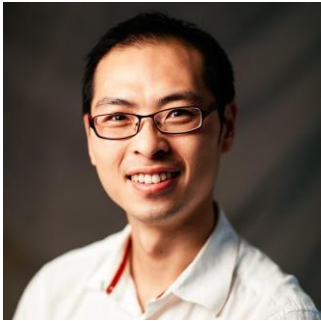


# Open Data Resources for Human Genomics Research



A/Prof Jason Wong  
Head, Computational Cancer Genomics  
School of Biomedical Sciences  
University of Hong Kong



Dr Rebecca Poulos  
NHMRC Early Career Fellow  
Children's Medical Research Institute  
University of Sydney

## Part One

### Human genomics data resources

1:30pm – 3pm

- UCSC Genome Browser + Exercise
- gnomAd + Exercise
- GTEx + Exercise
- ENCODE database + Exercise
- Exercise: Putting it all together

## Part Two

### Cancer genomics data resources

3:30pm – 5pm

- Brief introduction to TCGA
- cBioPortal
- Genomic Data Commons
- Xena Browser
- Cancer Genomics Cloud

# Part Two

## Cancer genomics data resources

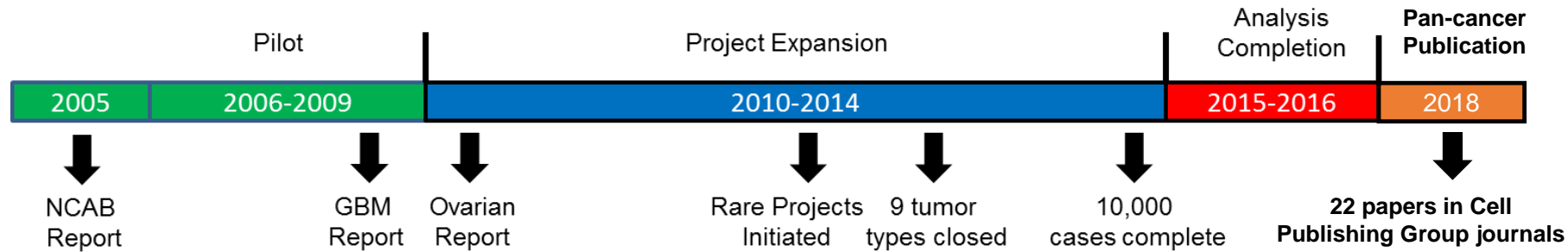
3:30pm – 5pm

- The Cancer Genome Atlas  
*Brief introduction to The Cancer Genome Atlas*
- cbioPortal  
*Looking up somatic mutations and any clinical associations*
- Genomic Data Commons  
*Accessing the raw and processed data*
- XENA browser  
*Downloading tidy processed data*
- Cancer Genomics Cloud  
*Accessing and analyzing data in the cloud*

*Some sections will be accompanied by an exercise to help you test your skills, plus an exercise at the end to put all your knowledge together.*

# The Cancer Genome Atlas

- Lunched in 2006 as a pilot and expanded in 2009



- 11,000 cancer patients across 33 cancer types

## NATIONAL CANCER INSTITUTE THE CANCER GENOME ATLAS

### TCGA BY THE NUMBERS

TCGA produced over  
**2.5**  
PETABYTES  
of data

To put this into perspective, 1 petabyte of data is equal to

**212,000**  
DVDs

TCGA data describes  
**33**  
DIFFERENT  
TUMOR TYPES

...including  
**10**  
RARE  
CANCERS

...based on paired tumor and normal tissue sets collected from  
**11,000**  
PATIENTS

...using  
**7**  
DIFFERENT  
DATA TYPES

### TCGA RESULTS & FINDINGS



MOLECULAR  
BASIS OF  
CANCER

Improved our understanding of the genomic underpinnings of cancer

For example, a TCGA study found the basal-like subtype of breast cancer to be similar to the serous subtype of ovarian cancer on a molecular level, suggesting that despite arising from different tissues in the body, these subtypes may share a common path of development and respond to similar therapeutic strategies.



TUMOR  
SUBTYPES

Revolutionized how cancer is classified

TCGA revolutionized how cancer is classified by identifying tumor subtypes with distinct sets of genomic alterations.\*



THERAPEUTIC  
TARGETS

Identified genomic characteristics of tumors that can be targeted with currently available therapies or used to help with drug development

TCGA's identification of targetable genomic alterations in lung squamous cell carcinoma led to NCI's Lung-MAP Trial, which will treat patients based on the specific genomic changes in their tumor.

### THE TEAM

**20**  
COLLABORATING  
INSTITUTIONS  
across the United States and Canada

### WHAT'S NEXT?

The Genomic Data Commons (GDC) houses TCGA and other NCI-generated data sets for scientists to access from anywhere. The GDC also has many expanded capabilities that will allow researchers to answer more clinically relevant questions with increased ease.



\*TCGA's analysis of stomach cancer revealed that it is not a single disease, but a disease composed of four subtypes, including a new subtype characterized by infection with Epstein-Barr virus.

[www.cancer.gov/ccg](http://www.cancer.gov/ccg)

# Types of TCGA data

- Core dataset:

- Pathology report
- Histology images
- Clinical data
- Whole exome-seq
- SNP 6.0 array
- mRNAseq
- miRNAseq
- Methylation array


- Optional datasets:

- 50x Whole-genome sequencing
- Bisulfide sequencing
- Protein Array
- Radiological images (MRI)

\* Also mRNA/methylation data for  
~10-20 matched normal tissue data

# cBioPortal

*Looking up somatic mutations in cancer samples*



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

[Query](#)
[Quick Search Beta!](#)
[Download](#)

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

Select Studies for Visualization & Analysis:

0 studies selected (0 samples)

Search...

PanCancer Studies

3

Cell lines

3

Adrenal Gland

3

Ampulla of Vater

1

Biliary Tract

9

Bladder/Urinary Tract

13

Bone

2

Bowel

10

Breast

15

CNS/Brain

18

Cervix

2

Esophagus/Stomach

14

Quick select:

TCGA PanCancer Atlas Studies

Curated set of non-redundant studies

PanCancer Studies

☐ MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017) 10945 samples
 ☐ Pan-Lung Cancer (TCGA, Nat Genet 2016) 1144 samples
 ☐ Pediatric Pan-cancer (Columbia U, Genome Med 2016) 103 samples

Cell lines

☐ Cancer Cell Line Encyclopedia (Broad, 2019) 1739 samples
 ☐ Cancer Cell Line Encyclopedia (Novartis/Broad, Nature 2012) 1020 samples
 ☐ NCI-60 Cell Lines (NCI, Cancer Res 2012) 67 samples

Adrenal Gland


Adrenocortical Carcinoma

☐ Adenoid Cystic Carcinoma Project (2019) 1049 samples
 ☐ Adrenocortical Carcinoma (TCGA, PanCancer Atlas) 92 samples
 ☐ Adrenocortical Carcinoma (TCGA, Provisional) 92 samples

Ampulla of Vater

What's New

@cbioportal



cBioPortal

@cbioportal

We have been busy recently releasing new cBioPortal features, thanks to our growing #opensource community: @sloan\_kettering @DanaFarber @ChildrensPhila @pmcancercentre @AstraZeneca @TheHyveNL. Waterfall plots, heatmaps, fusions, filters, quick links etc.: [cbioportal.org/news](http://cbioportal.org/news)

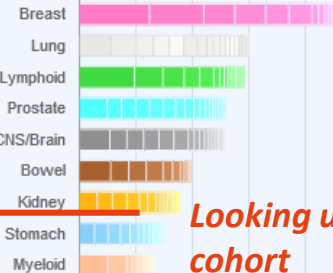
Sign up for low-volume email news alerts

Subscribe

Cancer Studies

The portal contains 273 cancer studies (details)

Cases by Top 20 Primary Sites




Looking up mutation profile of specific genes


Query By Gene


OR

Explore Selected Studies

Looking up mutation profile of a cohort


**CHILDREN'S MEDICAL RESEARCH INSTITUTE**  
 Jeans for Genes®


**THE UNIVERSITY OF SYDNEY**



香港大學  
**THE UNIVERSITY OF HONG KONG**

**AMSI BIOINFO SUMMER 19**  
 A SYMPOSIUM IN BIOINFORMATICS

QueryQuick Search Beta!Download

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

Select Studies for Visualization & Analysis:1 study selected (230 samples)Deselect all

Bone2

Bowel10

Breast15

CNS/Brain18

Cervix2

Esophagus/Stomach14

Eye3

Head and Neck13

Kidney17

Liver8

Lung21

Lymphoid19

Myeloid9

Other13

Ovary/Fallopian Tube4

Pancreas10

Peripheral Nervous System5

Thoracic PDX (MSK, Provisional)139 samples

Lung Neuroendocrine Tumor

→ SMALL CELL LUNG CANCER

Small Cell Lung Cancer (CLCGP, Nat Genet 2012)29 samples

Small Cell Lung Cancer (Johns Hopkins, Nat Genet 2012)51 samples

Small Cell Lung Cancer (U Cologne, Nature 2015)110 samples

Small-Cell Lung Cancer (Multi-Institute, Cancer Cell 2017)20 samples

Non-Small Cell Lung Cancer

Non-Small Cell Lung Cancer (MSK, Cancer Cell 2018)75 samples

Non-Small Cell Lung Cancer (MSKCC, J Clin Oncol 2018)240 samples

Non-Small Cell Lung Cancer (TRACERx, NEJM 2017)327 samples

Non-Small Cell Lung Cancer (University of Turin, Lung Cancer 2017)41 samples

Non-small cell lung cancer (MSK, Science 2015)16 samples

Pan-Lung Cancer (TCGA, Nat Genet 2016)1144 samples

→ LUNG ADENOCARCINOMA

Lung Adenocarcinoma (Broad, Cell 2012)183 samples

Lung Adenocarcinoma (MSKCC, Science 2015)35 samples

Lung Adenocarcinoma (TCGA, Nature 2014)230 samples

Lung Adenocarcinoma (TCGA, PanCancer Atlas)566 samples

Lung Adenocarcinoma (TCGA, Provisional)586 samples

Lung Adenocarcinoma (TSP, Nature 2008)163 samples

Non-Small Cell Cancer (MSKCC, Cancer Discov 2017)915 samples

→ LUNG SQUAMOUS CELL CARCINOMA

Lung Squamous Cell Carcinoma (TCGA, Nature 2012)178 samples

Lung Squamous Cell Carcinoma (TCGA, PanCancer Atlas)487 samples

Lung Squamous Cell Carcinoma (TCGA, Provisional)511 samples

1 study selected (230 samples)Deselect all

Query By GeneORExplore Selected Studies

Query
Quick Search **Beta!**
Download

Please cite: [Cerami et al., 2012](#) & [Gao et al., 2013](#)

Selected Studies: [Modify](#)
Lung Adenocarcinoma (TCGA, Nature 2014) (230 total samples)

Select Genomic Profiles:

☒ Mutations [?](#)
☐ Putative copy-number alterations from GISTIC [?](#)
☐ mRNA Expression z-Scores (RNA Seq V2 RSEM) [?](#)

