

GENOME DATA ANALYSIS

유전체 데이터 분석

II NGS편, 암과 질병 유전체

chapter 4 NGS를 이용한 암 유전체 데이터 분석

제13장 암 유전체 네트워크 분석과 생존 분석 ...255

chapter 6 NGS를 이용한 희귀질환과 복잡질환 유전체 데이터 분석

제24장 희귀 유전질환과 흔한 복잡질환의 유전체학 ...490

제25장 복잡 질환의 희귀변이 연관분석 ...516

목 차

1. 암유전체 네트워크 분석

- cBioPortal 소개 및 활용
- CCLE 소개 및 활용

2. 단일변이와 희귀변이 연관분석

- 단일변이 연관분석
- 희귀변이 분석

강의자료 다운로드

https://github.com/sun_snu/GDA_2024

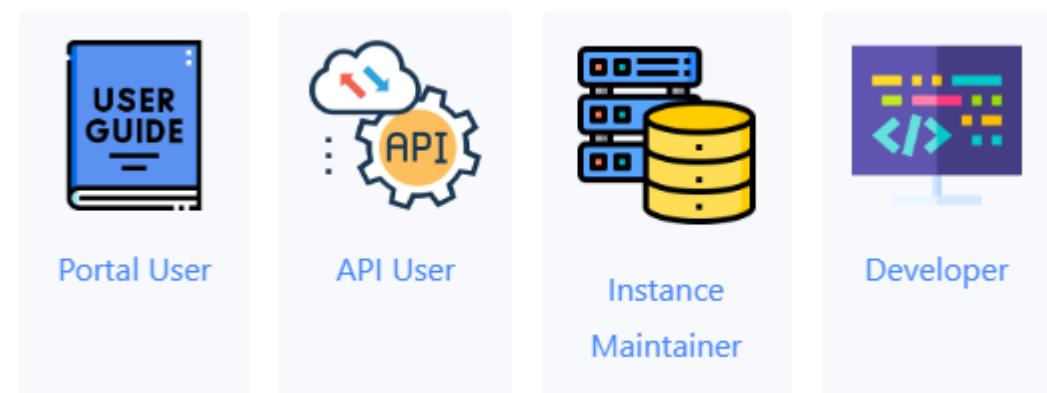
Virtualbox 이용 시 입력하기

source activate lecture2

R

1. cBioPortal

- The cBioPortal for Cancer Genomics provides visualization, analysis and download of large-scale cancer genomics data sets.
- The cBioPortal for Cancer Genomics was originally developed at [Memorial Sloan Kettering Cancer Center](#) (MSK). The [public cBioPortal site](#) is hosted by the [Center for Molecular Oncology](#) at MSK. The cBioPortal software is now available under an open source license via [GitHub](#). The software is now developed and maintained by a multi-institutional team, consisting of MSK, the Dana Farber Cancer Institute, Princess Margaret Cancer Centre in Toronto, Children's Hospital of Philadelphia, Caris Life Sciences, [The Hyve](#) and [SE4BIO](#) in the Netherlands, and Bilkent University in Ankara, Turkey.
- User-defined cohorts
- Cohort summaries
- Patient/sample level discovery
- Supports genomic and clinical data
- Annotations from UniProt, OncoKb and others



2. CCLE(Cancer Cell Line Encyclopedia)

- The Cancer Cell Line Encyclopedia (CCLE) project is an effort to conduct a detailed genetic characterization of a **large panel of human cancer cell lines**. The CCLE provides public access analysis and visualization of DNA copy number, mRNA expression, mutation data and more, for 1000 cancer cell lines.
- The CCLE is made possible through a collaboration between the [Broad Institute](#), the [Novartis Institutes for Biomedical Research](#), and the [Genomics Institute of the Novartis Research Foundation](#) to perform detailed genetic and pharmacologic characterization of a large number of human cancer models. The CCLE public project contains Open Access sequencing data (in the form of reads aligned to the hg19 broad variant reference genome) for nearly 1000 cancer cell line samples, as available from cgHub on May 11, 2016.

Phase I (2008) :

- goal 1) to conduct a detailed genetic and pharmacologic characterization of a large panel of human cancer models;
- 2) to develop integrated computational analyses that link distinct pharmacologic vulnerabilities to characteristic genetic, gene expression, and cell lineage patterns;
- 3) to translate cell line integrative genomics into cancer patient stratification.
- ->acquired 1000 cell lines including ATCC (American), DSMZ (Deutsche) and the KCLB (Korean Cell Line Bank).

Phase II

- applying the emerging Next-Gen sequencing to further expand and refine the characterization of expressed mRNAs through RNA-seq, by further characterizing genetic alterations through exome sequencing, by characterizing the miRNA content of all cell lines, by quantifying the metabolite abundance of 225 metabolites across the CCLE

Phase III

- we continued to strive towards an understanding of the protein content of cell lines. The vast majority of therapeutics act by interrupting or altering protein function and with the growing interest in antibody-drug conjugates, antibody mediated cellular cytotoxicity (ADCC), and CAR-T cells all directed at surface proteins we sought to try and define the CCLE proteome through mass spectrometry.
- To this end, the Gygi lab performed Tandem-mass tagging mass spectrometry to quantify the abundance of proteins in whole cell extracts derived from 375 of the CCLE cell lines.

cBioPortal Tutorial #1: Single Study Exploration

Explore all data in a dataset

Tutorial #1 Objectives

- Introduce cBioPortal main page
- Show two ways to select a study
 - From the Query box on the main page
 - From the Data Sets page
- Walk through the four possible tabs in the study view
 - Study Summary
 - Clinical Data
 - Heatmaps
 - CN Segments
- Show how to run a query from the study view

cBioPortal Main Page

Browse available datasets and select studies to explore or query

Number of studies for each tissue of origin (click to filter)

Search studies

List of all studies, organized by organ system

The screenshot shows the cBioPortal main page with several key features highlighted:

- Search studies:** A search bar at the top right.
- List of all studies, organized by organ system:** A large box containing a list of study categories and their sample counts:
 - PanCancer Studies: 10945 samples
 - Pediatric Cancer Studies: 2922 samples
 - Immunogenomic Studies: 261 samples
 - Cell lines: 1978 samples
 - Adrenal Gland: 72 samples
 - Ampulla of Vater: 657 samples
 - Biliary Tract: 1025 samples
 - Bladder/Urinary Tract: 1089 samples
 - Bone: 961 samples
 - Bowel: 103 samples
 - Breast: 73 samples
 - CNS/Brain: Primary vs. metastatic prostate cancer
 - Cervix: RAS/RAF alterations in colorectal cancer
 - Esophagus/Stomach: BRCA1 and BRCA2 mutations in ovarian cancer
 - POLE hotspot mutations in endometrial cancer
 - TP53 and MDM2/4 alterations in GBM
 - PTEN mutations in GBM in text format
 - Patient view of an endometrial cancer case
 - All TCGA Pan-Cancer
 - MSK-IMPACT clinical cohort, Zehir et al. 2017
 - Histone mutations across cancer types
- Study Selection:** A sidebar on the left lists various study categories with their respective sample counts and checkboxes for filtering.
- Footer:** Includes links for Data Sets, Web API, R/MATLAB, Tutorials/Webinars, FAQ, News, Visualize Your Data, About, cBio, Login, What's New (@cbioportal), and Local Installations (Host your own).

[Link to this page](#)

Selecting a study: from Query

The screenshot shows the cBioPortal 'Query' page. At the top, there's a search bar with the word 'glioma'. Below it, a section titled 'Select Studies for Visualization & Analysis:' lists various studies under categories like 'Immunogenomic Studies', 'CNS/Brain', and 'Soft Tissue'. A specific study, 'Brain Lower Grade Glioma (TCGA, PanCancer Atlas)', has its checkbox selected. At the bottom, there are two buttons: 'Query By Gene' and 'Explore Selected Studies'.

1. Filter the list of studies (optional)
2. Select the checkbox next to the study of interest and click "Explore Selected Studies"

Please cite: Cerami et al., 2012 & Gao et al., 2013

glioma

1 study selected (514 samples) Deselect all

Select Studies for Visualization & Analysis:

1 study selected (514 samples) Deselect all

1. Filter the list of studies (optional)

2. Select the checkbox next to the study of interest and click "Explore Selected Studies"

Query By Gene OR Explore Selected Studies

What's New @cbioportal

cBioPortal @cbioportal

Another How-To video is posted! This one shows how to use patient view to explore the longitudinal evolution of patients, for patients with multiple profiled samples or available clinical histories. youtube.com/watch?v=Hbbs-t...

YouTube @YouTube

Sign up for low-volume email news alerts

Example Qu

- Primary vs. me
- RAS/RAF altera
- BRCA1 and BF
- POLR hotspot
- TP53 and MDM2/4 alterations in GBM
- PTEN mutations in GBM in text format
- Patient view of an endometrial cancer case
- All TCGA Pan-Cancer
- MSK-IMPACT clinical cohort, Zehir et al. 2017
- Histone mutations across cancer types

Local Installations Host your own

Link to this page

Selecting a study: from Data Sets page

1. Use search functionality to find datasets of interest

Datasets

The table below lists the number of available samples per cancer study and data type. It also provides links to download the data for each study. For alternative ways of downloading, see the [Download Documentation](#).

Columns ▾

Name	Reference	All	Mutations	CNA	RNA-Seq
Acinar Cell Carcinoma of the Pancreas (JHU, J Pathol 2014)	Jia et al. J Pathol 2014	23	23	0	0
Acral Melanoma (TGEN, Genome Res 2017)	Liang et al. Genome Res 2017	38	38	38	36
Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)	Andersson et al. Nat Genet 2015	93	93	0	0
Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016)	Zhang et al. Nat Genet 2016	73	73	0	0
Acute Myeloid Leukemia (OHSU, Nature 2018)	Tyner et al. Nature 2018	672	622	0	451
Acute Myeloid Leukemia (TCGA, Firehose Legacy)		200	197	191	173
Acute Myeloid Leukemia (TCGA, NEJM 2013)	TCGA, NEJM 2013	200	200	191	173
Acute Myeloid Leukemia (TCGA, PanCancer Atlas)	TCGA, Cell 2018	200	200	191	173
Acute myeloid leukemia or myelodysplastic syndromes (WashU, 2016)	Welch et al. N Engl J Med. 2016	136	136	0	0
Adenoid Cystic Carcinoma (FMI, Am J Surg Pathl. 2014)	Ross et al. Am J Surg Pathl 2014	28	28	28	0
Adenoid Cystic Carcinoma (JHU, Cancer Prev Res 2016)	Rettig et al. Cancer Prev Res 2016	25	25	0	0
Adenoid Cystic Carcinoma (MDA, Clin Cancer Res 2015)	Mitani et al. Clin Cancer Res 2015	102	65	0	0
Adenoid Cystic Carcinoma (MGH, Nat Gen 2016)	Drier et al. Nature Genetics 2016	10	10	0	0
Adenoid Cystic Carcinoma (MSKCC, Nat Genet 2013)	Ho et al. Nat Genet 2013	60	60	60	0
Adenoid Cystic Carcinoma (Sanger/MDA, JCI 2013)	Stephens et al. JCI 2013	24	24	0	0
Adenoid Cystic Carcinoma of the Breast (MSKCC, J Pathol. 2015)	Martelotto et al. J Pathol 2015	12	12	12	0
Adenoid Cystic Carcinoma Project (J Clin Invest 2019)	Alien et al. J Clin Invest 2019	1049	1049	928	0
Adrenocortical Carcinoma (TCGA, Firehose Legacy)		92	90	90	79
Adrenocortical Carcinoma (TCGA, PanCancer Atlas)	TCGA, Cell 2018	92	91	89	78
Adult Soft Tissue Sarcomas (TCGA, Cell 2017)	TCGA, Cell 2017	206	206	206	206
Ampullary Carcinoma (Baylor College of Medicine, Cell Reports 2016)	Gingras et al. Cell Rep 2016	160	160	0	0
Anaplastic Oligodendrogloma and Anaplastic Oligoastrocytoma (MSKCC, Neuro Oncol 2017)	Thomas et al. Neuro Oncol 2017	22	22	22	0
Basal Cell Carcinoma (UNIGE, Nat Genet 2016)	Bonilla et al. Nat Genet 2016	293	293	0	0
Bladder Cancer (MSK/TCGA, 2020)		476	474	442	296
Bladder Cancer (MSKCC, Eur Urol 2014)	Kim et al. Eur Urol 2015	109	109	109	0
Bladder Cancer (MSKCC, J Clin Onco 2013)	Iyer et al. J Clin Oncol 2013	97	97	97	0
Bladder Cancer (MSKCC, Nat Genet 2016)	Al-Ahmadie et al. Nat Genet 2016	34	34	33	0
Bladder Cancer (TCGA, Cell 2017)	Robertson et al. Cell 2017	413	412	408	408
Bladder Urothelial Carcinoma (BGI, Nat Genet 2013)	Guo et al. Nat Genet 2013	99	99	0	0
Bladder Urothelial Carcinoma (DFCI/MSKCC, Cancer Discov 2014)	Van Allen et al. Cancer Discov 2014	50	50	0	0
Bladder Urothelial Carcinoma (TCGA, Firehose Legacy)		413	130	408	408
Bladder Urothelial Carcinoma (TCGA, Nature 2014)		131	130	128	129
Bladder Urothelial Carcinoma (TCGA, PanCancer Atlas)		411	410	408	407
Brain Lower Grade Glioma (TCGA, Firehose Legacy)		530	286	513	530
Brain Lower Grade Glioma (TCGA, PanCancer Atlas)		514	514	511	514
Brain Tumor PDXs (Mayo Clinic, 2019)		97	83	83	66

2. Or sort by number of samples with each data type

3. Click on data set of interest

Study Summary Tab: Overview

cBioPortal FOR CANCER GENOMICS Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data About cBioPortal Installations Login

Brain Lower Grade Glioma (TCGA, PanCancer Atlas) Click gene symbols below or enter here Query

Brain Lower Grade Glioma TCGA PanCancer data. The original data is [here](#). The publications are [here](#), [PubMed](#)

Summary Clinical Data CN Segments Selected: 514 patients | 514 samples Custom Selection Charts Groups

Cancer Type

	#	Freq
Diffuse Glioma	513	99.8%
Encapsulated Glioma	1	0.2%

Cancer Type Detailed

	#	Freq
Astrocytoma	194	37.7%
Oligodendrogloma	189	36.8%
Oligoastrocytoma	130	25.3%
Low-Grade Glioma (NOS)	1	0.2%

Genomic Profile Sample Counts

Molecular Profile	#	Freq
Fusions	514	100.0%
mRNA expression z-scores relativ...	514	100.0%
Mutations	514	100.0%
Putative arm-level copy-number fr...	514	100.0%
mRNA expression z-scores relativ...	514	100.0%
mRNA Expression, RSEM (Batch ...	514	100.0%
Microbiome Signatures (log RNA ...	513	99.8%
Log2 copy-number values	511	99.4%
Putative copy-number alterations ...	511	99.4%
Protein expression z-scores (RPPA)	428	83.3%
Protein expression (RPPA)	428	83.3%

KM Plot: Overall Survival (months)

KM Plot: Disease Free Survival (months)

Mutation Count

Fraction Genome Altered

Mutated Genes (514 profiled samples)

Gene	# Mut	#	Freq
IDH1	395	395	76.8%
TP53	319	249	48.4%
ATRX	218	194	37.7%
CIC	130	108	21.0%
TTN	117	63	12.3%
FUBP1	51	48	9.3%
PIK3CA	46	42	8.2%
NOTCH1	49	38	7.4%
MUC16	54	36	7.0%
EGFR	42	35	6.8%
NF1	47	31	6.0%

Structural Variant Genes (514 profiled samples)

Gene	# SV	#	Freq
CLU	6	6	1.2%
SEPTIN14	6	6	1.2%
EGFR	5	5	1.0%
PDGFRA	6	5	1.0%
QKI	4	4	0.8%
KIF5A	4	4	0.8%
FGFR3	5	4	0.8%
ABR	4	4	0.8%
FIP1L1	4	3	0.6%
MNAT1	3	3	0.6%
SHC2	3	3	0.6%

CNA Genes (511 profiled samples)

Gene	Cytoband	CNA	#	Freq
CDKN2B	9p21.3	HOMDEL	56	11.0%
CDKN2B...	9p21.3	HOMDEL	56	11.0%
CDKN2A	9p21.3	HOMDEL	55	10.8%
CDKN2A-DT	9p21.3	HOMDEL	51	10.0%
MTAP	9p21.3	HOMDEL	45	8.8%
EGFR	7p11.2	AMP	39	7.6%
EGFR-AS1	7p11.2	AMP	38	7.4%
ELDR	7p11.2	AMP	37	7.2%
SEC61G-DT	7p11.2	AMP	34	6.7%
SEC61G	7p11.2	AMP	30	5.9%
DMRTA1	9p21.3	HOMDEL	30	5.9%

Diagnosis Age

MSI MANTIS Score

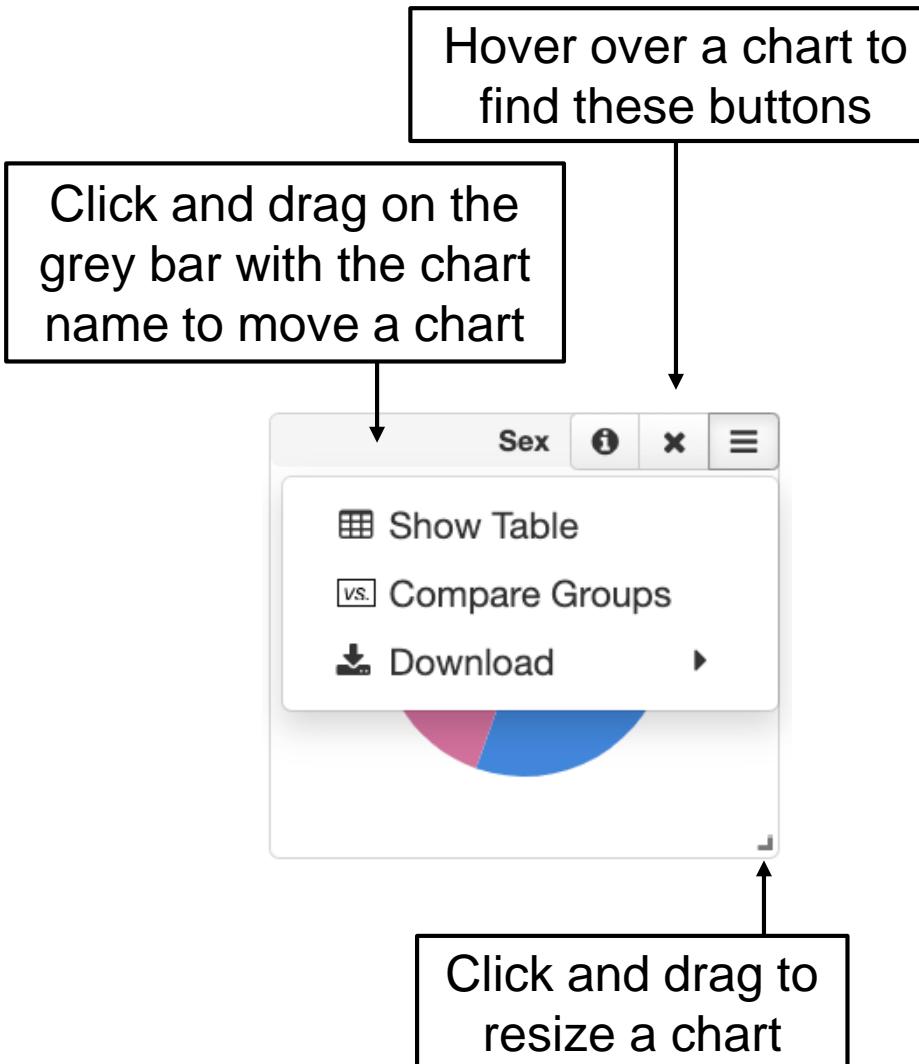
MSisensor Score

Overall Survival Status

Sex

Link to this page

Study Summary Tab: Charts



- i Hover for a description of the data in this chart.
- x Click to remove this chart from view.
- ≡ Hover over this button to bring up a menu with the options below:
 - grid Click to convert the pie chart to a table. Note that hovering over the chart will also bring up a tooltip with tabular data.
 - vs. Click to go to a group comparison session with groups based on these values. See the [group comparison tutorial](#) for more details.
 - download Click to download data (text file) or plot (PDF or SVG).

Study Summary Tab: Charts

Add charts using this button. Added charts can be used like any other chart to filter or define groups for comparison.

The Clinical tab displays a list of available data types. A callout box points to the 'Query' button at the top right of the header. The list includes:

- Selected: 514 patients | 514 samples
- Clinical Genomic Gene Specific Custom Data X vs Y Beta! Arm-level CNA Microbiome Signature
- Select all (55) Deselect all Search...
- Name Freq
- Buffa Hypoxia Score 100.0%
- Cancer Studies 100.0%
- Cancer Type 100.0%
- Cancer Type Detailed 100.0%
- Case Lists 100.0%
- Center of sequencing 100.0%
- In PanCan Pathway Analysis 100.0%
- MSIsensor Score 100.0%
- Number of Samples Per Patient 100.0%
- Oncotree Code 100.0%
- Ragnum Hypoxia Score 100.0%
- Sample Type 100.0%
- Somatic Status 100.0%
- TCGA PanCanAtlas Cancer Type Acronym 100.0%

Clinical: lists all patient- and sample-level data available for this study

The Genomic tab displays a list of charts summarizing genomic data. The header shows 'Selected: 514 patients | 514 samples'. The tabs include Clinical, Genomic (selected), Gene Specific, Custom Data, X vs Y Beta!, Arm-level CNA, and Microbiome Signature. The list of charts includes:

Name	Freq
Genomic Profile Sample Counts	100.0%
Mutated Genes	100.0%
Structural Variant Genes	100.0%
CNA Genes	99.4%
Fraction Genome Altered	99.4%
Mutation Count	99.0%

Genomic: lists charts summarizing genomic data

The Gene Specific tab displays a chart for the gene EGFR. The header shows 'Selected: 514 patients | 514 samples'. The tabs include Clinical, Genomic (selected), Gene Specific (selected), Custom Data, X vs Y Beta!, Arm-level CNA, and Microbiome Signature. The chart area shows:

- EGFR
- All gene symbols are valid.
- mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2) (514 sa...)
- Add Chart

Gene Specific: add charts for individual genes from any molecular profile with continuous data, e.g. mRNA expression.

Study Summary Tab: Charts

The screenshot shows the 'Custom Data' section of the 'Charts' tab. At the top, there are tabs for Clinical, Genomic, Gene Specific, Custom Data (which is selected), X vs Y Beta!, Arm-level CNA, and Microbiome Signature. Below these tabs, there are two radio buttons: 'By sample ID' (selected) and 'By patient ID'. Under 'currently selected', there is a list of TCGA samples grouped by ID: lgg_tcga_pan_can_atlas_2018:TCGA-CS-4938-01 group1, lgg_tcga_pan_can_atlas_2018:TCGA-CS-4941-01 group2, lgg_tcga_pan_can_atlas_2018:TCGA-CS-4942-01 group3, lgg_tcga_pan_can_atlas_2018:TCGA-CS-4943-01 group1, and lgg_tcga_pan_can_atlas_2018:TCGA-CS-4944-01 group3. There is also a 'Title (optional)' input field and a blue 'Add Chart' button.

Custom Data: add charts with new data for the existing samples, for example results of your own analysis that classifies TCGA samples into groups.

The screenshot shows the 'X vs Y' section of the 'Charts' tab. At the top, there are tabs for Clinical, Genomic, Gene Specific, Custom Data, X vs Y Beta! (selected), Arm-level CNA, and Microbiome Signature. Below these tabs, there are dropdown menus for 'X-Axis' and 'Y-Axis', both labeled 'Select x/y-axis clinical attribute'. A large blue 'Add Chart' button is at the bottom.

X vs Y: add charts comparing two clinical attributes. Note this feature is still under development.

The screenshot shows the 'Arm-level CNA' and 'Microbiome Signature' sections of the 'Charts' tab. At the top, there are tabs for Clinical, Genomic, Gene Specific, Custom Data, X vs Y Beta!, Arm-level CNA (selected), and Microbiome Signature. Below these tabs, there is a search bar 'Search for Arm-level CNAs...' and a dropdown menu 'Putative arm-level copy-number from GISTIC (514 samples)'. A blue 'Add Chart' button is at the bottom.

Arm-level CNA, Microbiome Signature and others: add charts for additional datatypes that are available for some studies. Not all studies will have these subtabs.

Study Summary Tab: Selecting subsets of data

Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 

Brain Lower Grade Glioma TCGA PanCancer data. The original data is [here](#). The publications are [here](#), [PubMed](#).

Click gene symbols below or enter here Query

IDH1  Clear All Filters 

Summary Clinical Data CN Segments

Selected: 395 patients | 395 samples    Custom Selection Charts Groups

3. After applying filters, this button will generate a shareable link with the filters applied.

2. All plots update to include just the samples with IDH1 mutations. You can also apply additional filters and plots will continue to update.

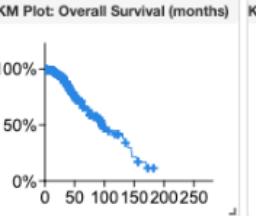
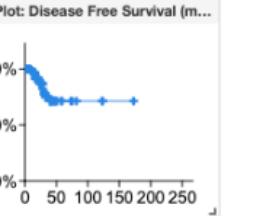
1. You can use any chart or plot to select a subset of samples. Here, IDH1 mutated samples are selected.

4. To remove filters, click on the blue arrow to remove the filter from a specific chart. Or, use the “Clear All Filters” button in the header.

Cancer Type	#	Freq
Diffuse Glioma	394	99.7%
Encapsulated Glioma	1	0.3%

Cancer Type Detailed	#	Freq
Oligodendrogloma	152	38.5%
Astrocytoma	133	33.7%
	109	27.6%
	1	0.3%

Genomic Profile Sample Counts	#	Freq
Fusions	395	100.0%
mRNA expression z-scores relativ...	395	100.0%
Mutations	395	100.0%
Putative arm-level copy-number fr...	395	100.0%
mRNA expression z-scores relativ...	395	100.0%

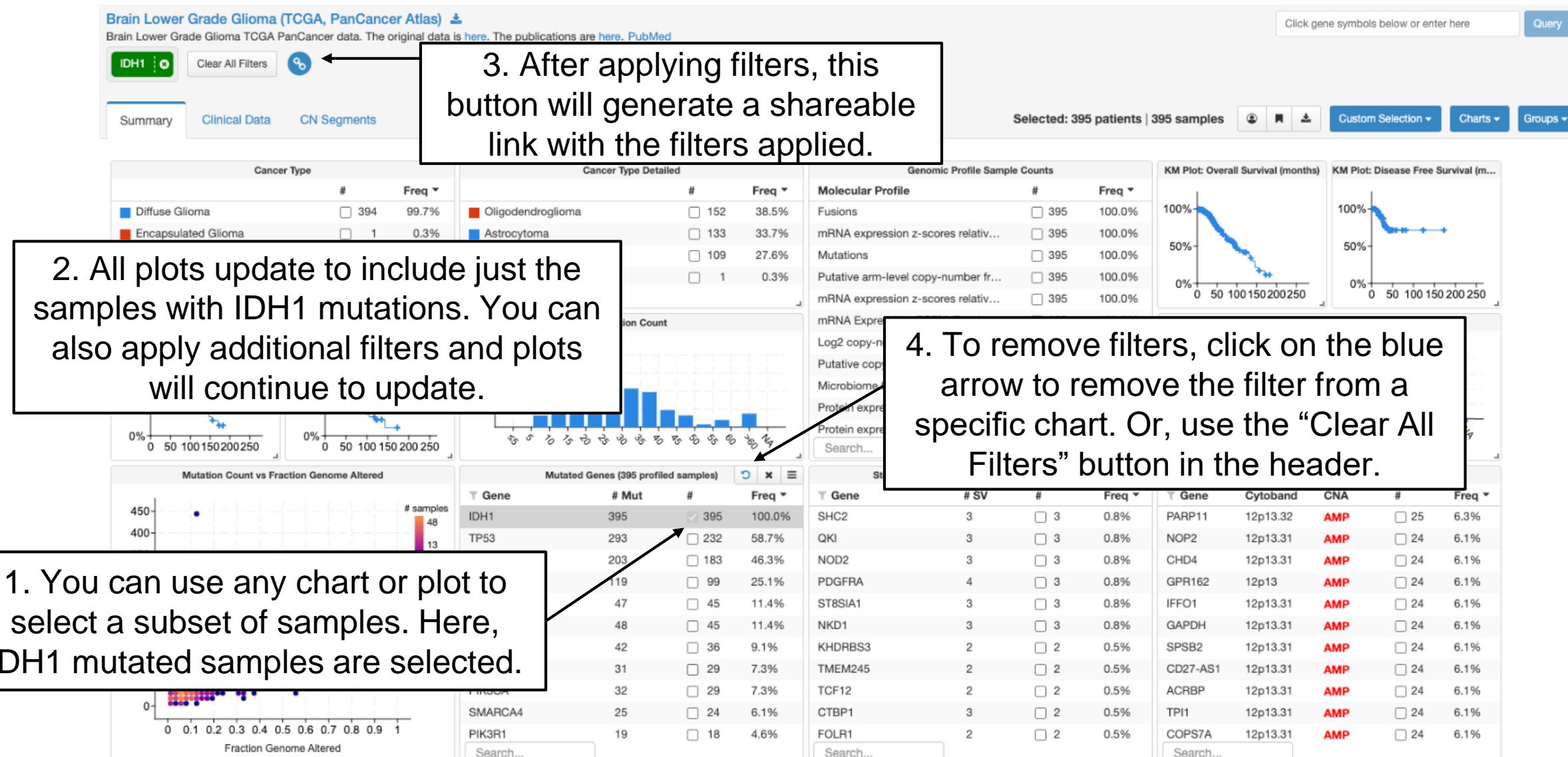
KM Plot: Overall Survival (months)	KM Plot: Disease Free Survival (m...
	

Mutated Genes (395 profiled samples)	Gene	# Mut	#	Freq
IDH1	395	395	100.0%	
TP53	293	232	58.7%	
	203	183	46.3%	
	119	99	25.1%	
	47	45	11.4%	
	48	45	11.4%	
	42	36	9.1%	
	31	29	7.3%	
	32	29	7.3%	
	SMARCA4	25	24	6.1%
	PIK3R1	19	18	4.6%

Gene	# SV	#	Freq
SHC2	3	3	0.8%
QKI	3	3	0.8%
NOD2	3	3	0.8%
PDGFRA	4	3	0.8%
ST8SIA1	3	3	0.8%
NKD1	3	3	0.8%
KHDRBS3	2	2	0.5%
TMEM245	2	2	0.5%
TCF12	2	2	0.5%
CTBP1	3	2	0.5%
FOLR1	2	2	0.5%

Gene	Cytoband	CNA	#	Freq
PARP11	12p13.32	AMP	25	6.3%
NOP2	12p13.31	AMP	24	6.1%
CHD4	12p13.31	AMP	24	6.1%
GPR162	12p13	AMP	24	6.1%
IFFO1	12p13.31	AMP	24	6.1%
GAPDH	12p13.31	AMP	24	6.1%
SPSB2	12p13.31	AMP	24	6.1%
CD27-AS1	12p13.31	AMP	24	6.1%
ACRBP	12p13.31	AMP	24	6.1%
TPI1	12p13.31	AMP	24	6.1%
COPS7A	12p13.31	AMP	24	6.1%

Search... Search... Search...



Clinical Data Tab

Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 

Brain Lower Grade Glioma TCGA PanCancer data. The original data is [here](#). The publications are [here](#), [PubMed](#)

IDH1  

Filters applied in the Summary tab apply to this table

Download clinical data table

Show additional data (available data will vary based on the study)

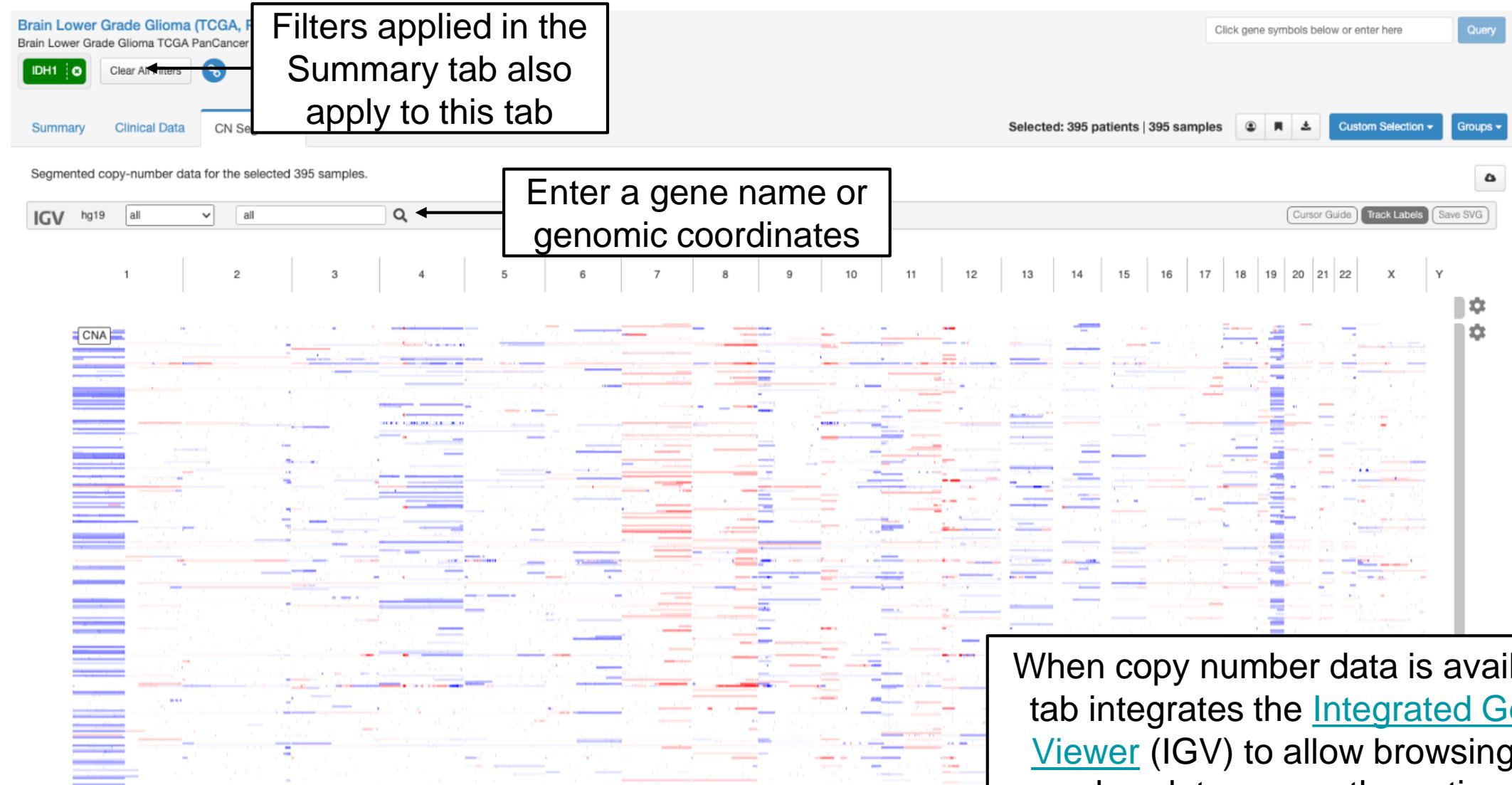
Selected: 395 patients | 395 samples     

Patient ID	Sample ID	Cancer Type	Cancer Type Detailed	Mutation Count	Fraction Genome Altered	Diagnosis Age	MSI MANTIS Score	MSIsensor Score	Overall Survival Status	Sex	Ethnicity Category	Race Category	Subtype	Tumor Type	Aneuploidy Score	Birth from Initial Pathologic Diagnosis Date
TCGA-CS-4938	TCGA-CS-4938-01	Diffuse Glioma	Astrocytoma	14	0.0518	31.0	0.303	0	0:LIVING	Female	Not Hispanic Or Latino	White	LGG_IDHmut-non-codel	Astrocytoma	1	-11509.0
TCGA-CS-4942	TCGA-CS-4942-01	Diffuse Glioma	Astrocytoma	26	0.0937	44.0	0.281	0.02	1:DECEASED	Female		Black or African American	LGG_IDHmut-non-codel	Astrocytoma	2	-16297.0
TCGA-CS-4943	TCGA-CS-4943-01	Diffuse Glioma	Astrocytoma	24	0.1625	37.0	0.2751	0.25	1:DECEASED	Male		White	LGG_IDHmut-non-codel	Astrocytoma	5	-13565.0
TCGA-CS-4944	TCGA-CS-4944-01	Diffuse Glioma	Astrocytoma	21	0.0603	50.0	0.2697	0.04	0:LIVING	Male		White	LGG_IDHmut-non-codel	Astrocytoma	1	-18494.0
TCGA-CS-5390	TCGA-CS-5390-01	Diffuse Glioma	Oligodendrogloma	44	0.0511	47.0	0.2623	0.1	0:LIVING	Female		White	LGG_IDHmut-codel	Oligodendrogloma	2	-17460.0
TCGA-CS-5393	TCGA-CS-5393-01	Diffuse Glioma	Astrocytoma	24	0.0569	39.0	0.2715	0	0:LIVING	Male		White	LGG_IDHmut-non-codel	Astrocytoma	0	-14418.0
TCGA-CS-5394	TCGA-CS-5394-01	Diffuse Glioma	Astrocytoma	22	0.0469	40.0	0.3295	0.16	0:LIVING	Male		White	LGG_IDHmut-non-codel	Astrocytoma	1	-14920.0
TCGA-CS-5396	TCGA-CS-5396-01	Diffuse Glioma	Oligodendrogloma	28	0.1259	53.0	0.2798	0	0:LIVING	Female	Not Hispanic Or Latino	White	LGG_IDHmut-codel			
TCGA-CS-6290	TCGA-CS-6290-01	Diffuse Glioma	Astrocytoma	20	0.0133	31.0	0.2888	0	1:DECEASED	Male			LGG_IDHmut-non-codel			
TCGA-	TCGA-	Diffuse	Astrocytoma	61	0.1319	51.0	0.2995	0.05	0:LIVING	Female	Not	White	LGG_IDHmut-			

Scroll to the right to see more columns. Each column can be sorted by clicking on the column header.

