

cBioPortal Tutorial #6: Group Comparison

Compare clinical and genomic features of
user-defined groups of samples/patients

Tutorial Objectives

- Explain what the new group comparison functionality enables
- Delineate the different ways to define groups and enter the group comparison view
- Highlight potential use cases for group comparison functionality

Group Comparison Overview

Group Comparison is a new suite of analysis features which allow you to compare the clinical and genomic features of user-defined groups of samples.

Groups can be defined in Study View based on any clinical or genomic features.

Groups can also be defined within the Group Comparison page based on the union/intersection of the compared groups.

In this tutorial, we will demonstrate two different ways to enter the group comparison view.

We begin by selecting a study of interest and viewing it in the Study Summary page.

Select a study

Query Quick Search **Beta!** Download

Select Studies: 0 studies selected (0 samples)

CNS/Brain 13

Soft Tissue 1

Select all listed studies matching filter (14)

CNS/Brain

Diffuse Glioma

- ☐ Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 514 samples
- ☐ Brain Lower Grade Glioma (TCGA, Provisional) 530 samples
- ☐ Glioma (MSK, 2018)
- ☐ Low-Grade Gliomas (UCSF, Science 2014)
- ☐ Merged Cohort of LGG and GBM (TCGA, Cell 2016) 1122 samples

→ **GLIOBLASTOMA**

- ☐ Glioblastoma (TCGA, Cell 2013) 585 samples
- ☐ Glioblastoma (TCGA, Nature 2008) 206 samples
- ☐ Glioblastoma Multiforme (TCGA, PanCancer Atlas) 592 samples
- ☐ Glioblastoma Multiforme (TCGA, Provisional) 604 samples

→ **OLIGODENDROGLIOMA**

- ☐ Anaplastic Oligodendroglioma and Anaplastic Oligoastrocytoma (MSK) 22 samples

Enter Genes:

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

User-defined List

Enter HUGO Gene Symbols, Gene Aliases, or OQL

Submit Query

Please select one or more cancer studies.

glioma

Start typing tumor type of interest...

View clinical and genomic data of this study

Click on "View study summary" button

Study View

Merged Cohort of LGG and GBM (TCGA, Cell 2016) [Download](#)
Whole-exome sequencing from TCGA LGG and GBM cases [PubMed](#)

Summary

Clinical Data

Heatmaps

Selected: 1,122 patients | 1,122 samples

Click gene symbols below or enter here

Query

Custom Selection

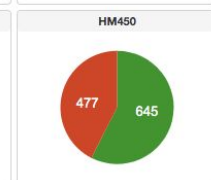
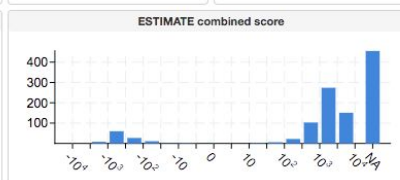
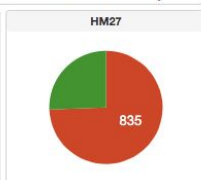
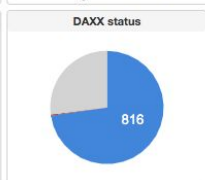
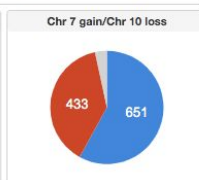
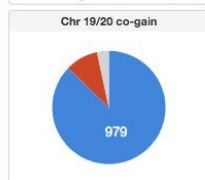
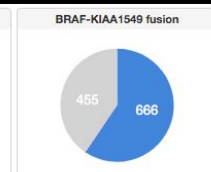
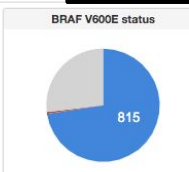
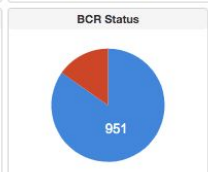
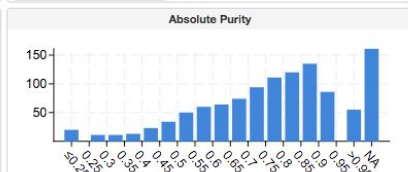
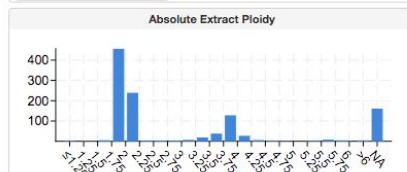
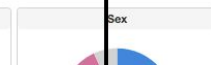
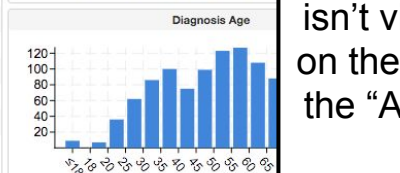
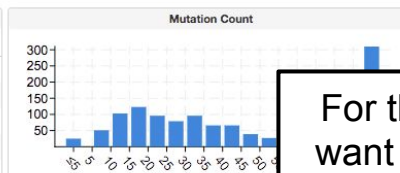
+ Add Chart

Groups

Quick Filters: ☐ 812 samples with mutation data ☐ 1,084 samples with CNA data

Mutated Genes (812 profiled samples)			
Gene	# Mut	#	Freq
IDH1	411	411	50.6%
TP53	398	322	39.7%
ATRX	216	208	25.6%
TTN	129	109	13.4%
PTEN	109	108	13.3%
EGFR	93	87	10.7%
CIC	85	79	9.7%
MUC16	73	69	8.5%
PIK3CA	67	63	7.8%
NF1	65	53	6.5%
RYR2	48	46	5.7%

CNA Genes (1084 profiled samples)				
Gene	Cytoband	CNA	#	Freq
CDKN2A	9p21.3	DEL	383	35.3%
CDKN2B	9p21.3	DEL	377	34.8%
CDKN2A-DT	9p21.3	DEL	372	34.3%
MTAP	9p21.3	DEL	348	32.1%
EGFR	7p11.2	AMP	289	26.7%
SEC61G	7p11.2	AMP	233	21.5%
RN7SL151P	9p21.3	DEL	220	20.3%
MIR31HG	9p21.3	DEL	207	19.1%
DMRTA1	9p21.3	DEL	203	18.7%
IFNE	9p21.3	DEL	197	18.2%
IFNA1	9p21.3	DEL	193	17.8%



Notice this new “Groups” button. We’ll use this in the second example.

For this first example, I want to use a chart that isn’t visible by default, so on the next slide we’ll use the “Add Chart” button to add it.

Study View

Merged Cohort of LGG and GBM (TCGA, Cell 2016) [PubMed](#)
Whole-exome sequencing from TCGA LGG and GBM cases

Summary

Clinical Data

Heatmaps

Selected: 1,122 patients | 1,122 samples

Click gene symbols below or enter here

Query

Custom Selection

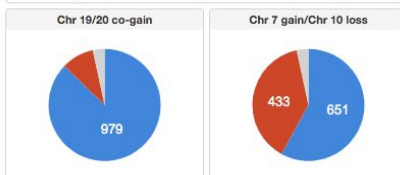
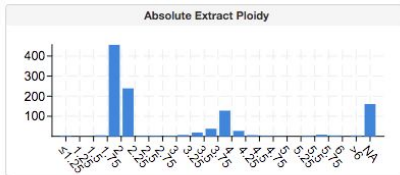
+ Add Chart

Groups

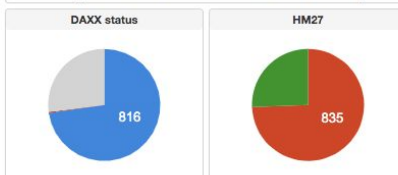
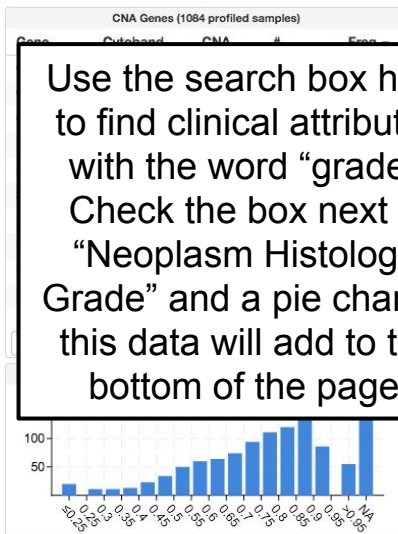
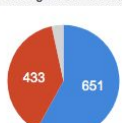
Quick Filters: ☐ 812 samples with mutation data ☐ 1,084 samples with

Mutated Genes (812 profiled samples)			
Gene	# Mut	#	Freq
IDH1	411	<input type="checkbox"/> 411	50.6%
TP53	398	<input type="checkbox"/> 322	39.7%
ATRX	216	<input type="checkbox"/> 208	25.6%
TTN	129	<input type="checkbox"/> 109	13.4%
PTEN	109	<input type="checkbox"/> 108	13.3%
EGFR	93	<input type="checkbox"/> 87	10.7%
CIC	85	<input type="checkbox"/> 79	9.7%
MUC16	73	<input type="checkbox"/> 69	8.5%
PIK3CA	67	<input type="checkbox"/> 63	7.8%
NF1	65	<input type="checkbox"/> 53	6.5%
RYR2	48	<input type="checkbox"/> 46	5.7%

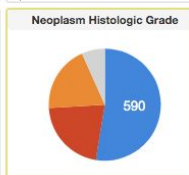
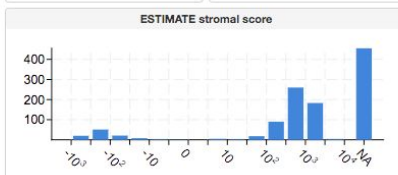
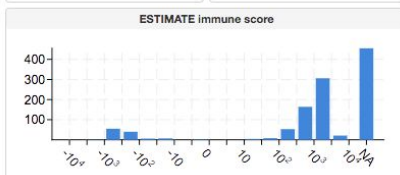
Search...