Select Patient/Case Set:

To build your own case set, try out our enhanced Study View.
Samples with mutation data (230)

Enter Genes:

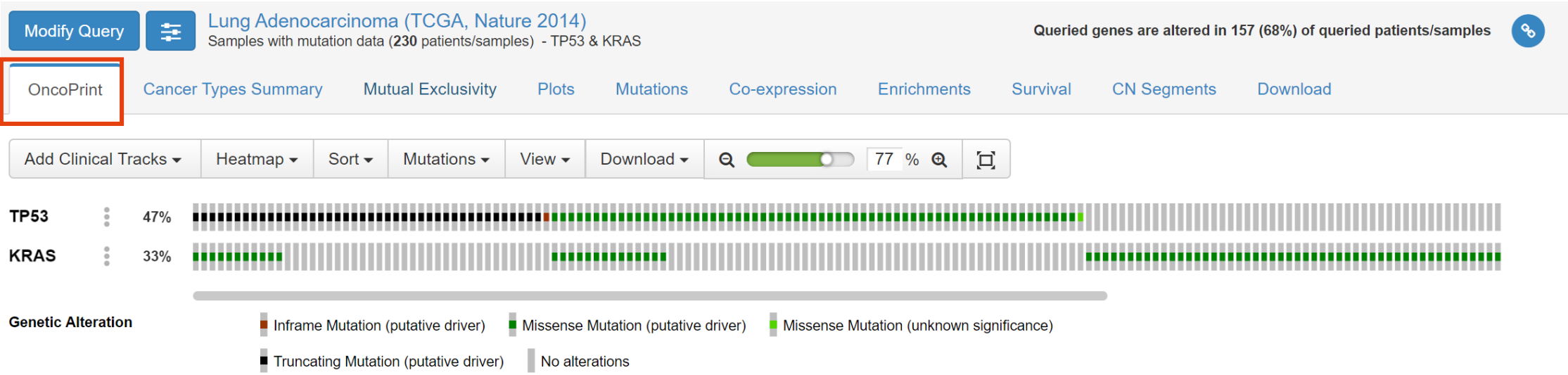
User-defined List

TP53  
KRAS

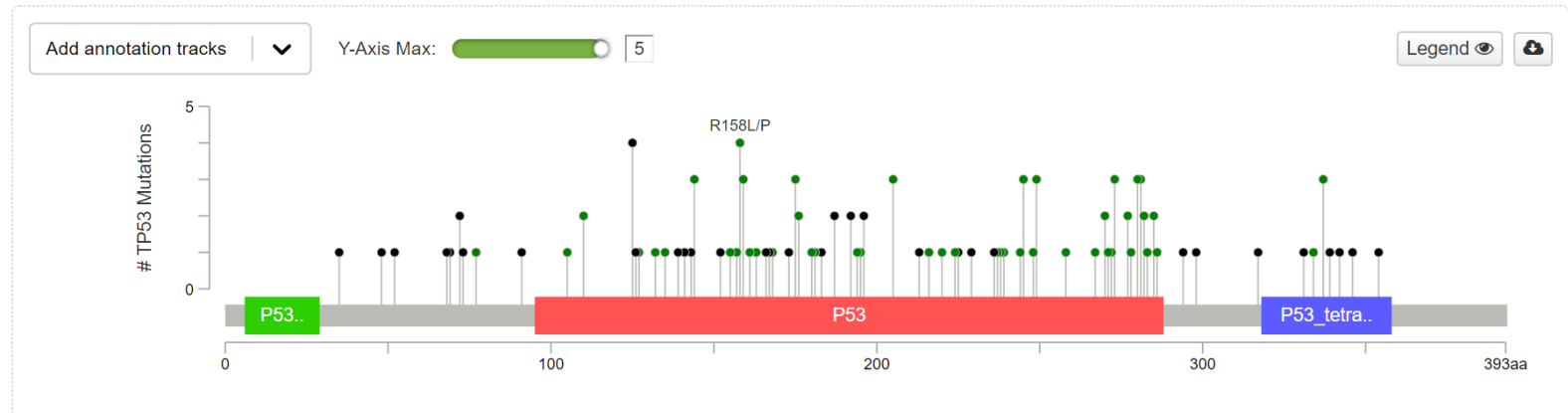
Type in "TP53" and "KRAS" one per line

All gene symbols are valid.

Submit Query



TP53 KRAS



**TP53**

RefSeq: [NM\\_000546](#)  
Ensembl: [ENST00000269305](#)  
CCDS: [CCDS11118](#)  
UniProt: [P53\\_HUMAN](#)

Somatic Mutation Frequency ⓘ 46.5%

73 Missense

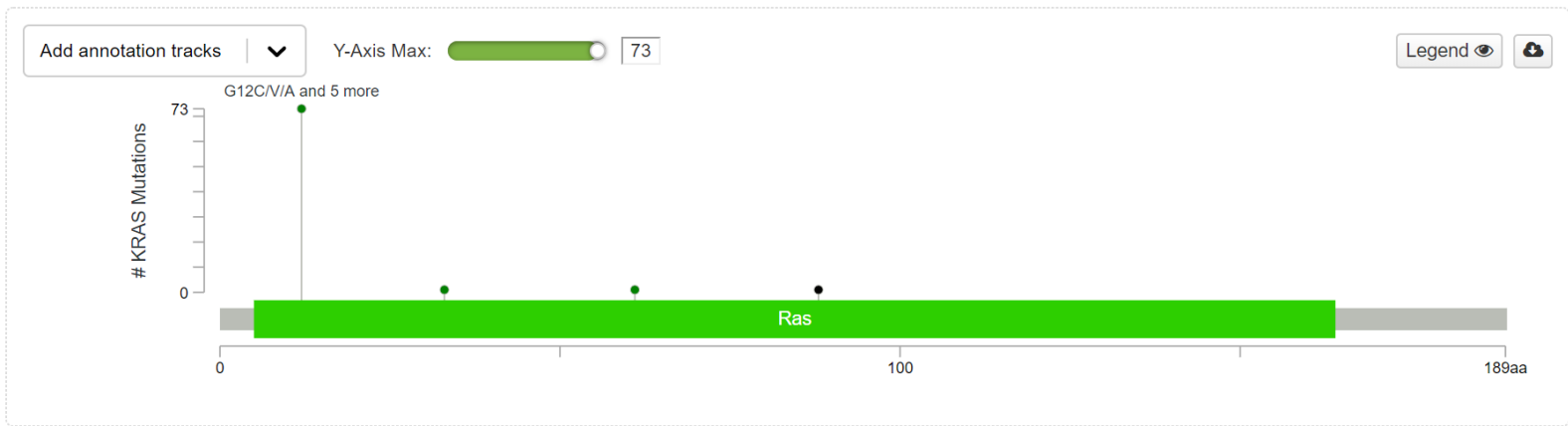
42 Truncating

1 Inframe

0 Other

View 3D Structure

TP53 KRAS



**KRAS**

RefSeq: [NM\\_033360](#)  
Ensembl: [ENST00000256078](#)  
CCDS: [CCDS8703](#)  
UniProt: [RASK\\_HUMAN](#)

Somatic Mutation Frequency ⓘ 32.6%

75 Missense

1 Truncating

0 Inframe

0 Other

View 3D Structure

Modify Query



Lung Adenocarcinoma (TCGA, Nature 2014)  
Samples with mutation data (230 patients/samples) - TP53 & KRAS

Queried genes are altered in 157 (68%) of queried patients/samples 

OncoPrint

Cancer Types Summary

Mutual Exclusivity

Plots

Mutations

Co-expression

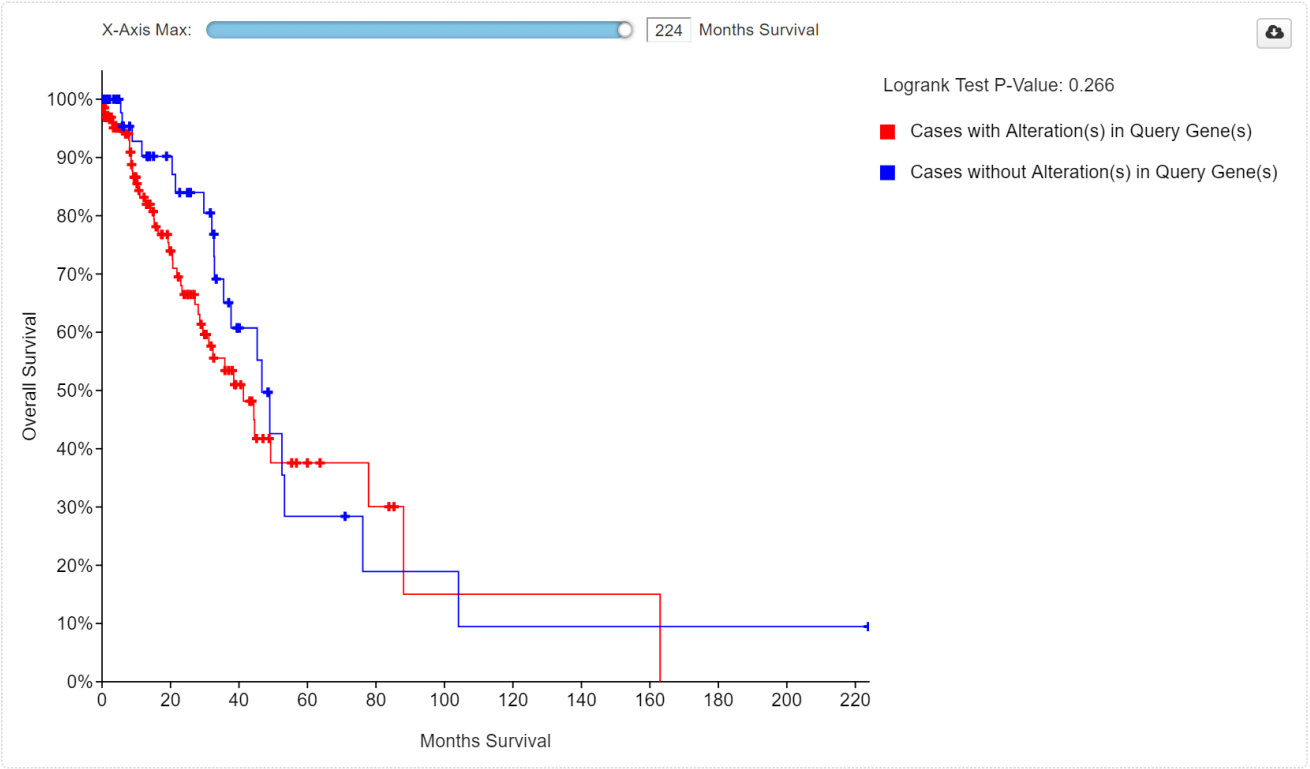
Enrichments

Survival

CN Segments

Download

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



	Number of Cases, Total	Number of Cases, Deceased	Median Months Survival
Cases with Alteration(s) in Query Gene(s)	140	44	41.2986093
Cases without Alteration(s) in Query Gene(s)	63	19	46.6868129

Select Studies for Visualization & Analysis:1 study selected (230 samples)Deselect all

Search...