CN Segments Tab



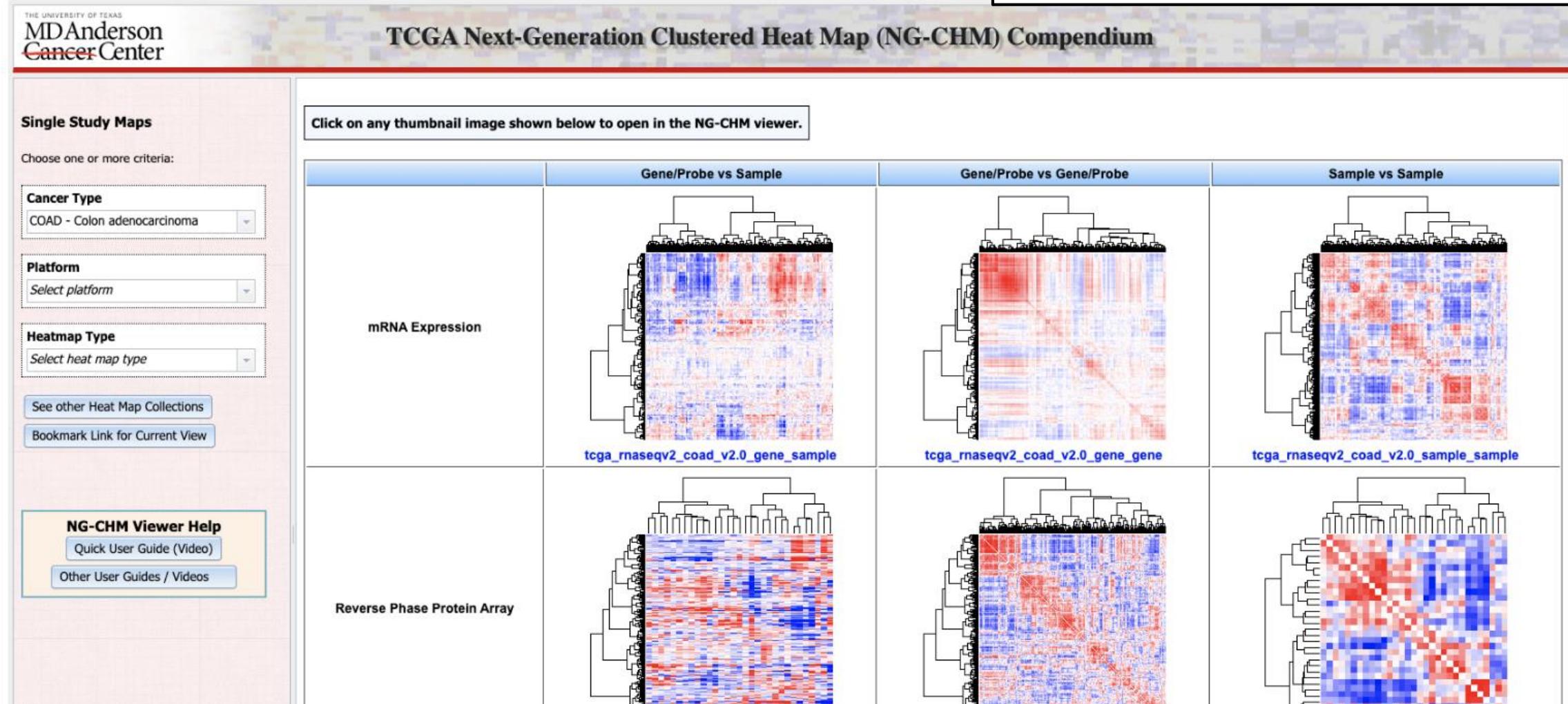
Additional Tabs: Heatmaps

Colorectal Adenocarcinoma (TCGA, Firehose Legacy) 

TCGA Colorectal Adenocarcinoma. Source data from GDAC Firehose. Previously known as TCGA Provisional.

Summary Clinical Data Heatmaps CN Segments

This tab will only appear for some TCGA studies. It is an embedding of the [Next-Generation Clustered Heat Map](#) interactive heatmap tool.



[Link to this page](#)

Study View: Additional Features

Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 

Brain Lower Grade Glioma TCGA PanCancer data. The original data is [here](#). The publications are [here](#). [PubMed](#)

Summary Clinical Data CN Segments Selected: 514 patients | 514 samples Custom Selection ▾ Charts ▾ Groups ▾

Click gene symbols below or enter here [Query](#)

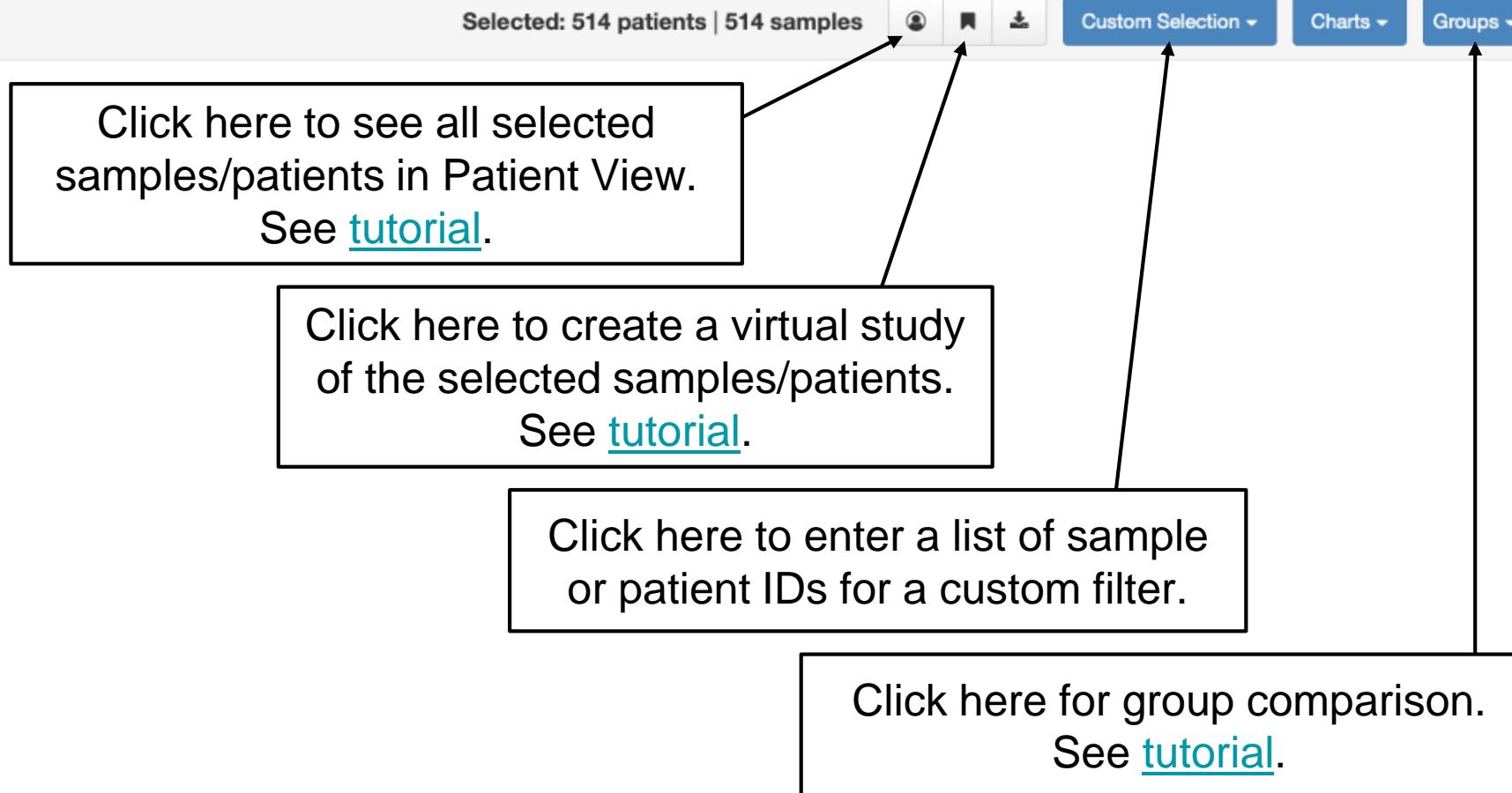
Use this box to run a query (see next slide).

Click here to see all selected samples/patients in Patient View.
See [tutorial](#).

Click here to create a virtual study of the selected samples/patients.
See [tutorial](#).

Click here to enter a list of sample or patient IDs for a custom filter.

Click here for group comparison.
See [tutorial](#).



[Link to this page](#)

Study Summary Tab: Run a query

3. Then click here to run the query

Brain Lower Grade Glioma (TCGA, PanCancer Atlas)

Brain Lower Grade Glioma TCGA PanCancer data. The original data is [here](#). The publications are [here](#). PubMed

Mutations and Putative copy-number alterations from GISTIC | Clear All Filters |

Summary Clinical Data CN Segments

Diffuse Glioma
Encapsulated Glioma

Search... KM Plot: Disease-specific Survival

KM Plot: Overall Survival (months)

Mutation Count vs Fraction Genome Altered

Mutated Genes (511 profiled samples)

IDH1 is included in the OncoKB Cancer Gene List as a oncogene.

CNA Genes (511 profiled samples)

Link to this page

1. Apply filters (optional). Here, we are filtering to samples that have both mutation and copy number data. The query will run in only these selected samples.

2. Type a gene name here

3. Or click on a gene to add it to the query

Genomic Profile Sample Counts

Molecular Profile	#	Freq
Mutations	511	100.0%
Putative copy-number alterations ...	511	100.0%
Fusions	511	100.0%
mRNA expression z-scores relativ...	511	100.0%
Putative arm-level copy-number fr...	511	100.0%
Log2 copy-number values	511	100.0%
mRNA expression z-scores relativ...	511	100.0%
mRNA Expression, RSEM (Batch ...	511	100.0%
Microbiome Signatures (log RNA ...	510	99.8%
Protein expression z-scores (RPPA)	425	83.2%
Protein expression (RPPA)	425	83.2%

KM Plot: Disease-Free Survival (months)

Fraction Genome Altered

Cytoband	CNA	#	Freq
9p21.3	HOMDEL	56	11.0%
9p21.3	HOMDEL	56	11.0%
9p21.3	HOMDEL	55	10.8%
9p21.3	HOMDEL	51	10.0%
9p21.3	HOMDEL	45	8.8%
7p11.2	AMP	39	7.6%
7p11.2	AMP	38	7.4%
7p11.2	AMP	37	7.2%
7p11.2	AMP	34	6.7%
7p11.2	AMP	30	5.9%
9p21.3	HOMDEL	30	5.9%

cBioPortal Tutorial #2: Single Study Query

Query one or multiple genes in a single dataset

Tutorial Objectives

- Show how to run a single-study query from the main page
- Walk through each of the data/analysis tabs in a single-study query
 - OncoPrint
 - Cancer Types Summary
 - Mutual Exclusivity
 - Plots
 - Mutations
 - Co-expression
 - Comparison/Survival
 - CN Segments
 - Pathways
 - Download
- Show how to modify and re-run a query

In this tutorial, blue boxes provide an overview of each tab on cBioPortal ...

... while green boxes ask a biological question that we can answer using cBioPortal.

Overview of Tabs in a Single Study Query

Note that depending on the query run and the data available for a particular study, not all of these will be present (e.g. a study without mRNA expression data will not have a Co-expression tab)

- **OncoPrint:** Overview of genetic alterations per sample in each query gene
- **Cancer Types Summary:** Frequency of alteration in each query gene in the detailed cancer types included in this study
- **Mutual Exclusivity:** Statistical analysis to determine if query genes are mutually exclusively altered
- **Plots:** explore the relationships among genetic alterations, gene expression, protein levels, DNA methylation and available clinical features
- **Mutations:** Details about mutations called in each query gene
- **Co-expression:** Explore which genes have mRNA/protein levels correlated with query genes
- **Comparison/Survival:** Explore overlaps, outcomes, clinical attributes and genomic data comparisons among groups of samples as defined by the query
- **CN Segments:** Explore copy number changes with the Integrated Genomics Viewer (IGV)
- **Pathways:** Explore queried genes in TCGA-defined pathways
- **Download:** Download data or copy sample lists

We're going to run a query in a TCGA Lower-Grade Glioma study.

Glioma is a growth of cells that starts in the brain or spinal cord. As a glioma grows it forms a tumor which can grow to press on brain or spinal cord tissue and cause symptoms.

Aberrant **epidermal growth factor receptor (EGFR)** signaling is common in cancer. EGFR gene amplification and overexpression are a particularly striking feature of glioblastoma (GBM), observed in approximately 40% of tumors.



Single study query

cBioPortal FOR CANCER GENOMICS

Data Sets Web API R/MATLAB Tutorials/Webinars FAQ News Visualize Your Data All

Query Quick Search Beta! Download

Please cite: Cerami et al. 2012 & Gao et al., 2013

glioma

Select Studies for Visualization & Analysis: 1 study selected (514 samples) Deselect all

Immunogenomic Studies 1 Select all listed studies matching filter (20)

CNS/Brain 18 Immunogenomic Studies

Soft Tissue 2 Glioblastoma (Columbia, Nat Med. 2019) 42 samples

CNS/Brain Integrated Proteogenomic Characterization across Major Histological T... 218 samples

Diffuse Glioma Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 530 samples

Brain Lower Grade Glioma (TCGA, PanCancer Atlas) 514 samples

Diffuse Glioma (GLASS Consortium, Nature 2019) 444 samples

Glioma (MSK, Nature 2019) 91 samples

Glioma (MSKCC, Clin Cancer Res 2019) 1004 samples

Low-Grade Gliomas (UCSF, Science 2014) 61 samples

Merged Cohort of LGG and GBM (TCGA, Cell 2016) 1102 samples

→ GLIOBLASTOMA Brain Tumor PDXs (Mayo Clinic, 2019) 97 samples

Glioblastoma (Columbia, Nat Med. 2019) 42 samples

Glioblastoma (TCGA, Cell 2013) 543 samples

Glioblastoma (TCGA, Nature 2008) 206 samples

Glioblastoma Multiforme (TCGA, Firehose Legacy) 619 samples

Glioblastoma Multiforme (TCGA, PanCancer Atlas) 592 samples

1 study selected (514 samples) Deselect all

Query By Gene OR **Explore Selected Studies**

1. Filter the list of studies (optional)

2. Check the box for study of interest.

3. Select “Query By Gene”

Login

What's New @cbioportal

cBioPortal @cbioportal Another How-To video is posted! This one shows how to use patient view to explore the longitudinal evolution of patients, for patients with multiple profiled samples or available clinical histories. youtube.com/watch?v=lbbs-t...

YouTube @YouTube

Sign up for low-volume email news alerts

Subscribe

Example Queries

- Primary vs. metastatic prostate cancer
- RAS/RAF alterations in colorectal cancer
- BRCA1 and BRCA2 mutations in ovarian cancer
- POLE hotspot mutations in endometrial cancer
- TP53 and MDM2/4 alterations in GBM
- PTEN mutations in GBM in text format
- Patient view of an endometrial cancer case
- All TCGA Pan-Cancer
- MSK-IMPACT clinical cohort, Zehir et al. 2017
- Histone mutations across cancer types

Local Installations Host your own



Link to this page

Single study query

4. This section lists all data types available for the selected study. Select data types to query. By default, Mutations and CNA will be selected (if available).

5. Select sample set. For most studies, an appropriate sample set will be automatically selected given the data types selected in Step 4.

The screenshot shows the cBioPortal interface for a 'Brain Lower Grade Glioma (TCGA, PanCancer Atlas)' study (514 total samples). The 'Query' tab is active. Step 4 highlights the 'Select Genomic Profiles' section, which includes checkboxes for 'Mutations', 'Structural Variant', 'Putative copy-number alterations from GISTIC', and a group for mRNA Expression with three options. Step 5 highlights the 'Select Patient/Case Set' and 'Enter Genes' sections. The 'Enter Genes' section shows a 'User-defined List' containing 'IDH1 EGFR' with a note that 'All gene symbols are valid.' Step 6 highlights the 'Submit Query' button. A callout box for Step 5 states: '5. Type gene(s) or select from pre-defined gene lists. cBioPortal will confirm that all entries are valid gene symbols.' A separate callout box on the right provides instructions: 'Refine your query: You can use Onco Query Language (OQL) to define which specific alterations to include. See [specifications](#) or [OQL tutorial](#).'

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.

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

User-defined List

IDH1 EGFR

All gene symbols are valid.

Submit Query

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

Please cite: Cerami et al., 2012 & Gao et al., 2013

Results View Header: General Information

The name of the study.
Click to view the full study in Study View.

The number (percentage) of samples/patients with an alteration in any of the query genes

Modify Query



Brain Lower Grade Glioma (TCGA, PanCancer Atlas)
Samples with mutation and CNA data (511 patients/samples) - IDH1 & EGFR

OncoPrint

Cancer Types Summary

Mutual Exclusivity

Plots

Mutations

Co-expression

Comparison/Survival

CN Segments

Pathways

Download

Queried genes are altered in 444 (87%) of queried patients/samples

The number of samples and patients included in the query. Note that these numbers can differ from each other if some patients have more than one tumor sample profiled.

Click on the number of patients/samples to go to Study View for just the queried samples.

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

Results View Header: Variant Settings

Use this menu to control how alterations are visualized. Changes made here are immediately reflected across Results View. However over the  to confirm how individual tabs reflect these selections.

↓

Brain Lower Grade Glioma (TCGA, PanCancer Atlas)
Samples with mutation and CNA data (511 patients/samples) - IDH1 & EGFR 

Queried genes are altered in 444 (87%) of queried patients/samples 

Annotate Data 
 Putative drivers vs VUS:
 Oncokb driver annotation
 Hotspots 🔥
 cBioPortal >= 0
 COSMIC >= 0

Filter Data
 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
 Exclude samples that are unprofiled in every queried gene and profile.

Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

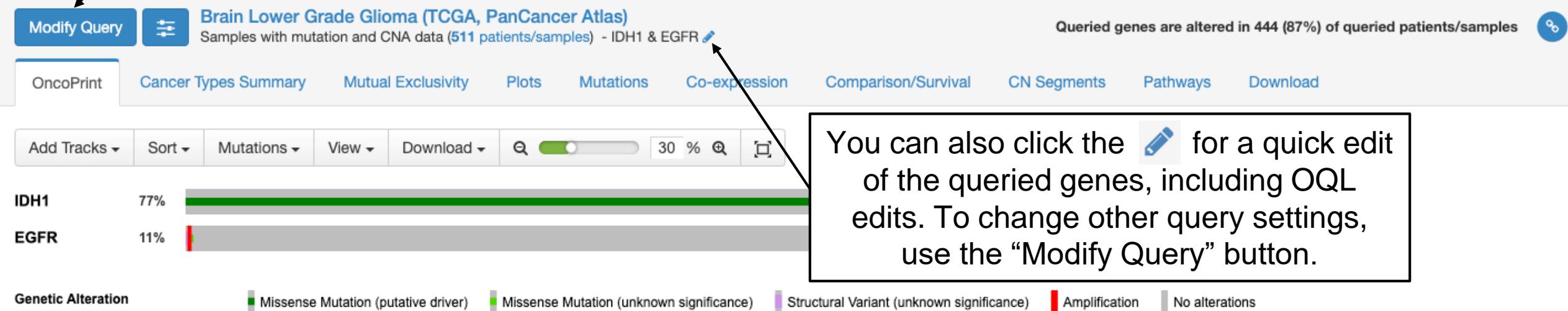
Set the definition of a putative driver vs variant of unknown significance (VUS).

Check boxes to exclude VUS (as defined above) or germline alterations. When checked, VUS or germline alterations are considered not present, so a sample with only VUS or germline alterations will be treated as an unaltered sample.

Check box to exclude samples where queried genes are not profiled or genomic profiles are not available.

Modify Query

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



You can also click the for a quick edit of the queried genes, including OQL edits. To change other query settings, use the “Modify Query” button.

Modify Query

The existing query is pre-populated for your convenience. You can change the study, the genomic profiles, the patient/case set or the gene set. Simply hit “Submit” when you are happy with the modified query.

Cancel Modify Query Brain Lower Grade Glioma (TCGA, PanCancer Atlas)
Samples with mutation and CNA data (511 patients/samples) - IDH1 & EGFR

Queried genes are altered in 444 (87%) of queried patients/samples

Please cite: Cerami et al., 2012 & Gao et al., 2013

Query

Select Studies for Visualization & Analysis: 1 study selected (514 samples) Deselect all Search...

PanCancer Studies 9 Quick select: TCGA PanCancer Atlas Studies Curated set of non-redundant studies

PanCancer Studies

- MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017)
- Metastatic Solid Cancers (UMich, Nature 2017)
- MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018)
- SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018)
- TMB and Immunotherapy (MSKCC, Nat Genet 2019)
- Tumors with TRK fusions (MSK, Clin Cancer Res 2020)
- Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020)
- China Pan-cancer (OriGenMed2020)
- Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020)

10945 samples

500 samples

249 samples

141 samples

1661 samples

106 samples

24146 samples

10194 samples

2922 samples

Pediatric Cancer Studies

- Pediatric Preclinical Testing Consortium (CHOP, Cell Rep 2019)
- Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018)

261 samples

1978 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: Hint: Learn Onco Query Language (OQL) to write more powerful queries

User-defined List

IDH1 EGFR IDH2

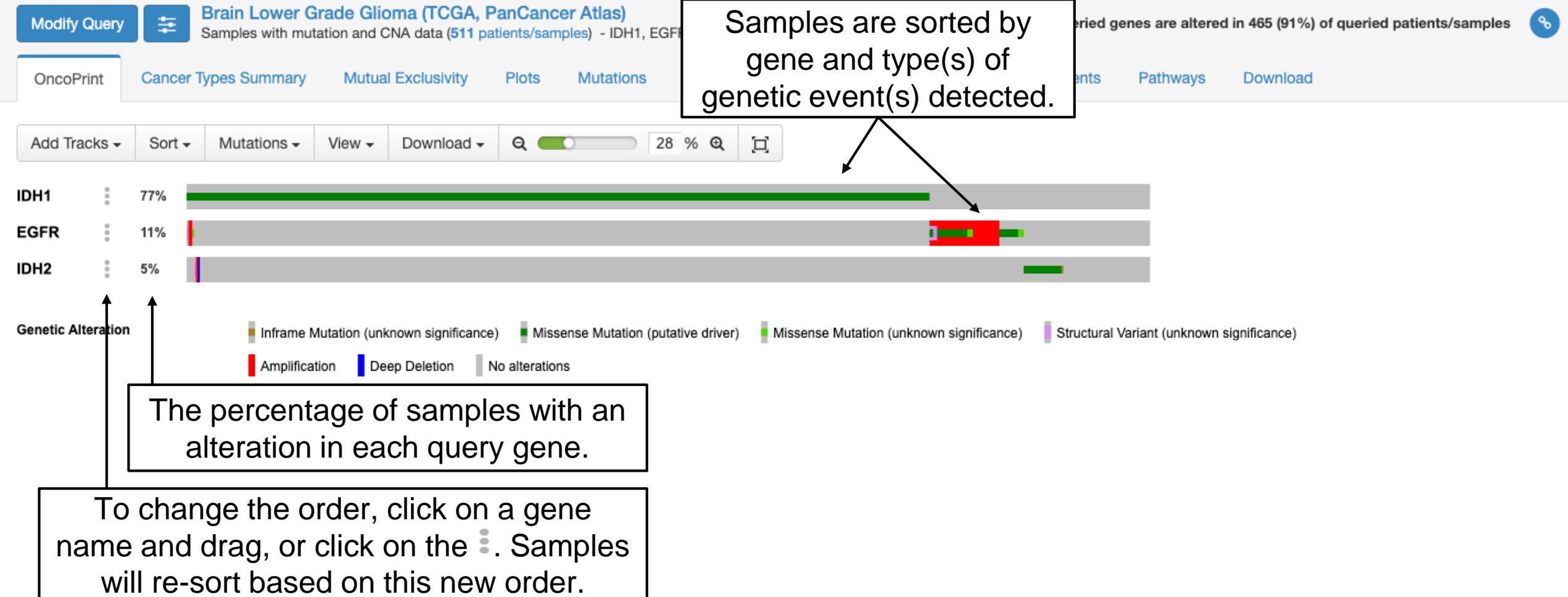
All gene symbols are valid.

In this case, I've added a third gene (IDH2) to the query.

Submit Query

OncoPrint

Summary of alterations per sample. Each sample is a column. Each gene is a row. Different kinds of genetic alterations are highlighted with different colors.



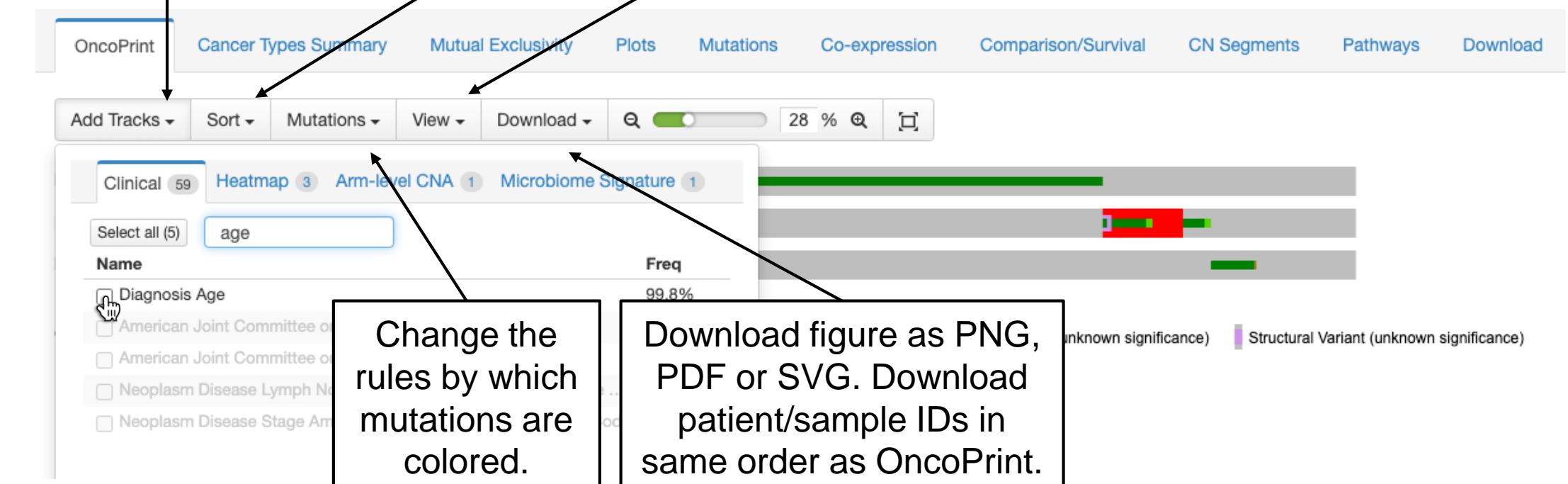
OncoPrint: Features

Add clinical tracks, heatmaps (eg RNA levels) or other data (eg Arm-level CNA).

Available data varies by study.

Change the sample sorting order

Customize visualization



OncoPrint: Zoom

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

Change the zoom by clicking the zoom in/out icons or moving the slider or typing a value

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

Add Tracks ▾ Sort ▾ Mutations ▾ View ▾ Download ▾

58 %

Click here to open “minimap” (see below)

Diagnosis Age



IDH1

77%



EGFR

11%



IDH2

5%



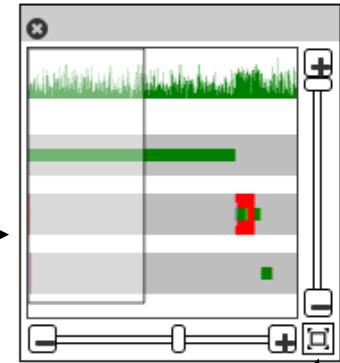
Genetic Alteration

Inframe Mutation (unknown significance) Missense Mutation (putative driver) Missense Mutation (unknown significance) Structural Variant (unknown significance)
Amplification Deep Deletion

Diagnosis Age

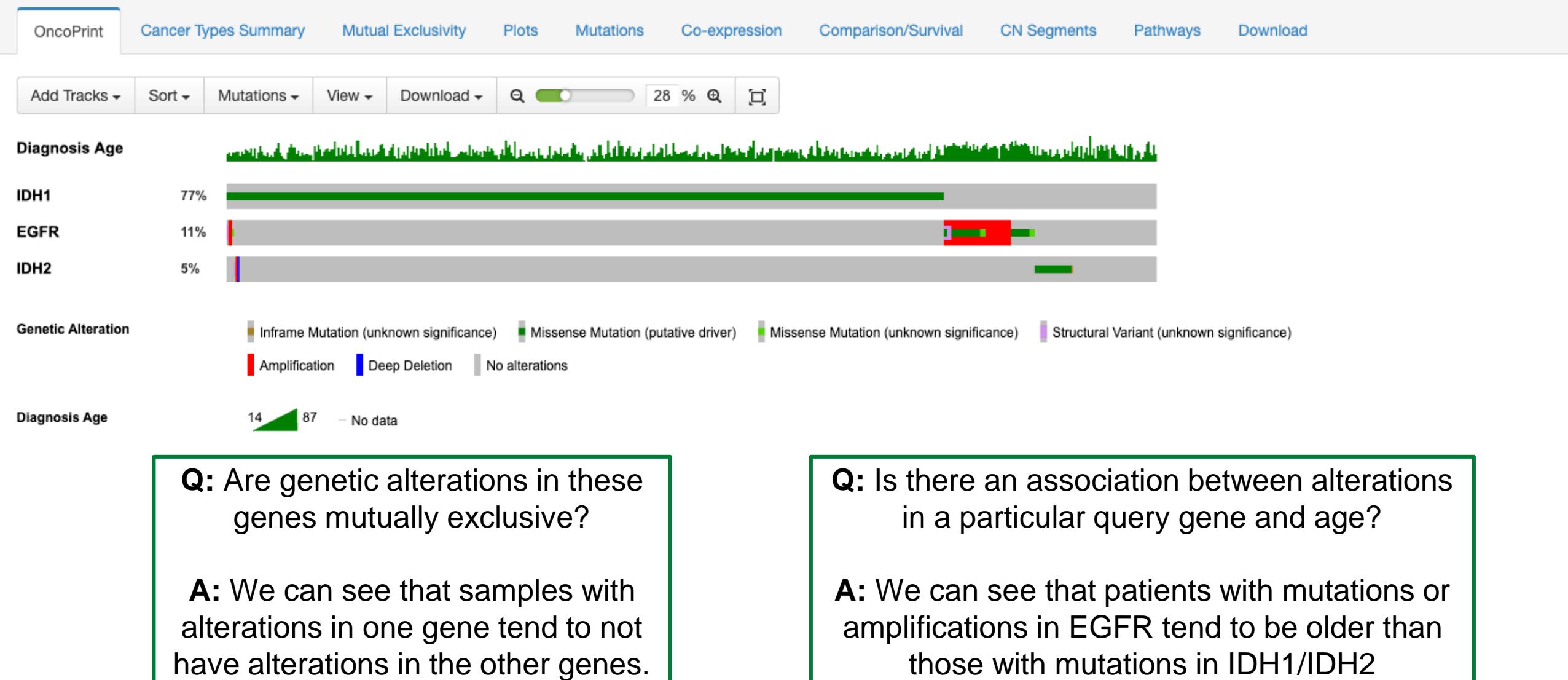
14 87 - No data

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



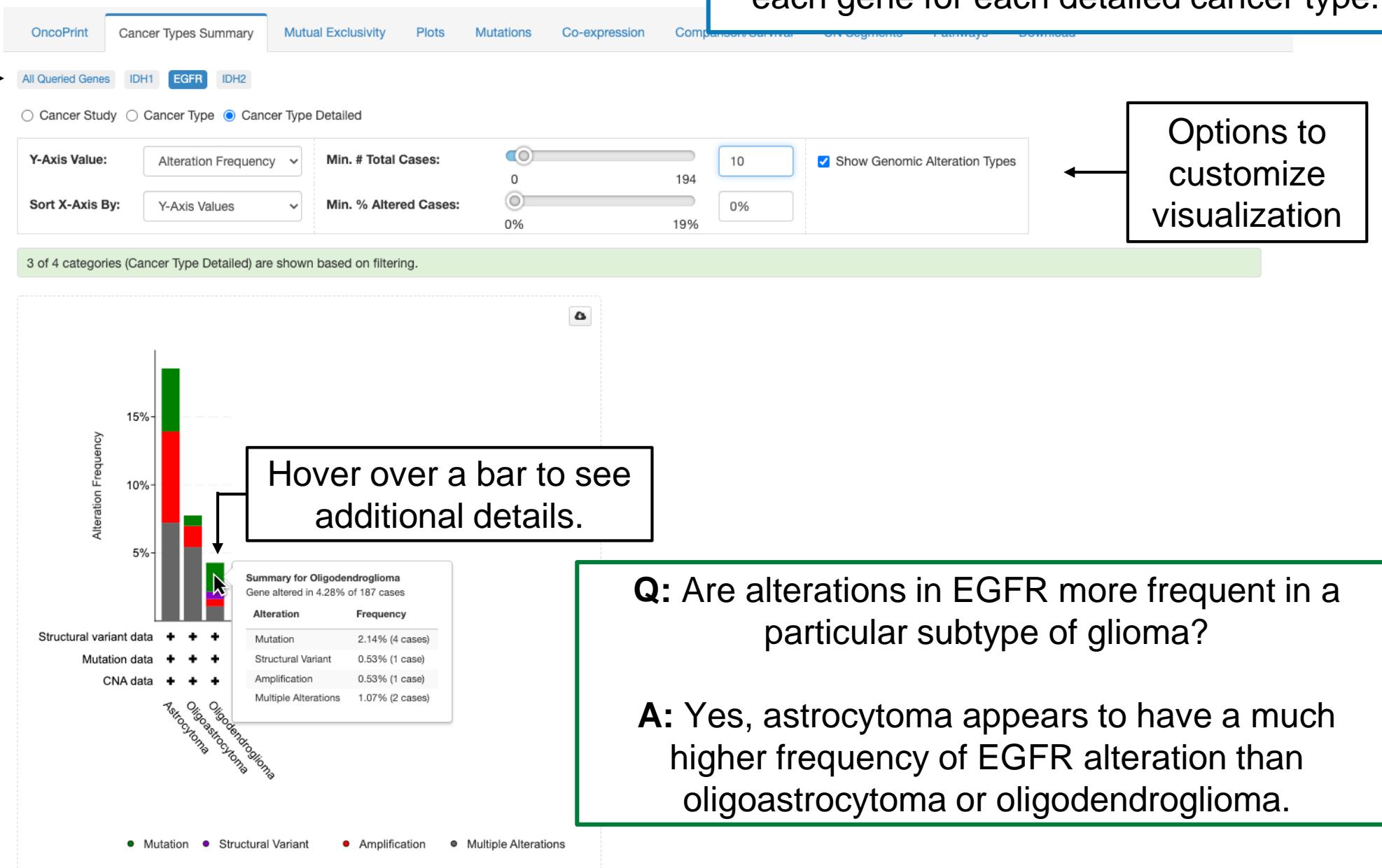
This button zooms OncoPrint to show all samples with alterations

OncoPrint: What can we learn?