Chr 7 gain/Chr 10 loss



HM27



Use the search box here to find clinical attributes with the word "grade". Check the box next to "Neoplasm Histologic Grade" and a pie chart of this data will add to the bottom of the page.

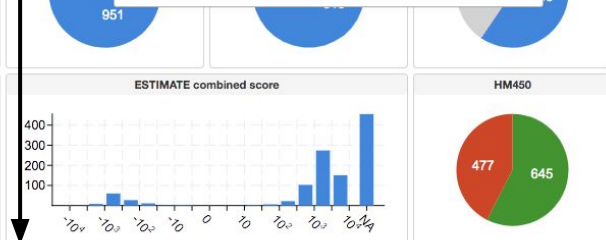
Clinical Genomic Custom Data

Select all (1) Deselect all

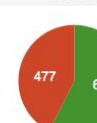
grade

Name	Freq
<input checked="" type="checkbox"/> Neoplasm Histologic Grade	93.3%

✓ Neoplasm Histologic Grade added as a pie chart

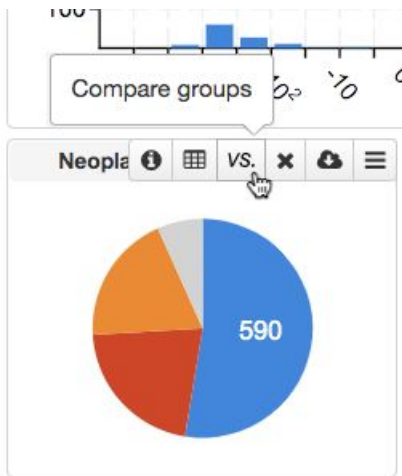


HM450



Study View

Let's compare samples of different histologic grades. Hover over the “Neoplasm Histologic Grade” pie chart and notice the new **vs.** button. We're going to click on this, and it will bring us to the new group comparison page where we can compare the clinical & genomic features of samples/patients by grade.



Group Comparison: Header

All group comparison pages share the same header:

The screenshot shows the header of a group comparison page. It includes a title, a subtitle, a list of groups, and a set of tabs. Annotations with arrows point to specific elements:

- The attribute used to create the groups.** Points to the subtitle "Groups from *Neoplasm Histologic Grade*".
- The original study. Click to return to study view.** Points to the title "Merged Cohort of LGG and GBM (TCGA, Cell 2016)".
- The available groups. Click on a group name to include or exclude it from analysis. Click the "x" to remove the group from the comparison session. Groups can also be reordered by dragging the group name.** Points to the group list: "(A) G2 (216)", "(B) G3 (241)", and "(C) G4 (590)".
- Each tab has specific functionality. We'll go through these one-by-one over the next few slides.** Points to the tabs: "Overlap", "Survival", "Clinical", "Mutations", and "Copy-number".

The header content is as follows:

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

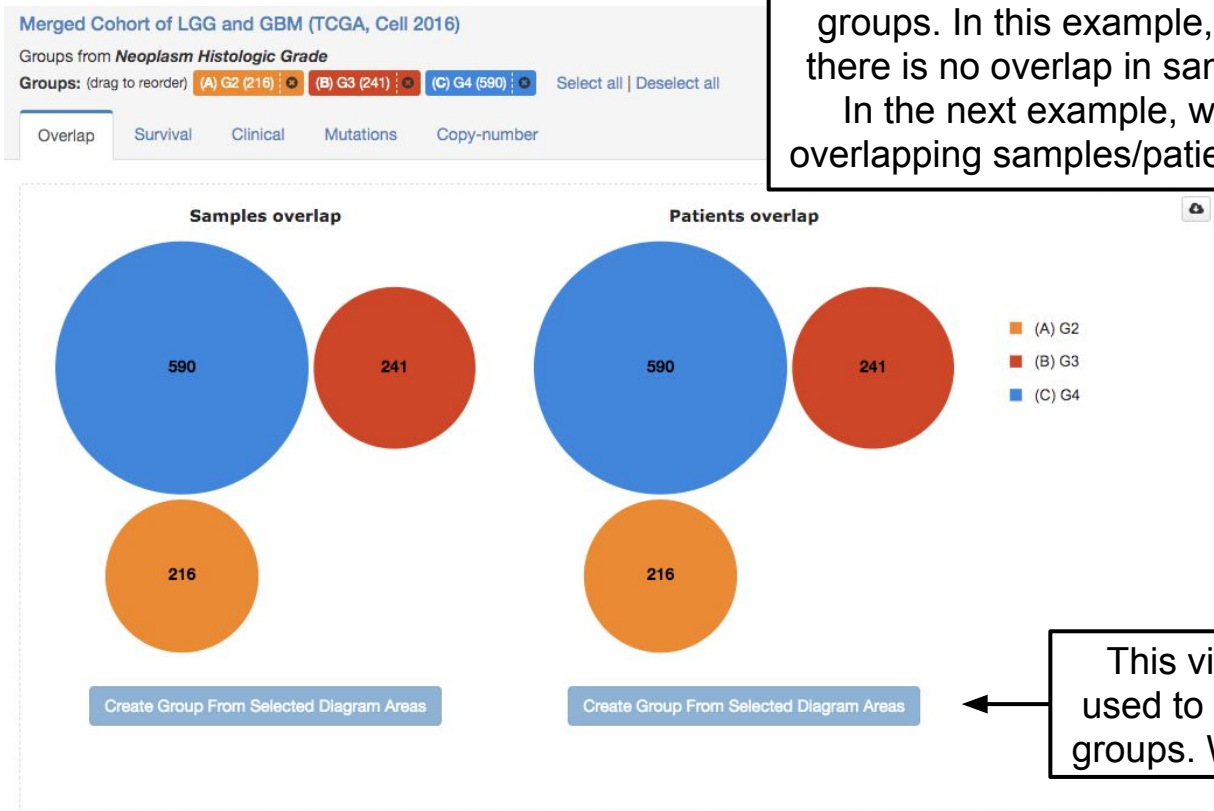
Groups from *Neoplasm Histologic Grade*

Groups: (drag to reorder) (A) G2 (216) (B) G3 (241) (C) G4 (590) Select all | Deselect all

Overlap Survival Clinical Mutations Copy-number

Group Comparison: Overlap

The Overlap tab shows which samples or patients may overlap among the selected groups. In this example, we can see that there is no overlap in samples or patients. In the next example, we'll look at how overlapping samples/patients are managed.



This view can also be used to create additional groups. We'll do this later.

Group Comparison: Survival

The Survival tab shows a Kaplan-Meier plot of Overall Survival or Disease/Progression-free Survival based on the selected groups. This tab will only be visible when the original study contains survival data.

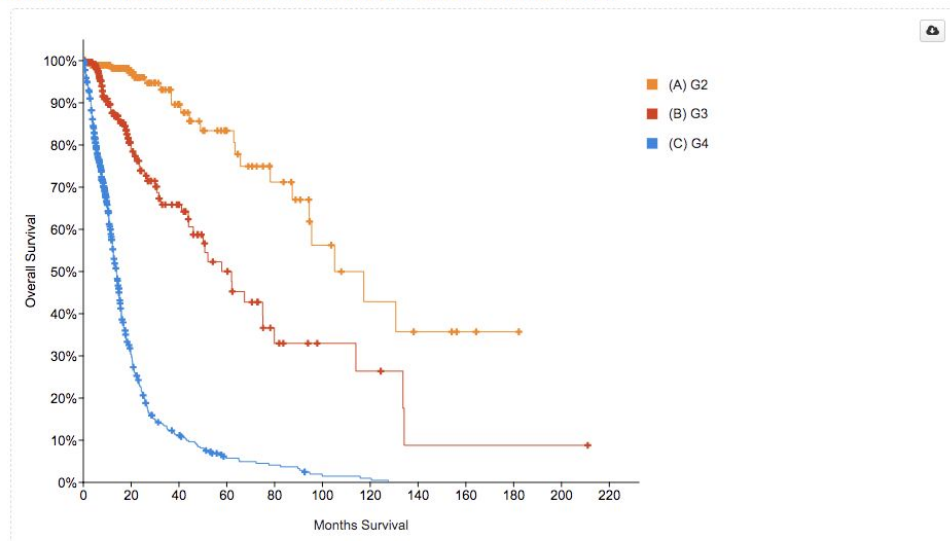
Merged Cohort of LGG and GBM (TCGA, Cell 2016)

Groups from *Neoplasm Histologic Grade*

Groups: (drag to reorder) (A) G2 (216) (B) G3 (241) (C) G4 (589) Select all | Deselect all