Bowel10

Breast15

CNS/Brain18

Cervix2

Esophagus/Stomach14

Eye3

Head and Neck13

Kidney17

Liver8

**Lung21**

Lymphoid19

Myeloid9

Other13

Ovary/Fallopian Tube4

Pancreas10

☐ Select all listed studies matching filter (21)

Lung

☐ Thoracic PDX (MSK, Provisional)139 samples

Lung Neuroendocrine Tumor

→ SMALL CELL LUNG CANCER

☐ Small Cell Lung Cancer (CLCGP, Nat Genet 2012)29 samples

☐ Small Cell Lung Cancer (Johns Hopkins, Nat Genet 2012)51 samples

☐ Small Cell Lung Cancer (U Cologne, Nature 2015)110 samples

☐ Small-Cell Lung Cancer (Multi-Institute, Cancer Cell 2017)20 samples

Non-Small Cell Lung Cancer

☐ Non-Small Cell Lung Cancer (MSK, Cancer Cell 2018)75 samples

☐ Non-Small Cell Lung Cancer (MSKCC, J Clin Oncol 2018)240 samples

☐ Non-Small Cell Lung Cancer (TRACERx, NEJM 2017)327 samples

☐ Non-Small Cell Lung Cancer (University of Turin, Lung Cancer 2017)41 samples

☐ Non-small cell lung cancer (MSK, Science 2015)16 samples

☐ Pan-Lung Cancer (TCGA, Nat Genet 2016)1144 samples

→ LUNG ADENOCARCINOMA

☐ Lung Adenocarcinoma (Broad, Cell 2012)183 samples

☐ Lung Adenocarcinoma (MSKCC, Science 2015)35 samples

☒ Lung Adenocarcinoma (TCGA, Nature 2014)230 samples

☐ Lung Adenocarcinoma (TCGA, PanCancer Atlas)566 samples

☐ Lung Adenocarcinoma (TCGA, Provisional)586 samples

☐ Lung Adenocarcinoma (TSP, Nature 2008)163 samples

1 study selected (230 samples)Deselect all

Query By Gene

OR

Explore Selected Studies

Whole exome sequencing, messenger RNA, microRNA, copy number, methylation and proteomic analysis of 230 lung adenocarcinoma tumor/normal pairs. [Raw data via the TCGA Data Portal.](#) [PubMed](#)

Click gene symbols below or enter here

Query

Summary Clinical Data Heatmaps CN Segments

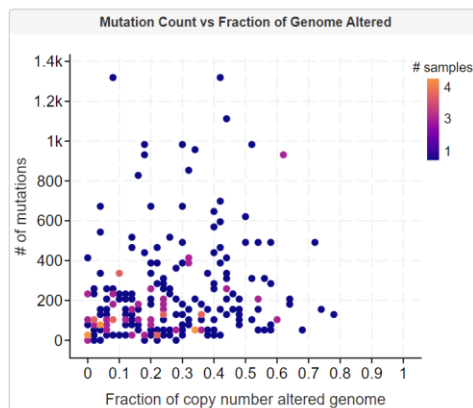
Selected: 230 patients | 230 samples

















Custom Selection ▾

Charts ▾

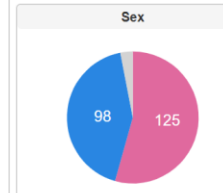
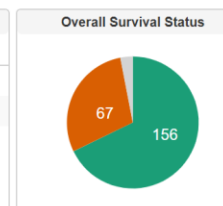
Groups ▾

Quick Filters: ☐ 230 samples with mutation data ☐ 230 samples with CNA data

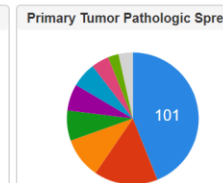
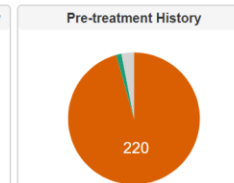
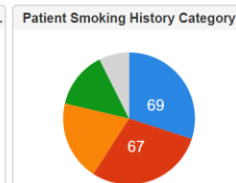
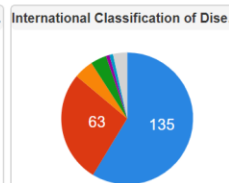
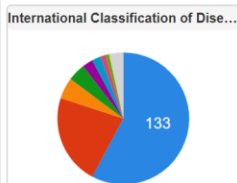
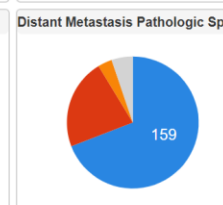
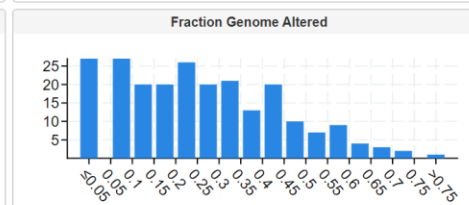
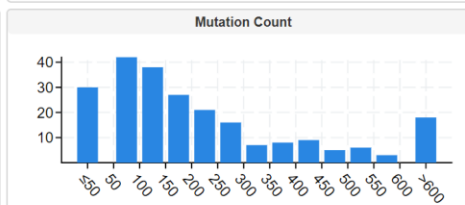


Mutated Genes (230 profiled samples)			
▼ Gene	# Mut	#	Freq ▼
TP53 	116	 107	46.5%
KRAS 	76	 75	32.6%
LRP1B	115	 68	29.6%
PCLO	56	 46	20.0%
KMT2C	50	 43	18.7%
STK11 	42	 40	17.4%
KEAP1 	40	 40	17.4%
RELN	45	 35	15.2%
FAT4	42	 34	14.8%
EGFR 	45	 33	14.3%
PTPRD	38	 33	14.3%

Fusion Genes (230 profiled samples)			
Gene	# Fusion	#	Freq
ROS1	4	<input type="checkbox"/> 4	1.7%
ALK	3	<input type="checkbox"/> 3	1.3%
RET	2	<input type="checkbox"/> 2	0.9%



CNA Genes (230 profiled samples)				
▼ Gene	Cytoband	CNA	#	Freq ▼
CDKN2A <a href="#">G</a>	9p21.3	DEL	<div><div></div></div> 45	19.6%
CDKN2B	9p21.3	DEL	<div><div></div></div> 42	18.3%
SDHA	5p15.33	AMP	<div><div></div></div> 41	17.8%
TERT	5p15.33	AMP	<div><div></div></div> 41	17.8%
TRIP13	5p15.33	AMP	<div><div></div></div> 41	17.8%
ARNT <a href="#">G</a>	1q21.3	AMP	<div><div></div></div> 37	16.1%
APH1A	1q21.2	AMP	<div><div></div></div> 37	16.1%
MCL1	1q21.2	AMP	<div><div></div></div> 36	15.7%
SETDB1 <a href="#">G</a>	1q21.3	AMP	<div><div></div></div> 36	15.7%
MLLT11 <a href="#">G</a>	1q21.3	AMP	<div><div></div></div> 35	15.2%
CKS1B	1q21.3	AMP	<div><div></div></div> 34	14.8%






Lung Adenocarcinoma (TCGA, Nature 2014) [Download](#)

Whole exome sequencing, messenger RNA, microRNA, copy number, methylation and proteomic analysis of 230 lung adenocarcinoma tumor/normal pairs. [Raw data via the TCGA Data Portal.](#) [PubMed](#)

Click gene symbols below or enter here

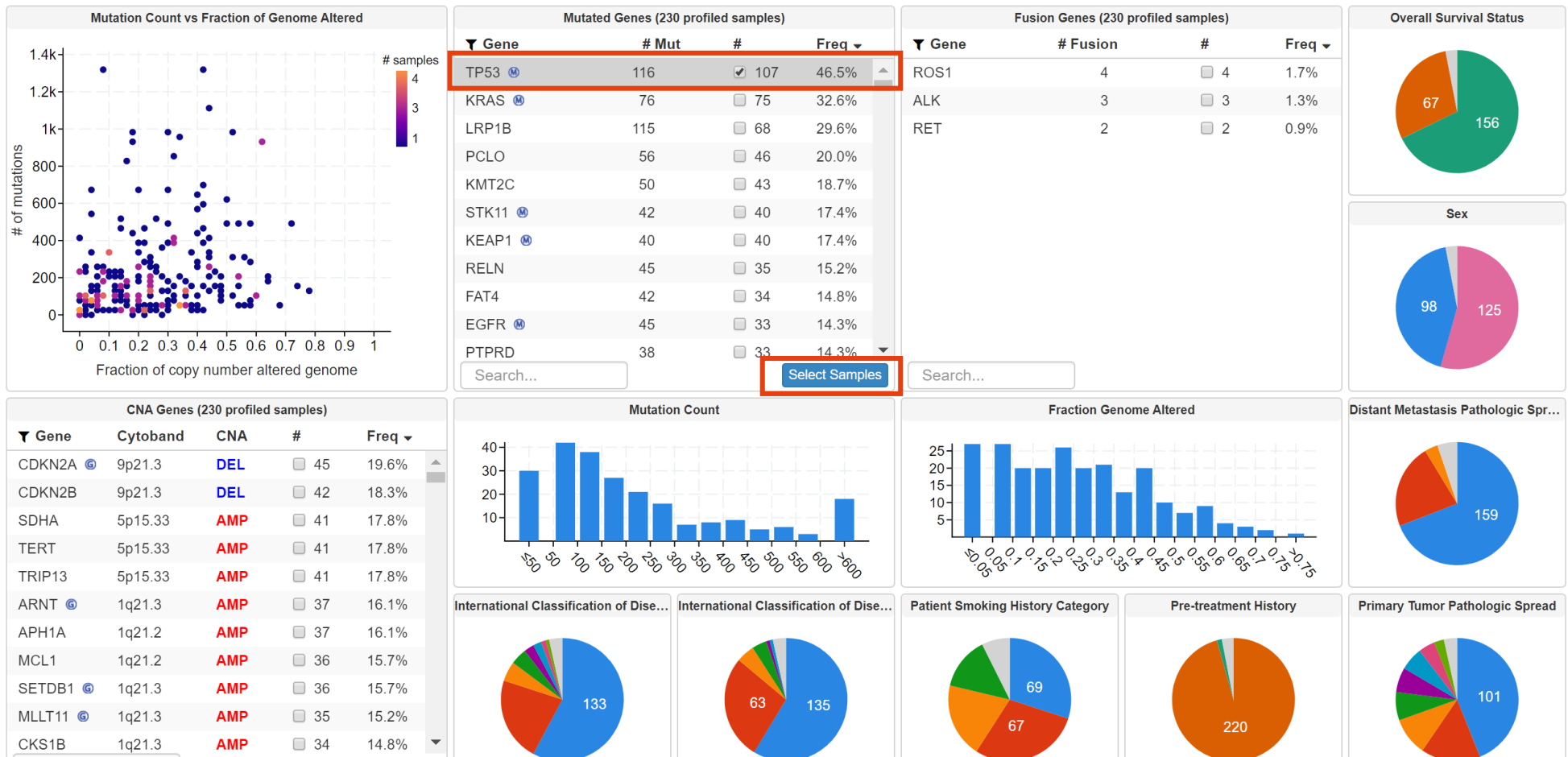
Query

Summary Clinical Data Heatmaps CN Segments

Selected: 230 patients | 230 samples   

Custom Selection Charts Groups

Quick Filters: ☐ 230 samples with mutation data ☐ 230 samples with CNA data



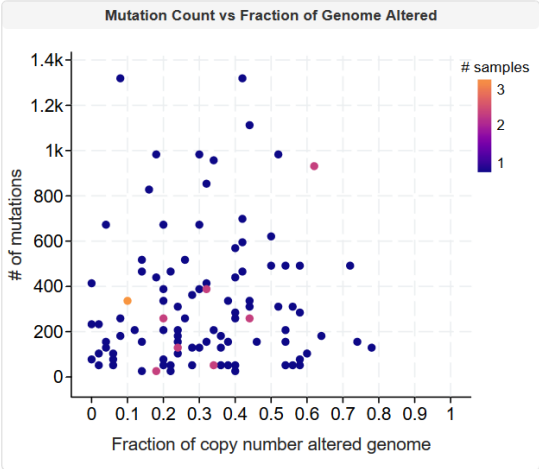
TP53 Clear All Filters

Summary Clinical Data Heatmaps CN Segments

Selected: 107 patients | 107 samples

Custom Selection Charts Groups

Quick Filters: 107 samples with mutation data 107 samples with CNA data



Mutated Genes (107 profiled samples)