Cancer Types Summary

Plots for all queried genes together and each individual gene are available as separate tabs.



[Link to this page](#)

Mutual Exclusivity

All pairwise combinations of query genes analyzed for mutual exclusivity or co-occurrence in the queried samples.

On the OncoPrint tab we could see visually that alterations in these three query genes tended to be mutually exclusive. Here we can address that same question with a statistical analysis.

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations CN Segr

The analysis tested 3 pairs between the 3 tracks in the OncoPrint.

Mutual exclusivity Co-occurrence Significant only

  Columns 

A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value	Tendency
IDH1	EGFR	67	390	50	4	<-3	<0.001	<0.001	Mutual exclusivity
IDH1	IDH2	96	391	21	3	<-3	<0.001	<0.001	Mutual exclusivity
EGFR	IDH2	433	54	24	0	<-3	0.064	0.064	Mutual exclusivity

Showing 1-3 of 3

A positive value here suggests that alterations in these genes co-occur in the same samples, while a negative value suggests that alterations in these genes are mutually exclusive and occur in different samples.

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

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

Plots

Example plot settings

Choose type of data

Select a query gene

Swap horizontal & vertical axis

OncoPrint Cancer Types Summary Mutual Exclusivity **Plots** Mutations

Examples: Mut# vs Dx FGA vs Dx Mut# vs FGA mRNA vs Dx mRNA vs mut type mRNA

Data Type: Copy Number **Horizontal Axis**

Copy Number Profile: Putative copy-number alteration...

Gene: EGFR

Filter categories: Select... Sort Categories by Median

↑ Swap Axes ↓

Data Type: mRNA **Vertical Axis**

mRNA Profile: mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeq_V2)

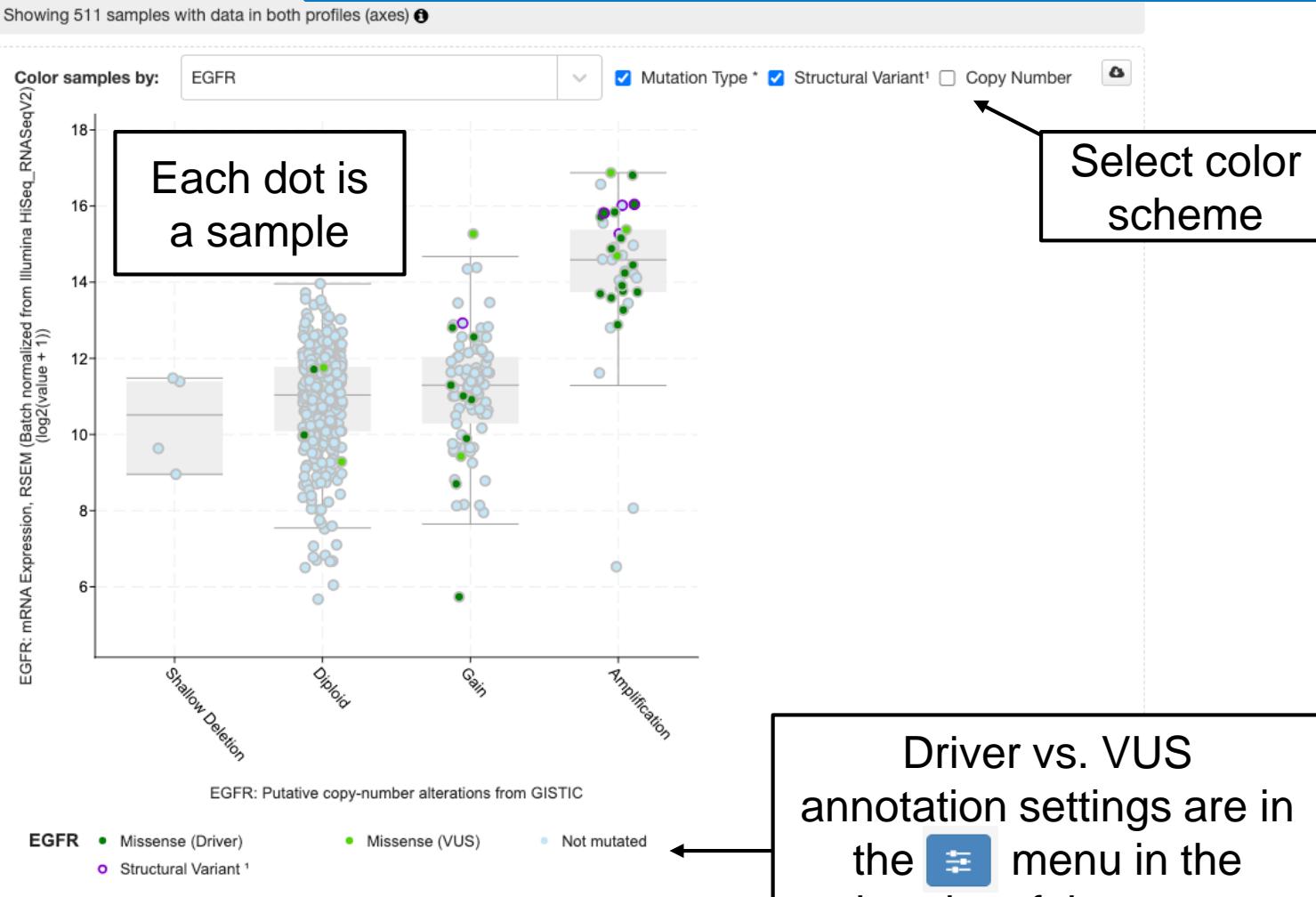
Log Scale:

Gene: Same gene (EGFR)

Search Case(s): Case ID..

Search Mutation(s): Protein Change..

Depending on available data types for a given study, this tab allows for plots comparing mutations, copy number, mRNA expression, protein levels and DNA methylation of query genes, along with any available clinical attributes.



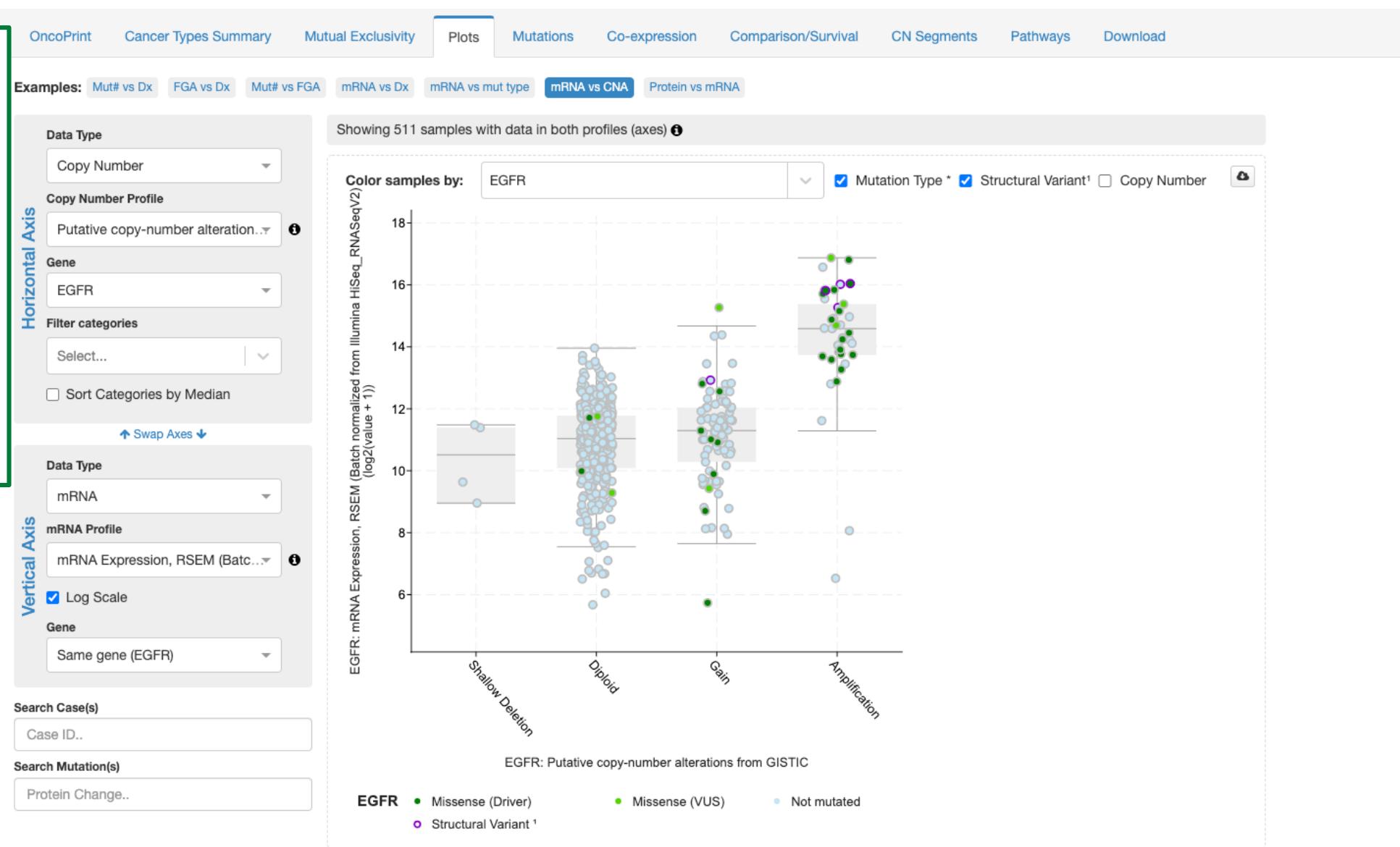
Select color scheme

Driver vs. VUS annotation settings are in the menu in the header of the page.

Plots

Q: Does amplification of EGFR alter gene expression?

A: Yes, we can see that higher copy number of EGFR (x-axis) is associated with increased expression (y-axis).



[Link to this page](#)

Mutations

This tab shows details about all mutations called in each query gene.

Each gene appears on a separate tab

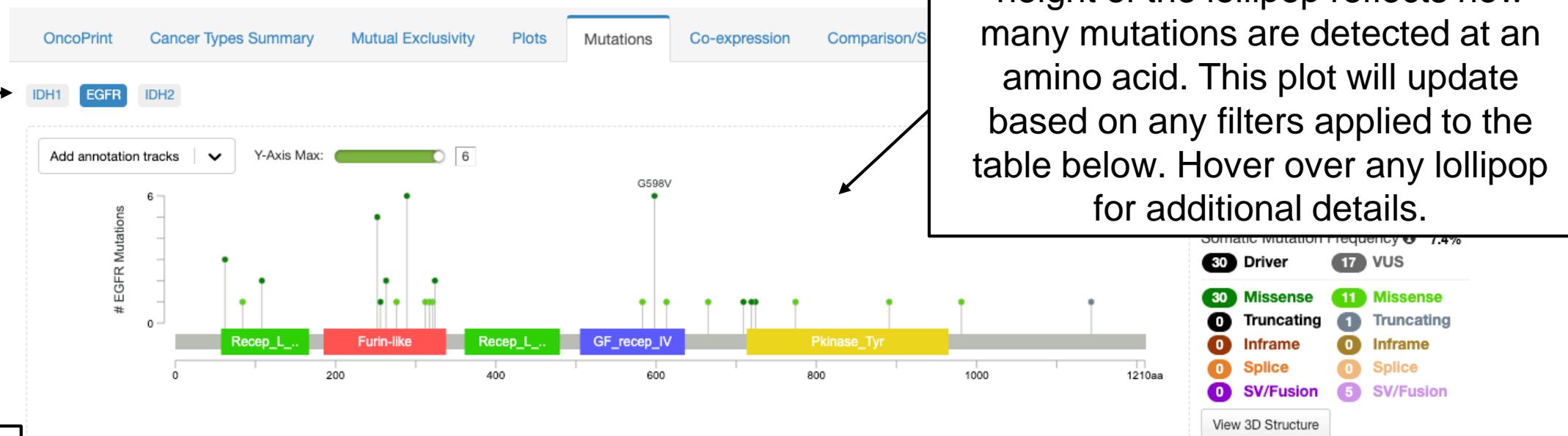


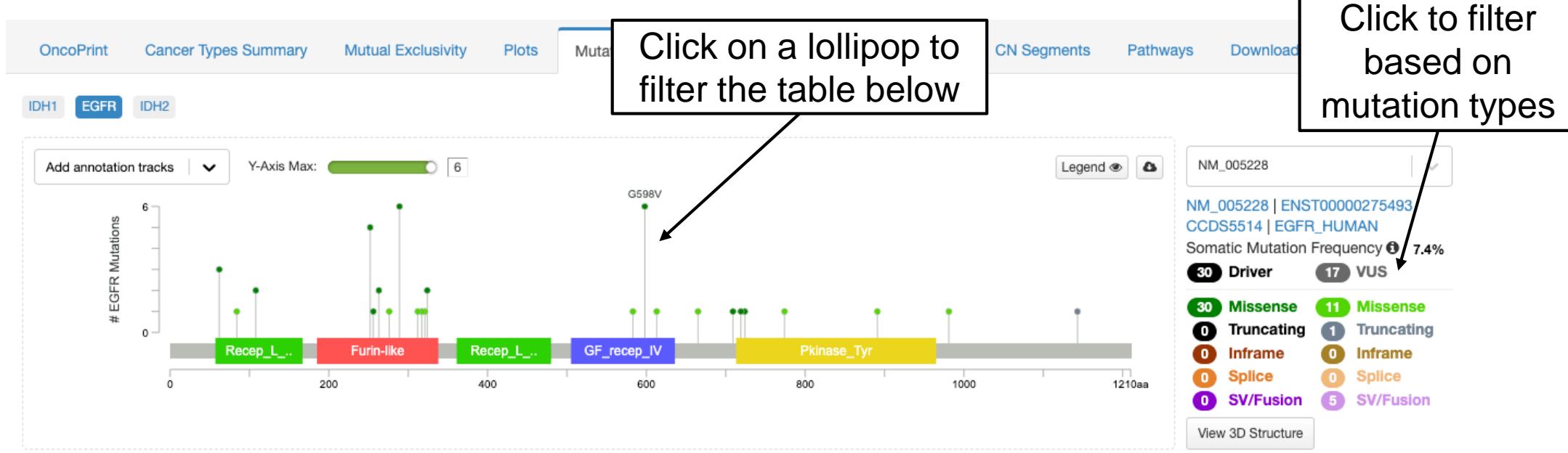
Table of all mutations with annotations

Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	G719D	● ○ ● ○ ○	Missense	Amp	125	0.18	39
TCGA-DU-A5TT...	Oligodendrogloma	A289V	● ○ ○ ○ ○	Missense	Gain			45
TCGA-E1-A7YM...	Astrocytoma	A289V	● ○ ○ ○ ○	Missense	Gain			35
TCGA-FG-6692-...	Oligodendrogloma	A289V	● ○ ○ ○ ○	Missense	Amp			63
TCGA-FG-A70Z...	Oligoastrocytoma	A289V	● ○ ○ ○ ○	Missense	Amp			47
TCGA-KT-A7W1...	Astrocytoma	A289V	● ○ ○ ○ ○	Missense	Amp			39
TCGA-TM-A7C3...	Astrocytoma	A289V	● ○ ○ ○ ○	Missense	Amp	50	0.71	56
TCGA-HT-8107-01	Oligodendrogloma	R108K	● ○ ○ ○ ○	Missense	Diploid	17	0.15	1
TCGA-HT-8110-01	Astrocytoma	R108K	● ○ ○ ○ ○	Missense	Amp	17	0.94	30
TCGA-HT-A61C...	Oligodendrogloma	T263P	● ○ ○ ○ ○	Missense	Gain	7	0.24	38
TCGA-DU-7013...	Astrocytoma	G598V	● ○ ○ ○ ○	Missense	Amp	36	0.96	36
TCGA-DU-8162...	Oligoastrocytoma	G598V	● ○ ○ ○ ○	Missense	Amp	36	0.57	22

Show additional columns

Link to this page

Mutations



47 Mutations (page 1 of 2)

Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	G719D	🕒 3B 🌸 🔥	Missense	Amp	125	0.18	39
TCGA-DU-A5TT...	Oligodendrogloma		🕒 3B 🌸 🔥	Missense	Gain	50	0.45	
TCGA-E1-A7YM...	Astrocytoma		🕒 3B 🌸 🔥	Missense	Gain	50	0.18	
TCGA-FG-6692-...	Oligodendrogloma		🕒 3B 🌸 🔥	Missense	Amp	50	0.95	
TCGA-FG-A70Z...	Oligoastrocytoma		🕒 3B 🌸 🔥	Missense	Amp	50	0.88	
TCGA-KT-A7W1...	Astrocytoma		🕒 3B 🌸 🔥	Missense	Amp	50	0.02	
TCGA-TM-A7C3...	Astrocytoma		🕒 3B 🌸 🔥	Missense	Amp	50	0.71	
TCGA-HT-8107-01	Oligodendrogloma	R108K	🕒 4 🌸 🔥	Missense	Diploid	17	0.15	1
TCGA-HT-8110-01	Astrocytoma	R108K	🕒 4 🌸 🔥	Missense	Amp	17	0.94	30
TCGA-HT-A61C...	Oligodendrogloma	T263P	🕒 4 🌸 🔥	Missense	Gain	7	0.24	38
TCGA-DU-7013...	Astrocytoma	G598V	🕒 4 🌸 🔥	Missense	Amp	36	0.96	36
TCGA-DU-8162...	Oligoastrocytoma	G598V	🕒 4 🌸 🔥	Missense	Amp	36	0.57	22

Click here (visible when you hover over a column) to filter on a specific column

Filter based on any visible text column

Mutations

View mutations in context of 3D protein structure

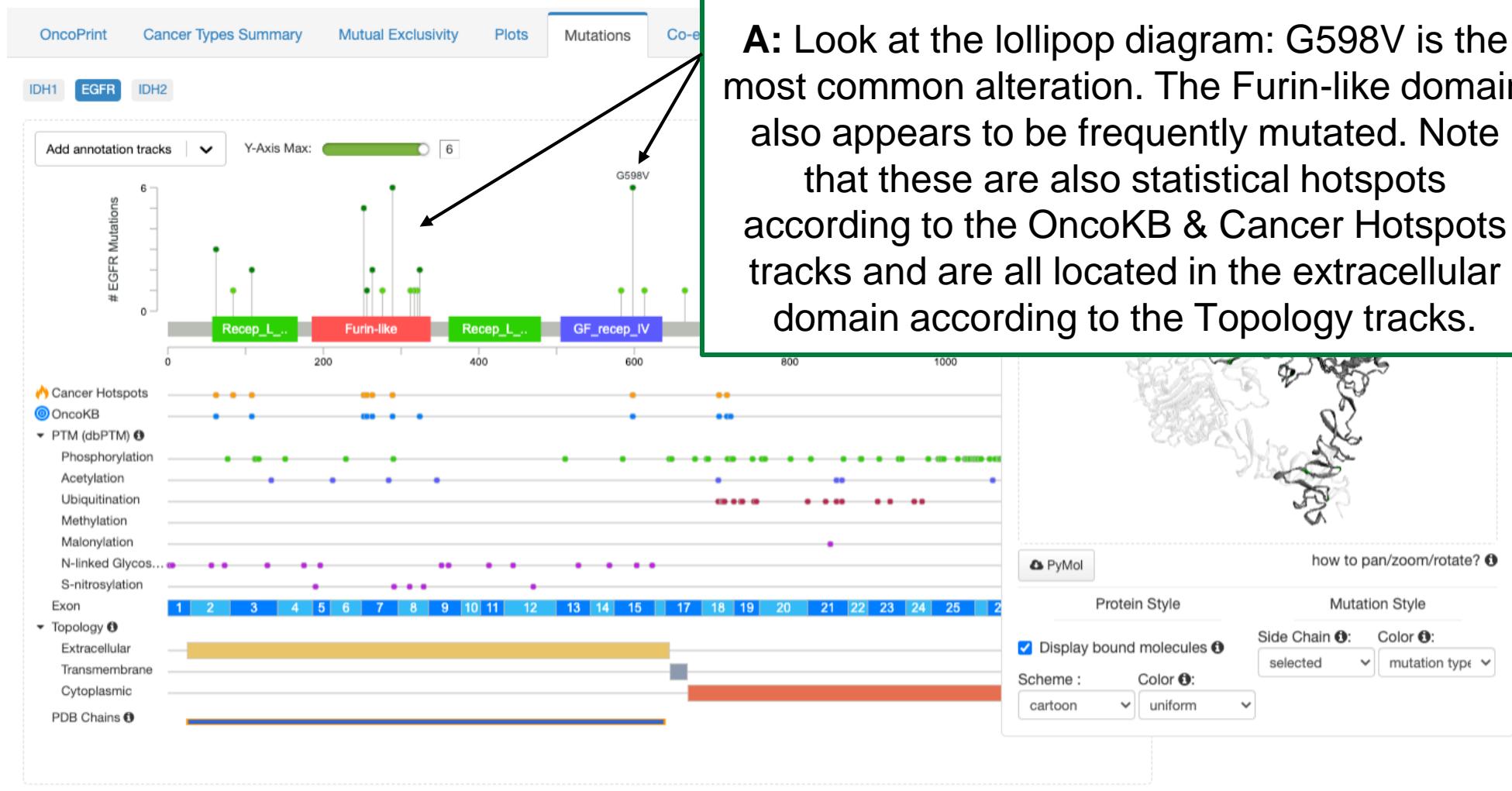


47 Mutations (page 1 of 2)

Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	G719D	● ● ● ● ● ● ● ●	Missense	Amp	125	0.18	39
TCGA-DU-A5TT...	Oligodendrogloma	A289V	● ● ● ● ● ● ● ●	Missense	Gain	50	0.45	45

Link to this page

Mutations

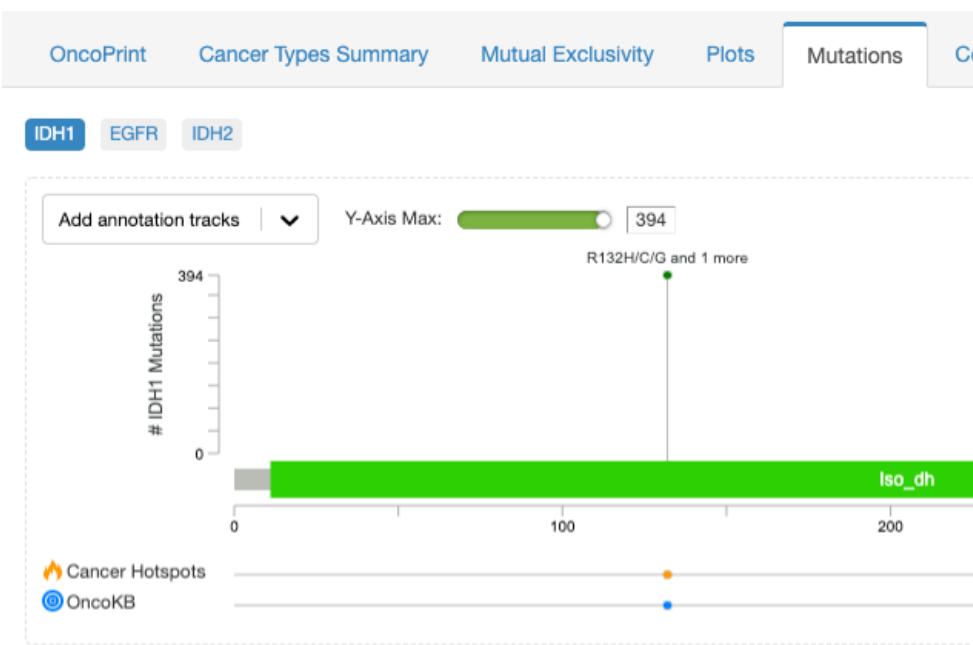


47 Mutations (page 1 of 2)

Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-KT-A7W1...	Astrocytoma	G719D	● ● ● ● ● ● ● ●	Missense	Amp	125	0.18	39
TCGA-DU-A5TT...	Oligodendrogloma	A289V	● ● ● ● ● ● ● ●	Missense	Gain	50	0.45	45

[Link to this page](#)

Mutations



3 This mutation is in [OncoKB](#) as a Level 3 variant. Hover over this symbol to see additional information, including that this is a known oncogenic mutation.

TCGA-DB-A4XF...	Astrocytoma	R132C
TCGA-DB-A64S...	Oligoastrocytoma	R132C
TCGA-DB-A75...		
TCGA-FG-8185...		
TCGA-HT-7479-0...		
TCGA-HT-7693-0...		
TCGA-HT-7855-0...		

This mutation is annotated in [CIViC](#). Hover over this symbol for additional information.

Q: The mutations in IDH1 appear to be highly recurrent. Are these mutations known hotspots? Known oncogenic drivers? Biomarkers for any drugs?

A: Look at the annotation tracks below the lollipop plot and the Annotation column in the table. Each mutation is annotated against 4 different databases with information about recurrence, oncogenicity and drugability.

Annotation	Mutation Type	Copy #	COSMIC	Allele	# Mut in
3A	Missense	Diploid	4964	0.40	29
3A	Missense	Diploid	4964	0.32	12
3A	Missense	Diploid	4964	0.20	25
3A	Missense	Diploid	4964	0.38	36
3A	Missense	Diploid	4964	0.25	22
3A	Missense	Diploid	4964	0.47	34
3A	Missense	Diploid	4964	0.47	36

Co-Expression

Compares mRNA/protein level expression of your query genes against all other genes.

Each gene appears on a separate tab

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

IDH1 EGFR IDH2

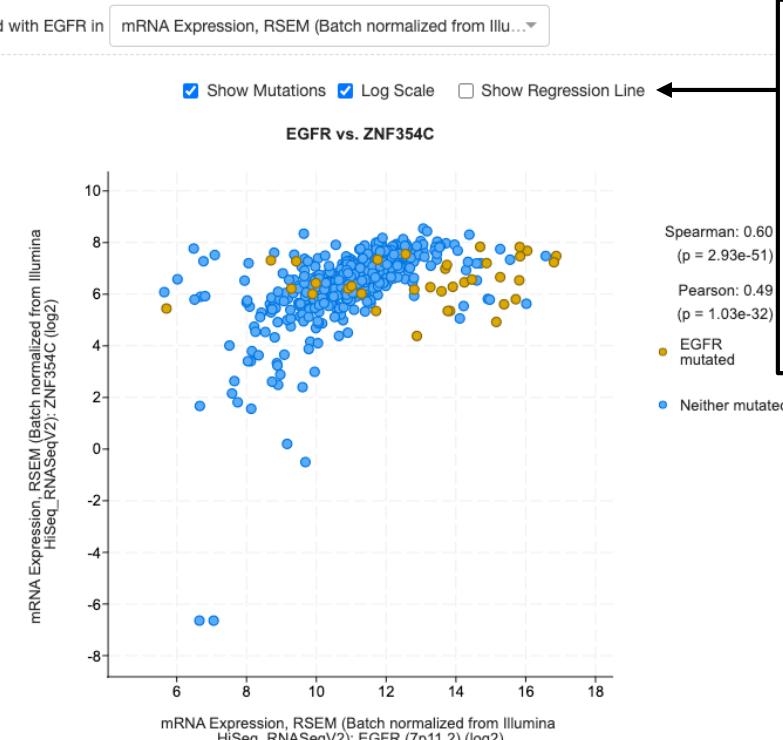
Select from available data types

Click on a gene name to see correlation plot

Correlated Gene	Cytoband	Spearman's Correlation	p-Value	q-Value
ZNF354C	5q35.3	0.600	2.93e-51	5.87e-47
ZKSCAN8	6p22.1	0.585	3.40e-48	2.34e-44
ZNF107	7q11.21	0.585	3.51e-48	2.34e-44
CHD9	16q12.2	0.582	9.65e-48	4.83e-44
ZNF426	19p13.2	0.582	1.33e-47	5.31e-44
SP4	7p15.3	0.580	2.43e-47	8.12e-44
TEAD1	11p15.3	0.580	3.00e-47	8.58e-44
KAT6A	8p11.21	0.574	4.60e-46	1.04e-42
SON	21q22.11	0.574	4.70e-46	1.04e-42
PYGO1	15q21.3	0.572	8.55e-46	1.71e-42
TRRAP	7q22.1	0.572	1.20e-45	2.19e-42
ZBTB20	3q13.31	0.570	2.17e-45	3.62e-42
RNASEK	17p13.1	-0.569	3.07e-45	4.66e-42
BAZ1B	7q11.23	0.569	3.26e-45	4.66e-42
ZNF699	19p13.2	0.569	3.83e-45	5.11e-42
MED13	17q23.2	0.568	6.64e-45	8.30e-42
KMT2C	7q36.1	0.566	1.08e-44	1.27e-41
DHX33	17p13.2	0.566	1.35e-44	1.50e-41
ZNF791	19p13.13	0.563	4.31e-44	4.54e-41
SMAD5	5q31.1	0.561	1.07e-43	1.07e-40
ECHS1	10q26.3	-0.560	1.33e-43	1.27e-40
RBL1	20q11.23	0.560	1.70e-43	1.55e-40
ZSCAN23	6p22.1	0.558	3.95e-43	3.35e-40
N4BP2	4p14	0.558	4.02e-43	3.35e-40
ZNF800	7q31.33	0.557	6.65e-43	5.32e-40

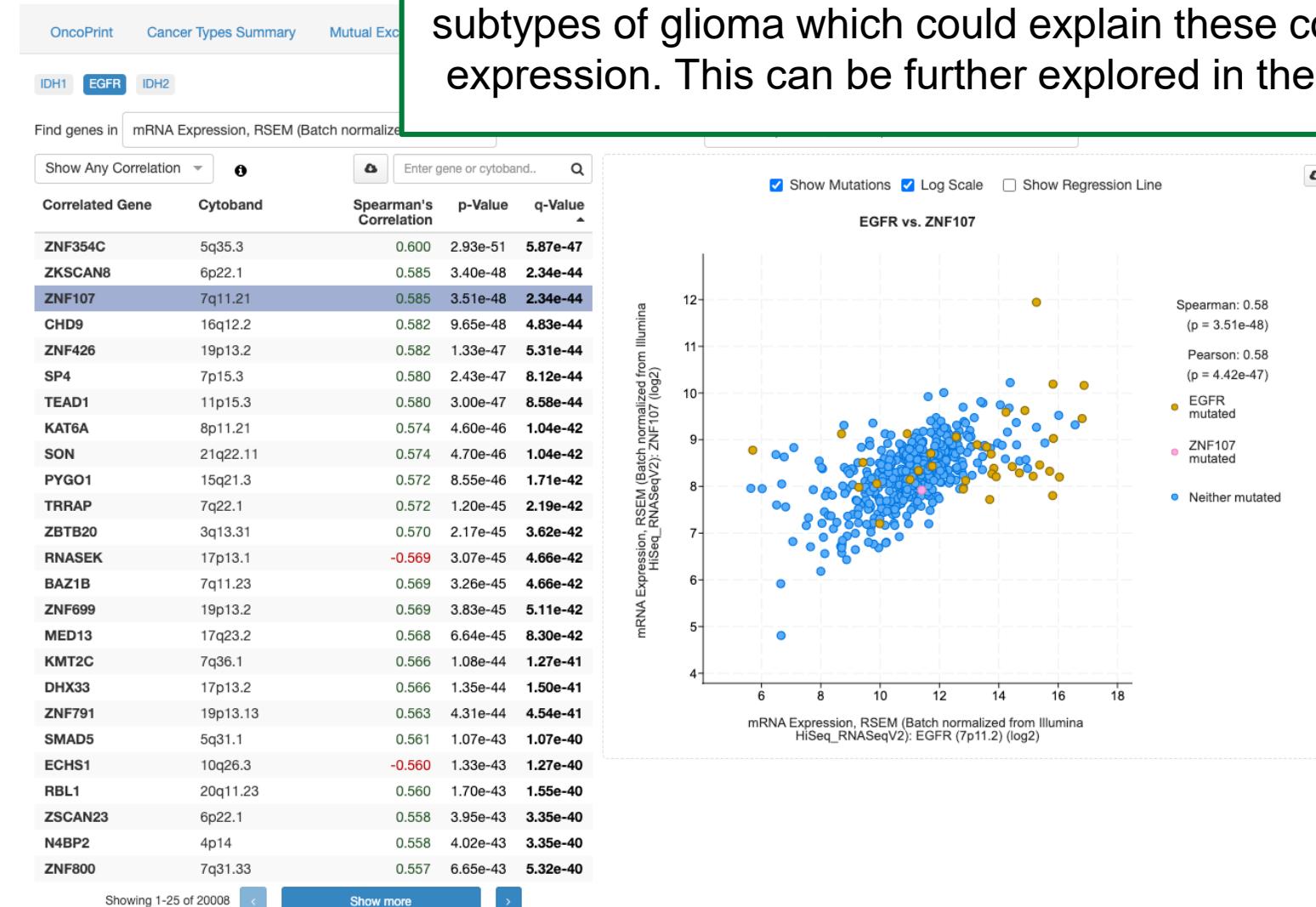
Showing 1-25 of 20008

Show more



Check boxes to color-code sample dots by mutation status, change x- or y-axis to log scale, or add a regression line.