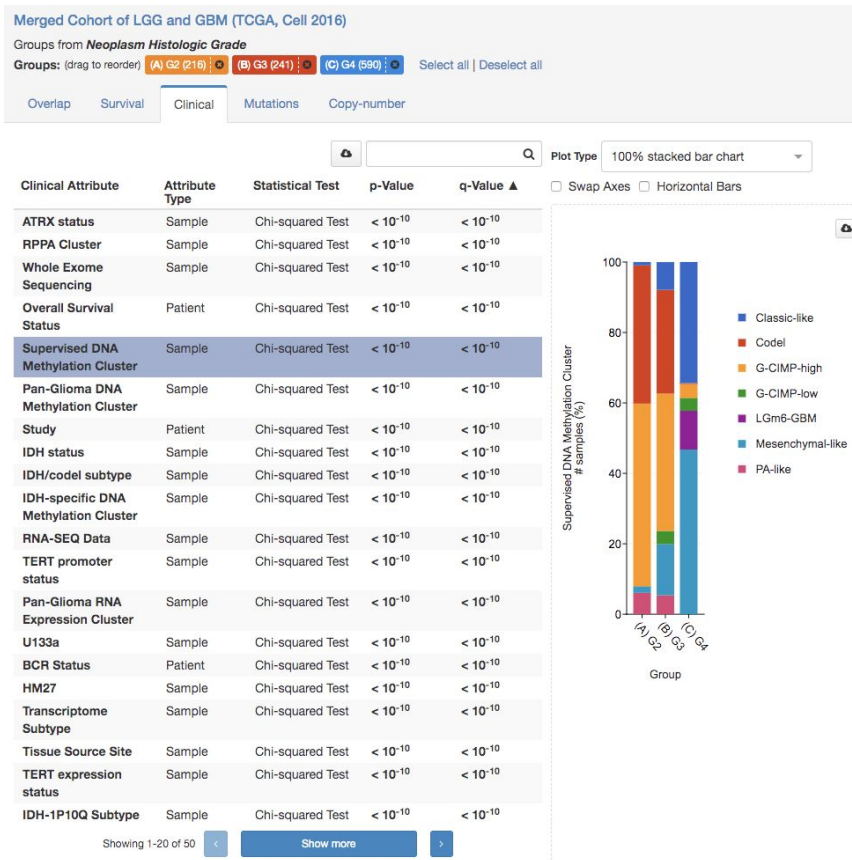
Overlap Survival Clinical Mutations Copy-number

Overall Survival Kaplan-Meier Estimate (Overall patient survival status.)



	Number of Cases, Total	Number of Cases, Deceased	Median Months Survival
(A) G2	216	22	105.1
(B) G3	241	58	62
(C) G4	589	444	13.9

Group Comparison: Clinical



The Clinical tab shows all the same clinical attributes that are present in Study View. Select a clinical attribute in the table (Supervised DNA Methylation Cluster is selected here) and a plot will appear to the right with the distribution of that clinical attribute across the selected groups.

Group Comparison: Mutations

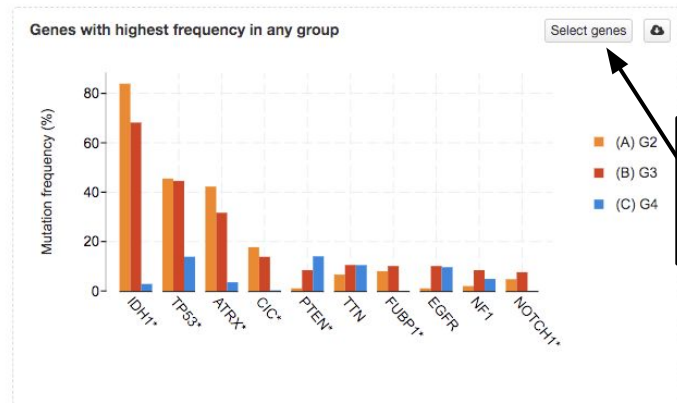
The Mutations tab compares the frequency of mutations in genes across the selected groups. The visible plots change depending on how many groups are selected. This screenshot shows the view with 3 or more groups selected.

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

Groups from *Neoplasm Histologic Grade*

Groups: (drag to reorder) (A) G2 (216) (B) G3 (241) (C) G4 (590) Select all | Deselect all

Overlap Survival Clinical Mutations Copy-number



Mutations

Select enriched groups ☐ Significant only

Columns

Gene	Cytoband	(A) G2	(B) G3	(C) G4	p-Value	q-Value ▲	Most enriched in
ATRX	Xq21.1	91 (42.13%)	76 (31.54%)	20 (3.39%)	0.00	0.00	(A) G2
CIC	19q13.2	38 (17.59%)	33 (13.69%)	1 (0.17%)	0.00	0.00	(A) G2
IDH1	2q34	181 (83.80%)	164 (68.05%)	16 (2.71%)	0.00	0.00	(A) G2
TP53	17p13.1	98 (45.37%)	107 (44.40%)	81 (13.73%)	0.00	0.00	(A) G2
FUBP1	1p31.1	17 (7.87%)	24 (9.96%)	0 (0.00%)	5.63e-13	1.04e-9	(B) G3
NOTCH1	9q34.3	10 (4.63%)	18 (7.47%)	0 (0.00%)	1.47e-9	2.277e-6	(B) G3

Group Comparison: Mutations

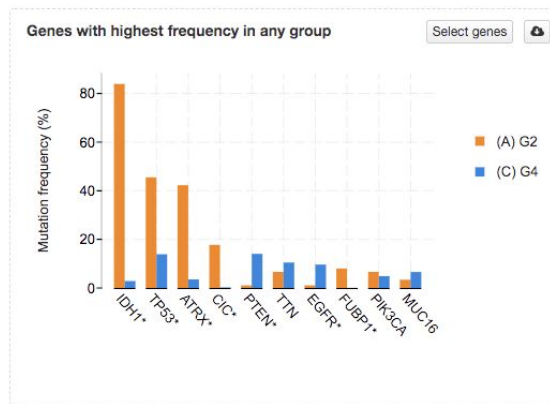
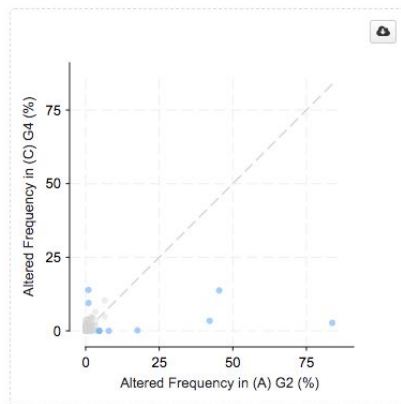
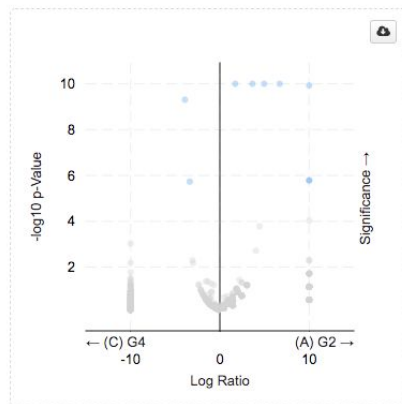
This screenshot shows the view with 2 groups selected (notice that I deselected group “(B) G3” by clicking on it). Two additional plots are shown when exactly 2 groups are compared.

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

Groups from *Neoplasm Histologic Grade*

Groups: (drag to reorder) (A) G2 (216) (B) G3 (241) (C) G4 (590) Select all | Deselect all

Overlap Survival Clinical Mutations Copy-number



Mutations

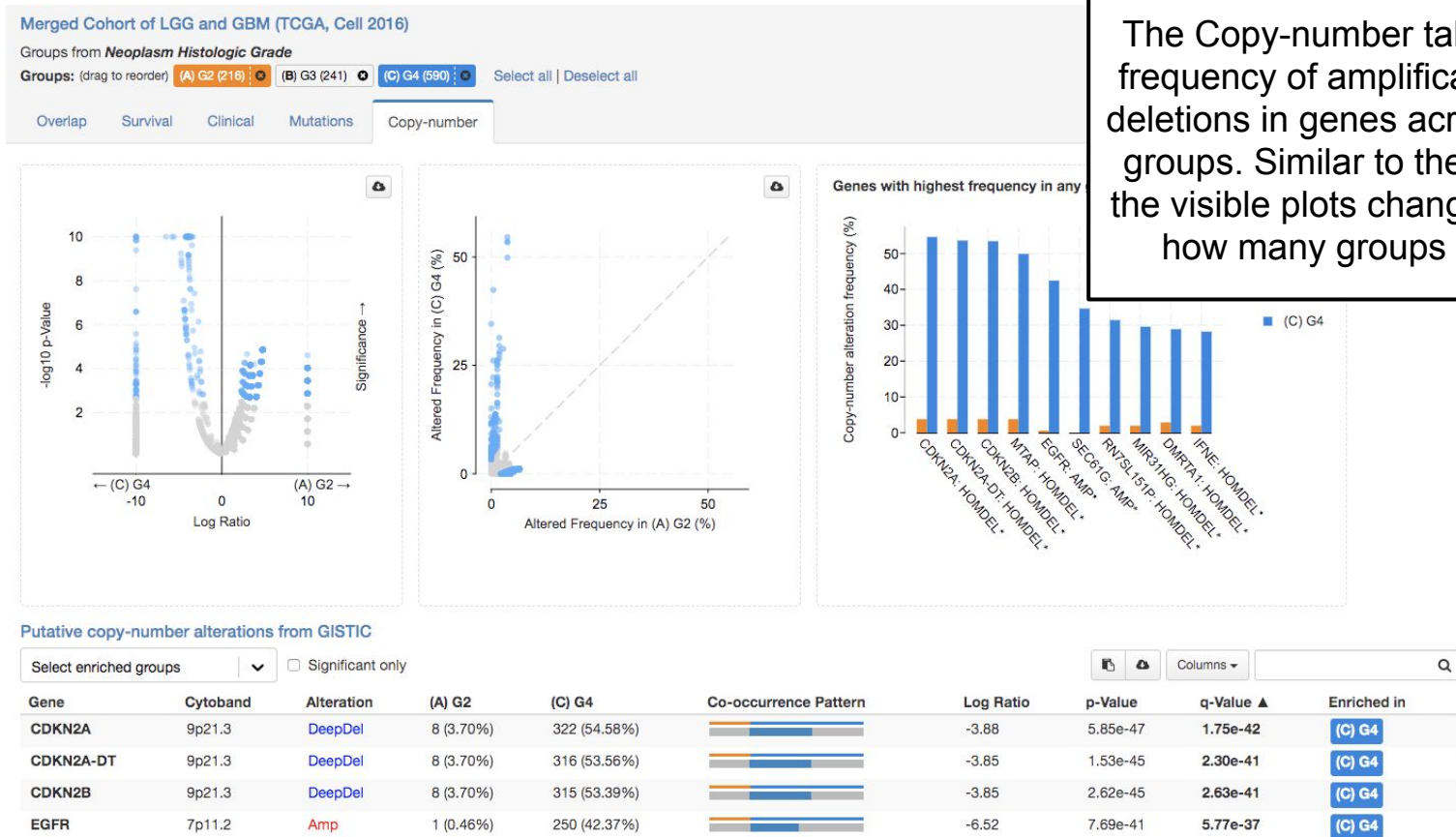
Select enriched groups ☐ Significant only