Gene	# Mut	#	Freq
TP53	116	107	100.0%
LRP1B	83	44	41.1%
PCLO	42	33	30.8%
KMT2C	35	30	28.0%
KRAS	26	25	23.4%
RELN	34	24	22.4%
NF1	25	21	19.6%
CPS1	21	20	18.7%
EGFR	30	20	18.7%
FAT4	24	20	18.7%
GRIN2A	23	19	17.8%

Fusion Genes (107 profiled samples)

Gene	# Fusion	#
ROS1	1	1

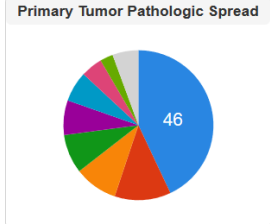
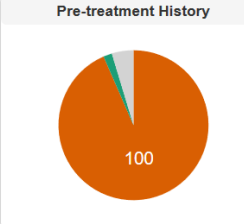
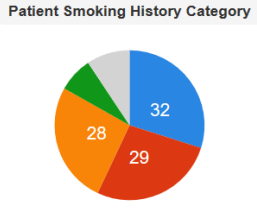
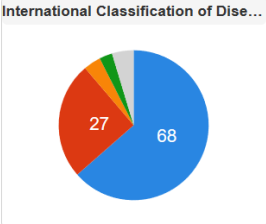
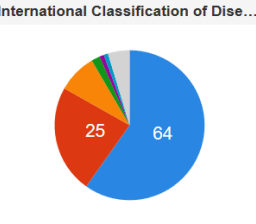
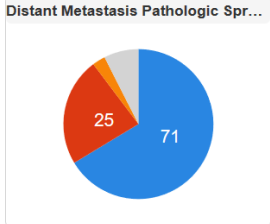
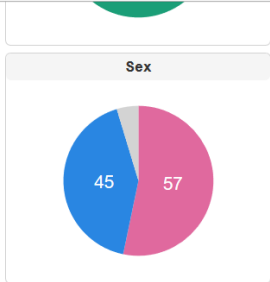
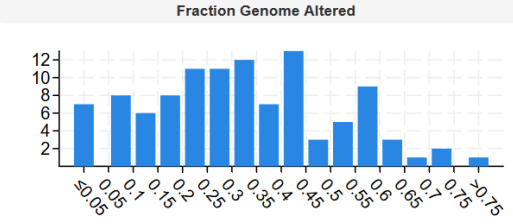
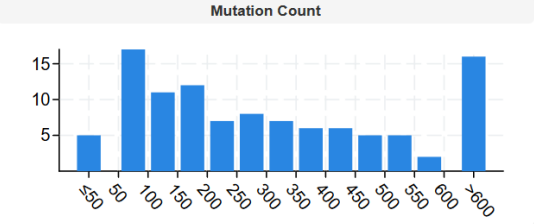
Select all (0) Deselect all Search..

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

TP53 mutant Create

CNA Genes (107 profiled samples)

Gene	Cytoband	CNA	#	Freq
SDHA	5p15.33	AMP	26	24.3%
TERT	5p15.33	AMP	26	24.3%
TRIP13	5p15.33	AMP	26	24.3%
CDKN2A	9p21.3	DEL	21	19.6%
CDKN2B	9p21.3	DEL	20	18.7%
IL7R	5p13.2	AMP	20	18.7%
DROSHA	5p13.3	AMP	20	18.7%
LIFR	5p13.1	AMP	19	17.8%
RICTOR	5p13.1	AMP	19	17.8%
FGF10	5p12	AMP	18	16.8%
MTAP	9p21.3	DEL	16	15.0%



Summary

Clinical Data

Heatmaps

CN Segments

Selected: 230 patients | 230 samples

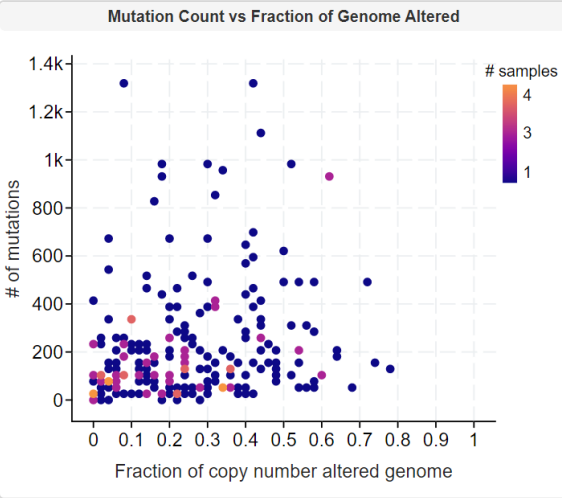


Custom Selection ▾

Charts ▾

Groups ▾

Quick Filters: ☐ 230 samples with mutation data ☐ 230 samples with CNA data



Mutated Genes (230 profiled samples)

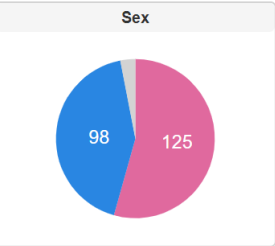
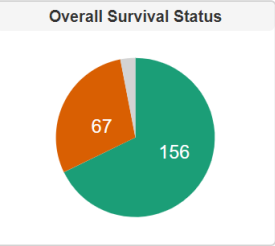
Gene	# Mut	#	Freq
TP53	116	107	46.5%
KRAS	76	75	32.6%
LRP1B	115	68	29.6%
PCLO	56	46	20.0%
KMT2C	50	43	18.7%
STK11	42	40	17.4%
KEAP1	40	40	17.4%
RELN	45	35	15.2%
FAT4	42	34	14.8%
EGFR	45	33	14.3%
PTPRD	38	33	14.3%

Search... [Select Samples](#)

Fusion Genes (230 profiled samples)

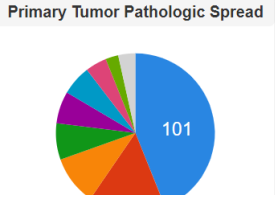
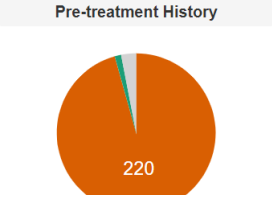
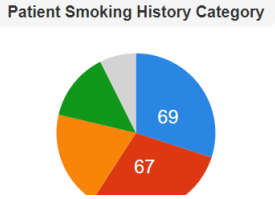
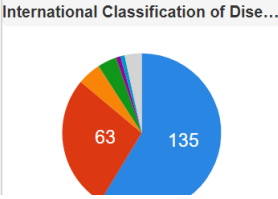
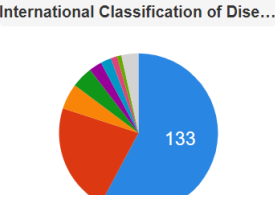
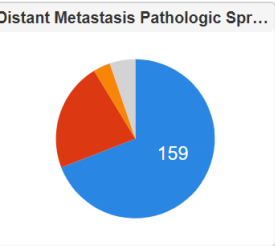
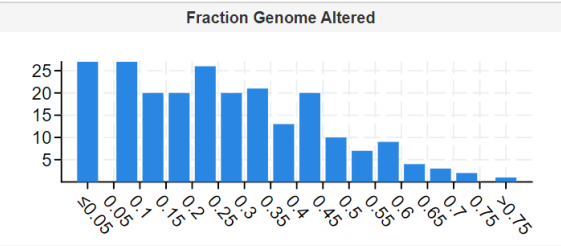
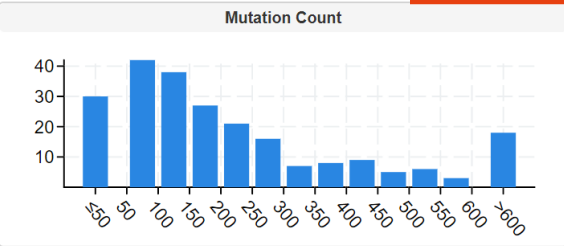
Gene	# Fusion	#	Freq
ROS1	4	4	1.7%
ALK	3	3	1.3%
RET	2	2	0.9%

Search...



CNA Genes (230 profiled samples)

Gene	Cytoband	CNA	#	Freq
CDKN2A	9p21.3	DEL	45	19.6%
CDKN2B	9p21.3	DEL	42	18.3%
SDHA	5p15.33	AMP	41	17.8%
TERT	5p15.33	AMP	41	17.8%
TRIP13	5p15.33	AMP	41	17.8%
ARNT	1q21.3	AMP	37	16.1%
APH1A	1q21.2	AMP	37	16.1%
MCL1	1q21.2	AMP	36	15.7%
SETDB1	1q21.3	AMP	36	15.7%
MLLT11	1q21.3	AMP	35	15.2%
CKS1B	1p21.3	AMP	34	14.8%



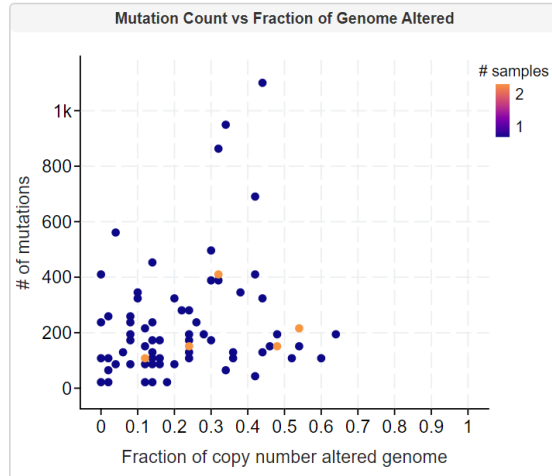
KRAS  [Clear All Filters](#)

Summary Clinical Data Heatmaps CN Segments

Selected: 75 patients | 75 samples

   Custom Selection Charts Groups

Quick Filters: ☐ 75 samples with mutation data ☐ 75 samples with CNA data






Mutated Genes (75 profiled samples)

Gene	# Mut	#	Freq
KRAS	76	75	100.0%
LRP1B	50	30	40.0%
TP53	29	25	33.3%
STK11	22	21	28.0%
RELN	18	15	20.0%
PCLO	17	15	20.0%
KEAP1	14	14	18.7%
KMT2C	14	13	17.3%
PTPRD	13	12	16.0%
RBM10	12	12	16.0%
FAT4	13	12	16.0%

Fusion Genes (75 profiled samples)

Select all (1) Deselect all Search..

