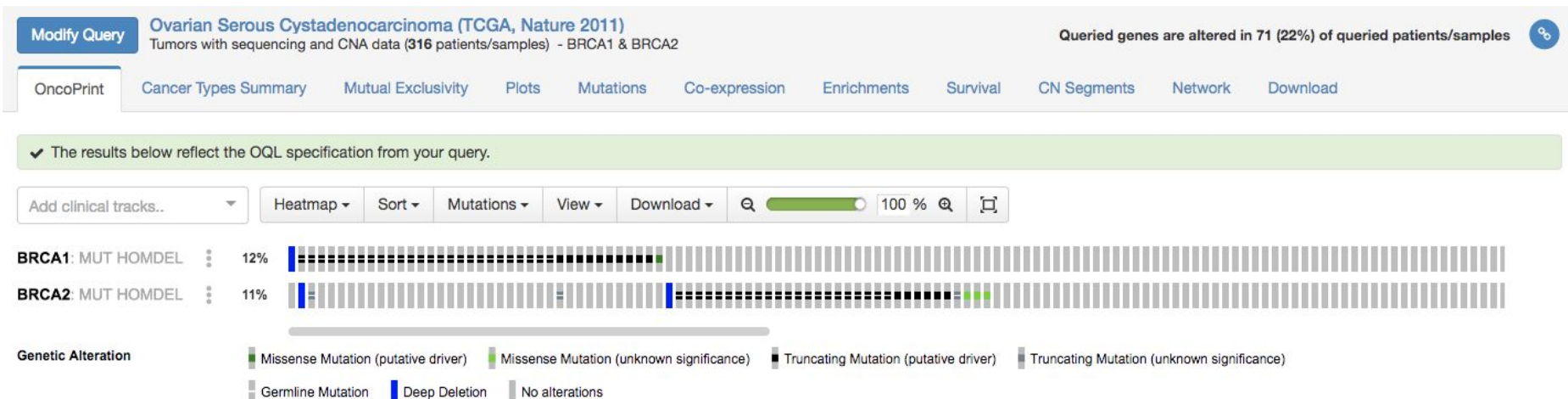
Advanced: Onco Query Language (OQL)

User-defined List

BRCA1: MUT HOMDEL  
BRCA2: MUT HOMDEL

✓ All gene symbols are valid.

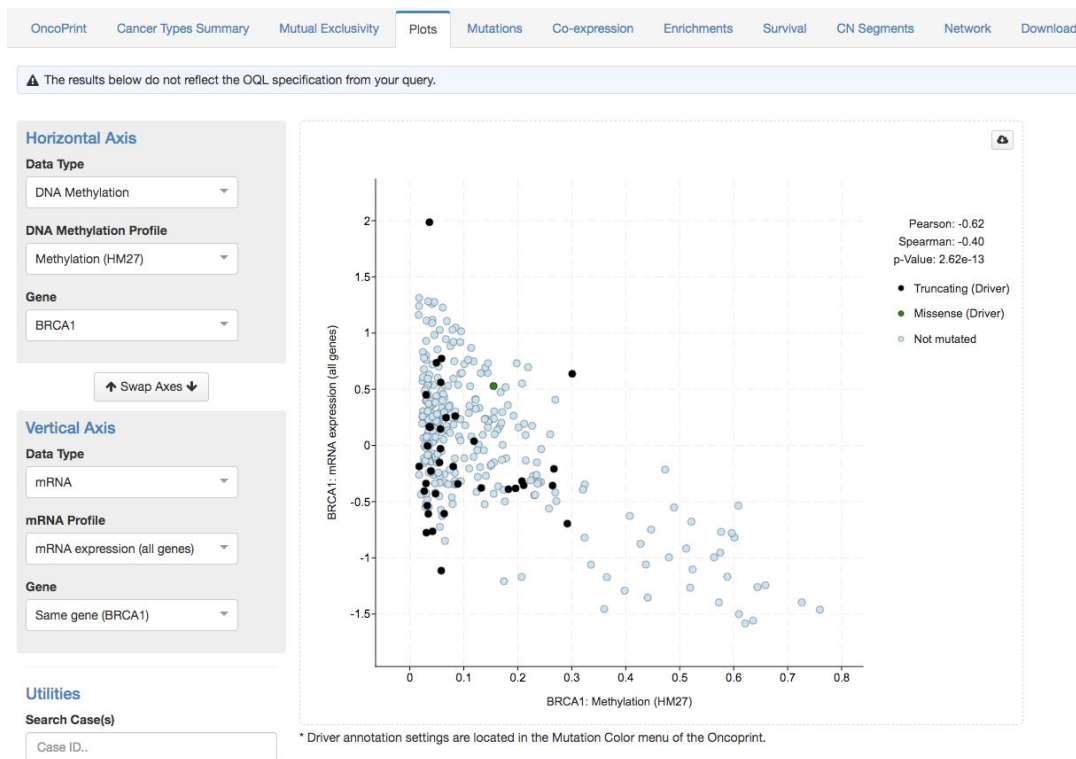
OncoPrint now shows a more accurate estimate of the frequency of BRCA1/2 loss:



# 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 cases.



# OQL Example: BRCA1/2 inactivation

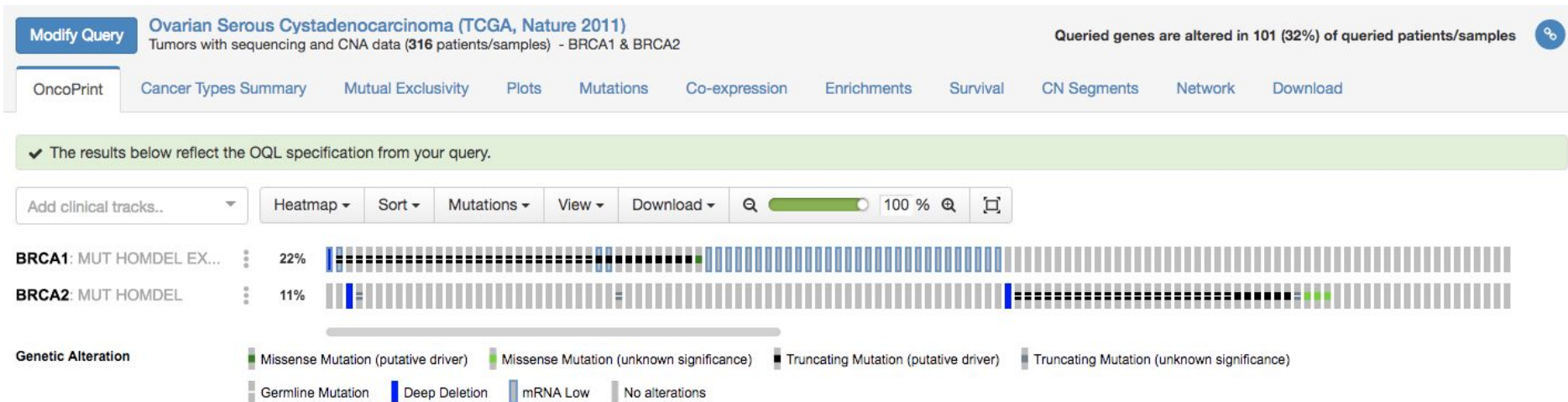
Modify the query to also include samples with decreased expression (don't forget to select "mRNA Expression" in the "Genomics Profiles" section):