Columns

Gene	Cytoband	(A) G2	(C) G4	Co-occurrence Pattern	Log Ratio	p-Value	q-Value ▲	Enriched in
IDH1	2q34	181 (83.80%)	16 (2.71%)		4.95	1.49e-122	1.13e-118	(A) G2
ATRX	Xq21.1	91 (42.13%)	20 (3.39%)		3.64	3.98e-40	1.51e-36	(A) G2
CIC	19q13.2	38 (17.59%)	1 (0.17%)		6.70	4.27e-22	1.08e-18	(A) G2
TP53	17p13.1	98 (45.37%)	81 (13.73%)		1.72	5.41e-20	1.03e-16	(A) G2
FUBP1	1p31.1	17 (7.87%)	0 (0.00%)		>10	1.18e-10	1.79e-7	(A) G2
PTEN	10q23.31	2 (0.93%)	82 (13.90%)		-3.91	4.93e-10	6.24e-7	(C) G4

Group Comparison: Copy-Number

The Copy-number tab compares the frequency of amplifications and deep deletions in genes across the selected groups. Similar to the Mutations tab, the visible plots change depending on how many groups are selected.



For the second approach to group comparison, we will define our own groups.

To do this, you must be logged in.

Logging in



[Data Sets](#) [Web API](#) [R/MATLAB](#) [Tutorials](#) [FAQ](#) [News](#) [Visualize Your Data](#) [About](#)

Login

This button is in the header on all pages.
Click on it. It will bring you to a page
where you can sign in (or register) using
any google account.



Welcome to cBioPortal - sign in with your Google account to store your [virtual studies](#) and [groups](#). This will allow you to access your studies and groups from any computer. Login is optional and not required to access any of the other features of cBioPortal.



Sign in with Google

Now that we're logged in,
let's go back to study view.

Study View: Defining Groups

Let's use this "Groups" button to define groups for comparison.

Merged Cohort of LGG and GBM (TCGA, Cell 2016) [Download](#)
Whole-exome sequencing from TCGA LGG and GBM cases [PubMed](#)

Click gene symbols below or enter here

Query

Summary

Clinical Data

Heatmaps

Selected: 1,122 patients | 1,122 samples



Custom Selection

+ Add Chart

Groups

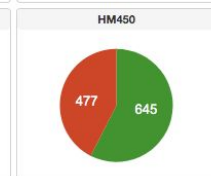
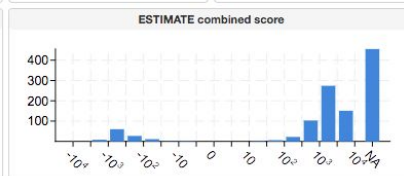
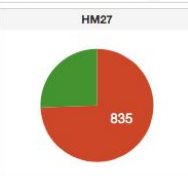
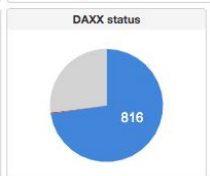
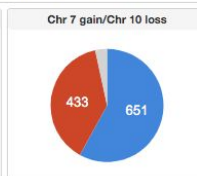
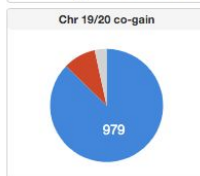
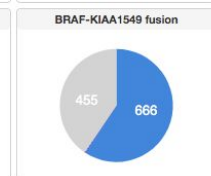
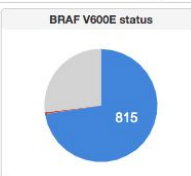
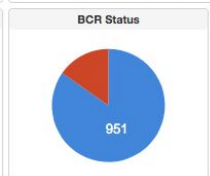
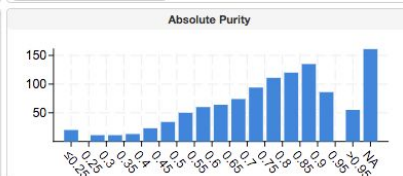
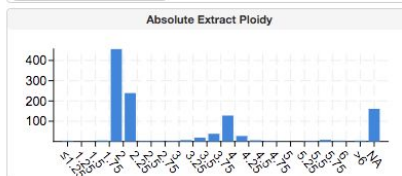
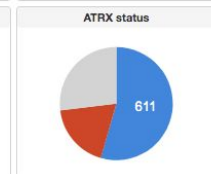
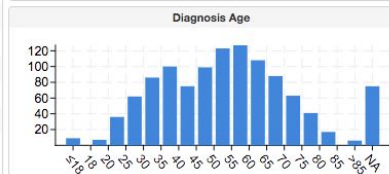
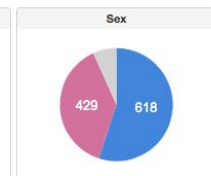
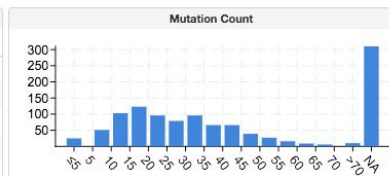
Quick Filters: ☐ 812 samples with mutation data ☐ 1,084 samples with CNA data

Mutated Genes (812 profiled samples)			
Gene	# Mut	#	Freq
IDH1	411	411	50.6%
TP53	398	322	39.7%
ATRX	216	208	25.6%
TTN	129	109	13.4%
PTEN	109	108	13.3%
EGFR	93	87	10.7%
CIC	85	79	9.7%
MUC16	73	69	8.5%
PIK3CA	67	63	7.8%
NF1	65	53	6.5%
RYR2	48	46	5.7%

Search...

CNA Genes (1084 profiled samples)				
Gene	Cytoband	CNA	#	Freq
CDKN2A	9p21.3	DEL	383	35.3%
CDKN2B	9p21.3	DEL	377	34.8%
CDKN2A-DT	9p21.3	DEL	372	34.3%
MTAP	9p21.3	DEL	348	32.1%
EGFR	7p11.2	AMP	289	26.7%
SEC61G	7p11.2	AMP	233	21.5%
RN7SL151P	9p21.3	DEL	220	20.3%
MIR31HG	9p21.3	DEL	207	19.1%
DMRTA1	9p21.3	DEL	203	18.7%
IFNE	9p21.3	DEL	197	18.2%
IFNA1	9p21.3	DEL	193	17.8%

Search...



Define Groups by Filtering

We define groups by applying filters in study view.
Here, I select samples with IDH1 mutations:

1. Use this table to select samples with IDH1 mutations
(Click the check box in the “#” column, then “Select Samples” at the bottom of the table)

2. Click the “Groups” button, and follow the instructions

Merged Cohort of LGG and GBM (TCGA, CPTAC)
Whole-exome sequencing from TCGA LGG and GBM

IDH1 ☒ Clear All Filters

Summary Clinical Data Heatmaps

Selected: 411

411 samples with mutation data 408 samples with CNA data

Mutated Genes (411 profiled samples)			
Gene	# Mut	#	Freq
IDH1	411	<input checked="" type="checkbox"/> 411	100.0%
TP53	290	<input type="checkbox"/> 237	57.7%
ATRX	198	<input type="checkbox"/> 190	46.2%
CIC	78	<input type="checkbox"/> 72	17.5%
FUBP1	42	<input type="checkbox"/> 42	10.2%
TTN	32	<input type="checkbox"/> 31	7.5%
NOTCH1	31	<input type="checkbox"/> 27	6.6%
PIK3CA	26	<input type="checkbox"/> 25	6.1%
MUC16	23	<input type="checkbox"/> 23	5.6%
SMARCA4	21	<input type="checkbox"/> 21	5.1%
PIK3R1	19	<input type="checkbox"/> 18	4.4%

CNA Genes (408 profiled samples)				
Gene	Cytoband	CNA	#	Freq
PARP11	12p13.32	AMP	<input type="checkbox"/> 26	6.4%
CD4	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
CD27	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
CHD4	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
GAPDH	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
GNB3	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
LAG3	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
NOP2	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
PTMS	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
VAMP1	12p13.31	AMP	<input type="checkbox"/> 25	6.1%
TPI1	12p13.31	AMP	<input type="checkbox"/> 25	6.1%

Mutation Count

Diagnosis Age

Group comparison allows you to create custom groups and compare their clinical and genomic features. Use the button below to create groups based on selections.

Create new group from selected samples (411)

Select all (0) Deselect all Search..

IDH1 mutant Create

Select all (1) Deselect all Search..

☒ IDH1 mutant (411 samples/patients)

Compare View

Create new group from selected samples (411)

[Link to this page](#)

Define Groups by Filtering

Clear that filter, and continue to apply new filters to define as many groups as you like. Here, I create groups for TP53 mutant samples and EGFR amplified samples.