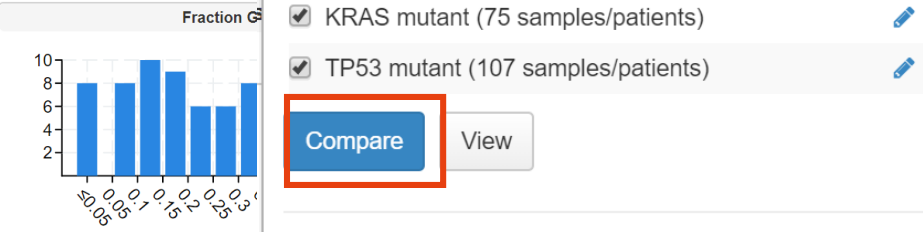
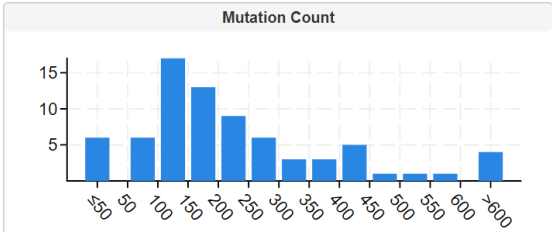
☒ TP53 mutant (107 samples/patients)   

Compare View

KRAS mutant Create

CNA Genes (75 profiled samples)



Gene	Cytoband	CNA	#	Freq
ARNT	1q21.3	AMP	12	16.0%
CKS1B	1q21.3	AMP	12	16.0%
MCL1	1q21.2	AMP	12	16.0%
SETDB1	1q21.3	AMP	12	16.0%
MLLT11	1q21.3	AMP	12	16.0%
CDKN2A	9p21.3	DEL	11	14.7%
LMNA	1q22	AMP	11	14.7%
MUC1	1q22	AMP	11	14.7%
PTEN	1p33	DEL	11	14.7%



International Classification of Disease... International Classification of Disease... Patient Smoking History Category

Select all (2) Deselect all Search..

☒ KRAS mutant (75 samples/patients)   

☒ TP53 mutant (107 samples/patients)   

Compare View

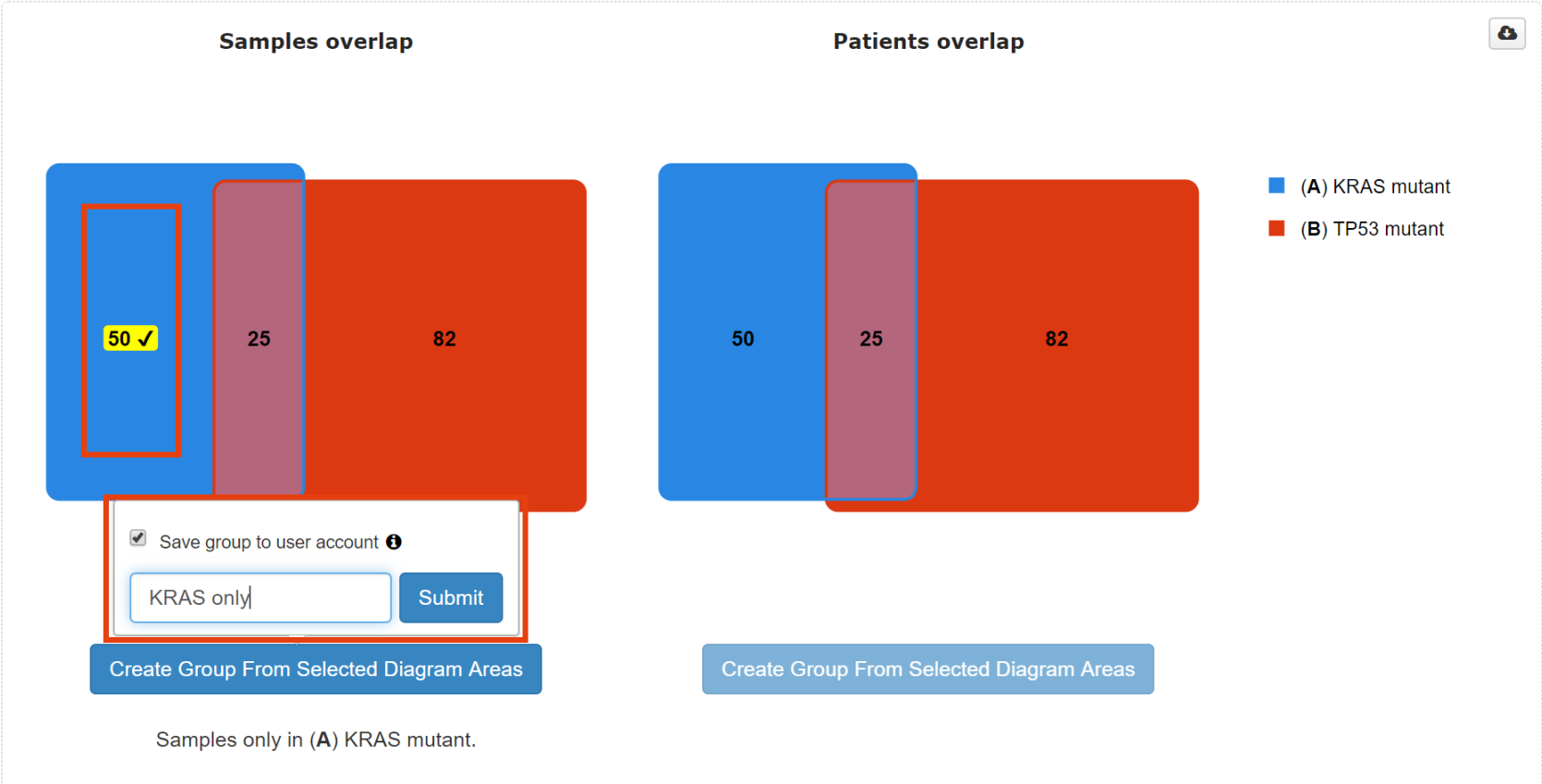
Create new group from selected samples (75)

Lung Adenocarcinoma (TCGA, Nature 2014)

Groups: (drag to reorder) (A) KRAS mutant (75) (B) TP53 mutant (107) Select all | Deselect all

- Overlap
- Survival
- Clinical
- Mutations
- Copy-number
- mRNA
- Protein

**i** Samples (25) that overlap in the selected groups are excluded from sample-level analysis in other tabs.  
**i** Patients (25) that overlap in the selected groups are excluded from patient-level analysis in other tabs.



Overlap	Survival	Clinical	Mutations	Copy-number	mRNA	Protein
---------	----------	----------	-----------	-------------	------	---------

### Patients overlap



AMSI  
**BIOINFO SUMMER**  
A SYMPOSIUM IN BIOINFORMATICS

Groups: (drag to reorder) (A) KRAS + TP53 (25) (B) KRAS mutant (75) (C) KRAS only (50) (D) TP53 mutant (107) Select all | Deselect all

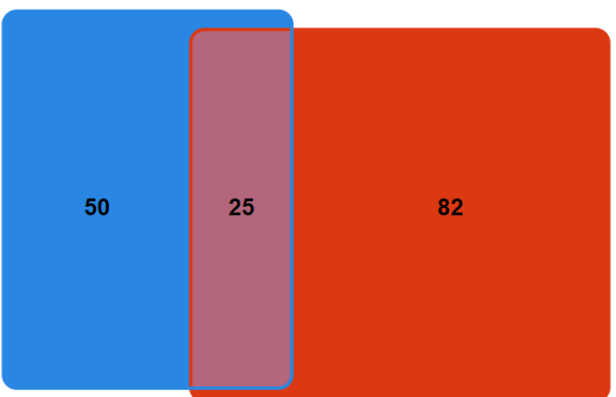
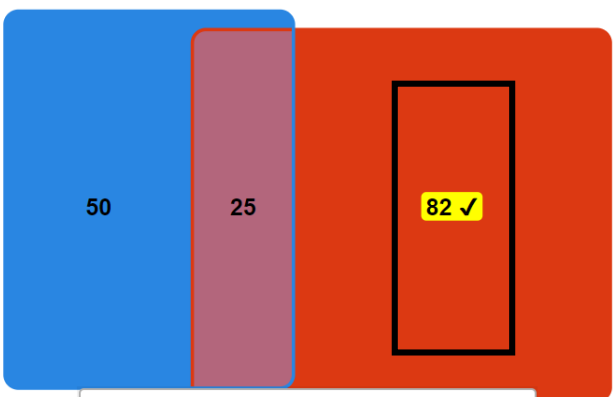
Overlap Survival Clinical Mutations Copy-number mRNA Protein

- ❗ Samples (25) that overlap in the selected groups are excluded from sample-level analysis in other tabs.
- ❗ Patients (25) that overlap in the selected groups are excluded from patient-level analysis in other tabs.