**Enter Genes:**  
Advanced: Onco Query Language (OQL)

User-defined List

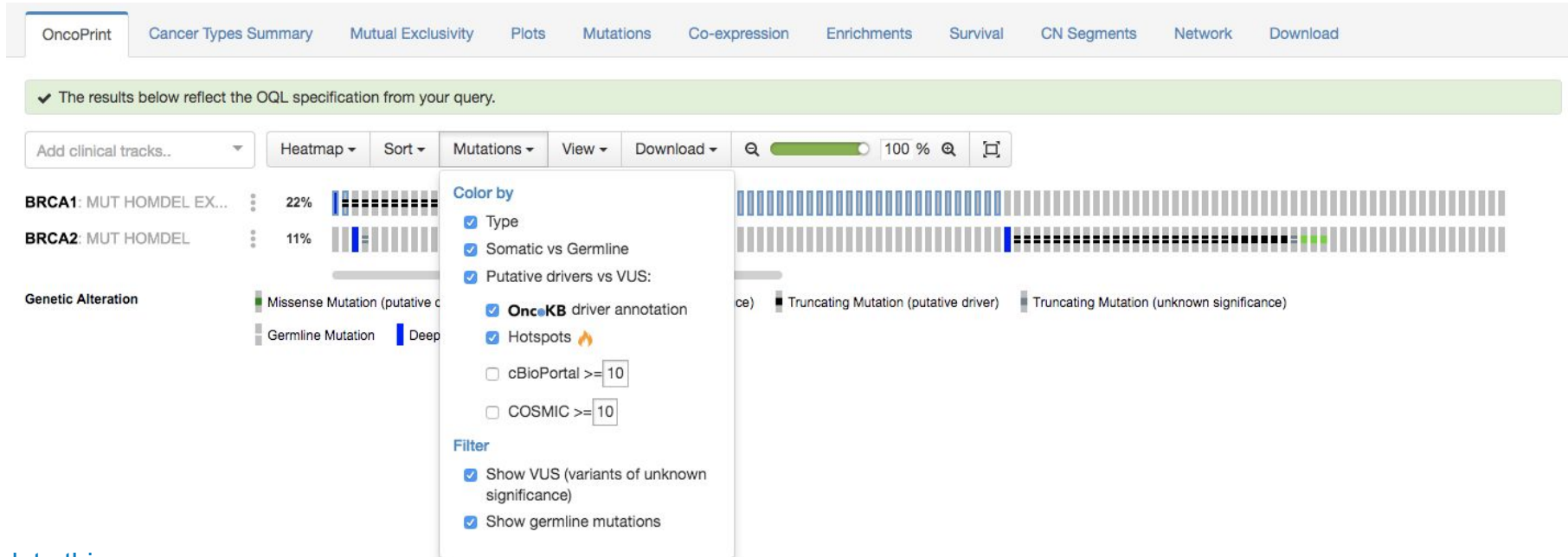
BRCA1: MUT HOMDEL EXP<-1.5  
BRCA2: MUT HOMDEL

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



# 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 “drivers” or “unknown significance” based on the “Mutations” dropdown.



# OQL Example: BRCA1/2 inactivation

We can further refine the query by only including those mutations which are putative drivers, as defined by the “Mutations” dropdown. 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 and copy number changes that are putative drivers (see BRCA2 below)

## Enter Genes:

Advanced: Onco Query Language (OQL)

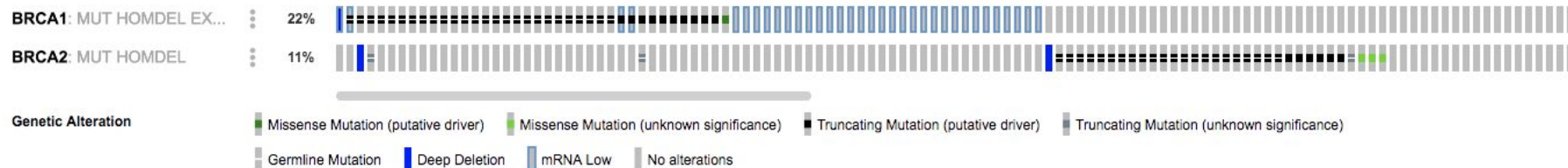
User-defined List



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

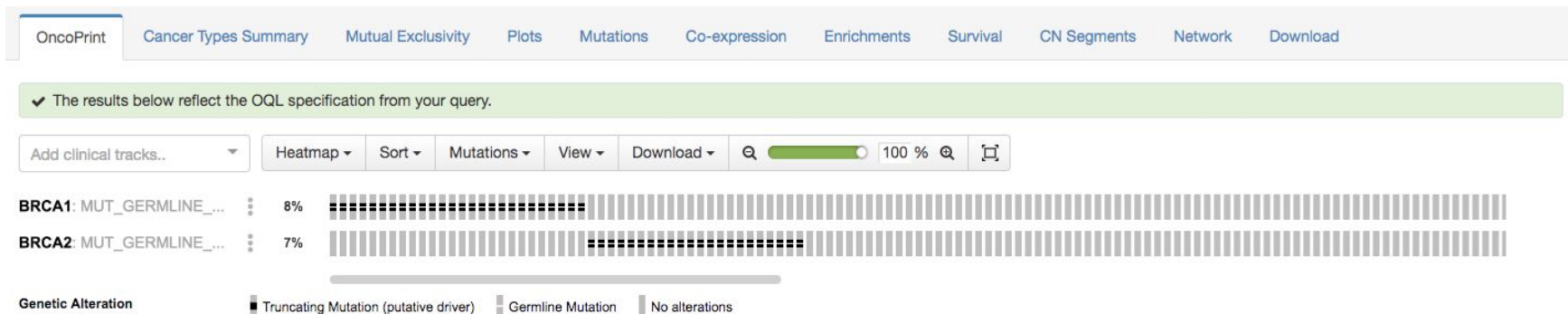
This study is one of the few in cBioPortal that include germline mutations. We can make one final adjustment to our query to ask a slightly different question: what percentage of samples have putative driver germline mutations in BRCA1/BRCA2? Note that the OQL for BRCA1 and BRCA2 are equivalent as the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