Co-Expression



Comparison

This tab enables the comparison of all available data types between samples with or without alterations in the query genes. This tab replaces and enhances the old “Enrichments” tab.

The Comparison tab is the same as the Group Comparison functionality that is accessible from Study View. See the [Group Comparison Tutorial](#) for more details about the functionality of this tab.

The screenshot shows the OncoPrint interface with the 'Comparison' tab selected. At the top, there's a navigation bar with tabs: OncoPrint, Cancer Types Summary, Mutual Exclusivity, Plots, Mutations, Co-expression, Comparison/Survival (which is active), CN Segments, Pathways, and Download. Below the navigation bar, there's a section for 'Groups' with a note '(drag to reorder)'. It lists 'Altered group (465)' (highlighted in red), 'Unaltered group (46)', and other specific groups like 'IDH1 (394)', 'EGFR (54)', and 'IDH2 (24)'. There are also 'Select all' and 'Deselect all' buttons. Below this, there's a 'Groups' dropdown menu with options: Overlap, Survival, Clinical, Genomic Alterations (which is selected and highlighted in blue), mRNA, Protein, and Microbiome Signature. A callout box points to the 'Genomic Alterations' option with the text: 'By default, the “Altered” (one or more alterations in one or more query genes) and “Unaltered” (no alterations in any query gene) groups are selected.' Another callout box points to the 'Select all | Deselect all' buttons with the text: 'Groups can be toggled on or off by clicking on them them. Analyses will update as the selections change.' A third callout box points to the list of specific groups with the text: 'Additional groups (deselected by default) correspond to each track shown in OncoPrint.'

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

Groups: (drag to reorder)

Altered group (465) Unaltered group (46) IDH1 (394) EGFR (54) IDH2 (24) Select all | Deselect all

Overlap Survival Clinical Genomic Alterations mRNA Protein Microbiome Signature

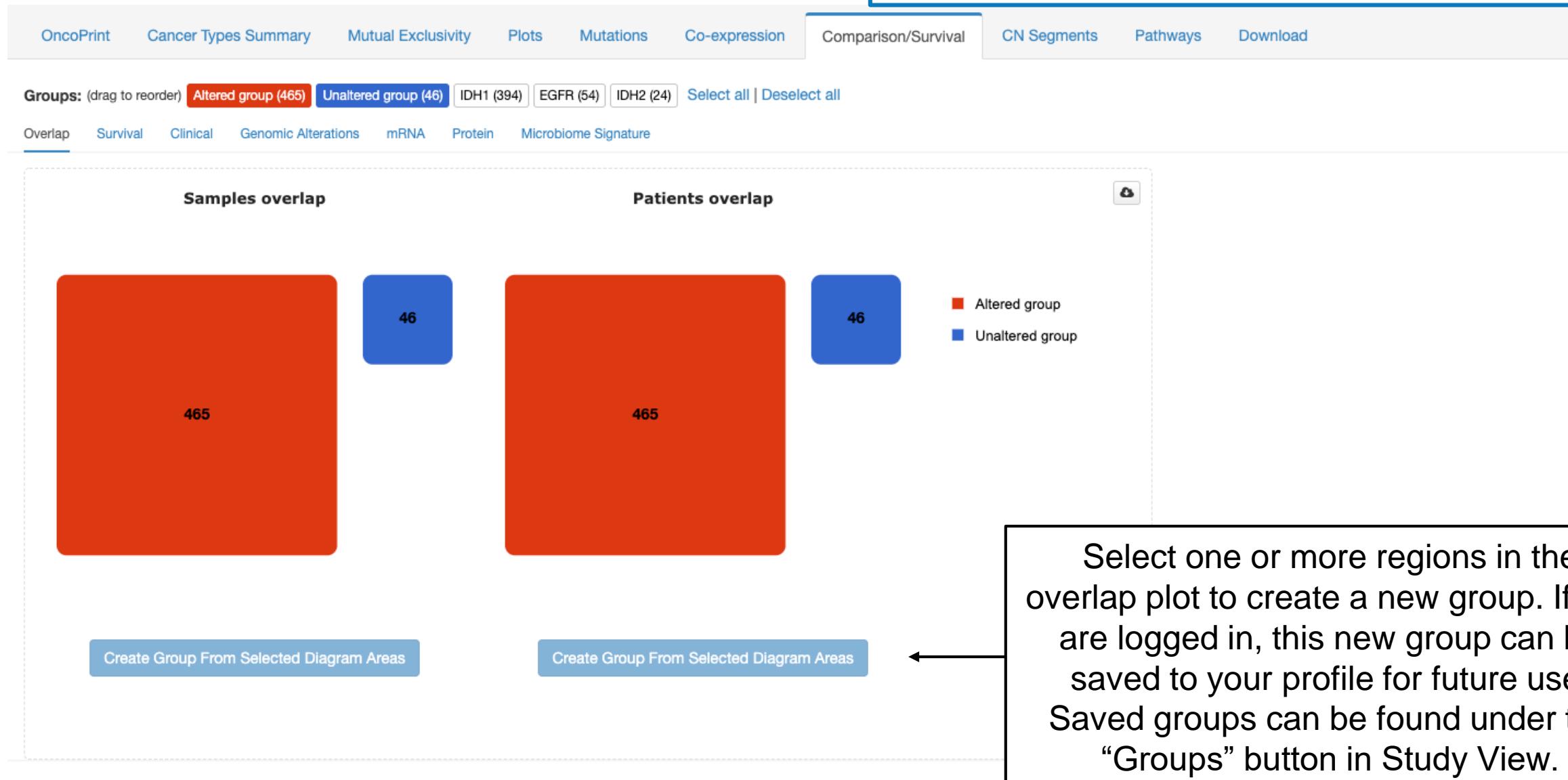
By default, the “Altered” (one or more alterations in one or more query genes) and “Unaltered” (no alterations in any query gene) groups are selected.

Groups can be toggled on or off by clicking on them them. Analyses will update as the selections change.

Additional groups (deselected by default) correspond to each track shown in OncoPrint.

Comparison: Overlap

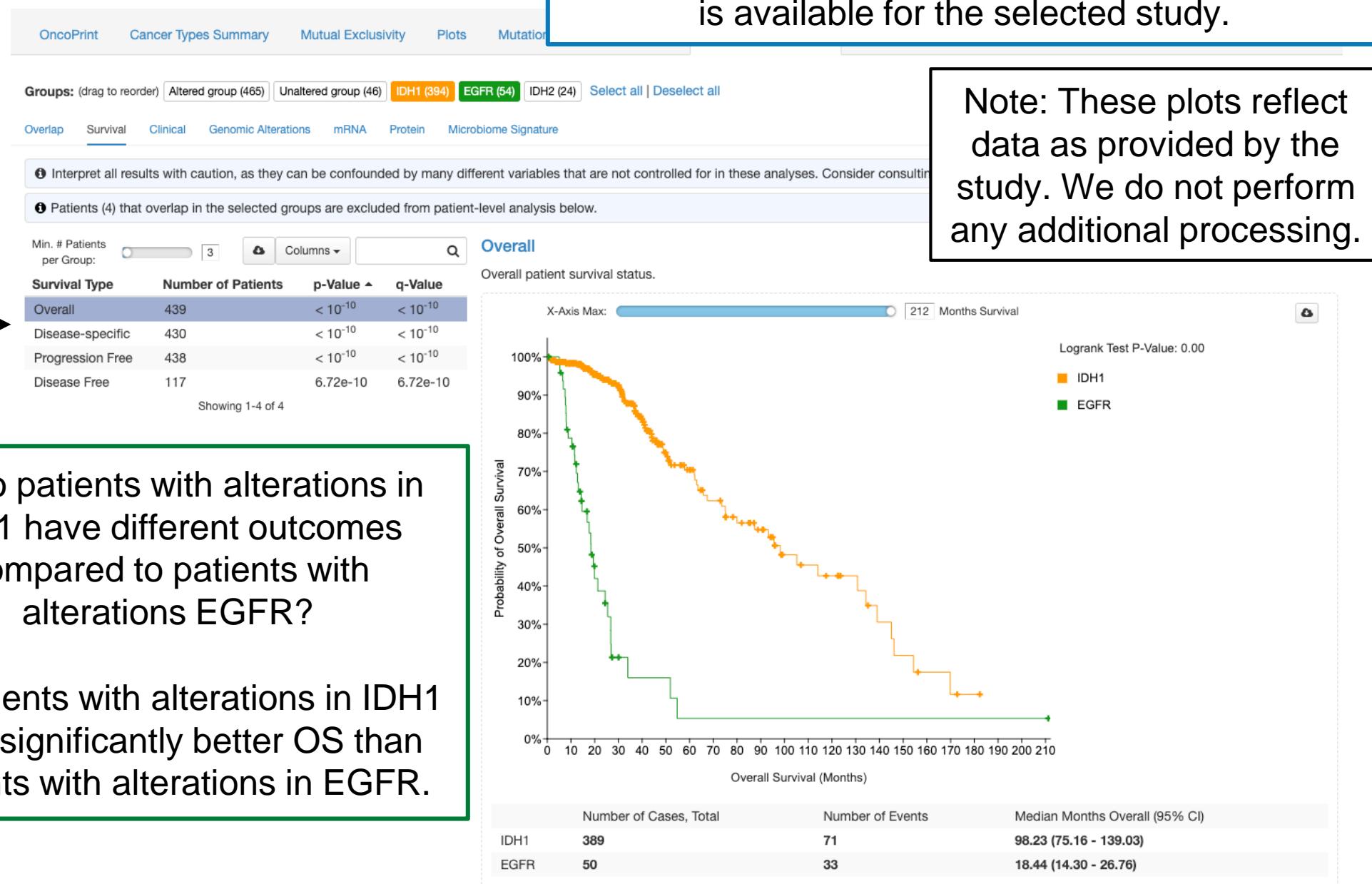
The Overlap subtab shows samples or patients that may overlap among the selected groups.



Comparison: Survival

The Survival subtab replaces the old “Survival” tab. This subtab will only be visible if outcome data is available for the selected study.

Select among different outcome measures. Options here depend on data availability for the study.



Q: Do patients with alterations in IDH1 have different outcomes compared to patients with alterations EGFR?

A: Patients with alterations in IDH1 have significantly better OS than patients with alterations in EGFR.

Comparison: Clinical

The Clinical subtab compares all available clinical data among the selected groups.

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

Groups: (drag to reorder) Altered group (465) Unaltered group (46) IDH1 (394) EGFR (54) IDH2 (24) Select all | Deselect all Exclude overlapping samples and patients

Overlap Survival Clinical Genomic Alterations mRNA Protein Microbiome Signature

ⓘ Interpret all results with caution, as they can be confounded by many different variables that are not controlled for in these analyses. Consider consulting a statistician.
 ⓘ Samples (4) that overlap in the selected groups are excluded from sample-level analysis below.
 ⓘ Patients (4) that overlap in the selected groups are excluded from patient-level analysis below.

Clinical Attribute Attribute Type Statistical Test p-Value q-Value ▲

Clinical Attribute	Attribute Type	Statistical Test	p-Value	q-Value
Subtype	Patient	Chi-squared Test	< 10 ⁻¹⁰	< 10 ⁻¹⁰
Aneuploidy Score	Sample	Wilcoxon Test	< 10 ⁻¹⁰	< 10 ⁻¹⁰
Diagnosis Age	Patient	Wilcoxon Test	< 10 ⁻¹⁰	< 10 ⁻¹⁰
Birth from Initial Pathologic Diagnosis Date	Patient	Wilcoxon Test	< 10 ⁻¹⁰	< 10 ⁻¹⁰
Fraction Genome Altered	Sample	Wilcoxon Test	< 10 ⁻¹⁰	< 10 ⁻¹⁰
Mutation Count	Sample	Wilcoxon Test	< 10 ⁻¹⁰	2.05e-10
TMB (nonsynonymous)	Sample	Wilcoxon Test	< 10 ⁻¹⁰	2.08e-10
International Classification of Diseases for Oncology, Third Edition ICD-O-3 Histology Code	Patient	Chi-squared Test	4.64e-10	1.82e-9
Neoplasm Histologic Grade	Sample	Chi-squared Test	2.09e-9	7.49e-9
Ragnum Hypoxia Score	Patient	Wilcoxon Test	2.71e-8	8.98e-8
Buffa Hypoxia Score	Patient	Wilcoxon Test	1.559e-6	4.790e-6
Radiation Therapy	Patient	Chi-squared Test	3.100e-5	8.330e-5
New Neoplasm Event Post Initial Therapy Indicator	Patient	Chi-squared Test	6.512e-5	1.474e-4
Oncotree Code	Sample	Chi-squared Test	8.642e-5	1.689e-4

Plot Type: 100% stacked bar chart Swap Axes Horizontal Bars

Subtype samples (%)

Group

LGG_IDHmut-codel
LGG_IDHmut-non-codel
LGG_IDHwt

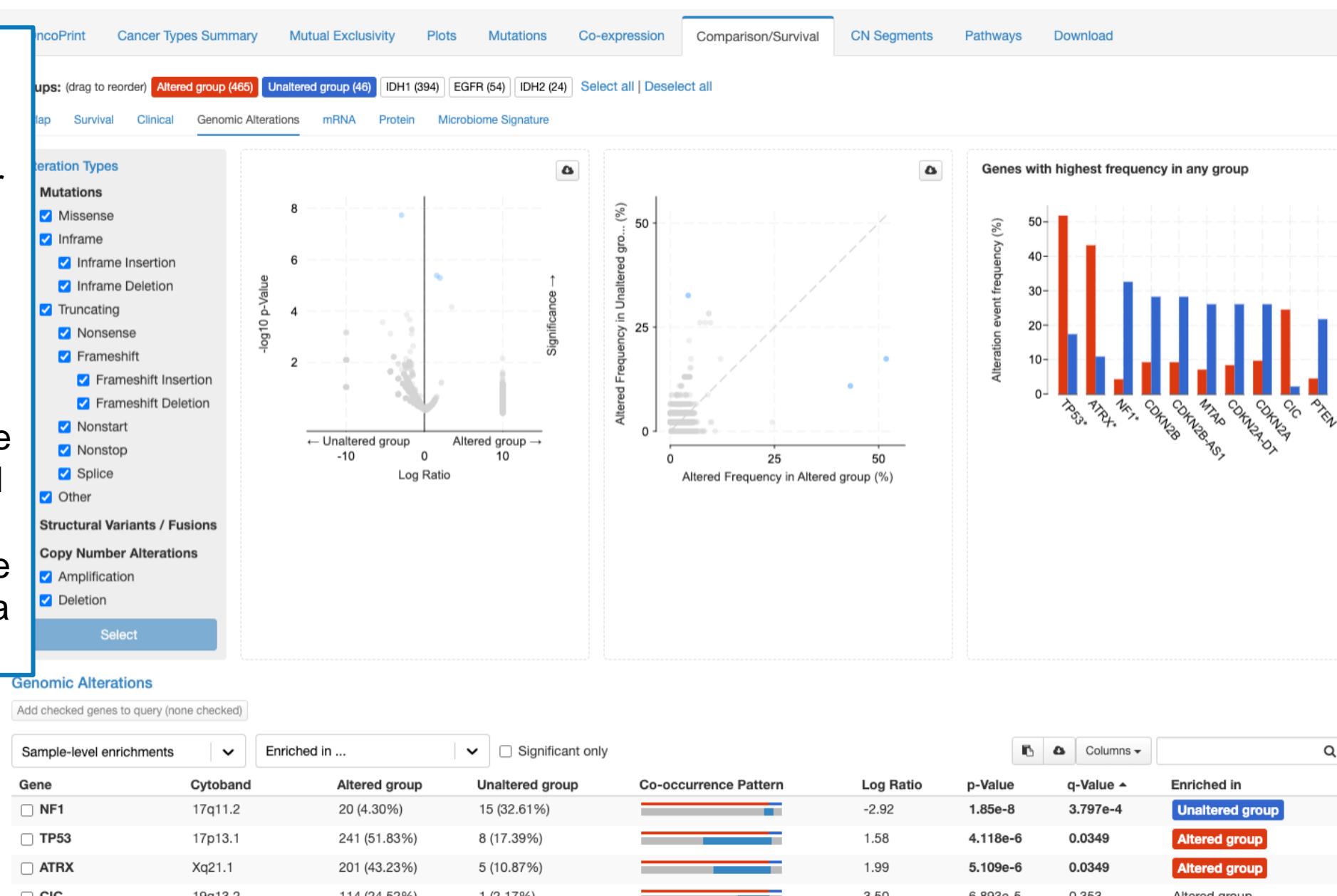
ⓘ Click on a clinical attribute to visualize the data in the plot on the right.

Link to this page

Comparison: Molecular Profiles

The molecular profiles subtabs replace the old “Enrichments” tab.

These analyses ask whether Genomic Alterations (mutations/copy-number alterations) or mRNA expression or protein expression in a particular gene is enriched in one of the selected groups. These, and additional subtabs like Microbiome Signature, will be visible depending on the data available for each study.



[Link to this page](#)

Comparison: Molecular Profiles

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

Groups: (drag to reorder) Altered group (465) Unaltered group (46) IDH1 (394) EGFR (54) IDH2 (24) Select all | Deselect all

Overlap Survival Clinical Genomic Alterations mRNA Protein Microbiome Signature

Select which types of alterations to include in the analysis

Select sample-level or patient-level analysis

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

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

Hover over a dot to see the gene name

Alteration Types

- Mutations
 - Missense
 - Inframe
 - Inframe Insertion
 - Inframe Deletion
 - Truncating
 - Nonsense
 - Frameshift
 - Frameshift Insertion
 - Frameshift Deletion
 - Nonstart
 - Nonstop
 - Splice
 - Other
- Structural Variants / Fusions
- Copy Number Alterations
 - Amplification
 - Deletion

Select

Genomic Alterations

Add checked genes to query (none checked)

Sample-level enrichments Enriched in ... Significant only

Gene	Cytoband	Altered group	Unaltered group	Co-occurrence Pattern	Log Ratio	p-Value	q-Value	Enriched in
<input type="checkbox"/> NF1	17q11.2	20 (4.30%)	15 (32.61%)		-2.92	1.85e-8	3.797e-4	Unaltered group
<input type="checkbox"/> TP53	17p13.1	241 (51.83%)	8 (17.39%)		1.58	4.118e-6	0.0349	Altered group
<input type="checkbox"/> ATRX	Xq21.1	201 (43.23%)	5 (10.87%)		1.99	5.109e-6	0.0349	Altered group
<input type="checkbox"/> CIC	19q13.2	114 (24.52%)	1 (2.17%)		3.50	6.893e-5	0.353	Altered group

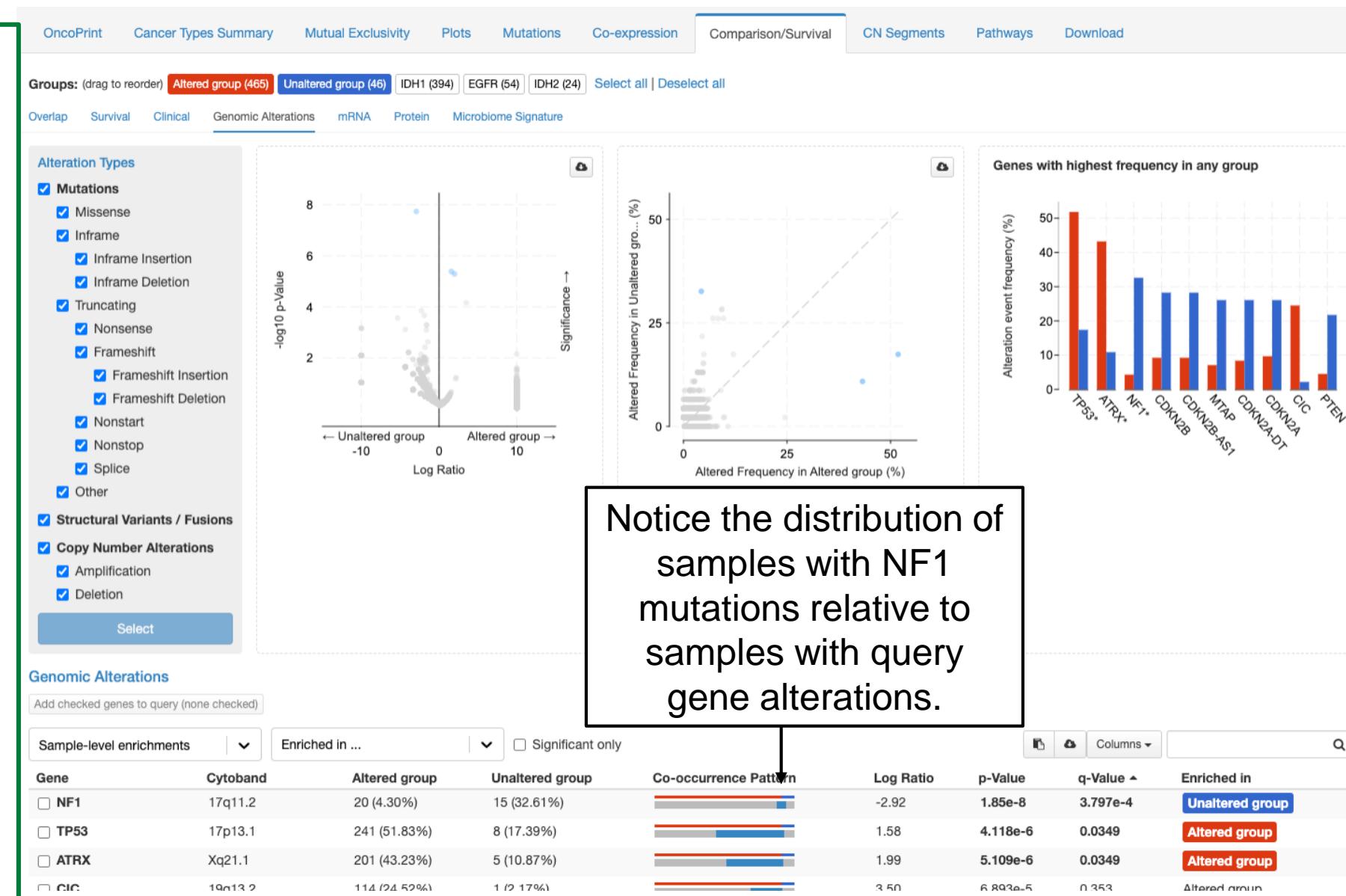
est frequency in any group

TP53*, ATRX*, NF1*, CDKN2B, CDKN2B-AS1, MTAP, CDKN2A-DT, CDKN2A, CIC, PTEN

Comparison

Q: Alterations in IDH1, IDH2 and EGFR are mutually exclusive but some samples have alterations in none of these genes. Do samples without IDH1, IDH2 or EGFR alterations commonly have genomic alterations in one or more other genes?

A: Alterations in NF1 are significantly mutually exclusive with alterations in IDH1, IDH2 and EGFR (see table). Try adding NF1 to the query (check the box next to NF1 and then click “Add checked genes to query”) and examine the OncoPrint and the Mutual Exclusivity tabs.



CN Segments

View copy number for each sample at each query gene via the [Integrated Genomics Viewer](#) (IGV).



Plots for each gene appear on a separate tab.

Toggle track labels, a vertical line marking the center of the viewing screen, and a vertical line that moves with your cursor. Use to zoom in or out.

Click for track settings, including expanding the height of each sample (see below)

Each row is a single sample

Gene structures

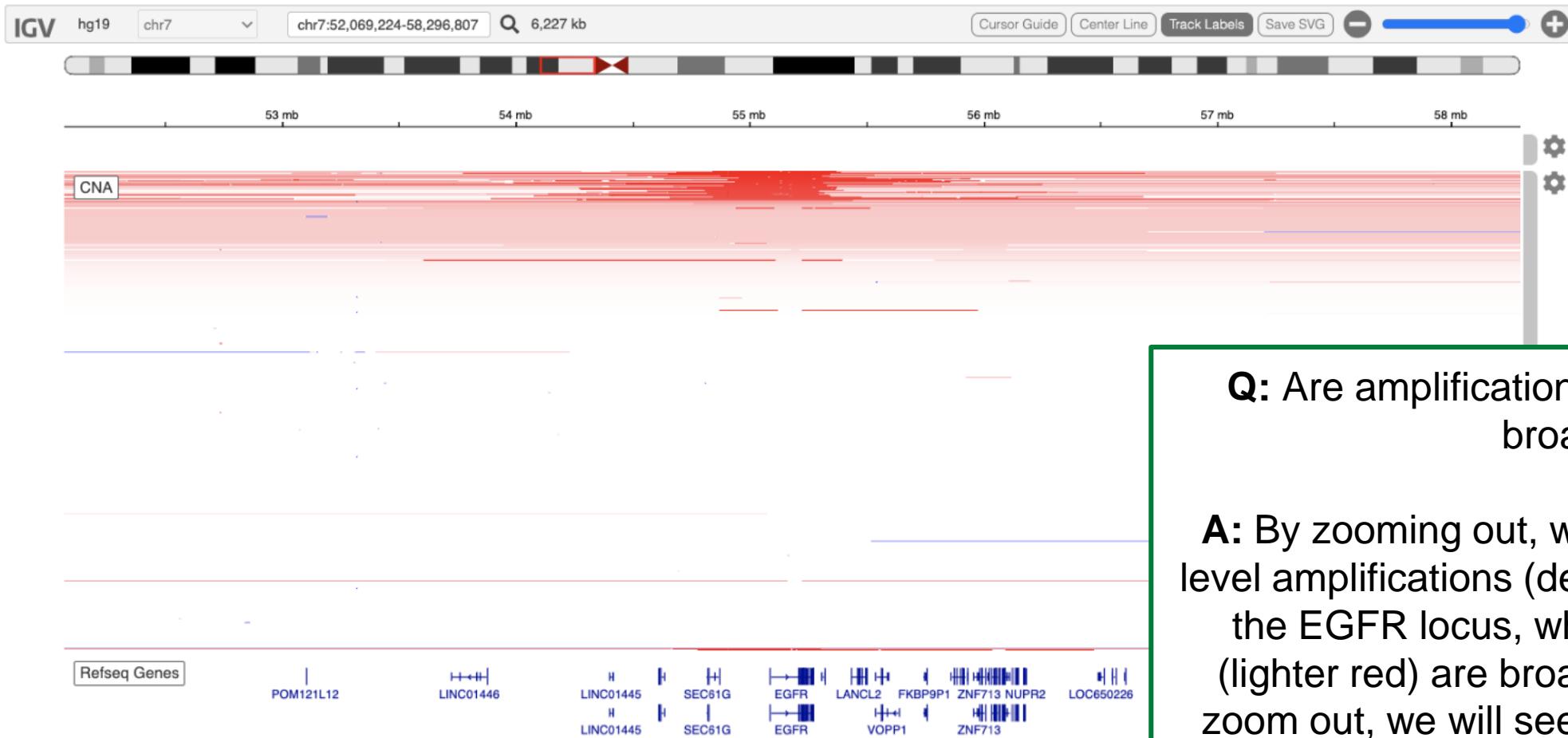


Click on a read for details

CN Segments

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

Whole Genome IDH1 EGFR IDH2

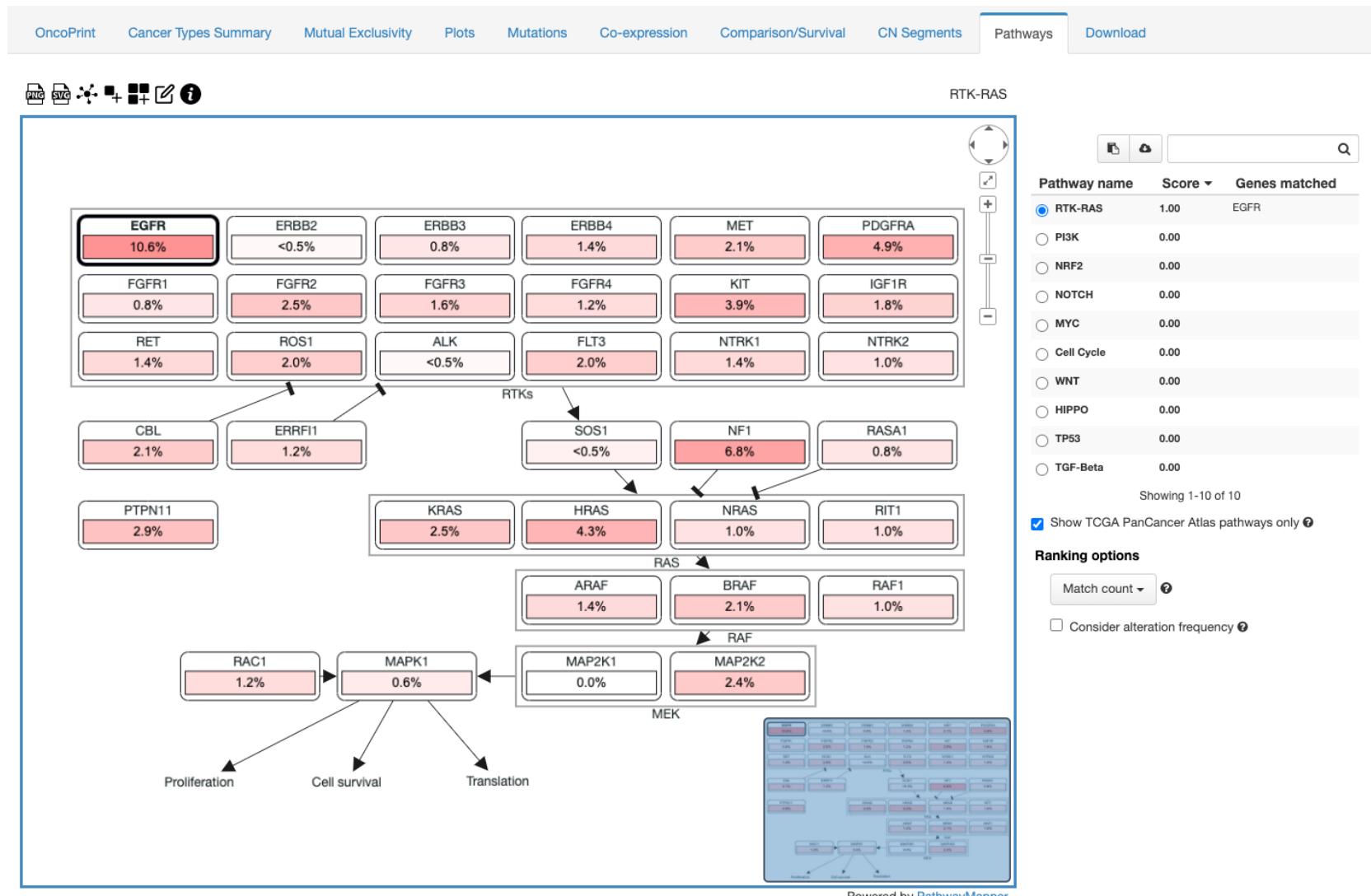


Q: Are amplifications of EGFR focal or broad?

A: By zooming out, we can see that high-level amplifications (deeper red) are focal at the EGFR locus, while low-level gains (lighter red) are broad. If we continue to zoom out, we will see that low-level gains often encompass the entire chromosome.

Pathways

The Pathways tab replaces the now retired “Network” tab. This tab is an integration with [PathwayMapper](#). The tab enables exploration of the queried genes in the context of Pathways defined by TCGA. For more detail on this tab, refer to the [Pathways Tutorial](#).



Link to this page

Pathways Tab in Results View

Pathways tab

TCGA pathways table, sorted by score using current ranking scheme

Toolbar for pathway operations

Pathway with alteration frequencies of selected genetic profiles of the chosen study overlaid

Ranking options

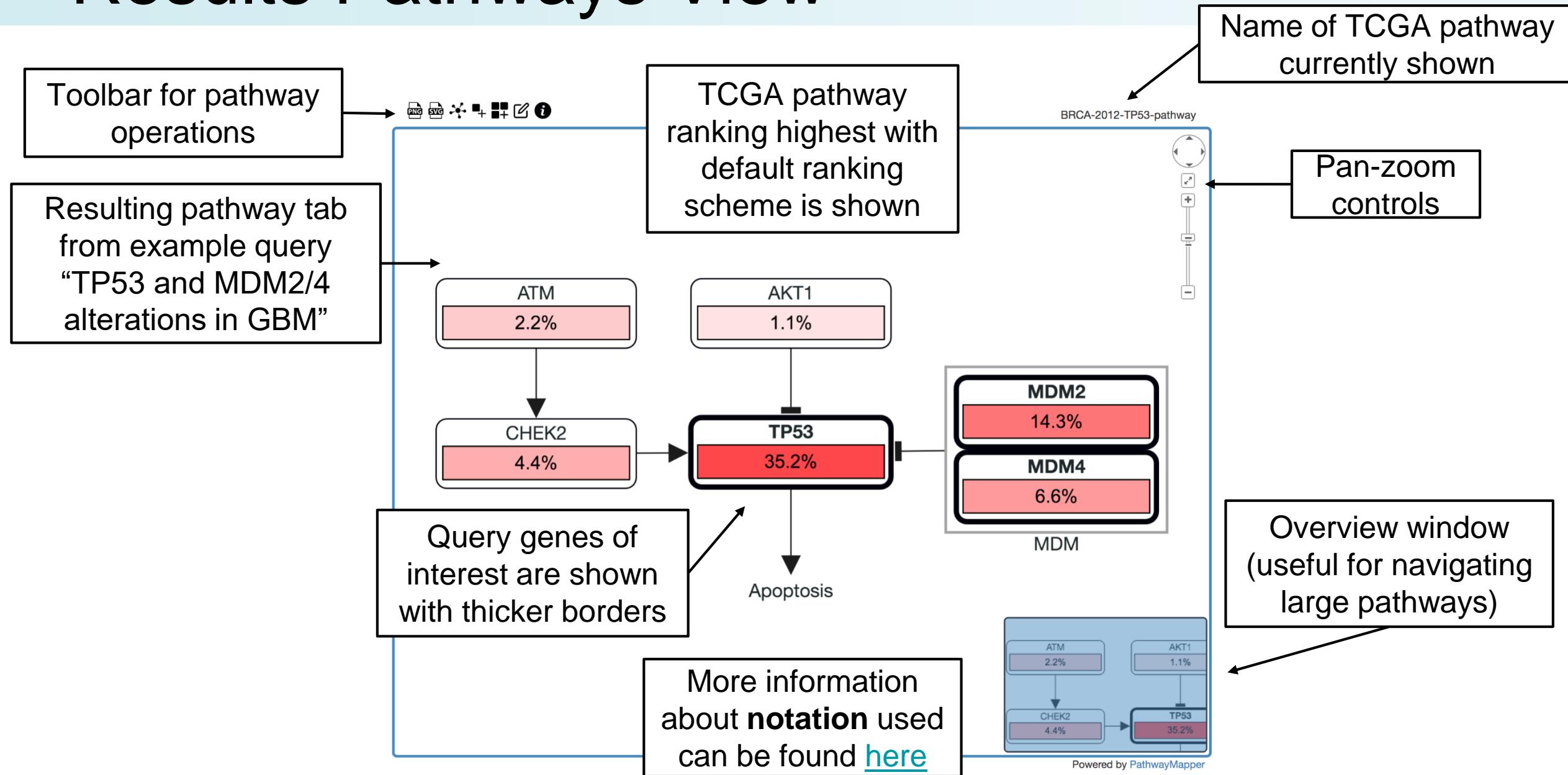
Toggle between pathways defined by [TCGA PanCancer Atlas](#) (default), or all TCGA publications

Link to this page

The screenshot shows the 'Pathways' tab in the Results View interface. At the top, there's a toolbar with various icons for saving and sharing. Below it is a navigation bar with tabs: OncoPrint, Cancer Types Summary, Mutual Exclusivity, Plots, Mutations, Co-expression, Comparison/Survival, CN Segments, Pathways (which is highlighted in blue), and Download. The main content area displays a pathway network diagram. In the center, a large red box labeled 'TP53' has a '35.2%' alteration frequency. Arrows point from other genes: 'ATM' (2.2%) to 'CHEK2' (4.4%), 'AKT1' (1.1%) to 'TP53', and 'MDM2' (14.3%) and 'MDM4' (6.6%) to 'TP53'. An arrow also points from 'TP53' down to a box labeled 'Apoptosis'. To the right of the diagram is a 'Ranking options' section with dropdown menus for 'Match count' and 'Consider alteration frequency'. Further right is a 'TCGA pathways table' showing 12 results, with one row for 'GBM-2013-TP53-pathway' highlighted in blue. At the bottom right, there's a note about toggling between TCGA PanCancer Atlas and all TCGA publications. A small 'Powered by' logo is at the very bottom.