Make sure both are selected

Samples overlap

Patients overlap



■ (B) KRAS mutant  
■ (D) TP53 mutant

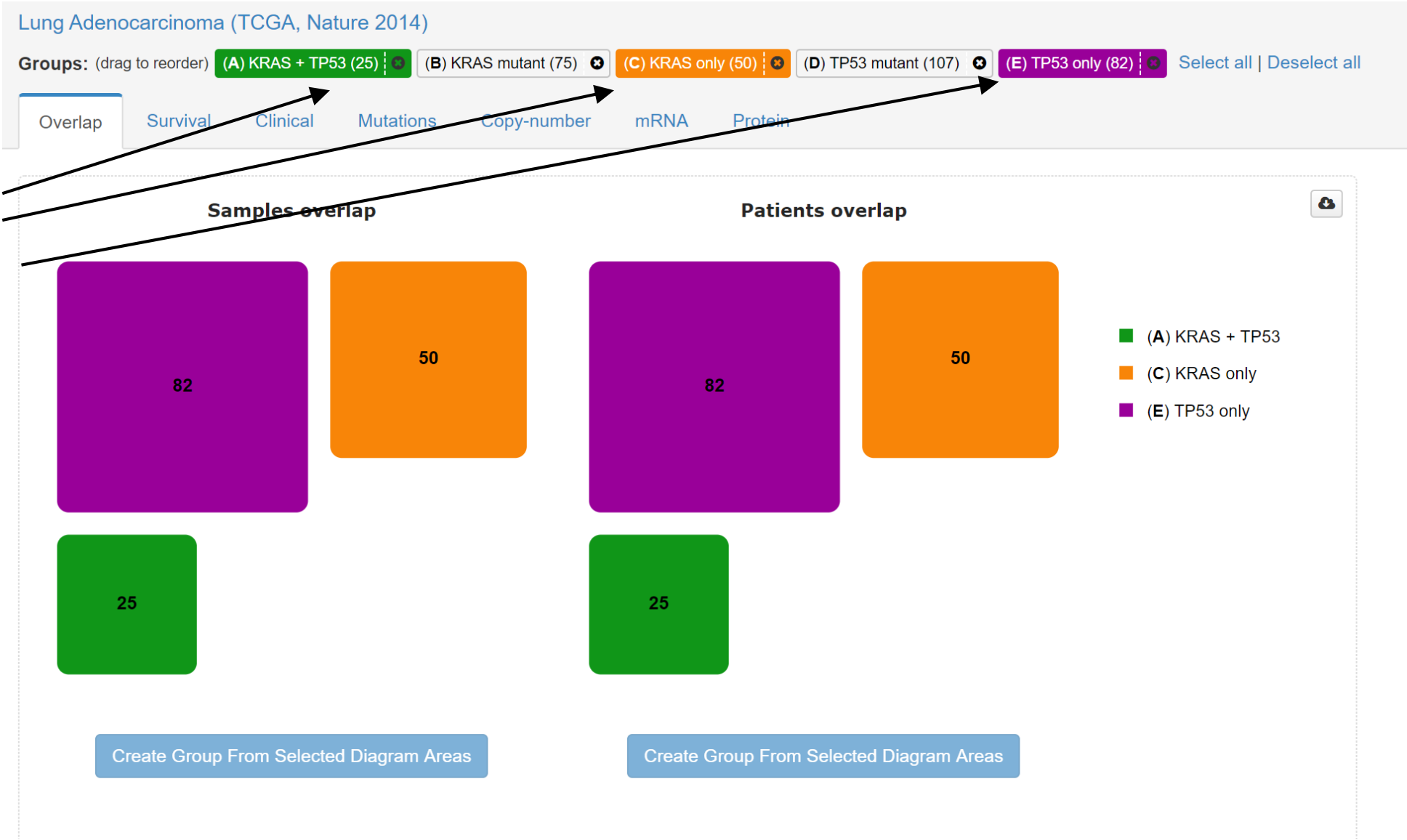
☒ Save group to user account ⓘ

TP53 only Submit

Create Group From Selected Diagram Areas

Create Group From Selected Diagram Areas

Samples only in (D) TP53 mutant.



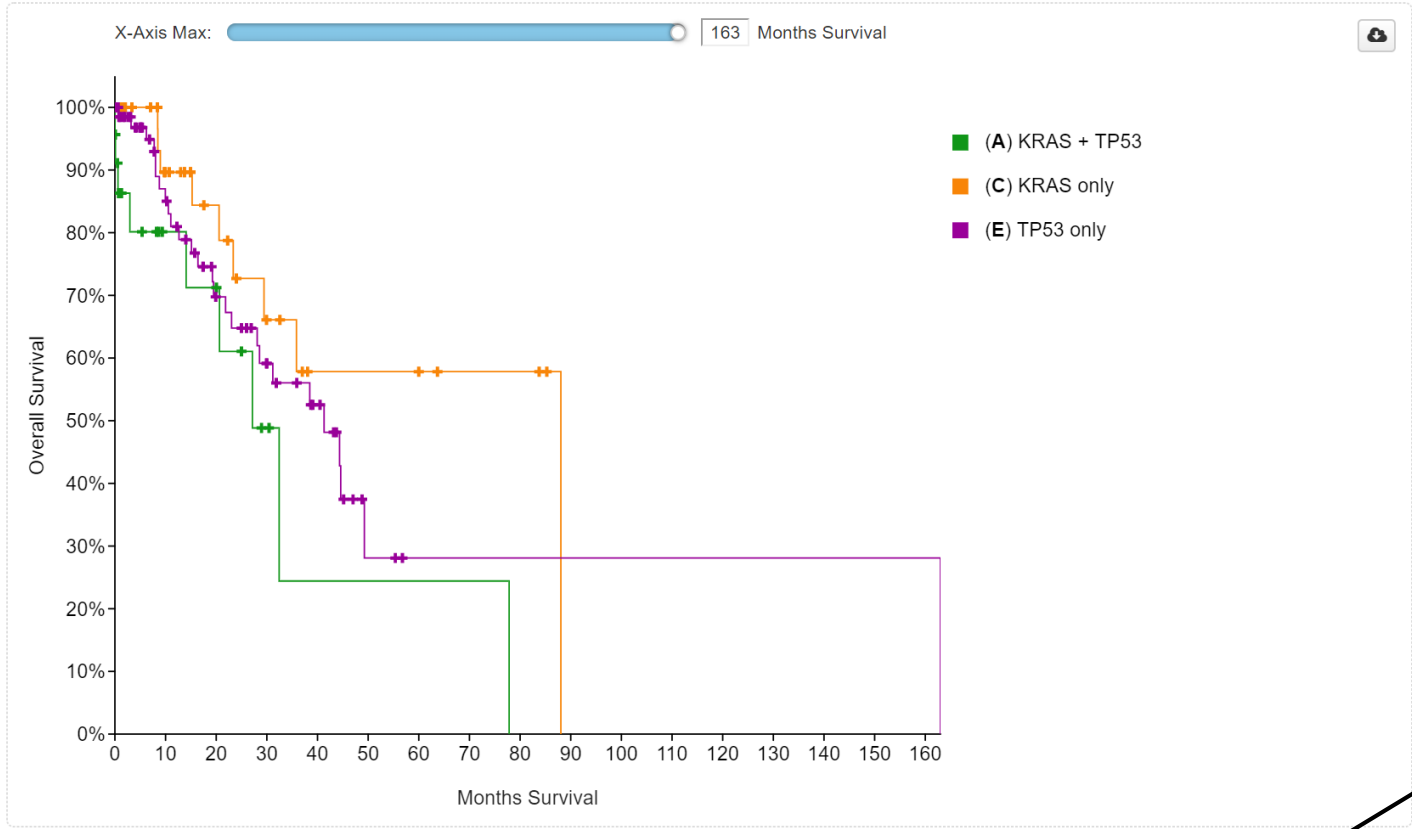
Make sure  
KRAS+TP53, KRAS  
only and TP53 only  
are selected.

Deselect KRAS  
mutant and TP53  
mutant

Groups: (drag to reorder) (A) KRAS + TP53 (25) (B) KRAS mutant (75) (C) KRAS only (50) (D) TP53 mutant (107) (E) TP53 only (82) [Select all](#) [Deselect all](#)

Overlap **Survival** Clinical Mutations Copy-number mRNA Protein

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



Double mutant has worst survival

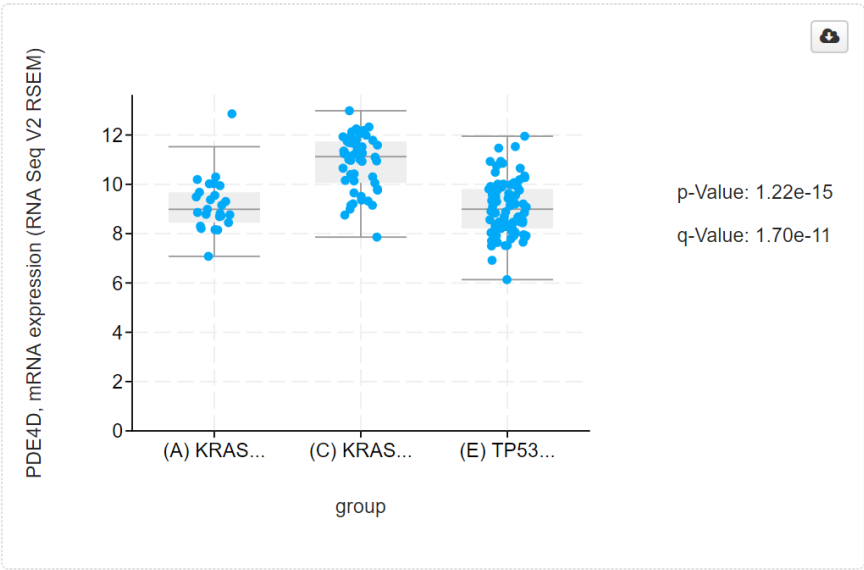
	Number of Cases, Total	Number of Cases, Deceased	Median Months Survival
(A) KRAS + TP53	23	9	27.1381474
(C) KRAS only	44	9	88.051132
(E) TP53 only	73	26	41.2986093

Lung Adenocarcinoma (TCGA, Nature 2014)

Groups: (drag to reorder) (A) KRAS + TP53 (25) (B) KRAS mutant (75) (C) KRAS only (50) (D) TP53 mutant (107) (E) TP53 only (82) [Select all](#) | [Deselect all](#)

Overlap Survival Clinical Mutations Copy-number mRNA Protein

Data Set: mRNA expression (RNA Seq V2 RSEM)



mRNA expression (RNA Seq V2 RSEM)

Select enriched groups		<input type="checkbox"/> Significant only		Columns						
Gene	Cyband	$\mu$ in (A) KRAS + TP53	$\mu$ in (C) KRAS only	$\mu$ in (E) TP53 only	$\sigma$ in (A) KRAS + TP53	$\sigma$ in (C) KRAS only	$\sigma$ in (E) TP53 only	p-Value	q-Value	Higher expression in
PDE4D	5q11.2-q12.1	9.19	10.85	9.04	1.09	1.13	1.11	1.22e-15	1.70e-11	(C) KRAS only
APOBEC3B	22q13.1	7.66	5.90	8.11	1.11	1.21	1.51	3.89e-15	2.71e-11	(E) TP53 only
GPSM2	1p13.3	9.09	8.23	9.27	0.68	0.63	0.72	1.01e-13	4.69e-10	(E) TP53 only

## Exercise...

*What is the most common hotspot point mutation in ERCC2 in bladder cancer?*

N238S


## Challenge...

*What is the most significantly higher expressed gene in ERCC2 mutant versus WT bladder cancer?*

SGO1

# Genomic Data Commons

*Accessing raw and processed data*



NATIONAL CANCER INSTITUTE  
GDC Data Portal

Home

Projects

Exploration

Analysis

Repository

Quick Search

Manage Sets

Login

Cart 0

GDC Apps

Harmonized Cancer Datasets

Genomic Data Commons Data Portal

Get Started by Exploring:

Projects

Exploration

Analysis

Repository

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary

Data Release 20.0 - November 11, 2019

PROJECTS

55

FILES

440,782

PRIMARY SITES

66

GENES

22,872

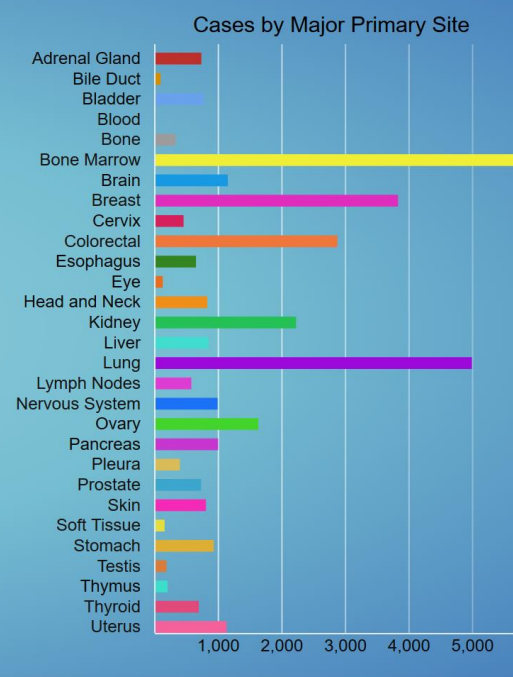
CASES

38,728

MUTATIONS

3,142,246

Cases by Major Primary Site



Primary Site	Cases
Adrenal Gland	~100
Bile Duct	~100
Bladder	~100
Blood	~100
Bone	~100
Bone Marrow	~100
Brain	~100
Breast	~3,800
Cervix	~100
Colorectal	~2,800
Esophagus	~100
Eye	~100
Head and Neck	~100
Kidney	~100
Liver	~100
Lung	~5,000
Lymph Nodes	~100
Nervous System	~100
Ovary	~100
Pancreas	~100
Pleura	~100
Prostate	~1,800
Skin	~100
Soft Tissue	~100
Stomach	~100
Testis	~100
Thymus	~100
Thyroid	~100
Uterus	~100

GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

Data Portal

Website

API

Data Transfer Tool

Documentation

Data Submission Portal

Legacy Archive

Publications

CHILDREN'S MEDICAL RESEARCH INSTITUTE  
Jeans for Genes®

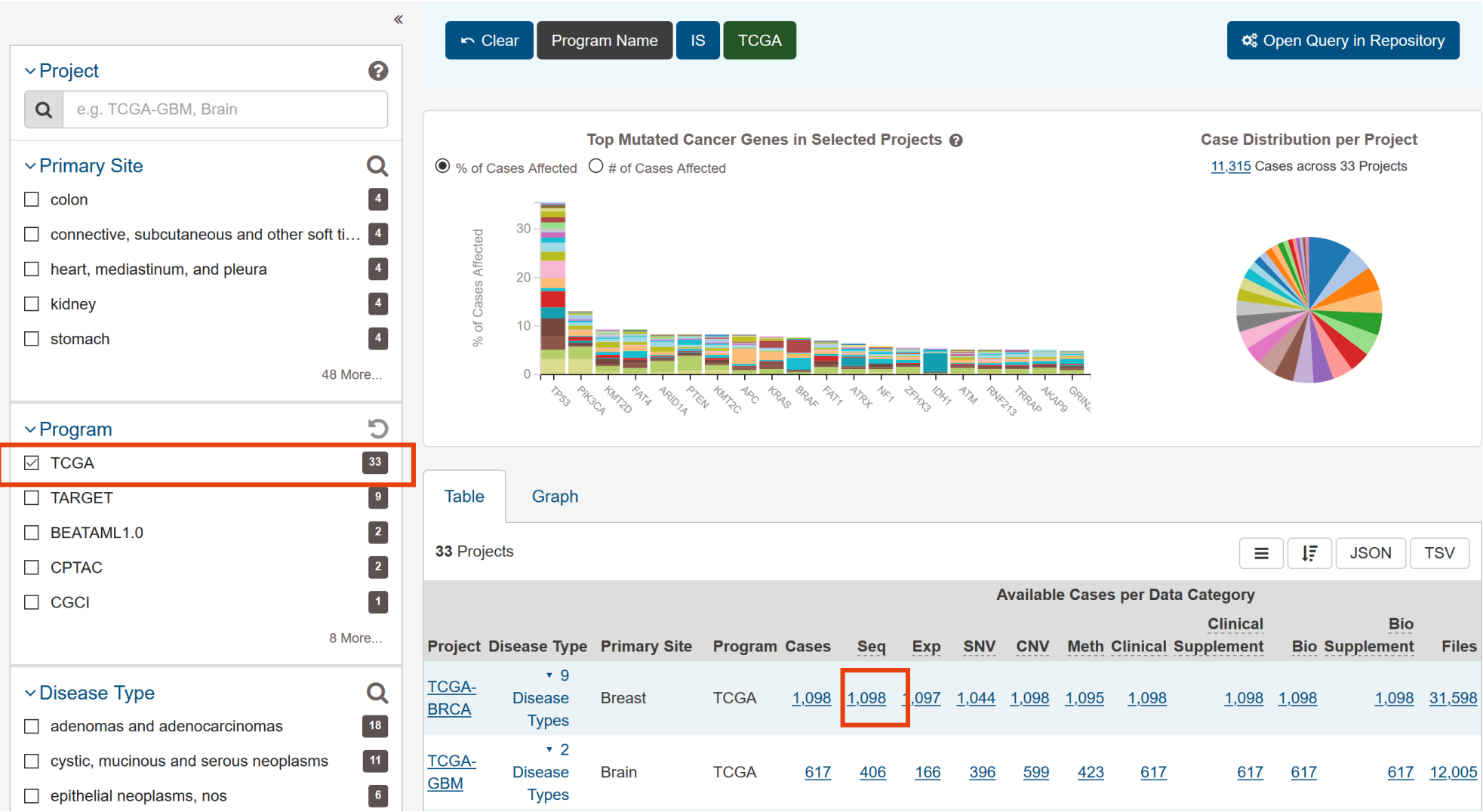
THE UNIVERSITY OF SYDNEY

THE UNIVERSITY OF HONG KONG

香港大學  
THE UNIVERSITY OF HONG KONG

AMSI  
BIOINFO SUMMER 19  
A SYMPOSIUM IN BIOINFORMATICS

# Genomic Data Commons



# Genomic Data Commons

NIH

NATIONAL CANCER INSTITUTE

GDC Data Portal

Home

Projects

Exploration

Analysis

Repository

Quick Search

Manage Sets

Login

Cart 0

GDC Apps

Files

Cases

Add a File Filter

File

e.g. 142682.bam, 4f6e2e7a-b...

Data Category

simple nucleotide variation

transcriptome profiling

biospecimen

sequencing reads

copy number variation

Data Type

Aligned Reads

Experimental Strategy

WXS

RNA-Seq

miRNA-Seq

ATAC-Seq

Workflow Type

Clear

Project Id

IS

TCGA-BRCA

AND

Data Category

IS

Sequencing Reads

Advanced Search

Add All Files to Cart

Manifest

View 1,098 Cases in Exploration

View Images

Browse Annotations

Files (4,679)

Cases (1,098)

55.29 TB

Primary Site

Project

Disease Type

Gender

Vital Status

Showing 1 - 20 of 1,098 cases

Biospecimen

Clinical

JSON

TSV

Save/Edit Case Set

Cart	Case ID	Project	Primary Site	Gender	Files	Available Files per Data Category							Annotations	Slides
						Seq	Exp	SNV	CNV	Meth	Clinical	Bio		
<input type="checkbox"/>	<a href="#">TCGA-A8-A08S</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">57</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">15</a>	<a href="#">2</a>	<a href="#">3</a>
<input type="checkbox"/>	<a href="#">TCGA-A8-A084</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">39</a>	<a href="#">2</a>	<a href="#">5</a>	<a href="#">0</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">15</a>	<a href="#">2</a>	<a href="#">3</a>
<input type="checkbox"/>	<a href="#">TCGA-A8-A08F</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">54</a>	<a href="#">3</a>	<a href="#">3</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">15</a>	<a href="#">2</a>	<a href="#">3</a>
<input type="checkbox"/>	<a href="#">TCGA-A8-A09E</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">57</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">15</a>	<a href="#">2</a>	<a href="#">3</a>
<input type="checkbox"/>	<a href="#">TCGA-A2-A0YT</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">57</a>	<a href="#">5</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">10</a>	<a href="#">15</a>	<a href="#">2</a>	<a href="#">3</a>
<input type="checkbox"/>	<a href="#">TCGA-A8-A09K</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">57</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">15</a>	<a href="#">2</a>	<a href="#">3</a>
<input type="checkbox"/>	<a href="#">TCGA-A8-A07E</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">39</a>	<a href="#">2</a>	<a href="#">5</a>	<a href="#">0</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">15</a>	<a href="#">1</a>	<a href="#">3</a>
<input type="checkbox"/>	<a href="#">TCGA-5L-AAT1</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">56</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">14</a>	<a href="#">1</a>	<a href="#">2</a>
<input type="checkbox"/>	<a href="#">TCGA-A7-A0CD</a>	<a href="#">TCGA-BRCA</a>	Breast	Female	<a href="#">58</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">16</a>	<a href="#">1</a>	<a href="#">4</a>

# Genomic Data Commons

Files

Cases

Add a File Filter

File

File

e.g. 142682.bam, 4f6e2e7a-b...

Data Category

transcriptome profiling3,666

☒ sequencing reads1,222

Data Type

Aligned Reads1,222

Experimental Strategy

WXS2,175

☒ RNA-Seq1,222

miRNA-Seq1,207

ATAC-Seq75

Workflow Type

STAR 2-Pass1,222

Data Format

bam1,222

ClearProject IdISTCGA-BRCAAND

Data CategoryISSequencing ReadsANDExperimental StrategyISRNA-Seq

Advanced Search

Add All Files to CartManifestView 1,092 Cases in ExplorationView ImagesBrowse Annotations

Files (1,222)Cases (1,092)10.39 TB

Primary Site

Project

Data Category

Data Type

Data Format

Show More

Showing 1 - 20 of 1,222 files

Access	File Name	Cases	Project	Data Category
<input checked="" type="checkbox"/> controlled	947e442-0010-4438-a69f-cc4b7881f2ad_gdc_realn_rehead.bam	1	TCGA-BRCA	Sequencing Reads E
<input checked="" type="checkbox"/> controlled	7686cd5-271a-4e24-9746-f93334fb86b5_gdc_realn_rehead.bam	1	TCGA-BRCA	Sequencing Reads E
<input checked="" type="checkbox"/> controlled	916abcd-61c0-4f23-96ac-be70aacb8dc1_gdc_realn_rehead.bam	1	TCGA-BRCA	Sequencing Reads E
<input checked="" type="checkbox"/> controlled	83e3574-d22c-4dd6-842f-656ee2ab3b32_gdc_realn_rehead.bam	1	TCGA-BRCA	Sequencing Reads E

To access raw data (i.e. control datasets), need to first apply through dbGAP. I won't go through how today but you are welcomed to ask if later if you are interested.

# Genomic Data Commons

Files

Cases

Add a File Filter

File

Q e.g. 142682.bam, 4f6e2e7a-b...