Merged Cohort of LGG and GBM (TCGA, Cell 2016) [PubMed](#)
Whole-exome sequencing from TCGA LGG and GBM cases [PubMed](#)

TP53

Clear All Filters

Summary Clinical Data Heatmaps

Selected: 322 patients | 322 samples

Custom Selection

+ Add Chart

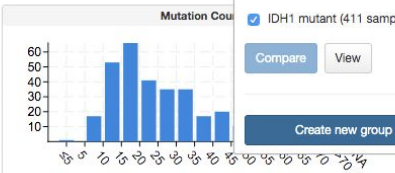
Groups

Quick Filters: ☐ 322 samples with mutation data ☐ 319 samples with CNA data

Mutated Genes (322 profiled samples)			
Gene	# Mut	#	Freq
TP53	398	<input checked="" type="checkbox"/> 322	100.0%
IDH1	237	<input type="checkbox"/> 237	73.6%
ATRX	193	<input type="checkbox"/> 185	57.5%
TTN	51	<input type="checkbox"/> 43	13.4%
PTEN	32	<input type="checkbox"/> 31	9.6%
MUC16	26	<input type="checkbox"/> 26	8.1%
RYR2	18	<input type="checkbox"/> 18	5.6%
PIK3CA	17	<input type="checkbox"/> 16	5.0%
EGFR	15	<input type="checkbox"/> 14	4.3%
FLG	14	<input type="checkbox"/> 14	4.3%
RB1	14	<input type="checkbox"/> 14	4.3%

Search...

CNA Genes (319 profiled samples)				
Gene	Cytoband	CNA	#	Freq
CDKN2A	9p21.3	DEL	<input type="checkbox"/> 43	13.5%
CDKN2A-DT	9p21.3	DEL	<input type="checkbox"/> 43	13.5%
CDKN2B	9p21.3	DEL	<input type="checkbox"/> 42	13.2%
MTAP	9p21.3	DEL	<input type="checkbox"/> 40	12.5%
RN7SL151P	9p21.3	DEL	<input type="checkbox"/> 29	9.1%



Merged Cohort of LGG and GBM (TCGA, Cell 2016) [PubMed](#)
Whole-exome sequencing from TCGA LGG and GBM cases [PubMed](#)

EGFR

Clear All Filters

Summary Clinical Data Heatmaps

Selected: 289 patients | 289 samples

Custom Selection

+ Add Chart

Groups

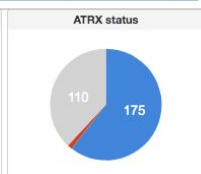
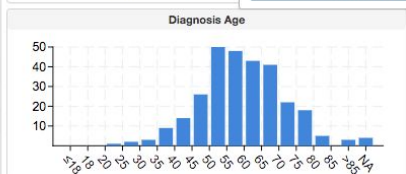
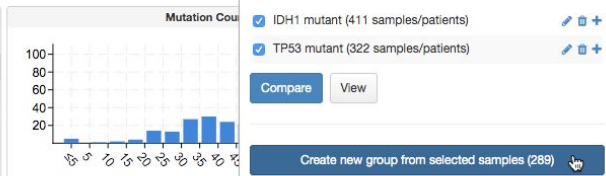
Quick Filters: ☐ 178 samples with mutation data ☐ 289 samples with CNA data

Mutated Genes (178 profiled samples)			
Gene	# Mut	#	Freq
EGFR	68	<input type="checkbox"/> 64	36.0%
PTEN	39	<input type="checkbox"/> 39	21.9%
TTN	44	<input type="checkbox"/> 39	21.9%
MUC16	28	<input type="checkbox"/> 25	14.0%
TP53	23	<input type="checkbox"/> 21	11.8%
RYR2	21	<input type="checkbox"/> 19	10.7%
PIK3CA	17	<input type="checkbox"/> 16	9.0%
HMCN1	15	<input type="checkbox"/> 14	7.9%
FLG	13	<input type="checkbox"/> 13	7.3%
PIK3R1	13	<input type="checkbox"/> 12	6.7%
PLO	13	<input type="checkbox"/> 12	6.7%

Search...

CNA Genes (289 profiled samples)				
Gene	Cytoband	CNA	#	Freq
EGFR	7p11.2	AMP	<input checked="" type="checkbox"/> 289	100.0%
SEC61G	7p11.2	AMP	<input type="checkbox"/> 232	80.3%
CDKN2A	9p21.3	DEL	<input type="checkbox"/> 201	69.6%
CDKN2B	9p21.3	DEL	<input type="checkbox"/> 199	68.9%
CDKN2A-DT	9p21.3	DEL	<input type="checkbox"/> 197	68.2%
MTAP	9p21.3	DEL	<input type="checkbox"/> 188	65.1%
LANCL2	7p11.2	AMP	<input type="checkbox"/> 174	60.2%
VOPP1	7p11.2	AMP	<input type="checkbox"/> 149	51.6%
VSTM2A	7p11.2	AMP	<input type="checkbox"/> 141	48.8%
RN7SL151P	9p21.3	DEL	<input type="checkbox"/> 121	41.9%
MIR31HG	9p21.3	DEL	<input type="checkbox"/> 115	39.8%

Search...



Compare User-Defined Groups

Once all groups are defined, select them and click the “Compare” button to reach the Group Comparison View.

Merged Cohort of LGG and GBM (TCGA, Cell 2016) [Download](#)
Whole-exome sequencing from TCGA LGG and GBM cases [PubMed](#)

Summary

Clinical Data

Heatmaps

Selected: 1,122 patients | 1,122 samples

Click gene symbols below or enter here

Query

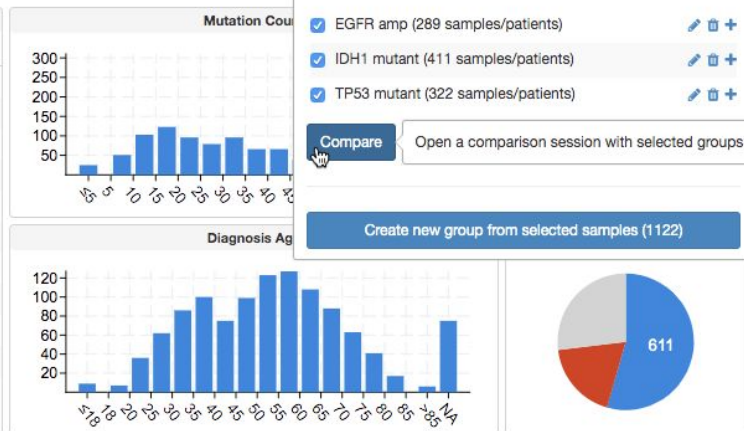
Custom Selection

+ Add Chart

Groups

Quick Filters: ☐ 812 samples with mutation data ☐ 1,084 samples with CNA data

Mutated Genes (812 profiled samples)				CNA Genes (1084 profiled samples)				
Gene	# Mut	#	Freq	Gene	Cytoband	CNA	#	Freq
IDH1	411	<input type="checkbox"/> 411	50.6%	CDKN2A	9p21.3	DEL	<input type="checkbox"/> 383	35.3%
TP53	398	<input type="checkbox"/> 322	39.7%	CDKN2B	9p21.3	DEL	<input type="checkbox"/> 377	34.8%
ATRX	216	<input type="checkbox"/> 208	25.6%	CDKN2A-DT	9p21.3	DEL	<input type="checkbox"/> 372	34.3%
TTN	129	<input type="checkbox"/> 109	13.4%	MTAP	9p21.3	DEL	<input type="checkbox"/> 348	32.1%
PTEN	109	<input type="checkbox"/> 108	13.3%	EGFR	7p11.2	AMP	<input type="checkbox"/> 289	26.7%
EGFR	93	<input type="checkbox"/> 87	10.7%	SEC61G	7p11.2	AMP	<input type="checkbox"/> 233	21.5%
CIC	85	<input type="checkbox"/> 79	9.7%	RN7SL151P	9p21.3	DEL	<input type="checkbox"/> 220	20.3%
MUC16	73	<input type="checkbox"/> 69	8.5%	MIR31HG	9p21.3	DEL	<input type="checkbox"/> 207	19.1%
PIK3CA	67	<input type="checkbox"/> 63	7.8%	DMRTA1	9p21.3	DEL	<input type="checkbox"/> 203	18.7%
NF1	65	<input type="checkbox"/> 53	6.5%	IFNE	9p21.3	DEL	<input type="checkbox"/> 197	18.2%
RYR2	48	<input type="checkbox"/> 46	5.7%	IFNA1	9p21.3	DEL	<input type="checkbox"/> 193	17.8%
Search...				Search...				



Select all (3) Deselect all

Search..