Pathway name	Score	Genes matched
BRCA-2012-TP53-pathway	3.00	TP53 MDM2 MDM4
TP53	3.00	TP53 MDM2 MDM4
GBM-2008-TP53-pathway	3.00	TP53 MDM2 MDM4
GBM-2013-TP53-pathway	3.00	TP53 MDM2 MDM4
SKCM-2015-TP53-pathway	2.00	TP53 MDM2
ACC-2016-TP53-RB-pathway	2.00	TP53 MDM2
BLCA-2014-TP53-RB-pathway	2.00	TP53 MDM2
Cell Cycle	2.00	TP53 MDM2
LUAD-2014-TP53-pathway	2.00	TP53 MDM2
COADREAD-2012-TP53-pathway	1.00	TP53

Results Pathways View



Results Pathways View Toolbar

Buttons on the toolbar provide useful operations



Save current pathway as PNG



Save current pathway as SVG



Perform incremental layout, respecting current positions



Add selected genes to query



Add all valid genes in this pathway to query



Edit pathway with [PathwayMapper](#) editor



Quick help with a link to [detailed documentation](#)

Results Pathways Table & Ranking Options

The screenshot shows a user interface for analyzing TCGA pathways. On the left, a sidebar displays a selected pathway (BRCA-2012-TP53-pathway) and allows switching between TCGA PanCancer Atlas and all TCGA publications. Below this are two sections: one about alteration frequency (MDM2 at 14.3%) and one about ranking options (Match count vs percentage).

Score of each TCGA pathway using current ranking scheme

Pathway name	Score	Genes matched
BRCA-2012-TP53-pathw...	3.00	TP53 MDM2 MDM4
TP53	3.00	TP53 MDM2 MDM4
GBM-2008-TP53-pathwa...	3.00	TP53 MDM2 MDM4
GBM-2013-TP53-pathwa...	2.00	TP53 MDM2 MDM4
SKCM-2015-TP53-pathw...	2.00	TP53 MDM2
ACC-2016-TP53-RB-pat...	2.00	TP53 MDM2
BLCA-2014-TP53-RB-pa...	2.00	TP53 MDM2
Cell Cycle	2.00	TP53 MDM2
LUAD-2014-TP53-pathw...	2.00	TP53 MDM2
COADREAD-2012-TP53-p...	1.00	TP53

Search pathway by name

Genes in current pathway matching those in query genes

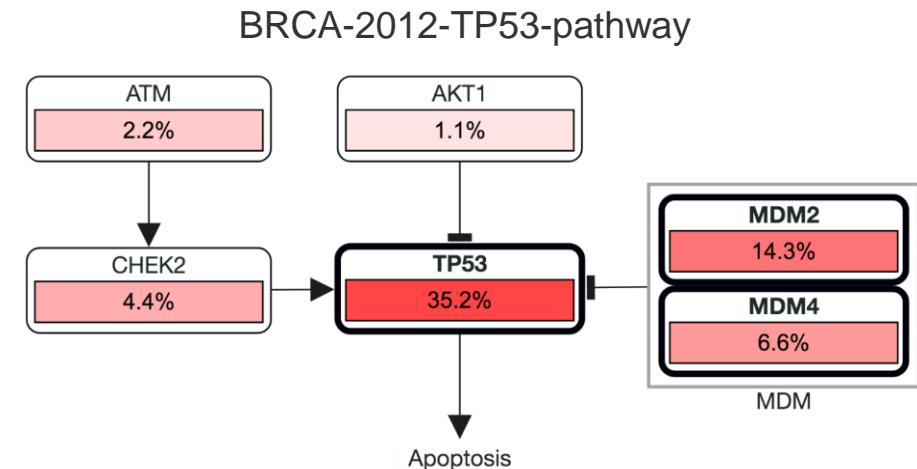
Match count vs percentage:
whether we should score pathways by the number of genes matched or by the percentage of genes matched

Consider alteration frequency:
whether we should take each matching gene with a count of 1 or with a weight of its alteration frequency in scoring

Link to this page

Results Pathways View Ranking Options

- When a query gene is in a particular pathway, we consider it “matching”.
- Example:
 - Query genes: TP53, MDM2, MDM4
 - Pathway: BRCA-2012-TP53-pathway (see on the right)
- Match count vs percentage:
 - Count the query genes matching and rank pathways based on this count. The score in our example is **3** as all three genes are in the pathway.
 - Take the ratio of query genes matching to total number of genes in the pathway. The score in our example is $3 / 6 = \mathbf{50\%}$.
- Consider alteration frequency:
 - When checked, each matching gene will not contribute to the score as 1 unit but with its alteration frequency of that gene. The score in our example is $35.2 + 14.3 + 6.6 = \mathbf{56.1}$.



[Modify Query](#)

Brain Lower Grade Glioma (TCGA, PanCancer Atlas)

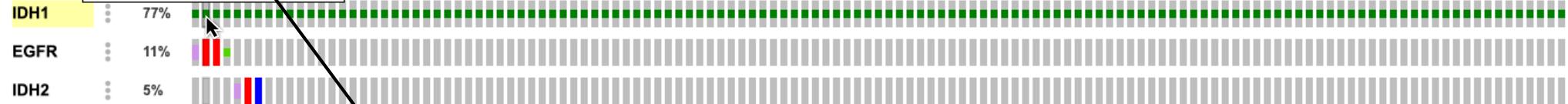
Samples with mutation and CNA data (511 patients/samples) - IDH1, EGFR & IDH2

Queried genes are altered in 465 (91%) of queried patients/samples

[OncoPrint](#)[Cancer Types Summary](#)[Mutual Exclusivity](#)[Plots](#)[Mutations](#)[Co-expression](#)[Comparison/Survival](#)[CN Segments](#)[Pathways](#)[Download](#)

TCGA-CS-5393
Mutation: IDH1 R132H 🔥 ⓘ
Profiled in all selected molecular profiles.

View ⏪ Download ⏪ 🔎 100 % 🔍



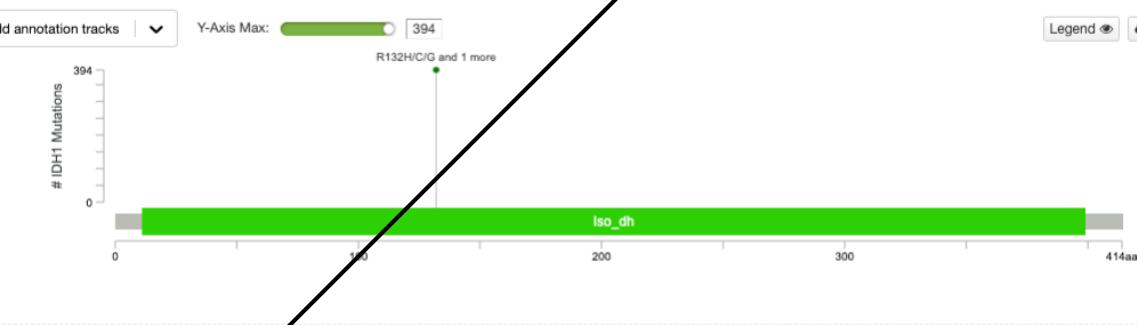
Genetic Alteration

Inframe Mutation (unknown significance) Missense Mutation (putative driver) Missense Mutation (unknown significance) Structural Variant (unknown significance)
Amplification Deep Deletion No alterations

Click on any of these sample/patient IDs

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression

IDH1 EGFR IDH2

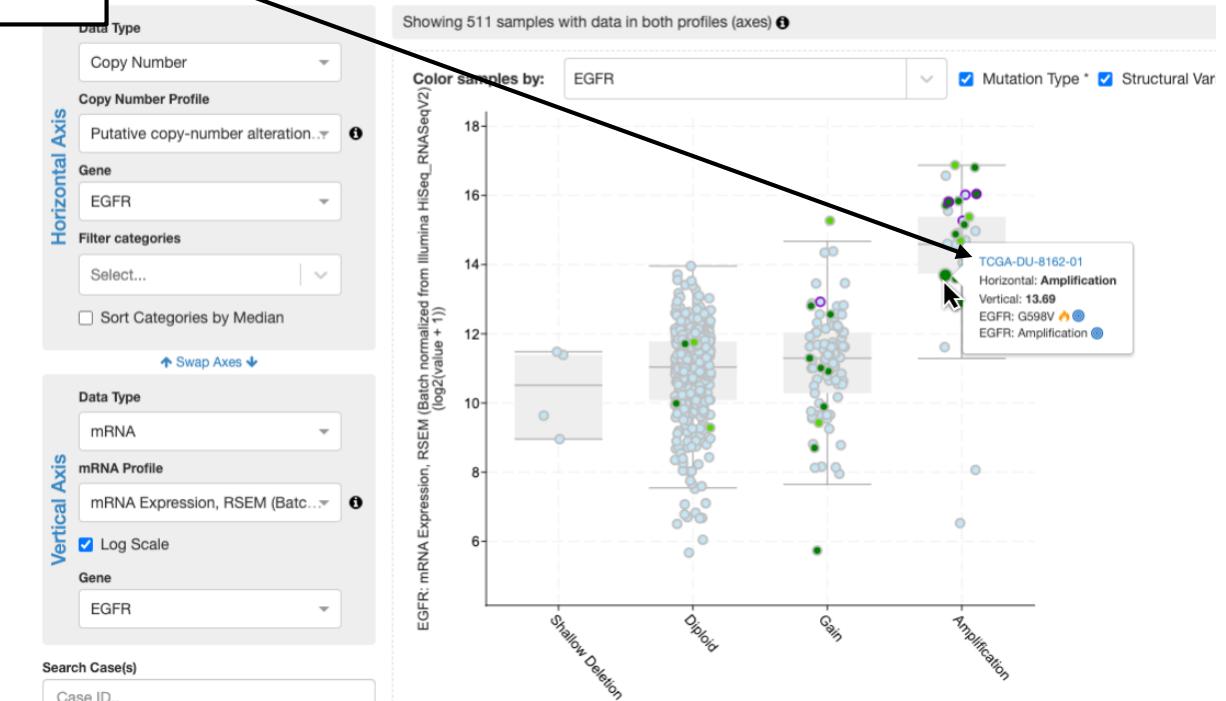


394 Mutations (page 1 of 16)

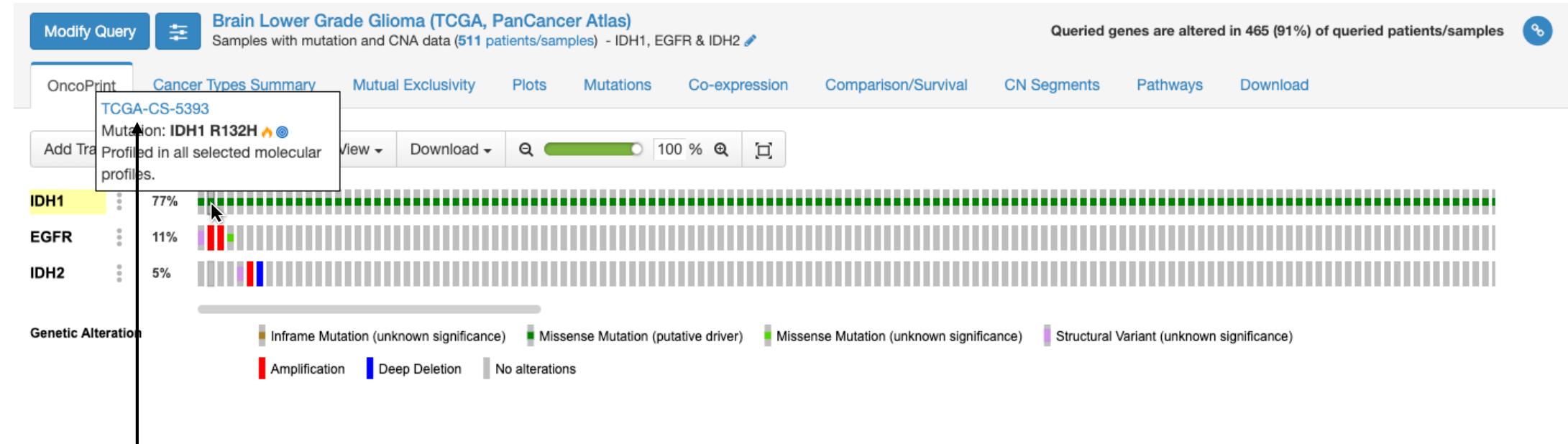
Sample ID	Cancer Type Detailed	Protein Change	Annotation	Mutation Type	Copy #
TCGA-DB-5276-01	Oligoastrocytoma	R132C	🕒 🌐 🔥	Missense	Diploid
TCGA-DB-A...	Oligoastrocytoma	R132C	🕒 🌐 🔥	Missense	Diploid
TCGA-DB-A...	Astrocytoma	R132C	🕒 🌐 🔥	Missense	Diploid
TCGA-DB-A...	Astrocytoma	R132C	🕒 🌐 🔥	Missense	Diploid
TCGA-DB-A...	Oligoastrocytoma	R132C	🕒 🌐 🔥	Missense	Diploid
TCGA-DB-A...	Astrocytoma	R132C	🕒 🌐 🔥	Missense	Diploid
TCGA-DB-A...	Astrocytoma	R132C	🕒 🌐 🔥	Missense	Diploid
TCGA-DB-A...	Astrocytoma	R132C	🕒 🌐 🔥	Missense	Diploid

Print Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways

Showing 511 samples with data in both profiles (axes) ⓘ



Patients view ex) Brain Lower Grade Glioma (TCGA, PanCancer Atlas)

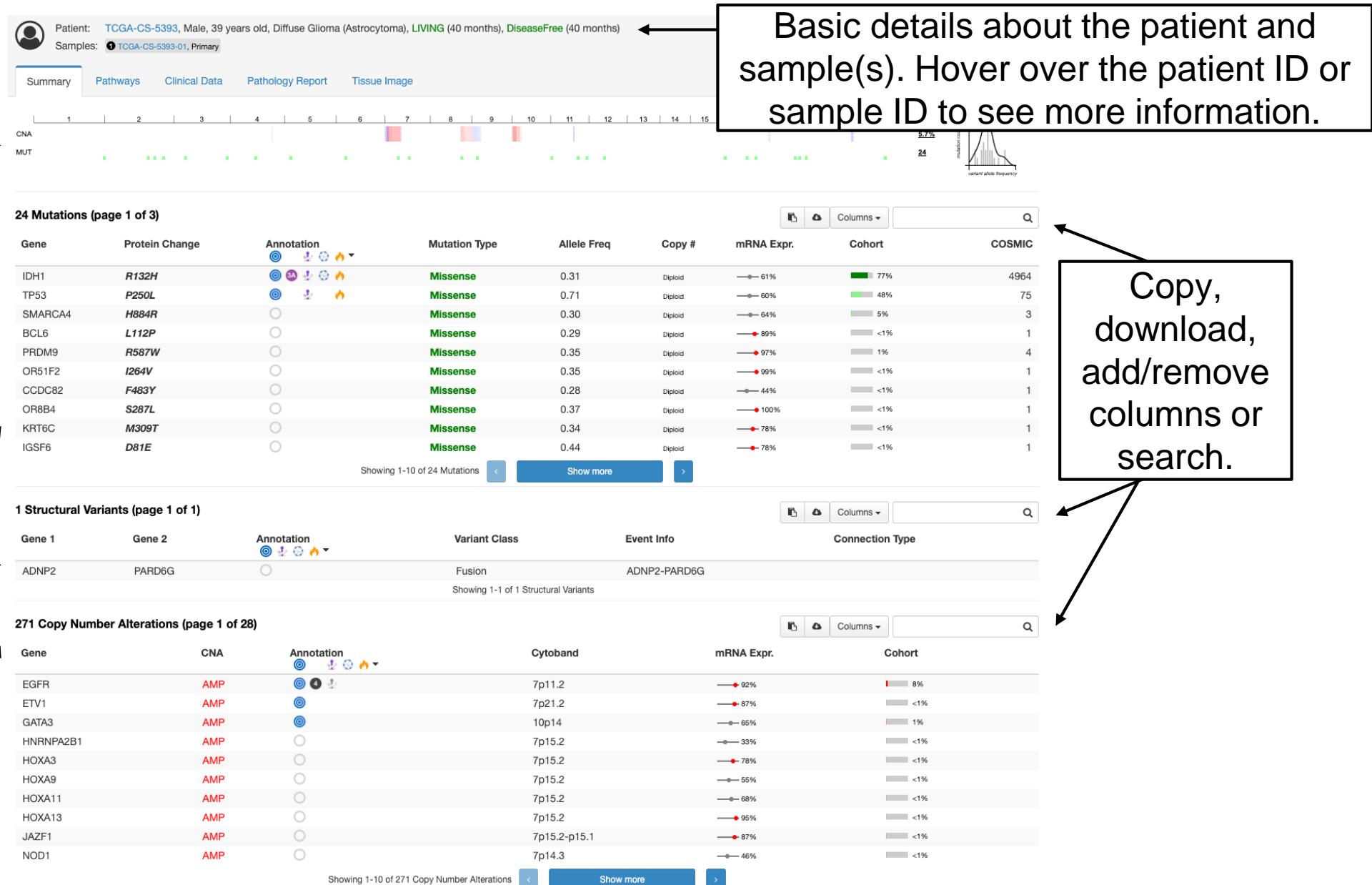


This is the same query that we used in the single study query tutorial.
Hover over a case of interest and then click on the patient ID.

Patient View, summary

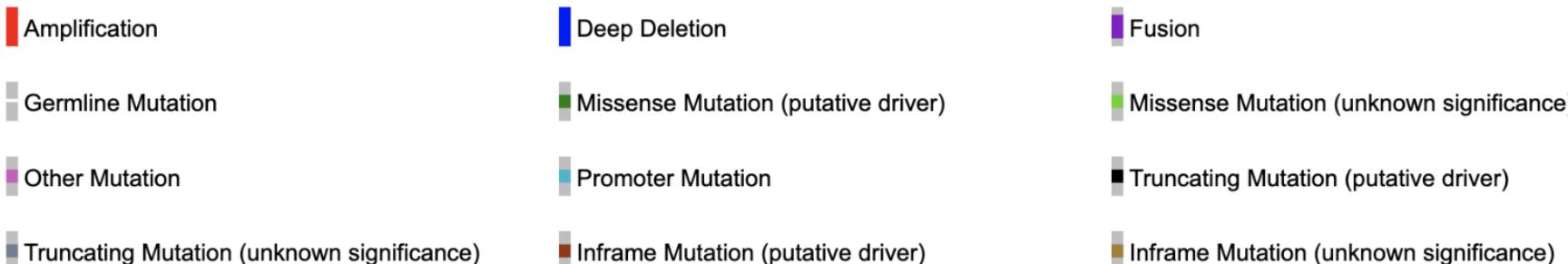
Figure showing where called CNA and mutations are across the genome. Hover over any of these for more details.

Lists of all called mutations, structural variants and CNAs (amplifications and deep deletions only).



Pathways Tab in Patient View

- One may be interested in viewing following types of genetic alterations of a *patient* in the context of pathways



Putative driver and unknown significance annotations are based on data from OncoKB and CancerHotspots.org.

- Start with “Patient view of an endometrial cancer case (TCGA, Nature 2013)” as an example

Patient: [TCGA-BK-A0CC](#), Female, 69 years old, Endometrial Cancer (Uterine Serous Carcinoma/Uterine Papillary Serous Carcinoma),
[LIVING](#) (10 months), [DiseaseFree](#) (10 months)

Samples: [1 TCGA-BK-A0CC-01](#), Primary, Stage III

[Uterine Corpus Endometrial Carcinoma \(TCGA, Nature 2013\)](#)

Not sure how to get to patient view? Review [Tutorial #3: Patient view](#)

Patient Pathways View

Pathways tab

Toolbar for pathway operations

More information about **notation** used can be found [here](#)

Pathway with genetic alterations of the patient

ACC-2016-TP53-RB-pathway

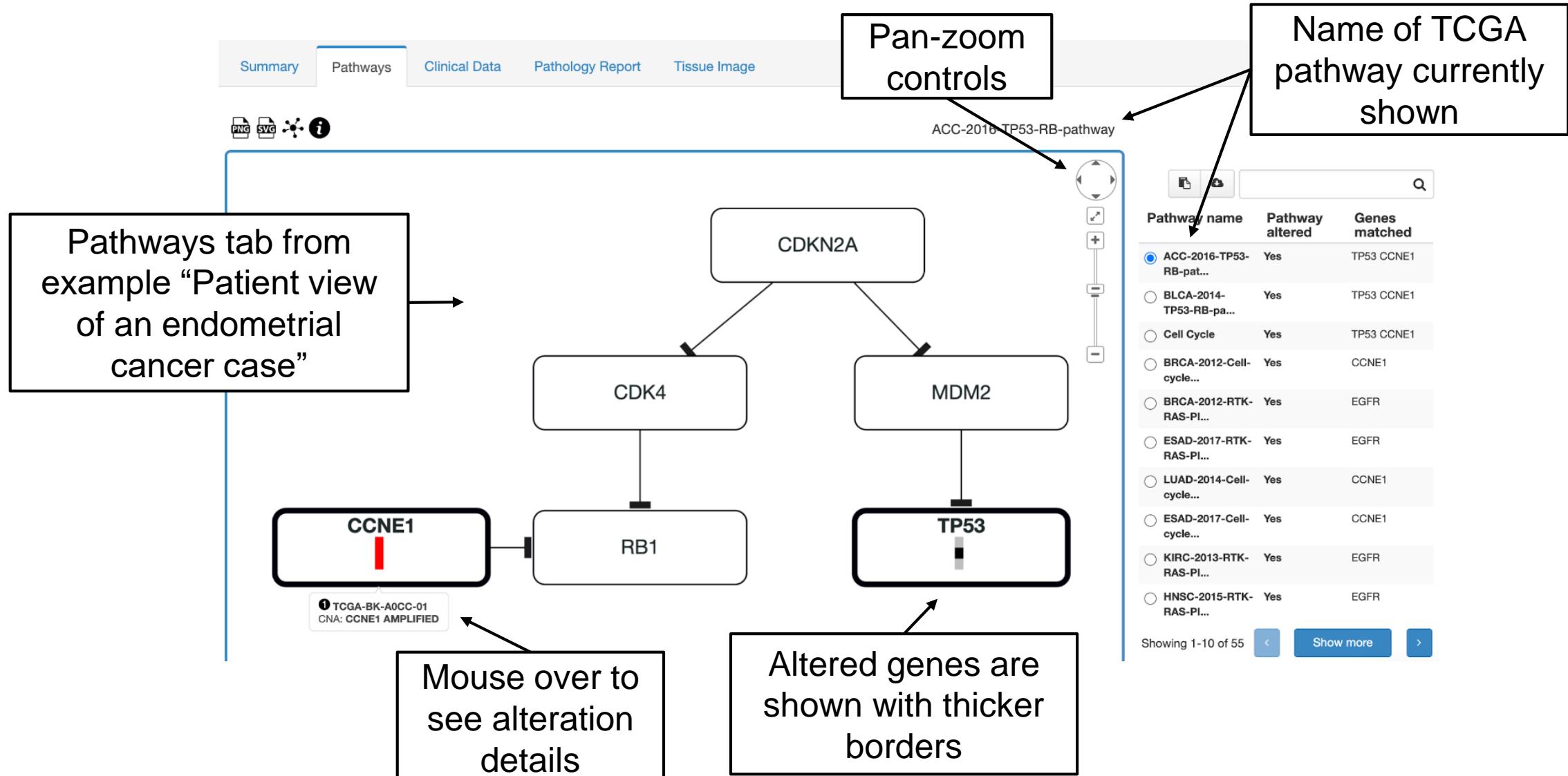
TCGA pathways table, where altered pathways are shown before non-altered ones by default

Pathway name	Pathway altered	Genes matched
ACC-2016-TP53-RB-pat...	Yes	TP53 CCNE1
BLCA-2014-TP53-RB-pa...	Yes	TP53 CCNE1
Cell Cycle	Yes	TP53 CCNE1
BRCA-2012-Cell-cycle...		
BRCA-2012-RTK-RAS-PI...		
ESAD-2017-RTK-RAS-PI...		
LUAD-2014-Cell-cycle...		
ESAD-2017-Cell-cycle...		
KIRC-2013-RTK-RAS-PI...		
HNSC-2015-RTK-RAS-PI...	Yes	

EGFR

Showing 1-10 of 55 < Show more >

Patient Pathways View



[Link to this page](#)

Download

Download data or copy lists of samples.

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

Downloadable Data Files

Copy-number Alterations (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Mutations (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Structural Variants (OQL is not in effect)	Tab Delimited Format Transposed Matrix
Altered samples: List of samples with alterations	Copy Download Query Virtual Study
Unaltered samples: List of samples without any alteration	Copy Download Query Virtual Study
Sample matrix: List of all samples where 1=altered and 0=unaltered	Copy Download
Log2 copy-number values ?	Tab Delimited Format Transposed Matrix
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2)	Tab Delimited Format Transposed Matrix
mRNA expression z-scores relative to diploid samples (RNA Seq V2 RSEM)	Tab Delimited Format Transposed Matrix
mRNA expression z-scores relative to all samples (log RNA Seq V2 RSEM)	Tab Delimited Format Transposed Matrix
Protein expression (RPPA) ?	Tab Delimited Format Transposed Matrix
Protein expression z-scores (RPPA) ?	Tab Delimited Format Transposed Matrix
Putative arm-level copy-number from GISTIC ?	Tracks added in the OncoPrint tab can be downloaded here.
Microbiome Signatures (log RNA Seq CPM) ?	Tracks added in the OncoPrint tab can be downloaded here.

Download queried data types for the queried genes.

Gene Alteration Frequency

Gene Symbol	Num Samples Altered	Percent Samples Altered
IDH1	394	77%
EGFR	54	11%
IDH2	24	5%

Frequency of gene alteration for each gene in the query

Type of Genetic Alterations Across All Samples

Study ID	Sample ID	Patient ID	Altered ▾	IDH1	EGFR	IDH2
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4938-01	TCGA-CS-4938	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4941-01	TCGA-CS-4941	1	no alteration	AMP (Driver)	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4942-01	TCGA-CS-4942	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4943-01	TCGA-CS-4943	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4944-01	TCGA-CS-4944	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-5390-01	TCGA-CS-5390	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-5393-01	TCGA-CS-5393	1	R132H (Driver)	AMP (Driver)	no alteration

List of all samples with status of each query gene.

Download

Download data or copy lists of samples.

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-expression Comparison/Survival CN Segments Pathways Download

Downloadable Data Files

Copy-number Alterations (OQL is not in effect)
Mutations (OQL is not in effect)
Structural Variants (OQL is not in effect)
Altered samples: List of samples with alterations
Unaltered samples: List of samples without any alteration
Sample matrix: List of all samples where 1=altered and 0=unaltered
Log2 copy-number values ⓘ
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2)
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 (RPPA) ⓘ
Protein expression z-scores (RPPA) ⓘ
Putative arm-level copy-number from GISTIC ⓘ
Microbiome Signatures (log RNA Seq CPM) ⓘ

Tab Delimited Format | Transposed Matrix
Tab Delimited Format | Transposed Matrix
Tab Delimited Format | Transposed Matrix
Copy | Download | Query | Virtual Study
Copy | Download | Query | Virtual Study
Copy | Download
Tab Delimited Format | Transposed Matrix
Tracks added in the OncoPrint tab can be downloaded here.
Tracks added in the OncoPrint tab can be downloaded here.

List of samples that have an alteration in one or more query genes

List of samples that have no alterations in any query genes

List of all samples with summary classification:
0 = no alteration in any query gene
1 = alteration in one or more query genes

Gene Alteration Frequency

Gene Symbol	Num Samples Altered	Percent Samples Altered
IDH1	394	77%
EGFR	54	11%
IDH2	24	5%

Showing 1-3 of 3

Type of Genetic Alterations Across All Samples

Study ID	Sample ID	Patient ID	Altered ▾	IDH1	EGFR	IDH2
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4938-01	TCGA-CS-4938	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4941-01	TCGA-CS-4941	1	no alteration	AMP (Driver)	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4942-01	TCGA-CS-4942	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4943-01	TCGA-CS-4943	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-4944-01	TCGA-CS-4944	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-5390-01	TCGA-CS-5390	1	R132H (Driver)	no alteration	no alteration
lgg_tcga_pan_can_atlas_2018	TCGA-CS-5393-01	TCGA-CS-5393	1	R132H (Driver)	AMP (Driver)	no alteration

Advanced feature: use these lists to build a custom sample list to run a new query, to create [virtual studies](#) or to build [custom groups](#).

Group Comparison_Select a study

1. Start typing tumor type of interest...

glioma

2. Select the checkbox next to the study of interest and click “Explore Selected Studies”

3. Or click on “View study summary” button

Query Quick Search Beta! Download

Select Studies for Visualization & Analysis: 1 study selected (1102 samples) Deselect all

CNS/Brain 14 Select all listed studies matching filter (15)

Soft Tissue 1

CNS/Brain

Diffuse Glioma

Brain Lower Grade Glioma (TCGA, PanCancer Atlas)

Brain Lower Grade Glioma (TCGA, Provisional)

Glioma (MSK, 2018)

Low-Grade Gliomas (UCSF, Science 2014)

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

GLIOBLASTOMA

Brain Tumor PDXs (Mayo Clinic, 2019)

Pilocytic Astrocytoma (ICGC, Nature Genetics 2013)

er Atlas)
al)
stic Oligoastrocytoma (MSK...
96 samples   

514 samples   
530 samples   
91 samples   
61 samples   
1102 samples   
95 samples   
585 samples   
206 samples   
592 samples   
604 samples   
22 samples   
Link to this page

1 study selected (1102 samples) Deselect all Query By Gene OR  Explore Selected Studies

Study View

Merged Cohort of LGG and GBM (TCGA, Cell 2016) [Download](#)
Whole-exome sequencing from 1,122 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

Click gene symbols below or enter here [Query](#)

Custom Selection Charts Groups ▾

Sex

Overall Survival

Surviving (%) vs Months Survival (0 to 240)

Mutated Genes (812 profiled samples)

T Gene	# Mut	#	Freq ▾
IDH1	411	411	50.6%
TP53	398	322	39.7%
ATRX	216	208	25.6%
PTEN	109	108	13.3%
EGFR	93	87	10.7%
CIC	85	79	9.7%
PIK3CA	67	63	7.8%
NF1	65	53	6.5%
PIK3R1	47	45	5.5%
FUBP1	44	44	5.4%
PCLO	34	33	4.1%

CNA Genes (1084 profiled samples)

T Gene	Cytoband	CNA	#	Freq ▾
CDKN2A	9p21.3	DEL	293	25.2%
CDKN2B	9p21.3	DEL		
MTAP	9p21.3	DEL		
EGFR	7p11.2	AMP		
CDK4	12q14.1	AMP		
FIP1L1	4q12	AMP		
PDGFRA	4q12	AMP		
CHIC2	4q12	AMP		
MLLT3	9p21.3	DEL		
KIT	4q12	AMP		
MDM4	1q32.1	AMP		

Mutation Count

Diagnosis Age

Absolute Extract Ploidy

BCR Status

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 "Charts" button to add it.

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

Study View

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

Whole-exome sequencing from 1,122 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

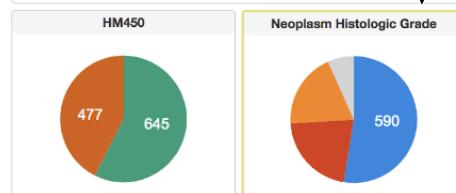
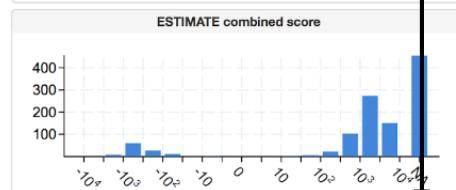
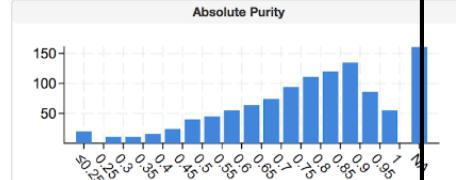
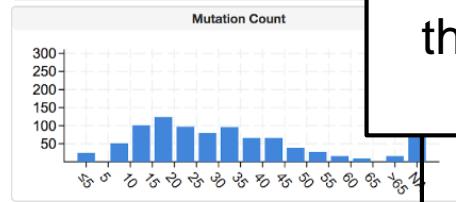
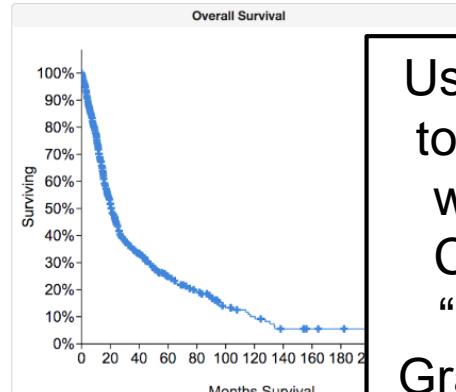


Custom Selection

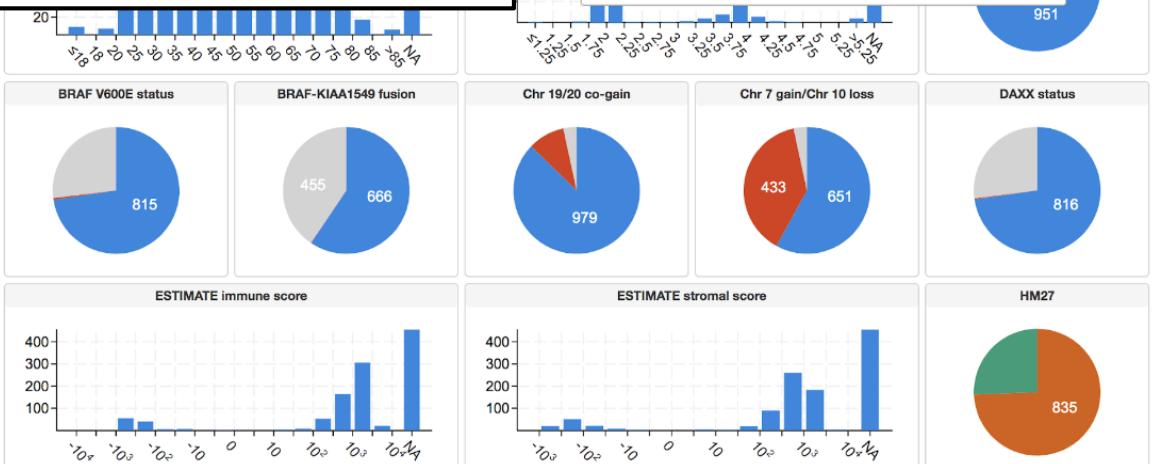
Charts

Groups

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



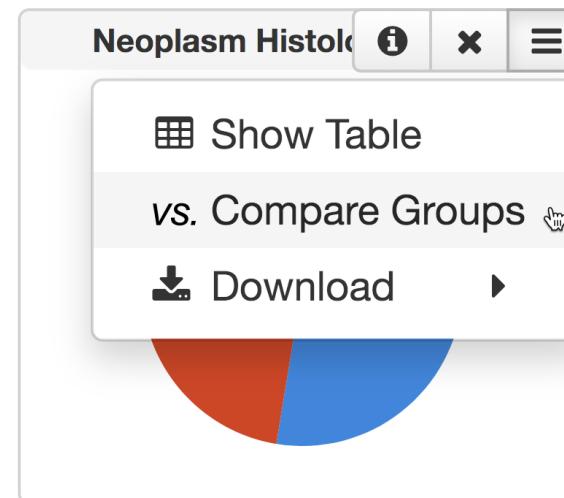
Use the search box here to find clinical attributes with the word "grade". Check the box next to "Neoplasm Histologic Grade" and a pie chart of this data will add to the bottom of the page.