**Enter Genes:**

Advanced: Onco Query Language (OQL)

### User-defined List

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





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

DOWNLOAD DATA

Select Studies:

1 studies selected (155 samples) Deselect all View summary

Search...

|                       |    |
|-----------------------|----|
| PanCancer Studies     | 3  |
| Cell lines            | 2  |
| Adrenal Gland         | 2  |
| Ampulla of Vater      | 1  |
| Biliary Tract         | 6  |
| Bladder/Urinary Tract | 12 |
| Bone                  | 2  |
| Bowel                 | 7  |
| Breast                | 14 |
| CNS/Brain             | 16 |

☐ Select all listed studies matching filter (7)

**Bowel**  
**Colorectal Adenocarcinoma**  
☐ Colorectal Adenocarcinoma (DFCI, Cell Reports 2016) 619 samples  
☐ Colorectal Adenocarcinoma (Genentech, Nature 2012) 72 samples  
☐ Colorectal Adenocarcinoma (TCGA, Nature 2012) 276 samples  
☒ Colorectal Adenocarcinoma (TCGA, PanCancer Atlas) 594 samples  
☐ Colorectal Adenocarcinoma (TCGA, Provisional) 640 samples  
☐ Colorectal Adenocarcinoma Triplets (MSKCC, Genome Biol 2014) 138 samples  
☐ Targeted sequencing of 1134 samples from metastatic colorectal cancer 1134 samples

Select Genomic Profiles:

☒ Mutations  
☒ Putative copy-number alterations from GISTIC  
☐ mRNA Expression Zscores, RSEM (Batch normalized from Illumina HiSeq\_RNASeqV2)

Select Patient/Case Set:

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

Tumor Samples with sequencing and CNA data (526)

Enter Genes:

Advanced: Onco Query Language (OQL)

User-defined List  
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:**  
*Advanced: Onco Query Language (OQL)*

User-defined List

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



# 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.

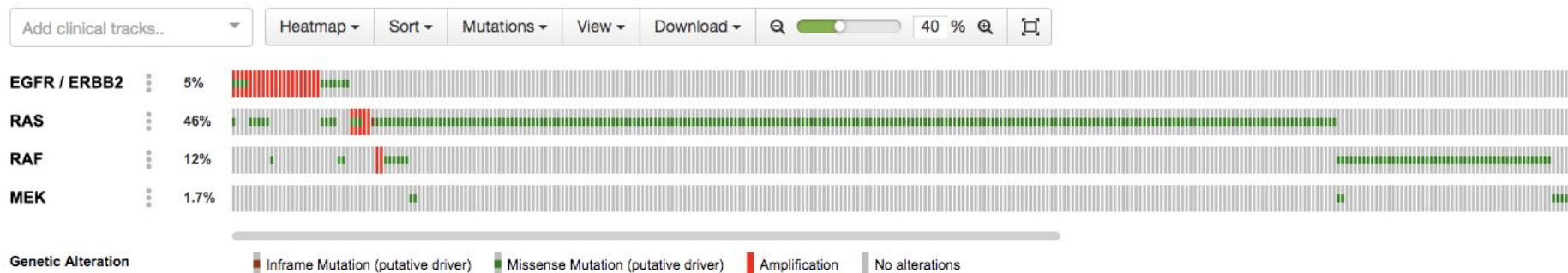
**Enter Genes:**

Advanced: Onco Query Language (OQL)


User-defined List

```
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.



# 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 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`