☒ EGFR amp (289 samples/patients)

☒ IDH1 mutant (411 samples/patients)

☒ TP53 mutant (322 samples/patients)

Compare

Open a comparison session with selected groups

Create new group from selected samples (1122)

Group Comparison: Overlapping Samples/Patients

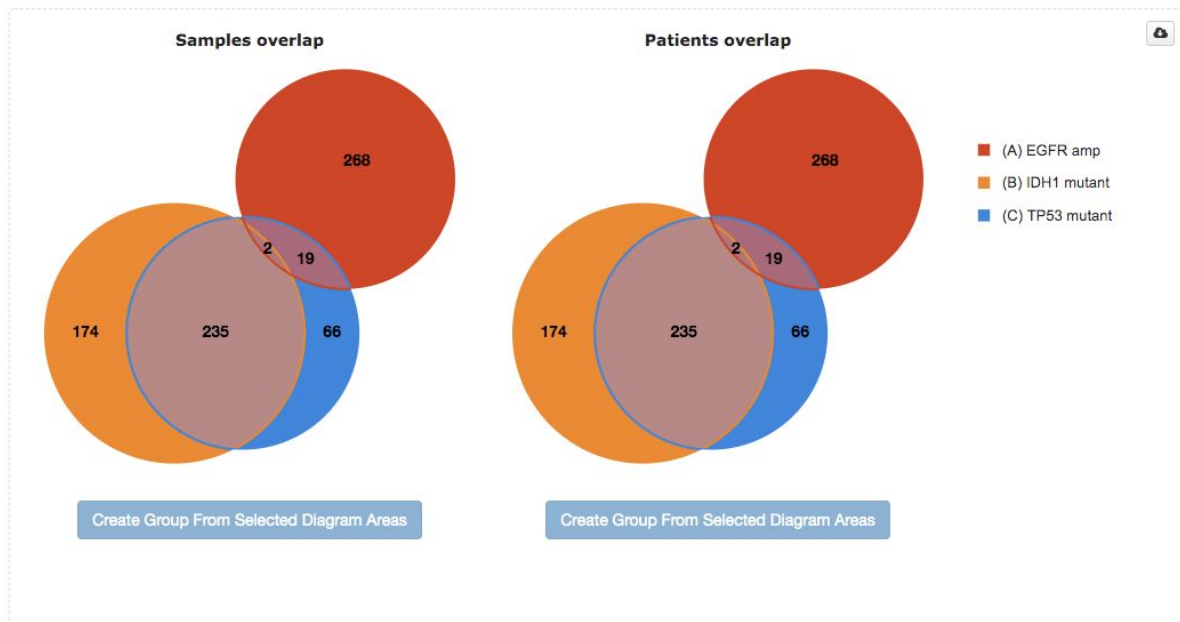
Merged Cohort of LGG and GBM (TCGA, Cell 2016)

Groups: (drag to reorder) (A) EGFR amp (289) (B) IDH1 mutant (411) (C) TP53 mutant (322) Select all | Deselect all

Exclude overlapping samples and patients

Overlap Survival Clinical Mutations Copy-number

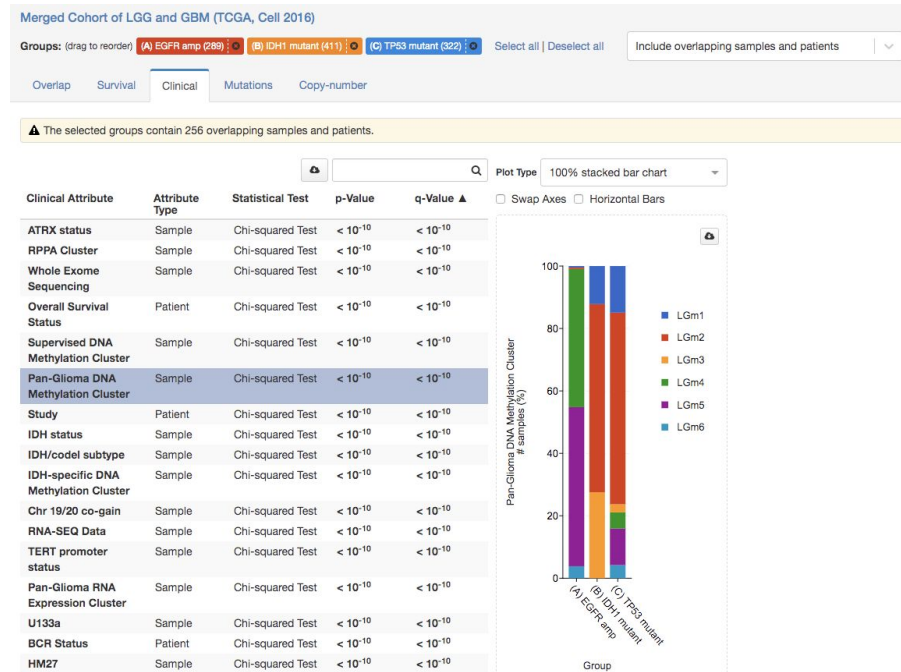
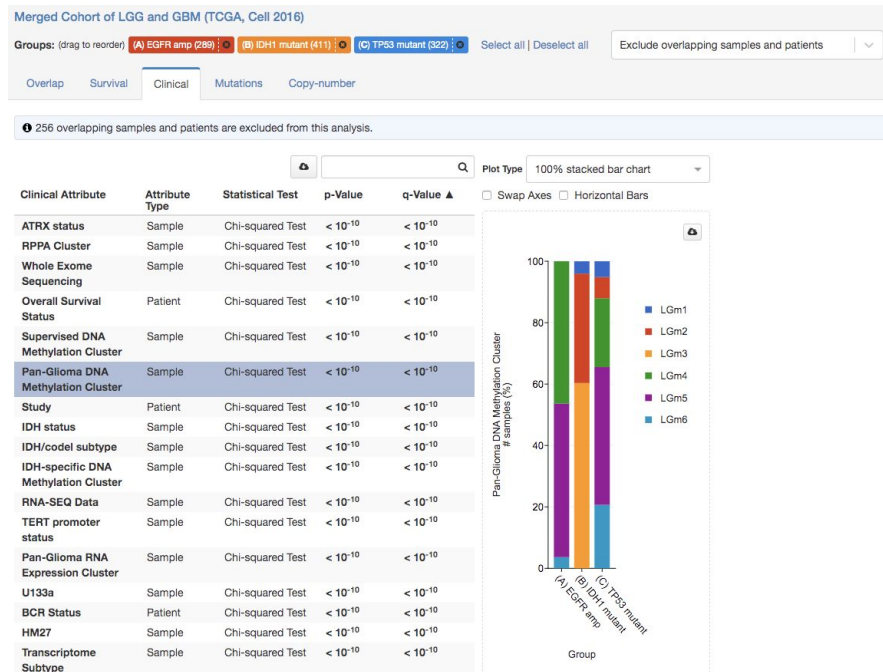
256 overlapping samples and patients are excluded from analysis in other tabs.



When samples/patients overlap among groups, a drop-down appears in the header which allows you to decide to exclude (default) or include those overlapping samples/patients in the analysis.

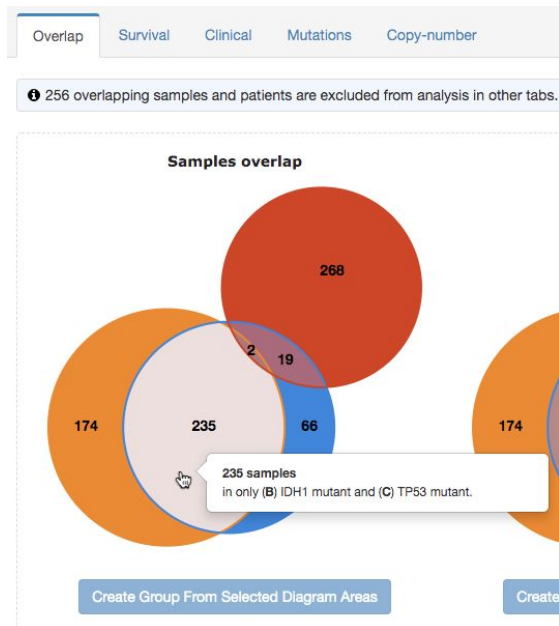
Group Comparison: Overlapping Samples/Patients

When samples/patients overlap among groups, each tab will include a warning message to make clear how those samples/patients are handled. This can be changed at any time by using the drop-down menu in the header of the page.

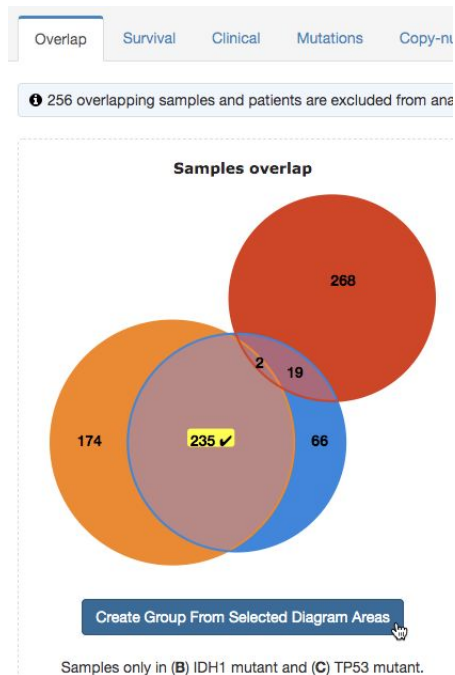


The Overlap tab can also be used to create new groups. For example, let's say I want to create a group of samples with mutations in both IDH1 and TP53, and without amplification of EGFR.

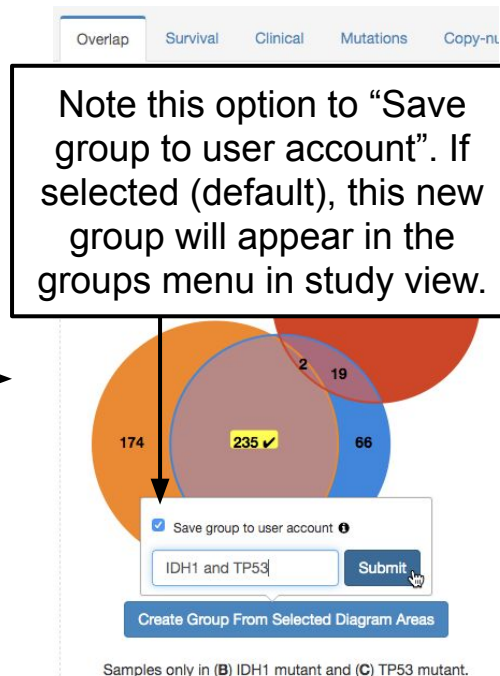
Custom Groups from Overlap Tab



Hover over the venn diagram to find the segment of interest. Click on it.