[Link to this page](#)

Study View

Let's compare samples of different histologic grades. Hover over the "Neoplasm Histologic Grade" pie chart menu icon (≡) and notice the vs. Compare Groups option. 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. At the top, it displays "Merged Cohort of LGG and GBM (TCGA, Cell 2016)". Below this, it says "Groups from **Neoplasm Histologic Grade**". Under "Groups:", there are three items: (A) G2 (216), (B) G3 (241), and (C) G4 (590). To the right of these are "Select all" and "Deselect all" buttons. Below the groups, there is a navigation bar with tabs: Overlap (selected), Survival, Clinical, Mutations, and Copy-number.

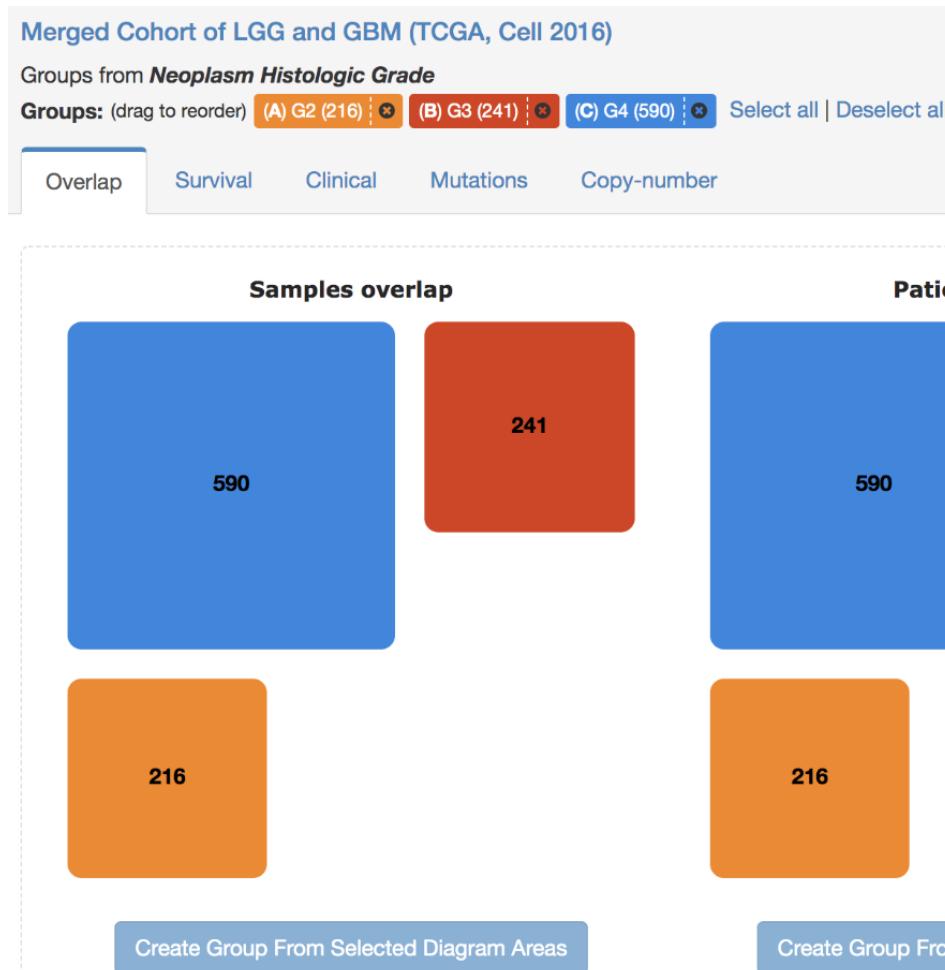
The attribute used to create the groups.

Each tab has specific functionality. We'll go through these one-by-one over the next few slides.

The original study. Click to return to study view.

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.

Group Comparison: Overlap

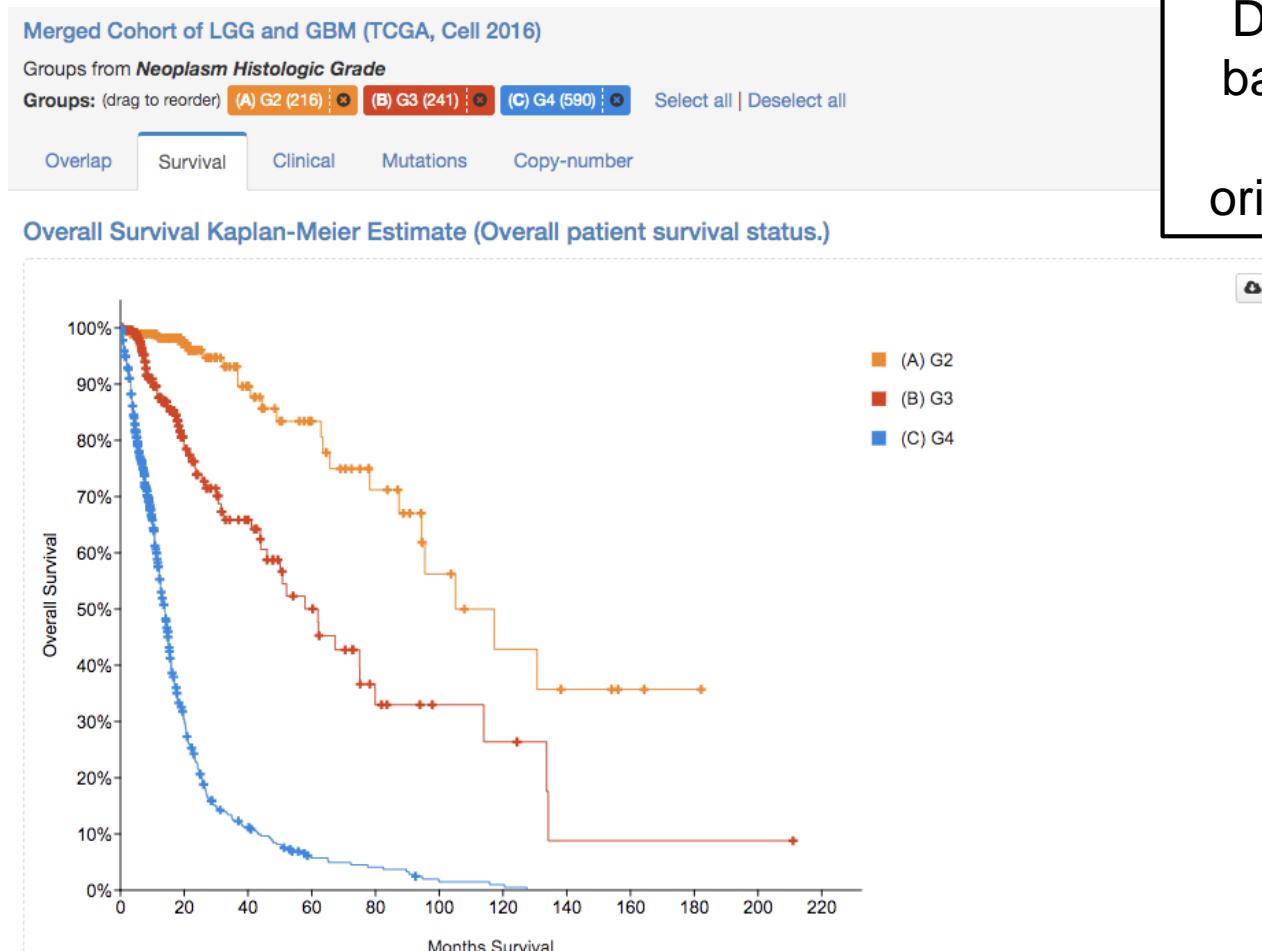


The Overlap tab shows which samples or patients may overlap among the selected groups. In this example, we can see that there is no overlap in samples or patients.

In the next example, we'll look at how overlapping samples/patients are managed.

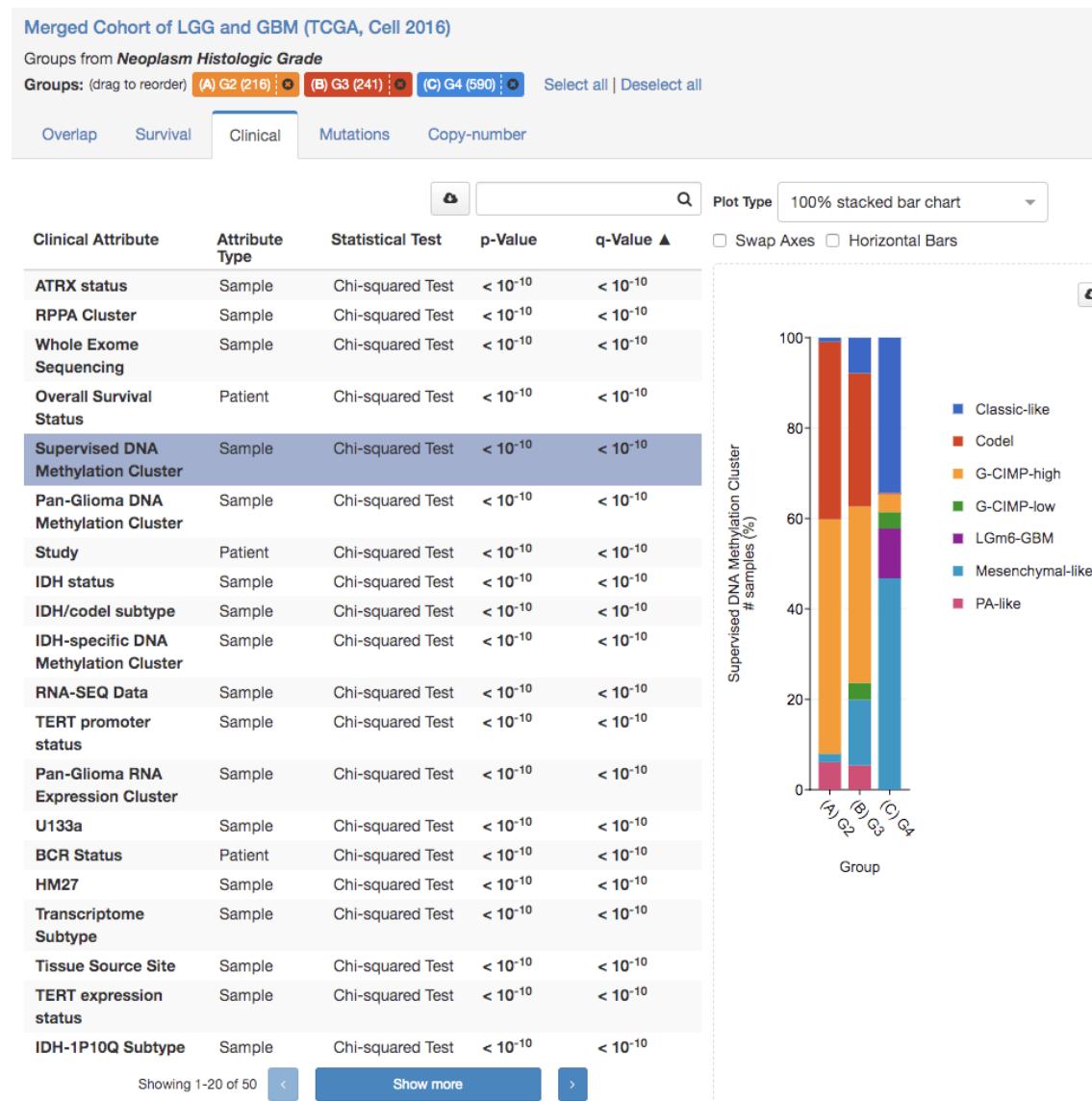
This view can also be used to create additional groups. We'll do this later.

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.

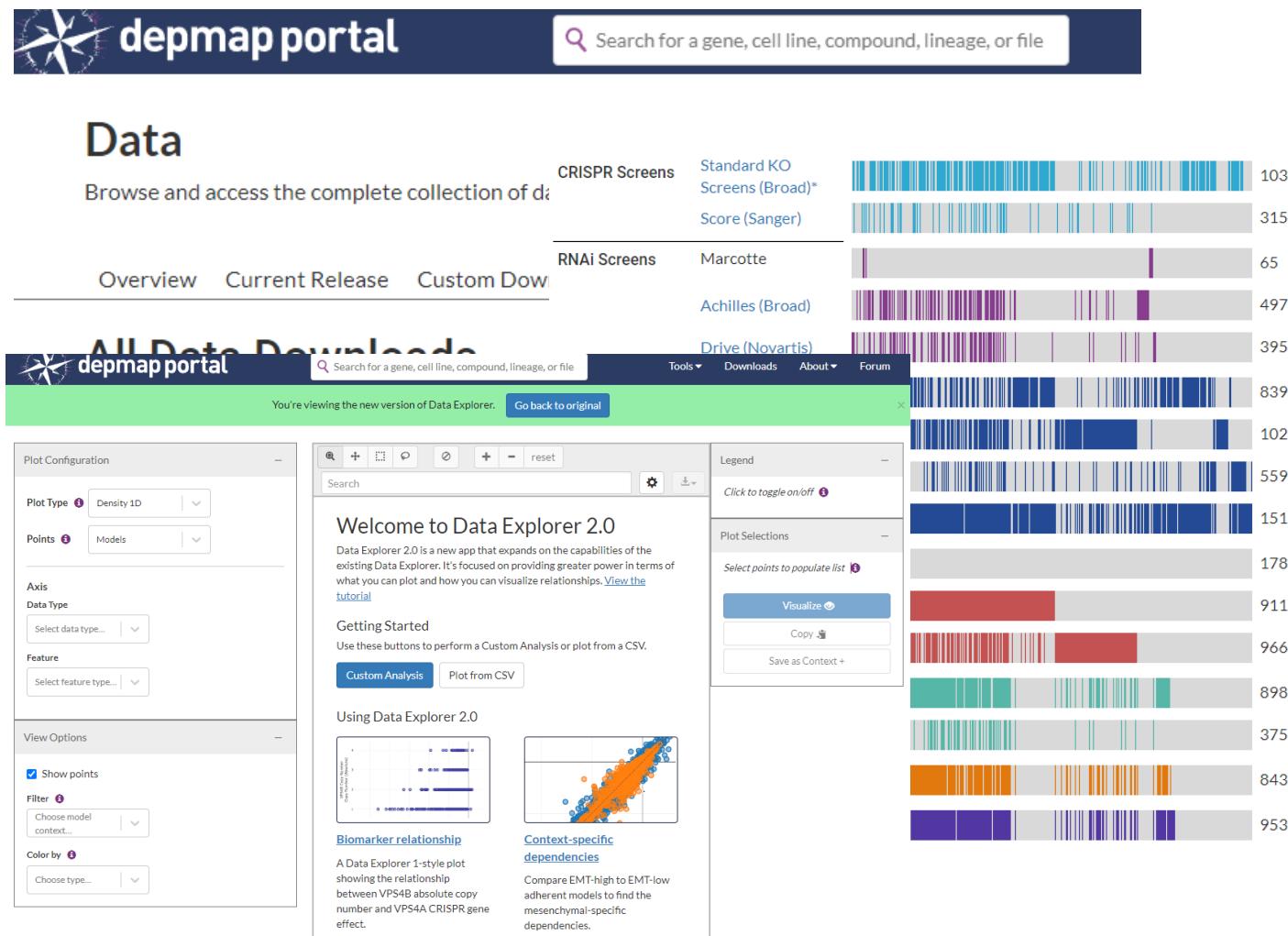
Group Comparison: Clinical



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.

2. CCLE(Cancer Cell Line Encyclopedia)

- DepMap project • https://depmap.org/portal/data_page/?tab=allData



- The DepMap project, building off of the original Cancer Cell Line Encyclopedia (CCLE) project, generates data and tools that can be used and shared by researchers. New DepMap data is released twice a year, in May and November.

- Data sources used in the DepMap portal
- Data in the DepMap portal is aggregated from many sources, including both DepMap Release data and collaborator datasets. This figure provides an overview of how many cell lines can be found in each data source.

<https://www.cbiportal.org/>
<https://github.com/cBioPortal/cbiportal>

CCLE_Depmap tutorial

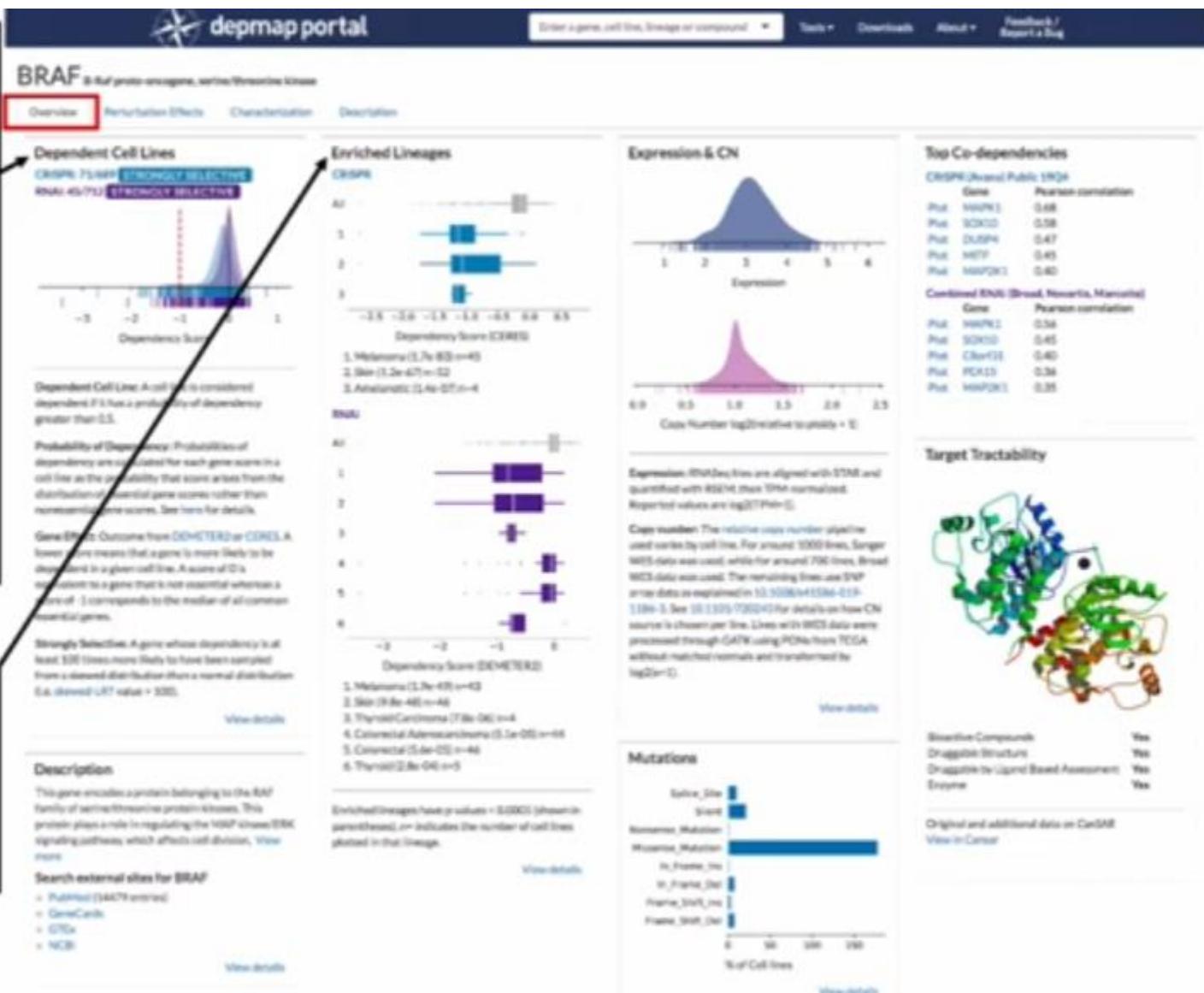
Tutorial #1: Single Gene Query

1. **UNDERSTAND** Gene essentiality across more than 600 cell lines.
2. **FIND** Detail genetic characterization of over 1000 cell lines.

Depmap tutorial

Dependent Cell lines:
Understand the dependency (essentiality) of the gene across more than 600 cell lines.
Dependency Score measures CRISPR knockout / RNAi effect on cell viability. A lower score indicates that the gene is more dependent (essential) in a given cell line.

Enriched Lineages:
Identify cell lineages that are preferentially dependent on the query gene.



Depmap tutorial

Expression & CN:

Understand the mRNA expression and copy number of the gene across cell lines.

Mutations:

Understand the mutation profile of the gene across cell lines.

BRAF B-Raf proto-oncogene, serine/threonine kinase

Overview Perturbation Effects Characterization Description

Dependent Cell Lines
CRISPR 75H9P STRONGLY SELECTIVE
RNAi 45/71L STRONGLY SELECTIVE

Enriched Lineages
CRISPR
AI 1 2 3 4 5 6
Dependency Score (CRISPR)
1. Melanoma (3.7e-03) n=45
2. Skin (7.3e-07) n=22
3. Adrenocortical (3.4e-07) n=4

RNAi
AI 1 2 3 4 5 6
Dependency Score (RNAi)
1. Melanoma (3.7e-03) n=45
2. Skin (7.3e-07) n=22
3. Adrenocortical (3.4e-07) n=4

Expression & CN
Expression
Copy Number log₂(relative to mode) + 1
Depression (75H9P) lines are aligned with TCGA and normalized with RSE414 cells. TFM normalized. Reported values are log₂(1/Pearson).

Copy number: The relative copy number pipeline used varies by cell line. For around 1000 lines, Singer WES data was used, while for around 700 lines, Broad WES data was used. The remaining lines use SNP array data as expressed in 13,3306/n1586-019-1386-3. See 10.1101/220043 for details on how CN source is chosen per line. Lines with WES data were processed through GATK using PCalls from TESCA without matched normals and transformed by log₂(n+1).

Top Co-dependencies
CRISPR (Broad, Novartis, Marquette)

Gene	Pearson correlation
Plk1	0.68
SORCI2	0.58
PLXNC1	0.47
MAP3K7	0.45
HARPPC1	0.40

Combined WES (Broad, Novartis, Marquette)

Gene	Pearson correlation
MAP3K1	0.74
SORCI1	0.45
Chezze	0.40
PCX23	0.34
HARPPC1	0.35

Target Tractability

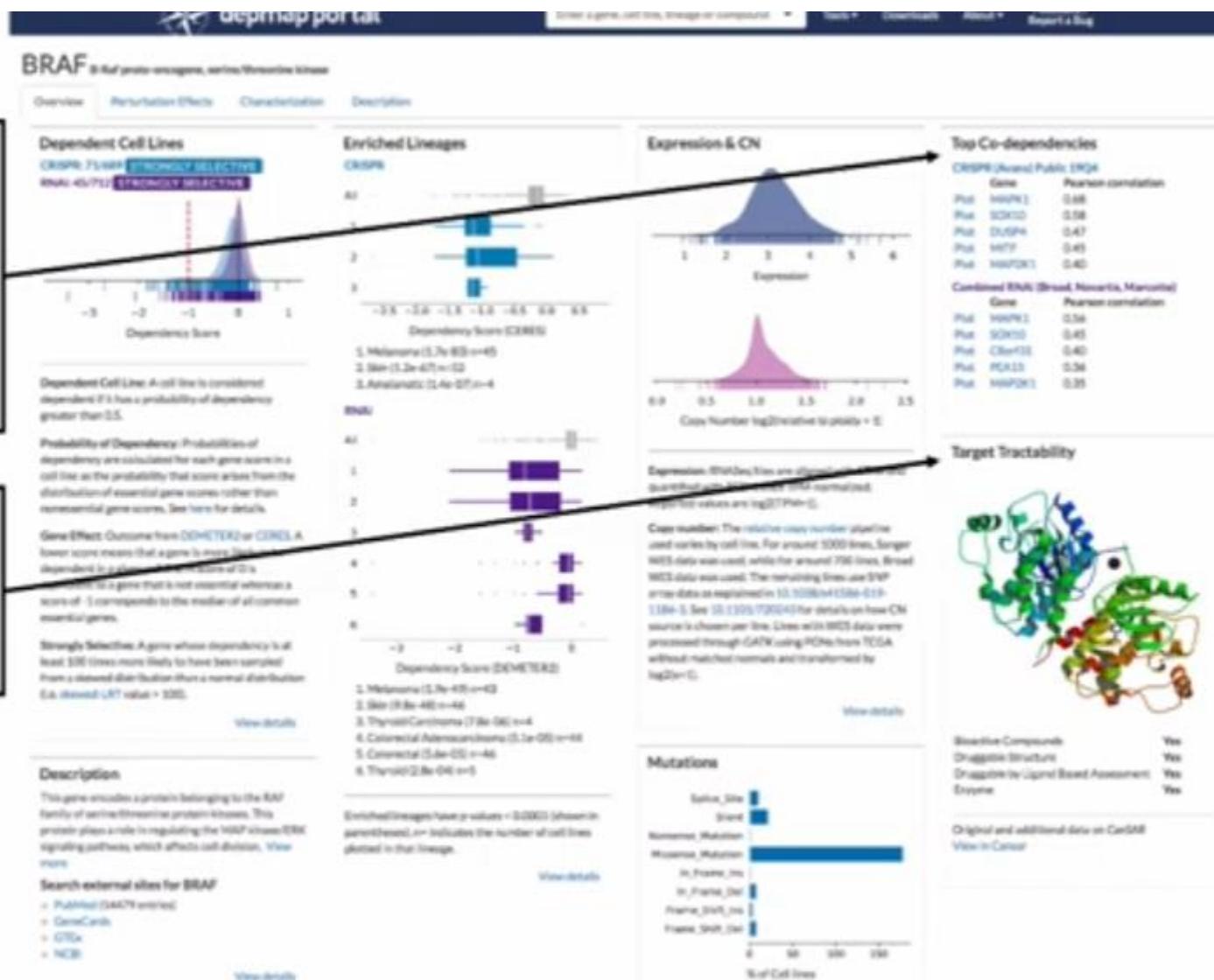
Mutations
Somatic_Mutations 1
Nonsense_Mutation 1
Missense_Mutation 1
In_Frame_Ins 1
In_Frame_Del 1
Frame_Shift_Ins 1
Frame_Shift_Del 1
%

Original and additional data on CanGAP
View in Cancer

Depmap tutorial

Top Co-dependencies:
Understand the top co-dependencies, whose dependency scores are highly correlated with that of the query gene.

Target Tractability:
Understand the target druggability of the gene encoding protein.



Depmap tutorial

1. Click the tab to view "Perturbation Effects"

Dependent Cell Lines

CRISPR: 751 HEP STRONGLY SELECTIVE RNAi: 45/711 STRONGLY SELECTIVE

Dependency Score

Dependent Cell Line: A cell line is considered dependent if it has a probability of dependency greater than 0.5.

Probability of Dependency: Probabilities of dependency are calculated for each gene score in a cell line as the probability that score arises from the distribution of essential gene scores rather than nonessential gene scores. See here for details.

Gene Effect: Outcome from ODHETER2 or CERES-A, lower score means that a gene is more likely to be dependent in a given cell line. A score of 0 is equivalent to a gene that is not essential whereas a score of -1 corresponds to the median of all common essential genes.

Strongly Selective: A gene whose dependency is at least 100 times more likely to have been sampled from a skewed distribution than a normal distribution (i.e. skewed LRT value > 300).

View details

Description

This gene encodes a protein belonging to the BRAF family of serine/threonine protein kinases. This protein plays a role in regulating the MAPK kinase/ERK signaling pathway, which affects cell division. View more

Search external sites for BRAF

View details

2. Or click "View details" under "Dependent Cell Lines"

Enriched Lineages

CRISPR

All 1 2 3

Dependency Score (ODHETER2)

1. Melanoma (3.7e-83) n=45
2. Skin (3.2e-47) n=52
3. Astrocytoma (3.4e-67) n=4

RNAi

All 1 2 3 4 5 6

Dependency Score (ODHETER2)

1. Melanoma (3.7e-49) n=43
2. Skin (3.8e-48) n=44
3. Thyroid Carcinoma (7.8e-06) n=4
4. Colorectal Adenocarcinoma (5.1e-03) n=44
5. Endometrial (5.8e-05) n=5
6. Thymid (2.8e-04) n=5

3. Or click "View details" under "Enriched Lineages"

Expression & CN

Expression

CN

Top Co-dependencies

CRISPR (Avana) Public RNAi

Gene	Pearson correlation
Plk1	0.68
SOR10	0.58
DUSP9	0.47
HIF1A	0.45
HAPOR1	0.40

Combined RNAi (Broad, Novartis, Mammalian)

Gene	Pearson correlation
Plk1	0.58
SOR10	0.45
Cdk12	0.40
PCX13	0.38
HAPOR1	0.35

Target Tractability

View details

Mutations

Silent Nonsense Missense In_Frame_Ins In_Frame_Del Frame_Shift_Ins Frame_Shift_Del

% of Cell Lines

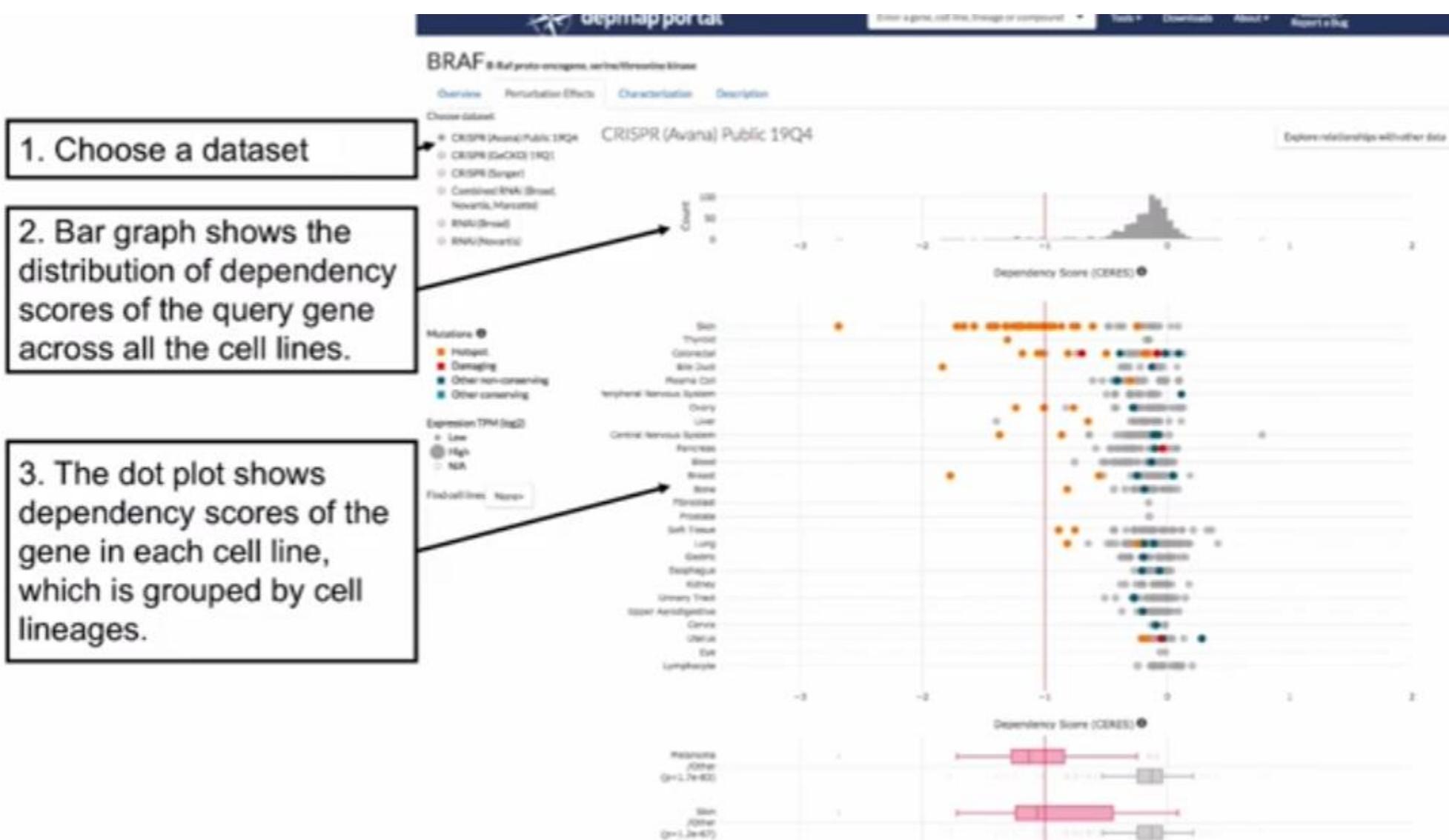
View details

Bioactive Compounds Druggable Structure Druggable by Ligand Based Assessment Enzyme

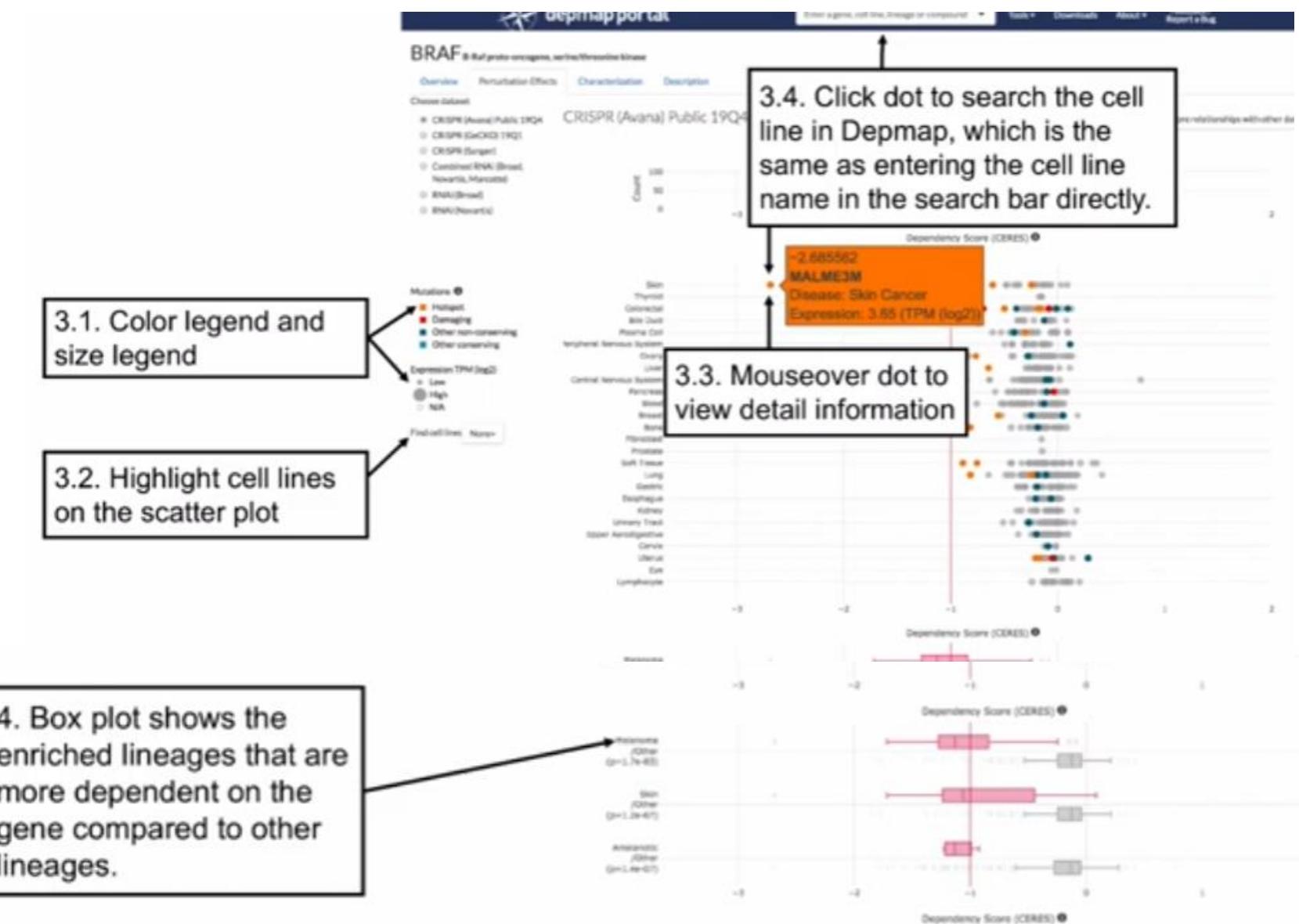
Yes Yes Yes Yes

Original and additional data on CanGAP View-in-Cancer

Depmap tutorial



Depmap tutorial



Depmap tutorial



Depmap tutorial

1. Click the tab to view "Characterization"

2. Or click "View details" under "Expression & CN"

3. Or click "View details" under "Mutations"

BRAF β₁-Araf proto-oncogene, serine/threonine kinase

Overview Perturbation Effects Characterization Description

Dependent Cell Lines

STRONGLY SELECTIVE RNAi: 45/734P
STRONGLY SELECTIVE RNAi: 45/734P

Dependency Score

Dependent Cell Line: A cell line is considered dependent if it has a probability of dependency greater than 0.5.

Probability of Dependency: Probabilities of dependency are calculated for each gene score in a cell line as the probability that score arises from the distribution of essential gene scores rather than nonessential gene scores. See [here](#) for details.

Gene-Effect: Outcome from DEHETERO or CINCH. A lower score means that a gene is more likely to be dependent in a given cell line. A score of 0 is equivalent to a gene that is not essential whereas a score of 1 corresponds to the median of all known essential genes.

Strongly Selective: A gene whose dependency is at least 100 times more likely to have been sampled from a skewed distribution than a normal distribution (i.e., skewed LRT value > 300).

Enriched Lineages

CRISPR

A1
1
2
3

Dependency Score (C3CRIS)
1. Melanoma (3.7e-63) n=45
2. Skin (1.2e-47) n=32
3. Adrenocortical (2.4e-27) n=4

RNAi

A1
1
2
3
4
5
6

Dependency Score (DEHETERO)
1. Melanoma (3.8e-49) n=43
2. Skin (2.8e-40) n=40
3. Thyroid/Lymphoma (7.8e-48) n=44
4. Colorectal Adenocarcinoma (5.2e-48) n=44
5. Colorectal (5.8e-45) n=44
6. Thyroid (2.8e-04) n=5

Expression & CN

Expression

Cosine Number log2(absolute value + 1)

Top Co-dependencies

CRISPR (Avanged Public CRISPR)

Gene	Pearson correlation
HAPK1	0.68
SCKO1	0.58
DUSP4	0.47
HPTF	0.45
HAPK3	0.40

Combined RNAi (Broad, Novartis, Marquette)

Gene	Pearson correlation
HAPK1	0.54
SCKO1	0.45
Cbr425	0.40
PCA23	0.36
HAPK3	0.35

Target Tractability

Small-molecule Compounds

Drugbank Structure Drugbank by Ligand-Based Assessment Enzmine

Original and additional data on CanGAP View in Cancer

Mutations

Enriched Mutations have ρ values < 0.0001 (shown in parentheses). $n =$ indicates the number of cell lines plotted in that lineage.

View details

Enriched Mutations

Silence_Mut
Silent
Nonsense_Mutation
Missense_Mutation

H₃ Frame_Shift
H₃ Frame_Delet
Frame_Shift_Inframe
Frame_Shift_Del

% of Cell Lines

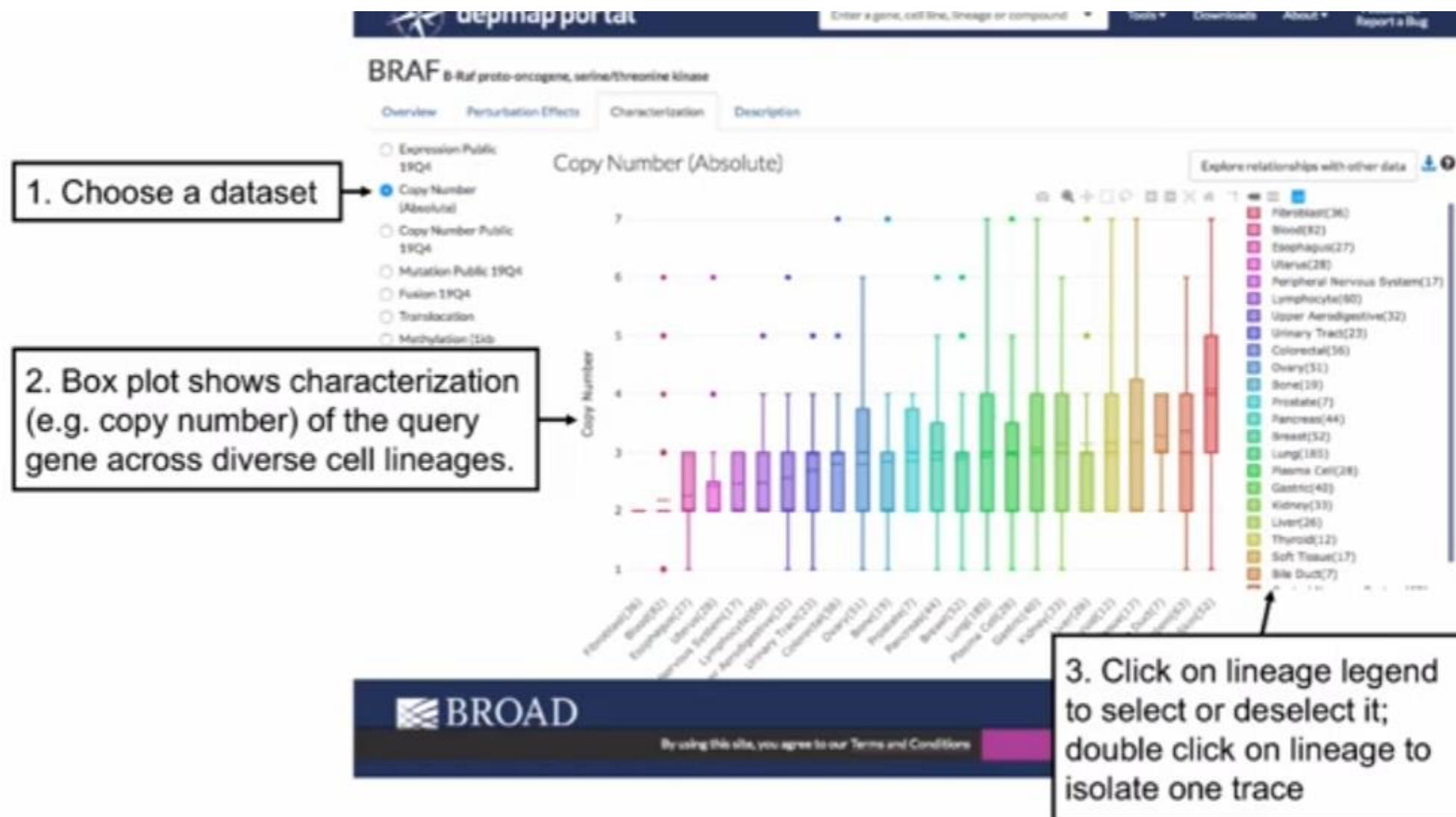
View details

View details

View details

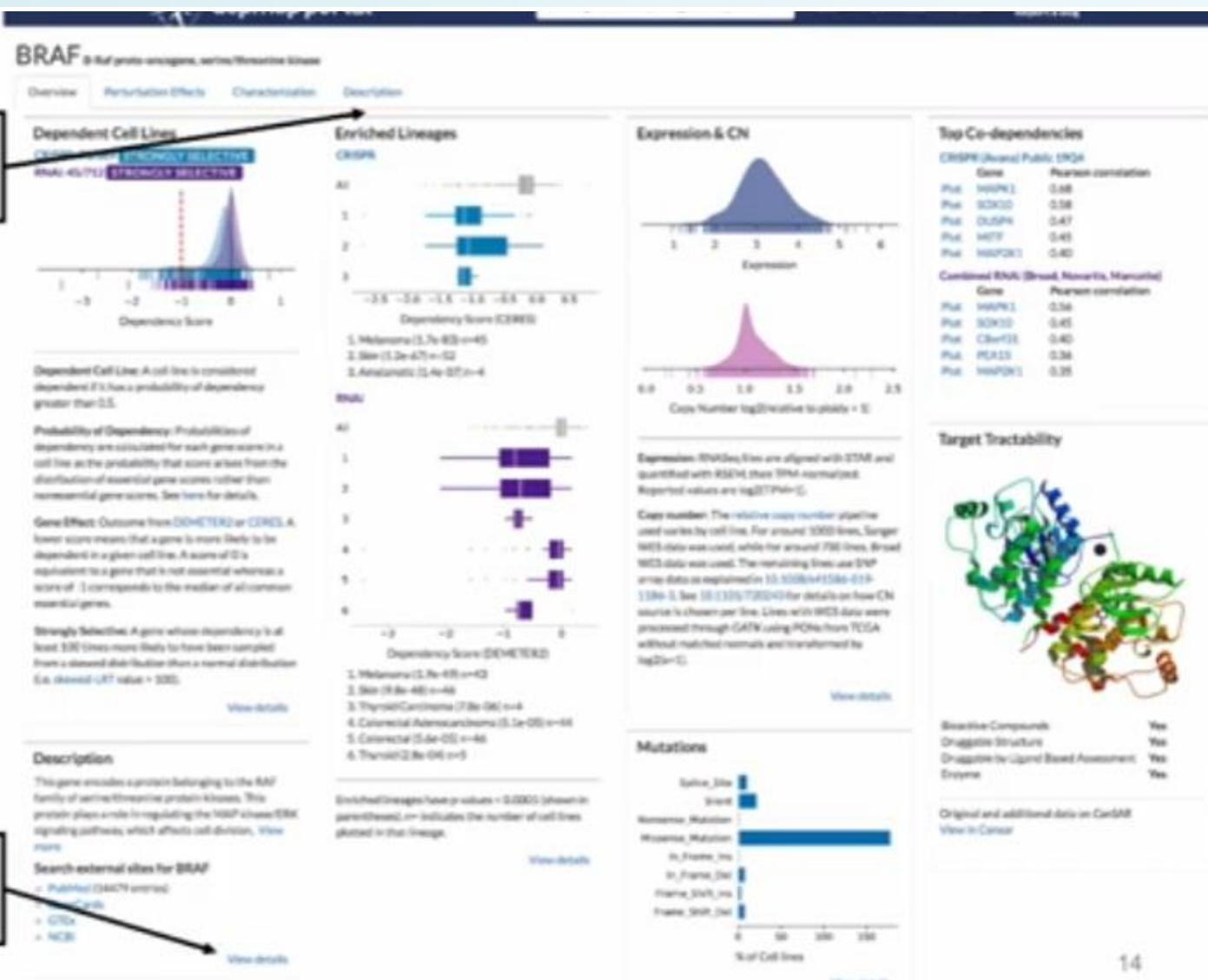
View details

Depmap tutorial



Depmap tutorial

1. View "Description" by clicking the tab



2. Or click "View details" under "Description"

Depmap tutorial

The screenshot shows the Depmap portal interface. At the top, there is a search bar with placeholder text "Enter a gene, cell line, lineage or compound" and a navigation bar with links for "Tools", "Downloads", "About", and "Feedback / Report a Bug". Below the header, the gene page for BRAF is displayed. The gene name "BRAF" is at the top, followed by its full name "B-Raf proto-oncogene, serine/threonine kinase". There are four tabs: "Overview", "Perturbation Effects", "Characterization", and "Description". The "Description" tab is selected, showing a detailed paragraph about the gene's function, mutations, and associated diseases. Below this, there is a list of external links under the heading "Search external sites for BRAF". The "External links for the gene" section is highlighted with a black border and an arrow pointing to it from the text "2. External links for the gene". The footer of the page includes the Broad Institute logo and copyright information, along with a "CONNECT" section featuring links to GitHub, Bitbucket, and Zenodo.

1. Description of the gene →

2. External links for the gene →

BRAF B-Raf proto-oncogene, serine/threonine kinase

Overview Perturbation Effects Characterization Description

This gene encodes a protein belonging to the RAF family of serine/threonine protein kinases. This protein plays a role in regulating the MAPK kinase/ERK signaling pathway, which affects cell division, differentiation, and secretion. Mutations in this gene, most commonly the V600E mutation, are the most frequently identified cancer-causing mutations in melanoma, and have been identified in various other cancers as well, including non-Hodgkin lymphoma, colorectal cancer, thyroid carcinoma, non-small cell lung carcinoma, hairy-cell leukemia, and adenocarcinoma of lung. Mutations in this gene are also associated with cardiofaciocutaneous, Noonan, and Costello syndromes, which exhibit overlapping phenotypes. A pseudogene of this gene has been identified on the X chromosome. [provided by RefSeq, Aug. 2017]

- Official symbol: BRAF
- Full name: B-Raf proto-oncogene, serine/threonine kinase
- Location: 7q34
- Alias known as: BRAF1
- Entrez ID: 473
- Ensembl ID: ENSG00000157764

Search external sites for BRAF

- PubMed (34482 entries)
- HGNC
- GeneCards
- Wikipedia
- FlyBase
- Tumor Portal
- GTEx
- COSMIC
- NCBI

BROAD INSTITUTE
Copyright © 2017 Broad Institute, Massachusetts General Hospital, Harvard University

CONNECT

- [GitHub](#)
- [Bitbucket](#)
- [Zenodo](#)

Depmap tutorial

Tutorial #2: Single Compound Query

FIND Detail pharmacologic characterization.

Depmap tutorial

depmap portal

PLX-4720

Overview Sensitivity Datasets

Sensitive Cell Lines

PLX-4720 (BMS-880-K344784PP-005-0H-2) Drug sensitivity (PRISM Repurposing Primary Screen) 3923 600 cell lines shown

Enriched Lineages

PLX-4720 (BMS-880-K344784PP-005-0H-2) Drug sensitivity (PRISM Repurposing Primary Screen) 3923

All

1

2

-3 -2 -1 0 1 2 3

log₂ fold change

1, Melanoma (2.7e-20) n=43
2, Skin (2.7e-20) n=43

Enriched lineages have p values < 0.001 (shown in parentheses). n= indicates the number of cell lines plotted in that lineage.

Top Correlated Expression

Gene Pearson correlation

PLX GAPDH -0.55
PLX TSHZS -0.54
PLX TSH -0.53
PLX CPV -0.52
PLX SLC4A2 -0.52

All viability profiles for a given compound were correlated against all expression features. The highest correlated viability profile for each gene reported in the table above. Clicking "View more" visualizes that comparison and shows which dataset was used for the viability profile.

Datasets with data for PLX-4720

Dataset	Cell Lines	Dose Range	Assay
GCNC1	167	1nM - 10μM	PRISM
Repurposing Primary	378	1nM - 10μM	CellTiterOne
GCNC2	409	1nM - 10μM	CellTiterOne
CTRP	840	1nM - 50μM	CellTiterOne

Datasets with data for the query compound

Top Correlated Expression: top genes whose expression is highly correlated with drug sensitivity.

Metadata of the query compound

Metadata from the Repurposing Hub

HQ4, RAF inhibitor
Phase: Preclinical
Target: BRAF, KRAS

CONNECT

- Depmap@Broad
- CCLE@Broad
- Feedback

INSTITUTE

Copyright © 2019 Broad Institute. All rights reserved. Privacy Policy | Cookies

Sensitive Cell lines:
Understand the drug sensitivity profile of the compound in ~600 cell lines. Negative log₂ fold change indicates drug response, while positive log₂ fold change indicates drug resistance.

Enriched Lineages:
Understand the drug sensitivity of the compound in different cell lineages (tissues and cancer types).

18

91

Depmap tutorial

depmap portal

PLX-4720

Overview Sensitivity Dose curves

1. View "Sensitivity" by clicking the tab

Sensitive Cell Lines

PLX-4720 (BIRC-BIRC-K3A4T764PF-001-0F-2) Drug sensitivity (PRISM Repurposing Primary Screen)

2RQ2

560 cell lines shown

AUC = 0.73 ± 0.01 (n=43)

log₂ fold change

Please note that AUC values depend on the dose range of the screen and are not comparable across different assays.

[View more](#)

2. Or click "View more" under "Sensitive Cell Lines"

Enriched Lineages

PLX-4720 (BIRC-BIRC-K3A4T764PF-001-0F-2) Drug sensitivity (PRISM Repurposing Primary Screen)

2RQ2

All

1

2

-3 -2 -1 0 1

log₂ fold change

1. Melanoma (2.7e 20 n=43)
2. Skin (2.7e 20 n=43)

Enriched lineages have p-value < 0.0001 (shown in parentheses), or indicates the number of cell lines plotted in that lineage.

[View more](#)

3. Or click "View more" under "Enriched Lineages"

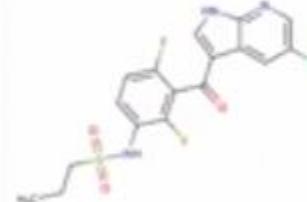
Top Correlated Expression

Gene	Pearson correlation
Pro CAPOH5	-0.55
Pro TSHZ3	-0.54
Pro TYR	-0.53
Pro CPN1	-0.52
Pro SLC45A2	-0.52

All viability profiles for a given compound were correlated against all expression features. The highest correlated viability profile for each gene reported in the table above. Clicking "plot" will visualize that comparison and show which dataset was used for the viability profile.

Datasets with data for PLX-4720

Dataset	Cell Lines	Dose Range	Assay
CCSP1	967	1nM - 30μM	Survival
Repurposing Primary	378	10μM	PRISM
CCSP2	809	1nM - 30μM	CellTiterGlo
CTRP	840	1nM - 30μM	CellTiterGlo



Metadata from the Repurposing Hub

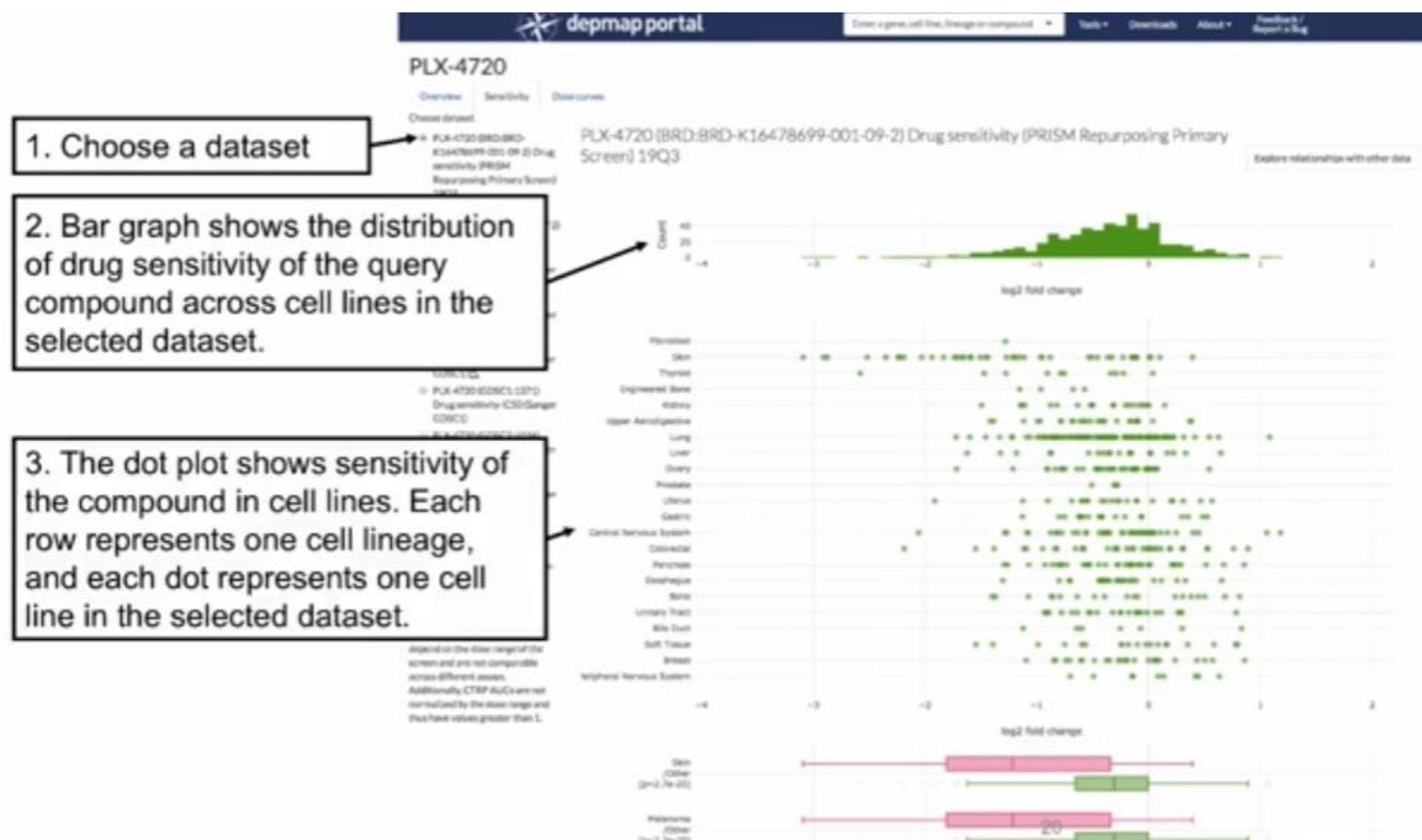
MOA: RAF Inhibitor
Phase: Preclinical
Target: BRAF, KRAS

BROAD INSTITUTE

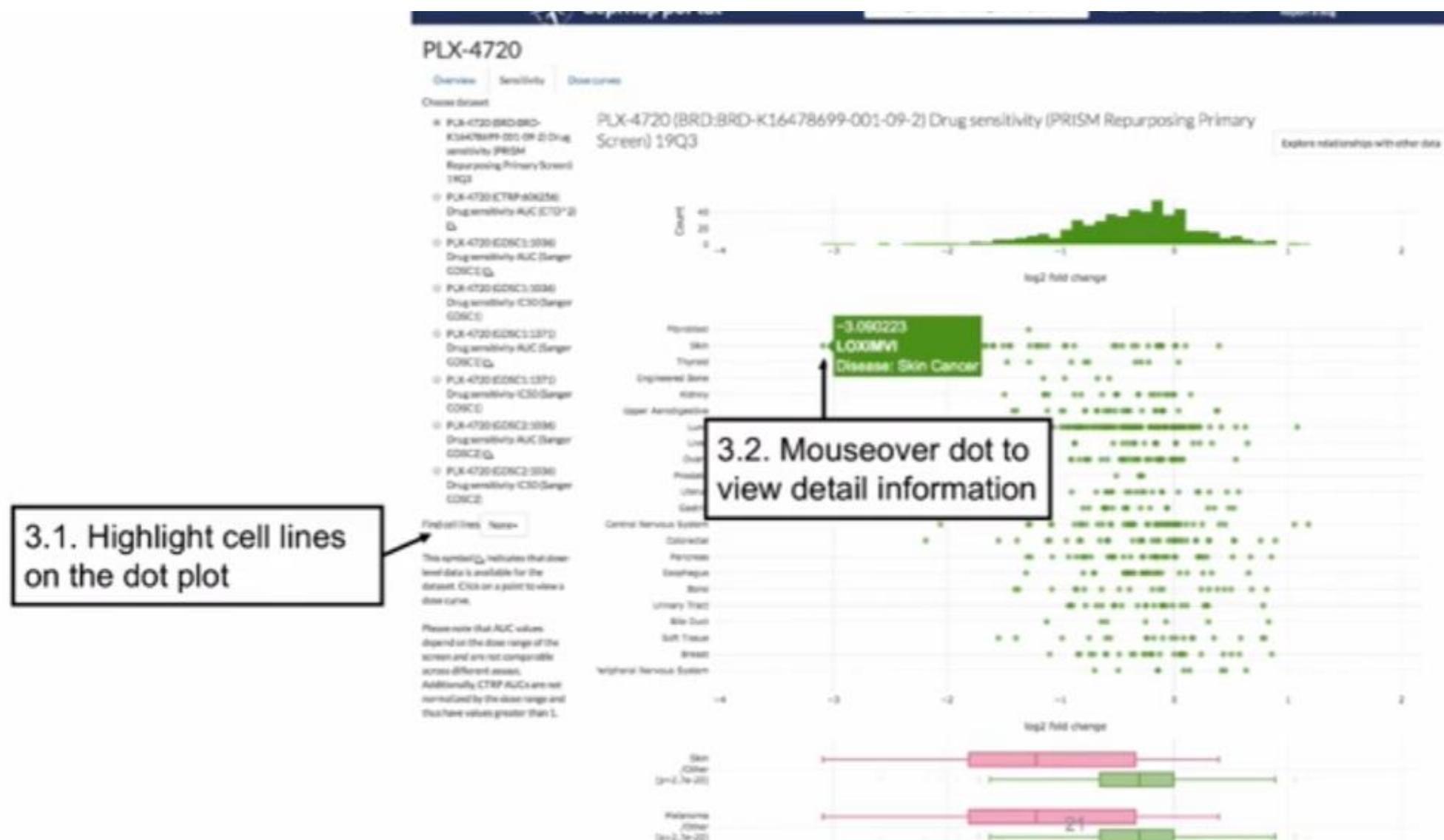
CONNECT

- [Email](#)
- [GitHub](#)
- [Twitter](#)
- [Facebook](#)

Depmap tutorial

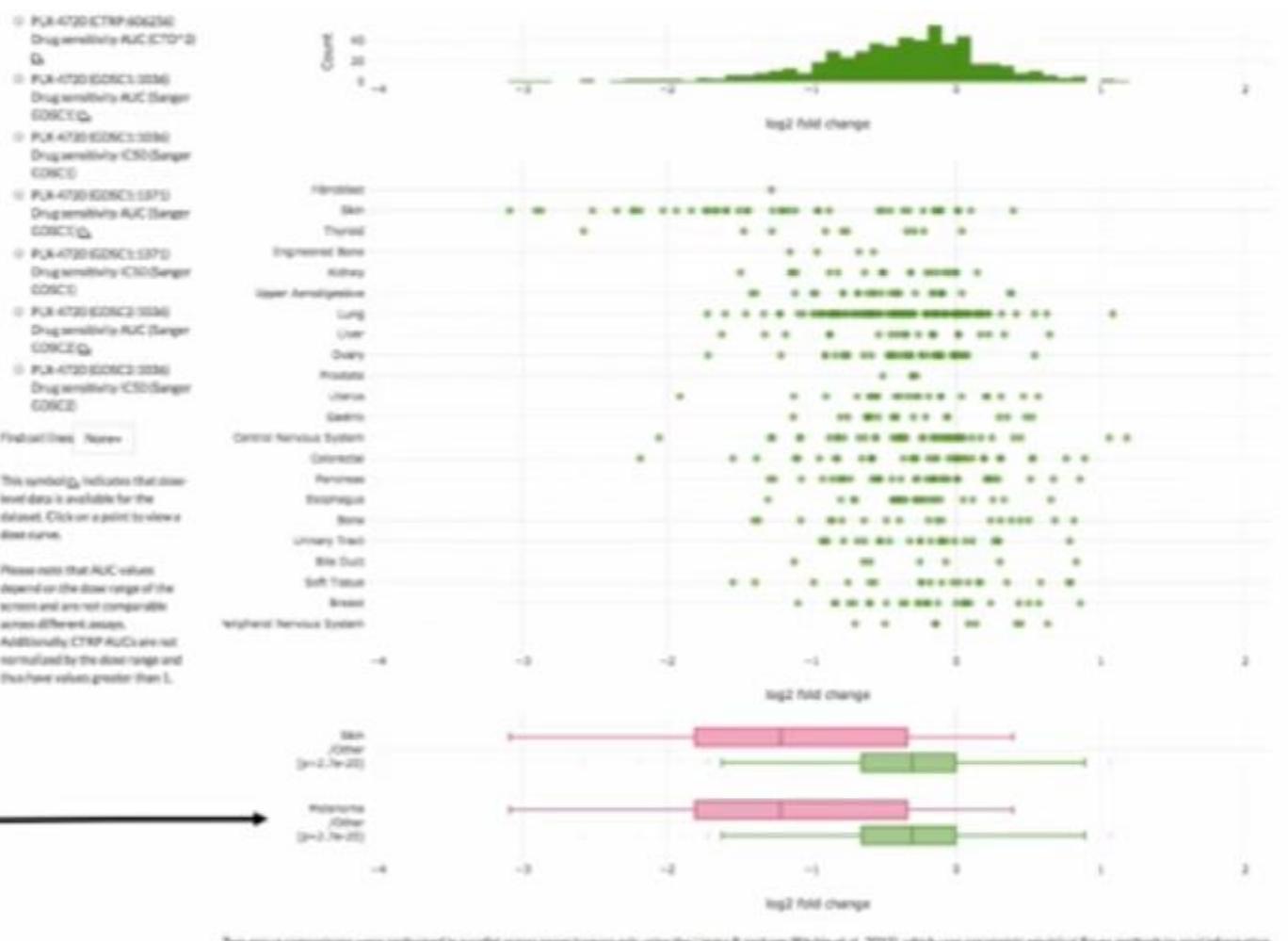


Depmap tutorial



Depmap tutorial

4. Box plot shows the enriched lineages that are more sensitive to the compound compared to other lineages.



Depmap tutorial

1. Explore relationships between data using “Data Explorer” tool

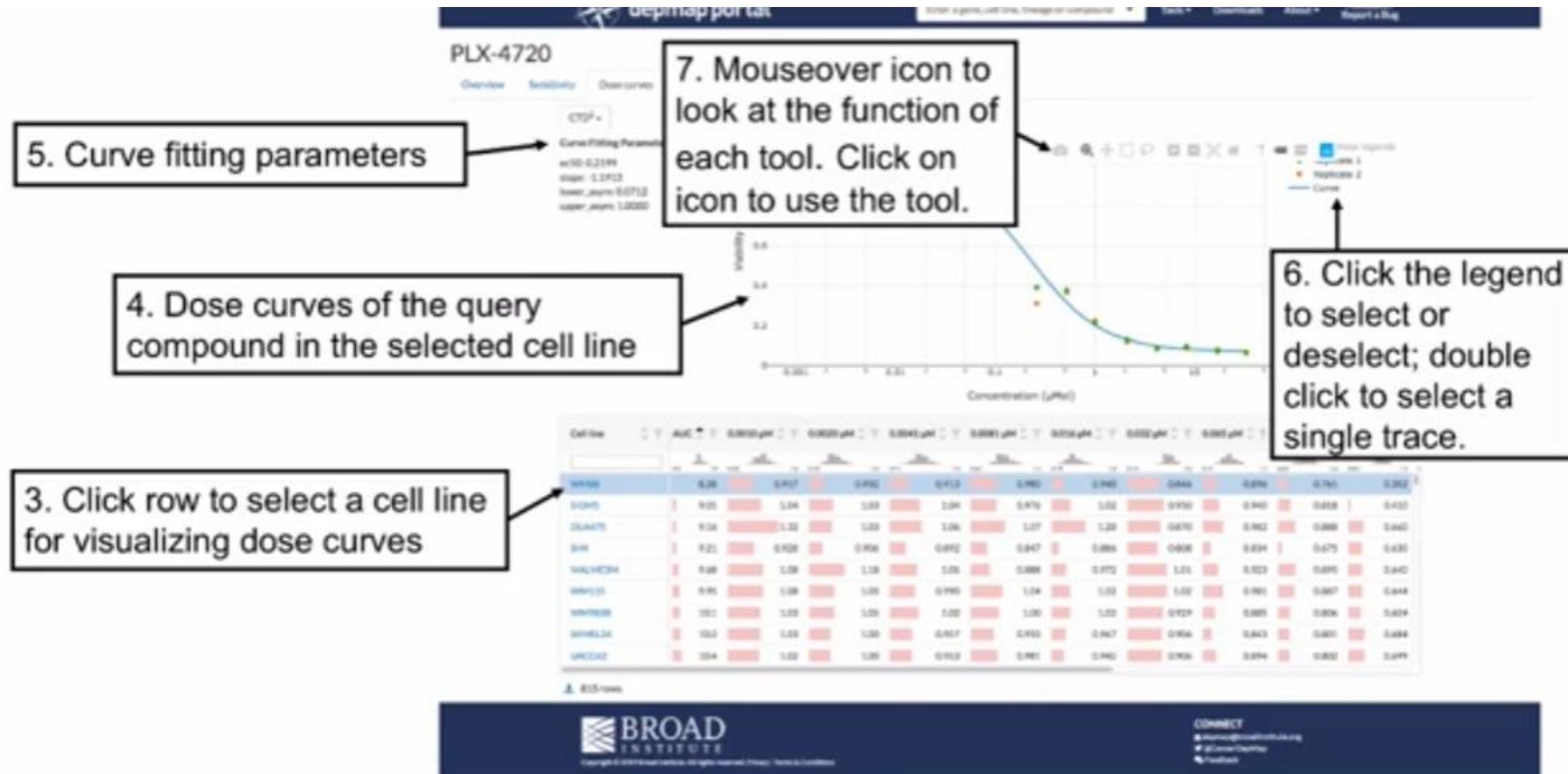
2. Mouseover icon to look at the function of each tool. Click on icon to use the tool.