The “Create Group...” button is now enabled. Click on it.



Enter a name for this new group and click “Submit”.

Custom Groups from Overlap Tab

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

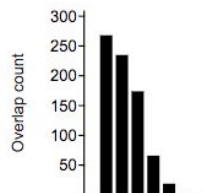
Groups: (drag to reorder) (A) EGFR amp (289) (B) IDH1 and TP53 (0) (C) IDH1 mutant (411) (D) TP53 mutant (322) Select all | Deselect all

Exclude overlapping samples and patients

Overlap Survival Clinical Mutations Copy-number

- (B) IDH1 and TP53 is completely overlapping with other selected groups, so is excluded from analysis in other tabs.
- 256 overlapping samples and patients between (A) EGFR amp, (C) IDH1 mutant, and (D) TP53 mutant are excluded from analysis in other tabs.

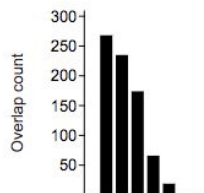
Samples overlap



(A) EGFR amp
(B) IDH1 and TP5...
(C) IDH1 mutant
(D) TP53 mutant

Create Group From Selected Diagram Areas

Patients overlap



(A) EGFR amp
(B) IDH1 and TP5...
(C) IDH1 mutant
(D) TP53 mutant

Create Group From Selected Diagram Areas

The Overlap tab looks different when 4+ groups are selected. Rather than a venn diagram, comparisons among 4+ groups are visualized with an [UpSet diagram](#).

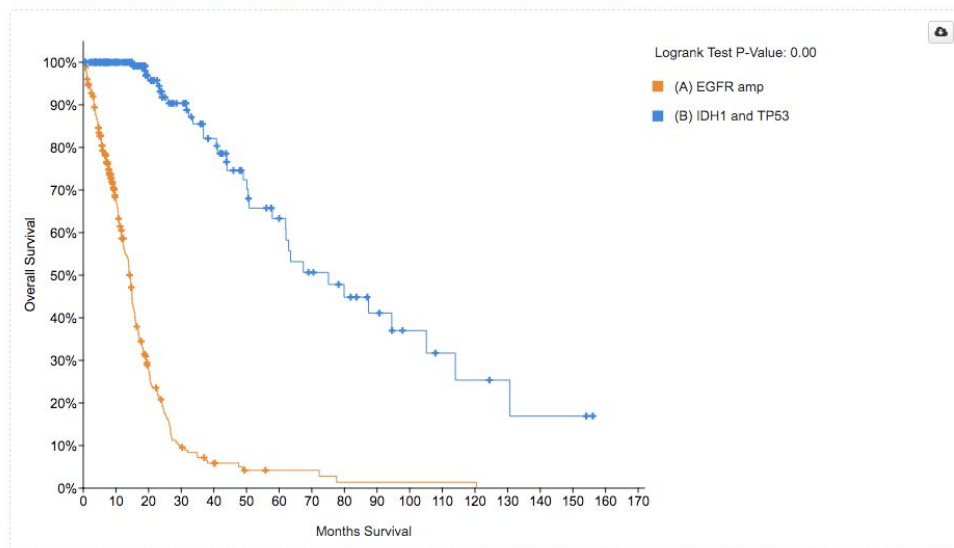
Custom Groups from Overlap Tab

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

Groups: (drag to reorder) (A) EGFR amp (289) (B) IDH1 and TP53 (235) (C) IDH1 mutant (411) (D) TP53 mutant (322) Select all | Deselect all

Overlap Survival Clinical Mutations Copy-number

Overall Survival Kaplan-Meier Estimate (Overall patient survival status.)



Now I can compare, for example, overall survival for patients with EGFR amplification vs IDH1 and TP53 mutations.

There are many powerful ways to take advantage of this new functionality.

What follows are a few possibilities...

Identify/Compare Samples Without a Specific Alteration

How can we compare samples with or without specific genomic features? As an example, let's take samples with IDH1 or IDH2 mutations and compare to samples without mutations in either gene.

1. Filter to samples with mutations in either gene. Create a group from these samples.

2. We can't directly identify samples without mutations in particular genes. But we can create a group of all samples with mutation data, and then use the overlap tab to find the samples with mutations data but without mutations in IDH1 or IDH2.

The screenshot displays the TCGA Data Portal interface for a merged cohort of LGG and GBM samples. The top navigation bar includes 'Summary', 'Clinical Data', and 'Heatmaps'. The main content area is divided into two panels. The left panel, titled 'Mutated Genes (812 profiled samples)', shows a table of genes with mutations. The right panel, titled 'CNA Genes (794 profiled samples)', shows a table of genes with copy number alterations. The 'Quick Filters' section indicates that 812 samples have mutation data and 794 samples have CNA data. The 'Selected: 812 patients | 812 samples' status is shown. The 'Mutated Genes' table lists genes such as IDH1, IDH2, CDKN2A, CDKN2B, and others, with columns for '# Mut', '#', and 'Freq'. The 'CNA Genes' table lists genes such as CDKN2A, CDKN2B, MTAP, TP53, and others, with columns for 'CNA', '#', and 'Freq'. A search bar is located at the bottom of the 'CNA Genes' table.

Gene	# Mut	#	Freq
IDH1	411	411	50.6%
IDH2	19	19	2.3%

Gene	CNA	#	Freq
CDKN2A	411	411	50.6%
CDKN2B	398	398	49.0%
MTAP	216	216	26.5%
TP53	129	129	15.9%
PTEN	109	109	13.4%
EGFR	93	87	10.7%
CIC	85	79	9.7%
MUC16	73	69	8.5%
PIK3CA	67	63	7.8%
NF1	65	53	6.5%
RYR2	48	46	5.7%

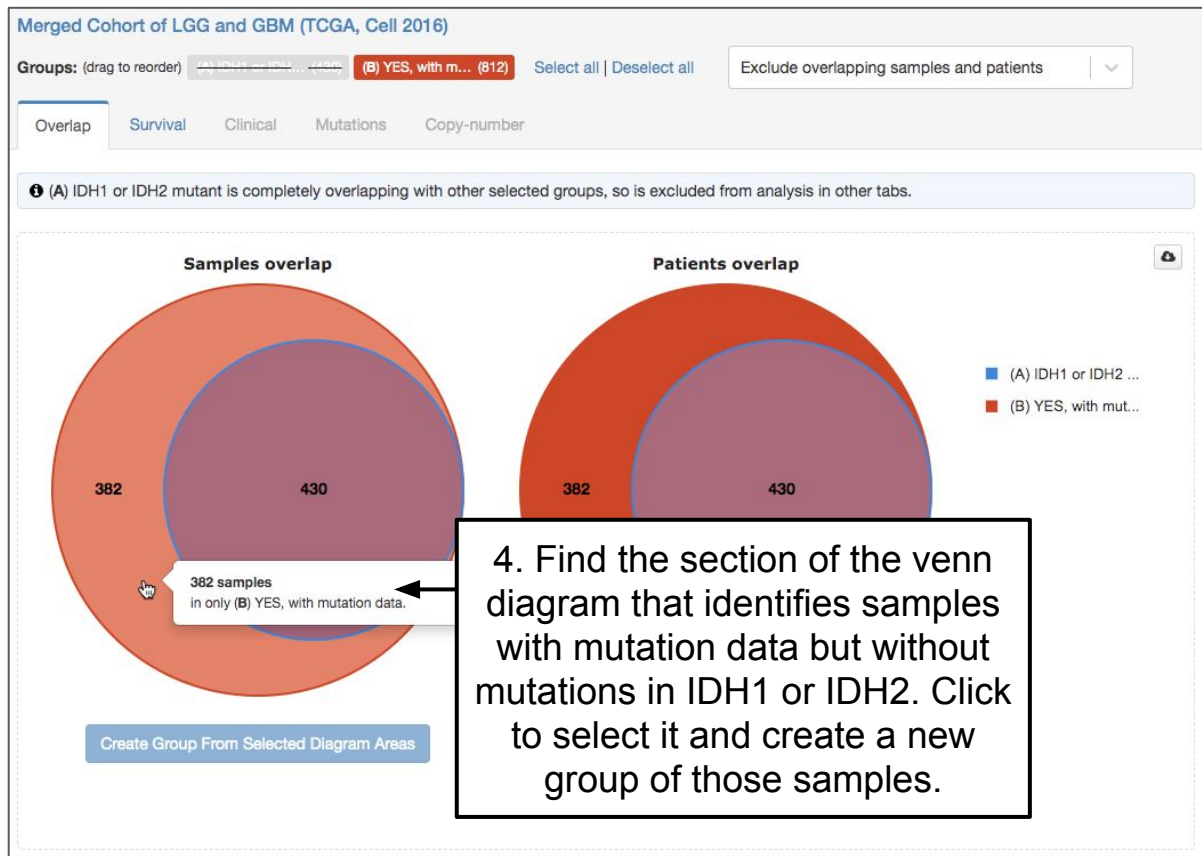
Identify/Compare Samples Without a Specific Alteration

Select all (6) Deselect all Search..

- ☐ EGFR amp (289 samples/patients)
- ☐ IDH1 and TP53 (235 samples/patients)
- ☐ IDH1 mutant (411 samples/patients)
- ☒ IDH1 or IDH2 m... (430 samples/patients)
- ☐ TP53 mutant (322 samples/patients)
- ☒ YES, with mutat... (812 samples/patients)

Compare View

3. Run a comparison between these two groups.



Identify/Compare Samples Without a Specific Alteration

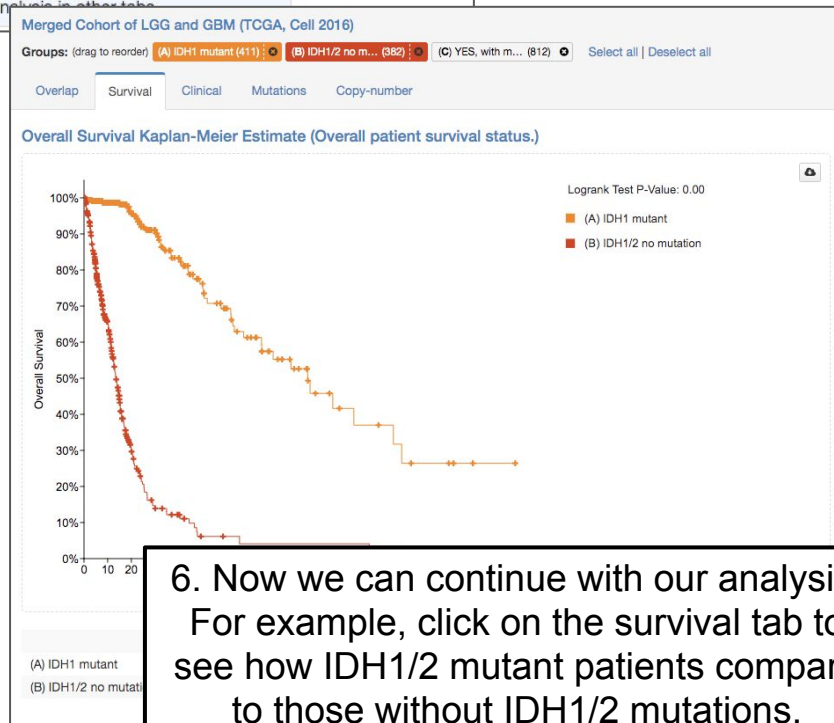
Merged Cohort of LGG and GBM (TCGA, Cell 2016)

Groups: (drag to reorder) (A) IDH1 mutant (411) (B) IDH1/2 no mutation (382) (C) YES, with m... (812) Select all | Deselect all Exclude overlapping samples and patients

Overlap Survival Clinical Mutations Copy-number

ⓘ (A) IDH1 mutant and (B) IDH1/2 no mutation are completely overlapping with other selected groups, so are excluded from analysis in other tabs.

5. When the page reloads, we see the newly created group and a warning that some groups are excluded from analysis due to all samples overlapping with another group. Deselect the “(C) YES, with m...” group.



6. Now we can continue with our analysis. For example, click on the survival tab to see how IDH1/2 mutant patients compare to those without IDH1/2 mutations.

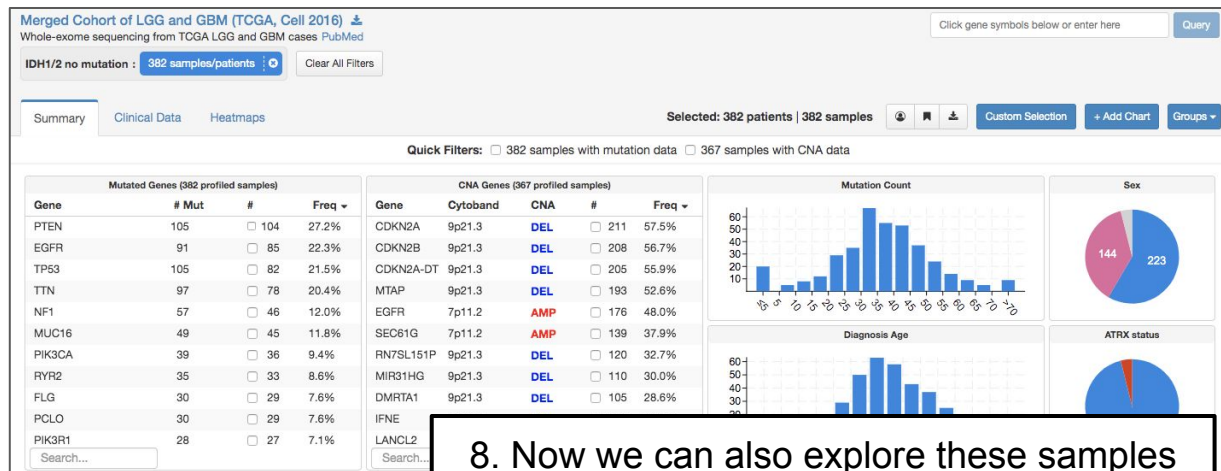
Identify/Compare Samples Without a Specific Alteration

Select all (7) Deselect all Search...

- ☐ EGFR amp (289 samples/patients)
- ☐ IDH1 and TP53 (235 samples/patients)
- ☐ IDH1 mutant (411 samples/patients)
- ☐ IDH1 or IDH2 m... (430 samples/patients)
- ☒ IDH1/2 no muta... (382 samples/patients)
- ☐ TP53 mutant (322 samples/patients)
- ☐ YES, with mutat... (812 samples/patients)

Compare View

7. Remember that these new groups get saved back to Study View. You can use the Groups dropdown to filter study view to this newly defined group.



8. Now we can also explore these samples without IDH1/2 mutations in study view.

Create a Group from a User-Defined List

What if you have identified a set of samples of interest based on your own analysis? You can create a group from that list in study view.

The screenshot displays the OncoPrint interface for a study titled "Merged Cohort of LGG and GBM (TCGA, Cell 2016)". The interface includes a "Modify Query" button and a navigation bar with tabs: "OncoPrint", "Cancer Types Summary", "Plots", "Mutations", "Enrichments", "Survival", "Network", and "Download". A green banner indicates that the results reflect the OQL specification from the query. Below this, there is a section for "Add clinical tracks" and a "Sort" dropdown menu. A specific query is shown: "IDH1: MUT=R132C", which has resulted in 2.1% of the samples being affected. A bar chart visualizes the distribution of mutations. A callout box explains that the user is identifying samples with IDH1 R132C mutations by running a query with OQL. The "Downloadable Data Files" section lists various data types: "Copy-number Alterations", "Mutations", "Samples affected: Only samples with an alteration are included", and "Sample matrix: 1 = Sample harbors alteration in one of the input genes". Each data type has options to download in "Tab Delimited Format" or "Transposed Matrix" format, and a "Copy" button. A callout box points to the "Copy" button for the "Sample matrix" data type, indicating that the user should use the "Download" tab to copy a list of the altered samples.

1. Define your samples of interest. Here, I will identify samples with IDH1 R132C mutations by running a query with OQL...

2. ...and then use the Download tab to copy a list of the altered samples.

Create a Group from a User-Defined List

Merged Cohort of LGG and GBM (TCGA, Cell 2016) Whole-exome sequencing from TCGA LGG and GBM cases PubMed

Summary Clinical Data Heatmaps

Selected: 1,122 patients | 1,122 samples

Quick Filters: ☐ 812 samples with mutation data ☐ 1,084 samples with CNA data

By sample ID ☒ By patient ID

Use currently selected samples/patients Data Format

lgggbm_tcga_pub:TCGA-T5-A5F4-01
lgggbm_tcga_pub:TCGA-S9-A6U1-01
lgggbm_tcga_pub:TCGA-S9-A6U8-01
lgggbm_tcga_pub:TCGA-S9-A6U9-01
lgggbm_tcga_pub:TCGA-S9-A6W1-01
lgggbm_tcga_pub:TCGA-VM-A8CH-01

Select

Sex

Diagnosis Age

ATRX status

Gene # Mut

Gene	# Mut
IDH1	411
TP53	398
ATRX	216
TTN	129
PTEN	109
EGFR	93
CIC	85
MUC16	73
PIK3CA	67
NF1	65

3. Go to study view, click the “Custom Selection” button, and paste in that copied list of samples. Click “Select” to filter to that sample list.

Merged Cohort of LGG and GBM (TCGA, Cell 2016) Whole-exome sequencing from TCGA LGG and GBM cases PubMed

Custom Selection: 17 samples Clear All Filters

Summary Clinical Data Heatmaps

Selected: 17 patients | 17 samples

Quick Filters: ☐ 17 samples with mutation data ☐ 17 samples with CNA data

Select all (7) Deselect all Search...

EGFR amp (289 samples/patients) ☐
IDH1 and TP53 (235 samples/patients) ☐
IDH1 mutant (411 samples/patients) ☐
IDH1 or IDH2 m... (430 samples/patients) ☐
☒ IDH1/2 no muta... (382 samples/patients)
TP53 mutant (322 samples/patients) ☐
YES, with mutat... (812 samples/patients) ☐

Compare View

IDH1 R132C Create

Gene # Mut # Freq

Gene	# Mut	#	Freq
IDH1	17	17	100.0%
ATRX	14	12	70.6%
TP53	14	12	70.6%
APOB	2	2	11.8%
FAM47A	2	2	11.8%
ERICH5			
SERPINA3			
ACAT2			
ACAN			
ABCD1			
ANK2			

Gene Cytoband CNA # Freq

Gene	Cytoband	CNA	#	Freq
AGXT	2q37.3	DEL	3	17.6%
ALPI	2q37.1	DEL	3	17.6%
ALPP	2q37.1	DEL	3	17.6%
ALPG	2q37.1	DEL	3	17.6%
KIF1A	2q37.3	DEL	3	17.6%

4. Then go ahead and create a group of just those filtered samples. Now you can use this group to run a group comparison.

Questions?

Check out our other tutorials
or email us at:

cbioportal@googlegroups.com