Depmap tutorial



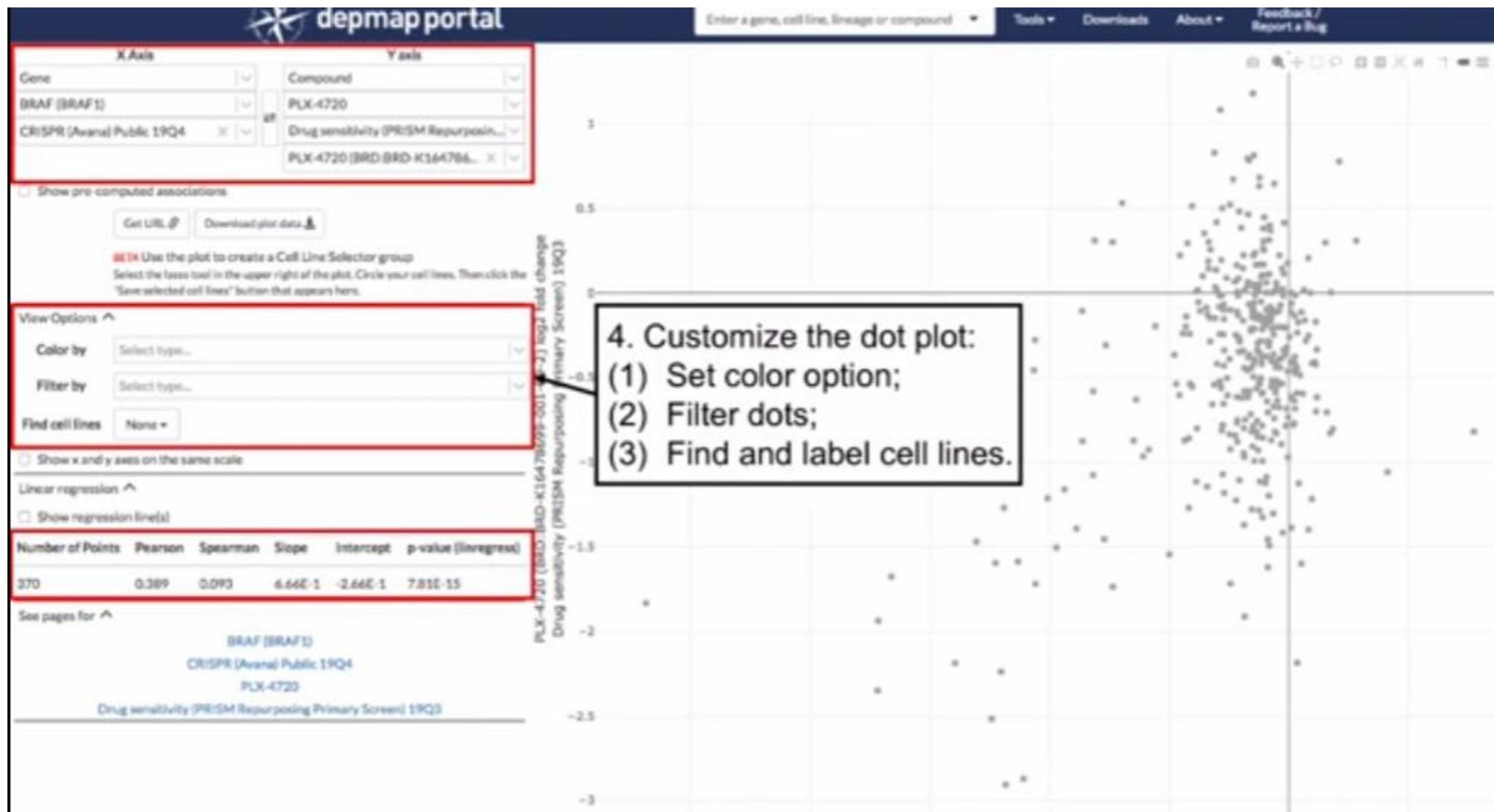
Depmap tutorial



Tutorial #3: Tool – Data Explorer

1. EXPLORE Relationships between characterizations

Depmap tutorial



목 차

1. 암유전체 네트워크 분석

- cBioPortal 소개 및 활용
- CCLE 소개 및 활용

2. 단일변이와 희귀변이 연관분석

- 단일변이 연관분석
- 희귀변이 분석

GENOME DATA ANALYSIS

유전체 데이터 분석

II NGS편, 암과 질병 유전체

chapter 4 NGS를 이용한 암 유전체 데이터 분석

제13장 암 유전체 네트워크 분석과 생존 분석 ...255

chapter 6 NGS를 이용한 희귀질환과 복잡질환 유전체 데이터 분석

제24장 희귀 유전질환과 흔한 복잡질환의 유전체학 ...490

제25장 복잡 질환의 희귀변이 연관분석 ...516

목 차

1. 암유전체 네트워크 분석

- cBioPortal 소개 및 활용
- CCLE 소개 및 활용

2. 단일변이와 희귀변이 연관분석

- 단일변이 연관분석
- 희귀변이 분석

연관분석(Association study)

- 표현형-유전자형 연관성에 대해 통계적 유의성을 판단하는 분석법.
- 유전체의 SNP 좌위에 대해 둘 이상의 표현형을 갖는 집단(population)에 대하여 모든 개체로부터 얻은 유전자형(genotype) 정보를 이용해 집단의 표현형과 유전자형의 연관성을 검정. 크게 single SNP level, haplotype에 대한 검정이 가능
- SNP은 질병의 진단목적 이외에도 유전적 질환 추적에 이용될 수 있으며, SNP이 유전자 내 regulatory region에 존재하는 경우 유전자의 기능, 단백질의 생물학적 기능을 변화시켜 직접적으로 질병을 유발시킬 수 있음.

연관분석(Association study)

1. Statistical tests for association

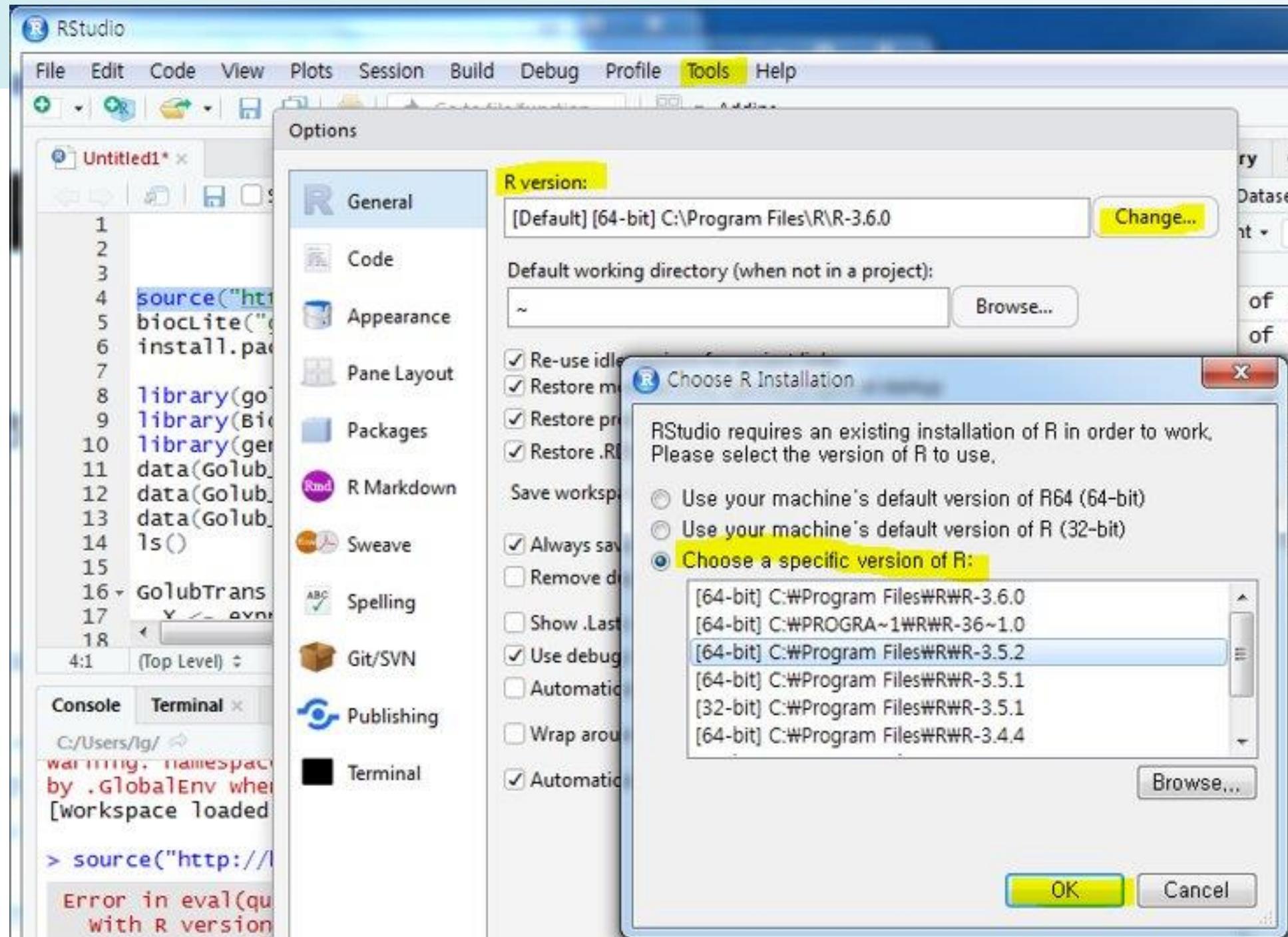
- 1) Fisher's exact test and Chi-squared test
- 2) Odds Ratio (OR)
- 3) Cochran-Armitage Trend Test (CATT)
- 4) Regression methods
- 5) Hardy-Weinberg Equilibrium (HWE) test
- 6) MANHATTAN Plots

2. Rare variant association tests

- 1) Burden tests
- 2) SKAT test
- 3) SKAT-O test

3. LD calculation based on 1000 Genomes data

4. HLA imputation



강의자료 다운로드

https://github.com/sun_snu/GDA_2024

Virtualbox 이용 시 입력하기

source activate lecture2

R

패키지 설치

```
install.packages("SNPassoc")
install.packages("coin")
install.packages("qqman")
install.packages("devtools")
install.packages("seqMeta")
library(SNPassoc)
library(coin)
library(qqman)
data(SNPs)
library(seqMeta)
data(seqMetaExample)
```

R STUDIO 단축키

- # 1. 코드실행
- # ctrl + enter
- # 2. 소스 저장
- # ctrl + s
- # 3. command 창 지우기
- # Ctrl+L
- # 4. 주석 처리/해제
- # 해당 라인에 커서를 두고 ctrl + shift + c
- # 5. 함수 또는 R 소스파일의 내용보기
- # 확인하려는 함수 또는 R 소스파일에 커서를 올리고 F2
- # 6. 실행중인 명령어 중지
- # ESC
- # 7. 이전 실행 명령어 창에 띄우기
- # UP / DOWN
- # 8. 이전 실행 명령어 확인
- # CTRL + UP

Single Nucleotide Polymorphisms (SNPs)

- 인간의 유전자는 99.9% 유사하고, 나머지의 0.1%가 특정의 질환에 관한 감수성이나 약제에 대한 부작용 등의 개체 차이에 관여하고 있다고 추정됨. 이 0.1%의 부분에 포함되는 유전자 다형의 대부분을 차지하는 것이 SNP.
- SNPs는 2 개의 대립유전자형 (bi-allele)이 서로 조합을 이루어 존재하는 유전변이형으로, 출현빈도가 높고 genome을 탐색하는데 필요한 신뢰성이 높은 유전자변형으로 알려져 있음.
- 290 base pair (bp)당 하나의 SNP가 존재하는 것으로 추정-> 인간 유전체 전체 염기서열 30억 개에 존재하는 SNP는 약 1,000만 개가 있을 것으로 추정. (Kruglyak and Nickerson)
- “Polymorphism”: 인구의 1% 이상에서 발생하는 변이 (1% 이하: “mutation”)

Statistical tests for association

1) Fisher's Exact Test

- Fisher's exact test는 실험군과 대조군의 allele count를 비교하여 연관성을 검증.

2) Chi-square Test

- Fisher's exact test뿐만 아니라 Pearson's Chi-squared test도 많이 사용됨. Chi-squared test는 2×2 contingency table에서 기대빈도 값이 최소 5를 넘어야 함.

3) Cochran-Armitage Trend Test (CATT)

- 성공률이 증가하거나 혹은 감소하는 일정 방향이 있다는 대립가설에 대해서 모집단의 성공 확률이 같은지 다른지를 검정.

Statistical tests for association

4) Regression Methods

(1) Linear regression

$y = a_1x_1 + a_2x_2 + \dots + b$ 꼴로 나타내는 선형방정식 구하기

coefficient: 기울기, intercept : y절편

ex) x (snp10001의 변이) 1개 추가-> y(blood.pre) 값이

coefficient(0.10222)만큼 증가

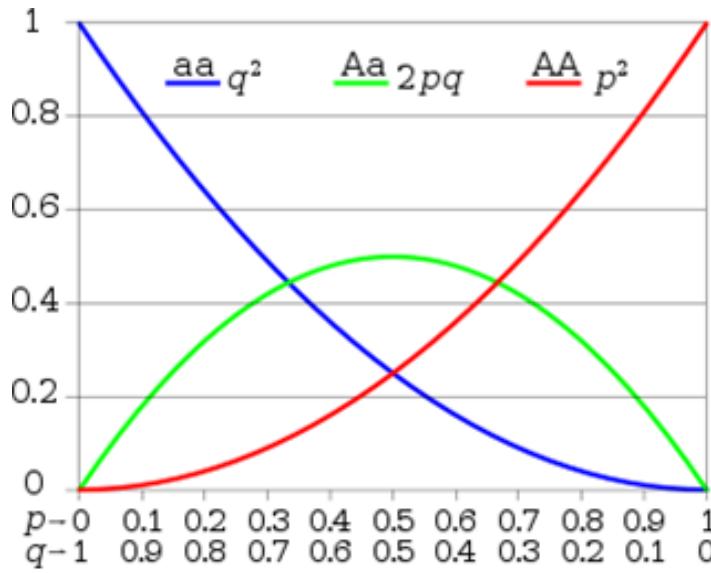
(2) Logistic regression

$$\log_e\left(\frac{y}{1-y}\right) = a_1x_1 + a_2x_2 + \dots + b$$

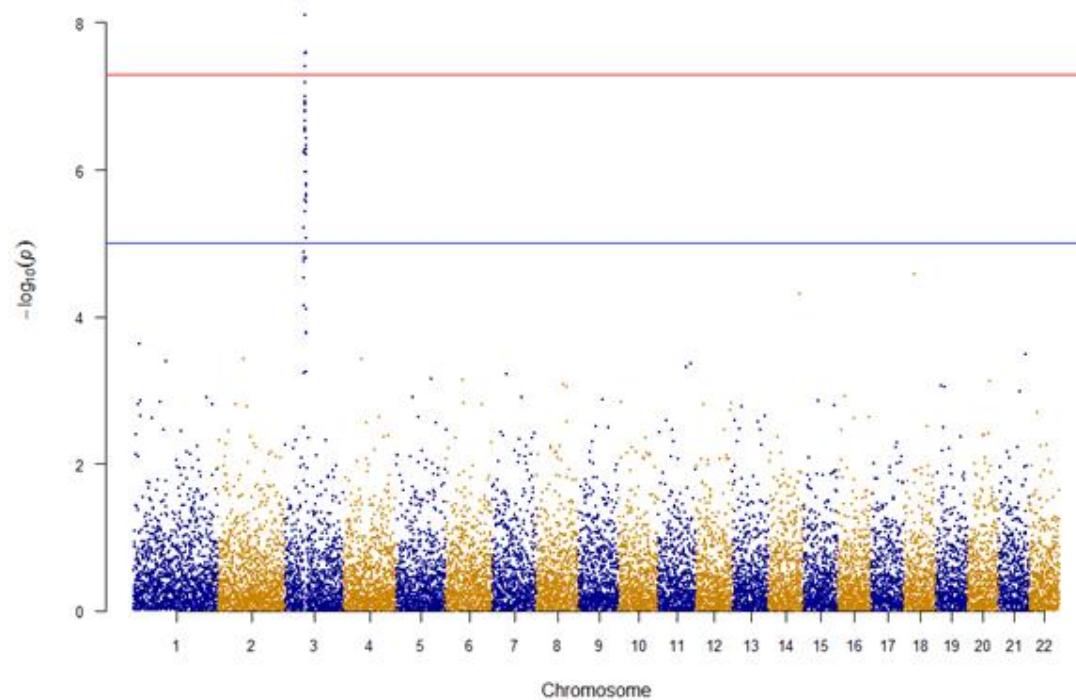
ex) x (snp10001의 변이) 1개 추가 -> case가 될 odds ratio가 coefficient(-0.1447)의 exponential 값(0.8653)

Statistical tests for association

5) Hardy-Weinberg Equilibrium (HWE)



6) MANHATTAN plot for GWAS result



Rare variant association test

- 지난 10년 간 복잡한 질병과 양적 특성의 유전적 구조를 해부하기 위해 GWAS (genome-wide association study)가 광범위하게 사용. 이러한 연구는 MAF (minor allele frequency)가 5% 이상인 일반적인 유전 변이 (Common variant or SNP)를 평가하기 위해서 앞서 살펴본 통계분석을 사용.
- GWAS는 주로 SNP에 초점을 맞추기 때문에, 희귀변이가 앞서 GWAS에서 설명하지 못했던 질병요인 및 특성을 설명해줄 것이라고 여겨지고 있으나 희귀변이에 대한 single-variant test는 샘플의 수를 확보하기 매우 어려워 통계적인 파워가 낮은 편.
- 따라서 희귀 변이는 주로 여러 변이를 유전자 혹은 특정 지역을 기반으로 통합하여 접근하는 방식이 주로 사용.

Burden Test

- A test collapses the variant data within a region by summing the minor allele counts for each marker in the region, and testing this against the phenotype.
- By contrast to [Count Number of Variants \(Per Gene\)](#), the counts are usually weighted by a function of each marker's minor allele frequency (MAF), so as to establish a contrast between rare and common variants.

$$Q_{\text{burden}} = \left(\sum_{j=1}^m w_j S_j \right)^2$$

Sequence Kernel Association Test (SKAT)

- SKAT is a SNP-set (e.g., a gene or a region) level test for association between a set of rare (or common) variants and dichotomous or quantitative phenotypes, SKAT aggregates individual score test statistics of SNPs in a SNP set and efficiently computes SNP-set level p-values.
- It collapses the variant data within a region by summing the squares of score statistics for testing individual markers. Weights based on each marker's MAF are usually used to establish a contrast between rare and common variants.

$$Q_{\text{SKAT}} = \sum_{j=1}^m w_j^2 S_j^2$$

Optimized SKAT (SKAT-O)

- a procedure which optimizes Generalized SKAT over a grid of N values of p between zero and 1, inclusive, in such a way as to count as only one test for multiple testing purposes instead of as N tests. (In Golden Helix SVS, seven grid points are used ($N = 7$), so we are talking about avoiding having to multiply the number of tests by 7 to get a proper multiple testing correction.)

$$Q_\rho = (1 - \rho)Q_{\text{SKAT}} + \rho Q_{\text{burden}}, \quad 0 \leq \rho \leq 1$$

참 고 자 료

Fisher's Exact test

- Fisher's exact test는 모수적 방법으로 실험군과 대조군의 allele count를 비교하여 연관성을 검증한다. Reference allele을 'A', alternative allele을 'G'라고 가정했을 때, 아래와 같이 allele별 실험군과 대조군의 빈도를 2×2 테이블로 구성하여 검증한다
- -> `fisher.test(matrix(c(12, 10, 26, 66), nrow=2))`

Allele	G	A	Total
Case	12	10	22
Control	26	66	92
Total	38	76	114

PLINK를 이용한 연관분석

PLINK

- 표현형-유전자형 연관분석을 위해 일반적으로 사용되는 프로그램. Case/control 비교와 같은 일반적인 분석부터 셋 이상의 표현형 집단에 대해 정량, 정성데이터와 유전자형 간 연관성을 볼 수 있음. 병렬처리를 통한 빠른 속도를 장점

SNP2HLA

- SNP2HLA 알고리즘: HLA 유전자 사이에 있는 intergenic SNP 정보를 이용하여, reference data와 비교함으로써, HLA 유전형의 정보를 예측.

연관불평형(Linkage disequilibrium)

- 한 염색체에 위치하는 2개의 SNPs간 거리가 아주 멀다면, 감수분열 과정에서 교차(Crossover)가 빈번히 발생하는데 이때 재조합은 단일염기수준이 아니라 큰 블록의 단위로 일어나게 되며 이렇게 함께 유전되는 단위를 Haplotype
- 재조합의 일어남으로써 서로 다른 조합의 haplotype이 독립적으로 발생하게 되는데 이처럼 독립적으로 가능한 조합이 모두 발생하는 경우 두 SNPs는 linkage equilibrium(LE, 연관평형) 상태. 그러나 SNPs간의 거리가 매우 가깝다면 2개의 SNPs는 서로 연관되어 다음 세대에 같이 전달되게 되고(Gabriel SB, 2002), 이때 haplotype의 조합이 모두 존재하지 않는다면 linkage disequilibrium(LD, 연관불평형) 상태. 이런 이론을 바탕으로 LD는 한 집단에서 2개 이상의 Loci에 존재하는 유전자 간의 통계적 관련성을 나타내는 하나의 지표로 사용.
- 양적형질 유전자 발굴을 위한 LD mapping 연구 흐름: 유전체 상의 표지인자의 유전자형을 이용하여 LD block 구조를 분석. 각 LD block에 존재하는 haplotype의 규명. 대표적인 마커가 되는 htSNP (Haplotype-tagging SNP) 혹은 다른 tagging 표지인자를 찾아내어 통계모형으로 표현형과 유전자형간의 통계모형으로 연관분석을 수행하여 형질에 관련된 유의한 LD의 위치를 찾기

*.PED 파일

Column 1 = Family ID

Column 2 = Individual ID

Column 3 = Paternal ID (0 인 경우는 missing)

Column 4 = Maternal ID (0 인 경우는 missing)

Column 5 = Sex

Column 6 = Phenotype (1,2, 또는 0 으로 구분. [1=unaffected, 2=affected, 0=missing])

Column 7+8 = SNP1의 genotype pair (0 인 경우는 missing)

Column 9+10 = SNP2의 genotype pair (0 인 경우는 missing)

....

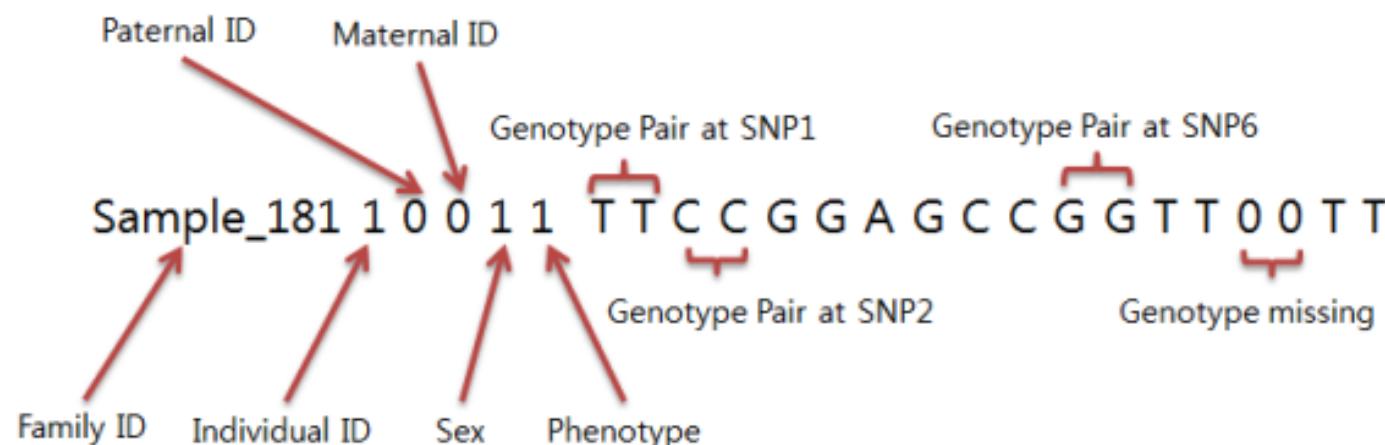


그림 PED 파일 형식

*.MAP

Column 1 = chromosome number

Column 2 = SNP ID

Column 3 = Genetic Distance (morgans)

Column 4 = physical base-pair position (bp)

따라서 2개의 파일은 아래와 같은 관계를 갖고 있다.

genotype.ped

Sample_181 1 0 0 1 1 T T C C G G A G C C G G T T 0 0 T T

genotype.map

chrNo, marker id, genetic distance(morgan), physical base-pair position(bp)

1 rs6681049 0 789870
1 rs4074137 0 1016570
1 rs7540009 0 1050098
1 rs1891905 0 1090080
1 rs9729550 0 1125105
1 rs3813196 0 1159244
1 rs6704013 0 1187454
1 rs307347 0 1250623
1 rs9439440 0 1441632

<--- *normal.ped* --->

1	1	0	0	1	1	A A	G T
2	1	0	0	1	1	A C	T G
3	1	0	0	1	1	C C	G G
4	1	0	0	1	2	A C	T T
5	1	0	0	1	2	C C	G T
6	1	0	0	1	2	C C	T T

<--- *normal.map* --->

1	snp1	0	5000650
1	snp2	0	5000830

would be represented as TPED/TFAM files:

<--- *trans.tped* --->

1	snp1	0	5000650	A A A C C C A C C C C C
1	snp2	0	5000830	G T G T G G T T G T T T

<--- *trans.tfam* --->

1	1	0	0	1	1
2	1	0	0	1	1
3	1	0	0	1	1
4	1	0	0	1	2
5	1	0	0	1	2
6	1	0	0	1	2

- TPED: one row -> SNP / TFAM: one row -> individual

*.BIM 파일

- Chromosome • Marker ID • Genetic distance • Physical position • Allele 1 • Allele 2

Example of a BIM file of the binary PLINK format:

• 21	rs115116470	26765	A	T
• X	rs3883674	0	C	G
• X	rs122188820	48172	T	T
• 9	rs109040450	48426	A	T
• 9	rs107519310	49949	C	T
• 8	rs112521270	52087	A	C
• 10	rs127752030	52277	A	A
• 8	rs122556190	52481	G	T

*.FAM 파일

- Family ID • Sample ID • Paternal ID • Maternal ID • Sex
(1=male; 2=female; other=unknown) • Affection (Phenotype)
(0, -9= missing, 1=unaffected, 2=affected)

EAS_JPT_NA18939_F	EAS_JPT_NA18939_F	0	0	0	-9
EAS_JPT_NA18940_M	EAS_JPT_NA18940_M	0	0	0	-9
EAS_JPT_NA18941_F	EAS_JPT_NA18941_F	0	0	0	-9

Linkage disequilibrium 지수

(1) R-square : Linkage disequilibrium 지수

- Pairwise SNP 간에 계산된 correlation coefficient($=r$)을 제곱한 값이 지정한 값보다 큰 SNP만을 tagging SNP로 간주
- 0에서 1 사이의 범위로 0일 때 perfect equilibrium, 1일 때 완전 LD상태를 의미, 보통 1/3 이상의 값이면 강한 LD 상태.

(2) D' : Linkage disequilibrium 지수

- D' 값이 1이면 완전 LD 상태, $D' < 1$ 이면 이전 세대 어디에선가 유전자 재조합이 일어났음을 의미, D' 값이 0이면 연관 평형 상태.
- D' 값이 낮을수록 두 site 간에 강한 유전자 재조합과 돌연변이가 빈번하게 나타났음을 의미.

R² vs D'

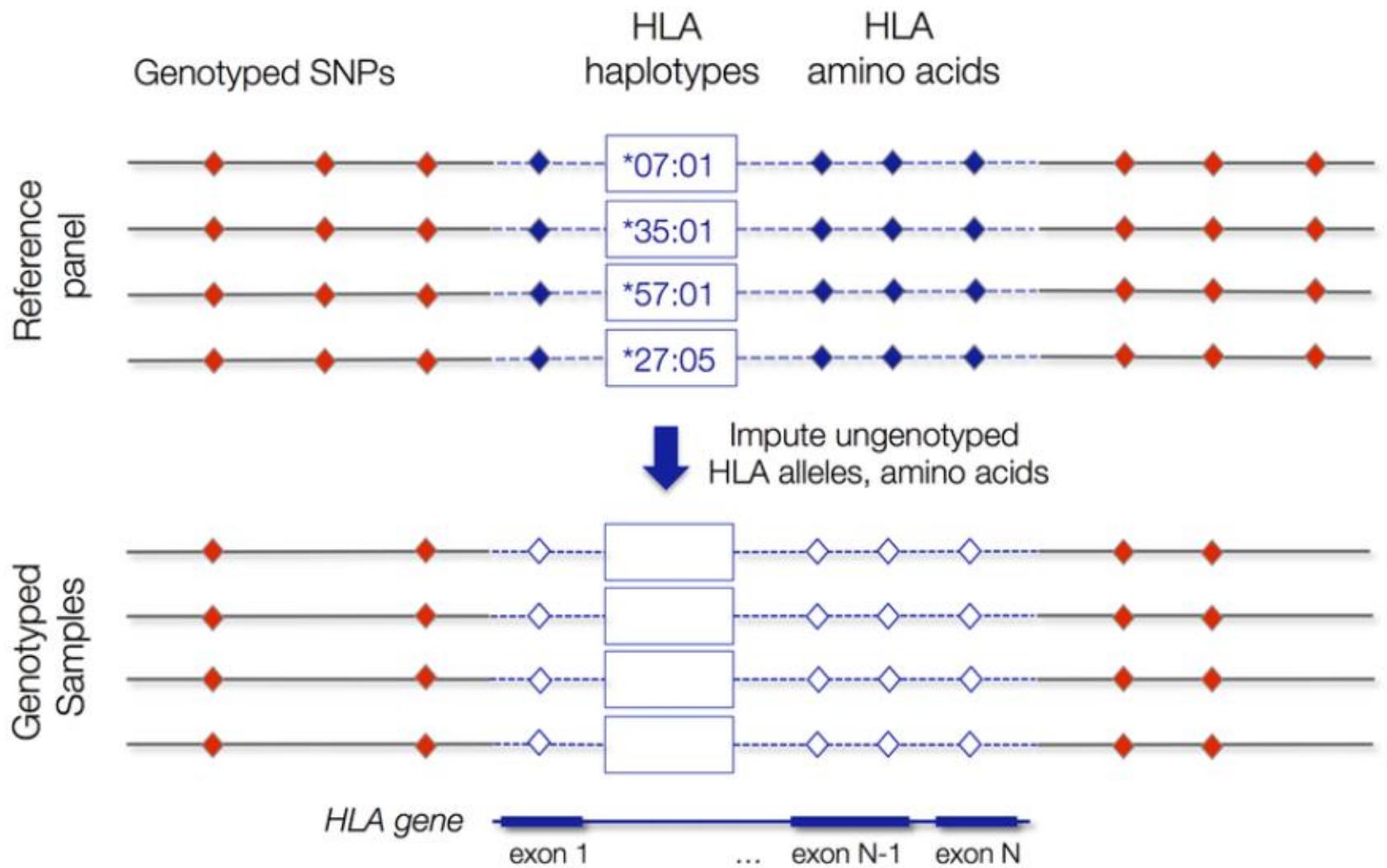
- R² -> successfully identify LD for common variant
- But drop with low allele frequencies -> unlikely deal with rare alleles
- D' -> theoretically can have better chance of identifying LD among rare variants (>R²)
- But may accumulate more false positives

```
plink --bfile JPT --chr 16 --from-bp 54000000 --to-bp 54050000 --  
ld-snp rs11646260 --r2 --ld-window-r2 0 --ld-window 999999 --ld-  
window-kb 99999 --noweb --out ./JPT
```

- --bfile : 이전 단계에서 생성한 binary file을 입력/ --chr: chromosome 위치
- --from-bp : 시작 base pair의 위치/ --to-bp : 끝 base pair의 위치
- --ld-snp : ld를 계산할 기준이 되는 snp를 입력
- --r2 : phased haplotypes을 이용하여 r2값을 계산
- --ld-window-r2 : 결과를 filtering 할 때 사용하는 r2값을 의미
- --ld-window : 결과를 filtering 할 때 사용하는 LD값을 계산하기 위해
사용된 SNP 들 사이의 최대 SNP 수
- --ld-window-kb : 결과를 filtering 할 때 사용하는 LD값을 계산하기 위해
사용된 SNP 사이의 최대 물리적 거리 (kb)

SNP2HLA: Imputation of Amino Acid Polymorphisms in Human Leukocyte Antigens

- SNP2HLA is a tool to impute amino acid polymorphisms and single nucleotide polymorphisms in human leukocyte antigens (HLA) within the major histocompatibility complex (MHC) region in chromosome 6.
- The unique feature of SNP2HLA is that it imputes not only the classical HLA alleles but also the amino acid sequences of those classical alleles, so that individual amino acid sites can be directly tested for association. This allows for facile amino-acid focused downstream analysis.
- SNP2HLA 알고리즘: HLA 유전자 사이에 있는 intergenic SNP 정보를 이용하여, reference data와 비교함으로써, HLA 유전형의 정보를 예측.
- 실습데이터: 1958년생 영국인 10명 샘플 (1958BC), reference: 1000 Genomes 의 CEU (HM_CEU_REF; Utah residents)



Overview of the SNP2HLA imputation procedure. The reference panel (top) contains SNPs in the MHC, classical HLA alleles at the class I and class II loci, and amino acid sequences corresponding to the 4-digit HLA types at each locus. For a data set with genotyped SNPs across the MHC (bottom), we use the reference panel to impute classical alleles and their corresponding amino acid polymorphisms.

./SNP2HLA.csh 1958BC HM_CEU_REF 1958BC_IMPUTED plink
2000 1000

-> CEU reference data 에 맞춰 10명 샘플 (1958BC) imputation

- 1958BC_IMPUTED.dosage: 모든 markers (HLA alleles, amino acids, SNPs)로부터 top statistical peak를 찾기 위한 imputation을 수행한 allele dosage data 이다. Dosage 파일은 Beagle 프로그램의 포맷으로, 행은 markers, 열은 marker의 정보와 개인별 imputation 값 나타냄
- 1958BC_IMPUTED.bgl.phased: genotype의 imputation 수행한 파일. Beagle 포맷으로, 행은 markers 정보를 나타내며 2개의 열이 한 개인의 genotype 정보를 나타냄
- 1958BC_IMPUTED.bgl.gprobs: markers (HLA alleles, amino acids, SNPs)에 대한 imputation posterior probabilities를 계산한 결과
- 1958BC_IMPUTED.bgl.r2: genotype을 이용하여 impute predicted r2 (correlation)를 계산한 결과

```
plink --noweb --dosage 1958BC_IMPUTED.dosage  
noheader format=1 --fam 1958BC_IMPUTED.fam --  
logistic --out 1958_IMPUTED
```

--noweb: 웹 접속을 차단하고 local 컴퓨터에서 실행.

--dosage OUTPUT.dosage: OUTPUT.dosage 파일을 생성.

noheader: header가 존재하지 않을 경우 사용하는 옵션

format=N: Dosage, two probabilities or three (N=1,2,3)

--fam OUTPUT.fam: OUTPUT.fam 파일을 생성.

--logistic: disease traits에 대해 logistic regression model 계산

--out OUTPUT.assoc: 결과 파일 OUTPUT.assoc 파일을 생산

.assoc.dosage 필드정보

SNP	A1	A2	FRQ	INFO	OR	SE	P
rs13207673	A	G	0.5251	0.0132	NA	NA	NA
rs9356991	T	G	0.7026	0.0216	NA	NA	NA

- SNP: SNP identifier,
- A1: Tested allele (minor allele by default),
- A2: Major allele,
- FRQ : Frequency of A1 from dosage data
- INFO : R-squared quality metric/ Information content
- OR: Odds ratio for association
- SE: Standard error of effect estimate
- P: P-value for association test