Data Category

simple nucleotide variation

33

Data Type

Masked Somatic Mutation

33

Experimental Strategy

WXS

33

Workflow Type

MuTect2 Variant Aggregation and Masking

33

Data Format

maf

33

Clear

Workflow Type

IS

MuTect2 Variant Aggregation and Masking

AND

Data Category

IS

simple nucleotide variation

AND

Data Type

IS

Masked Somatic Mutation

Advanced Search

Add All Files to Cart

Manifest

View 10,418 Cases in Exploration

View Images

Browse Annotations

Files (33)

Cases (10,418)

758.06 MB

Primary Site

Project

Data Category

Data Type

Data Format

Show More

Showing 1 - 20 of 33 files

JSON

TSV

Access	File Name	Cases	Project	Data Category	Data Format	File Size	Annotations
<div>open</div>	<div><div>TCGA.LUAD.mutect.0458c57f-316c-4a7c-9294-ccd11c97c2f9.DR-10.0.somatic.maf.gz</div></div>	569	<a href="#">TCGA-LUAD</a>	Simple Nucleotide Variation	MAF	50.3 MB	84
<div>open</div>	<div><div>TCGA.READ.mutect.faa5f62a-2731-4867-a264-0e85b7074e87.DR-10.0.somatic.maf.gz</div></div>	158	<a href="#">TCGA-READ</a>	Simple Nucleotide Variation	MAF	15.19 MB	36

FL TCGA.LUAD.mutect.0458c57f-316c-4a7c-9294-ccd11c97c2f9.DR-10.0.somatic.maf.gz

Add to Cart

Download

### File Properties

Name	TCGA.LUAD.mutect.0458c57f-316c-4a7c-9294-ccd11c97c2f9.DR-10.0.somatic.maf.gz
Access	open
UUID	0458c57f-316c-4a7c-9294-ccd11c97c2f9
Data Format	MAF
Size	50.3 MB
MD5 Checksum	fd7e2e6086332b2152626264ab6d6d5d
Archive	--
Project	<a href="#">TCGA-LUAD</a>

### Data Information

Data Category	Simple Nucleotide Variation
Data Type	Masked Somatic Mutation
Experimental Strategy	WXS
Platform	--

Showing 1 - 10 of 1,223 associated cases/biospecimen

### Associated Cases/Biospecimen



Entity ID

eg. TCGA-13\*, \*13\*, \*09

Entity ID	Entity Type	Sample Type	Case UUID	Annotations
<a href="#">TCGA-91-6829-01A-21D-1855-08</a>	aliquot	Primary Tumor	<a href="#">1427cd18-5ad3-491a-9981-908e31ae49db</a>	0
<a href="#">TCGA-50-5932-01A-11D-1753-08</a>	aliquot	Primary Tumor	<a href="#">ebcba7f2-ce13-4bae-97cd-91a6b1dcd465</a>	0
<a href="#">TCGA-49-6743-01A-11D-1855-08</a>	aliquot	Primary Tumor	<a href="#">a391d49f-a822-460b-981c-6fbe1868ee38</a>	0
<a href="#">TCGA-86-8669-01A-11D-2393-08</a>	aliquot	Primary Tumor	<a href="#">7b89166e-c5b8-481b-aa70-495141499b91</a>	0

# Genomic Data Commons

FILES  
1



## File Counts by Project

Project	Cases (n=569)	Files (n=1)	File Size (Σ=50.3 MB)
TCGA-LUAD	569 <div></div>	1 <div></div>	50.3 MB <div></div>

CASES  
569



FILE SIZE  
50.3 MB



## File Counts by Authorization Level

Level	Files (n=1)	File Size (Σ=50.3 MB)
Authorized	1 <div></div>	50.3 MB <div></div>

How to download files in my Cart?

### Download Manifest:

Download a manifest for use with the [GDC Data Transfer Tool](#). The GDC Data Transfer Tool is recommended for transferring large volumes of data.

### Download Cart:


Download Files in your Cart directly from the Web Browser.

 Biospecimen

 Clinical

 Sample Sheet

 Metadata

 Download ▾

 Remove From Cart ▾

## Cart Items

Showing 1 - 1 of 1 files

 Manifest

 Cart



TSV

Access	File Name	Cases	Project	Data Category	Data Format	File Size	Annotations
 open	<a href="#">TCGA.LUAD.mutect.0458c57f-316c-4a7c-9294-ccd11c97c2f9.DR-10.0.somatic.maf.gz</a>	569	<a href="#">TCGA-LUAD</a>	Simple Nucleotide Variation	MAF	50.3 MB	84

Show 20 ▾ entries

« < 1 > »

<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool>

## GDC Data Transfer Tool



The GDC provides a standard client-based mechanism in support of high-performance data downloads and submission.

The raw sequence files, typically stored as BAM or FASTQ, make up the bulk of data. The size for a single file can vary greatly depending on the specific analysis; However, some of the whole genome BAM files in The Cancer Genome Atlas (TCGA) reach sizes of 200-300 GB. In such cases, a high-performance data download and submission tool is essential.

Below are basic instructions and links for downloading the GDC Data Transfer Tool client-based interface for data downloads and submission and user interface (beta version) for data downloads. For additional instructions, please visit the [GDC Data Transfer Tool User's Guide](#).

# Genomic Data Commons – download WXS BAM files for all ERCC2 N238S mutants

NIH

NATIONAL CANCER INSTITUTE  
GDC Data Portal

HomeProjectsExplorationAnalysisRepository

Manage Sets0

CasesClinicalGenesMutations

▼ Mutation

ERCC2 N238S

adc5ca97-e44e-5f9e-8f28-87117415fd32  
ERCC2 N238S  
chr19:g.45364429T>C

☐ modifier116,194

☐ moderate76,158

☐ low30,426

☐ high20,391

▼ SIFT Impact

☐ deleterious38,669

☐ tolerated27,536

☐ deleterious\_low\_confidence6,910

☐ tolerated\_low\_confidence4,976

▼ Polyphen Impact

☐ probably\_damaging35,559

☐ benign34,726

☐ possibly\_damaging23,301

☐ unknown5,756

ClearIs Cancer Gene CensusIStrue

View Files in Repository

Cases (10,843)Genes (576)Mutations (144,131)OncoGrid

Primary Site

Project

Disease Type

Gender

Vital Status

Showing 1 - 20 of 10,843 cases

≡

🔍

📄 Biospecimen

📄 Clinical

JSON

TSV

Save/Edit Case Set

Case ID	Project	Primary Site	Gender	Files	Available Files per Data Category							# Mutations	# Genes	Slides
					Seq	Exp	SNV	CNV	Meth	Clinical	Bio			
<input type="checkbox"/> <a href="#">TCGA-A5-A0G2</a>	<a href="#">TCGA-UCEC</a>	Corpus uteri	Female	<a href="#">58</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">10</a>	<a href="#">17</a>	<a href="#">1,856</a>	<a href="#">481</a>	<a href="#">📄 (3)</a>
<input type="checkbox"/> <a href="#">TCGA-EO-A22U</a>	<a href="#">TCGA-UCEC</a>	Corpus uteri	Female	<a href="#">57</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">10</a>	<a href="#">16</a>	<a href="#">1,234</a>	<a href="#">445</a>	<a href="#">📄 (2)</a>
<input type="checkbox"/> <a href="#">TCGA-FI-A2D5</a>	<a href="#">TCGA-UCEC</a>	Corpus uteri	Female	<a href="#">58</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">11</a>	<a href="#">16</a>	<a href="#">1,103</a>	<a href="#">437</a>	<a href="#">📄 (2)</a>
<input type="checkbox"/> <a href="#">TCGA-B5-A3FC</a>	<a href="#">TCGA-UCEC</a>	Corpus uteri	Female	<a href="#">57</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">10</a>	<a href="#">16</a>	<a href="#">1,180</a>	<a href="#">436</a>	<a href="#">📄 (2)</a>
<input type="checkbox"/> <a href="#">TCGA-AX-A2HC</a>	<a href="#">TCGA-UCEC</a>	Corpus uteri	Female	<a href="#">65</a>	<a href="#">6</a>	<a href="#">10</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">2</a>	<a href="#">10</a>	<a href="#">16</a>	<a href="#">1,095</a>	<a href="#">434</a>	<a href="#">📄 (2)</a>
<input type="checkbox"/> <a href="#">TCGA-EO-A22R</a>	<a href="#">TCGA-UCEC</a>	Corpus uteri	Female	<a href="#">59</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">2</a>	<a href="#">10</a>	<a href="#">17</a>	<a href="#">1,100</a>	<a href="#">429</a>	<a href="#">📄 (3)</a>
<input type="checkbox"/> <a href="#">TCGA-IB-7651</a>	<a href="#">TCGA-PAAD</a>	Pancreas	Female	<a href="#">56</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">8</a>	<a href="#">17</a>	<a href="#">1,062</a>	<a href="#">422</a>	<a href="#">📄 (3)</a>

# Genomic Data Commons – download WXS BAM files for all ERCC2 N238S mutants

NIH

NATIONAL CANCER INSTITUTE

GDC Data Portal

Home

Projects

Exploration

Analysis

Repository

Manage Sets

0

Cases

Clinical

Genes

Mutations

▼ Mutation

adc5ca97-e44e-5f9e-8f28-87117415fd32

Q e.g. BRAF V600E, chr7:g.140753336A>T

Upload Mutation Set

▼ VEP Impact

modifier

1

moderate

1

▼ SIFT Impact

deleterious

1

▼ Polyphen Impact

probably\_damaging

1

▼ Consequence Type

missense\_variant

1

intron\_variant

1

upstream\_gene\_variant

1

downstream\_gene\_variant

1

▼ Type

Clear

Is Cancer Gene Census

IS

true

AND

Ssm Id

IS

adc5ca97-e44e-5f9e-8f28-87117415fd32

View Files in Repository

Cases (9)

Genes (1)

Mutations (1)

OncoGrid

Primary Site

Project

Disease Type

Gender

Vital Status

Showing 1 - 9 of 9 cases

Biospecimen

Clinical

JSON

TSV

Save/Edit Case Set

Case ID	Project	Primary Site	Gender	Files	Available Files per Data Category							# Mutations	# Genes	Slides
					Seq	Exp	SNV	CNV	Meth	Clinical	Bio			
<input type="checkbox"/> TCGA-E7-A67Z	<a href="#">TCGA-BLCA</a>	Bladder	Male	<a href="#">54</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">9</a>	<a href="#">14</a>	<a href="#">1</a>	<a href="#">1</a>	<a href="#">(2)</a>
<input type="checkbox"/> TCGA-GC-A3RC	<a href="#">TCGA-BLCA</a>	Bladder	Male	<a href="#">52</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">3</a>	<a href="#">1</a>	<a href="#">9</a>	<a href="#">14</a>	<a href="#">1</a>	<a href="#">1</a>	<a href="#">(2)</a>
<input type="checkbox"/> TCGA-E7-A85H	<a href="#">TCGA-BLCA</a>	Bladder	Male	<a href="#">54</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">9</a>	<a href="#">14</a>	<a href="#">1</a>	<a href="#">1</a>	<a href="#">(2)</a>
<input type="checkbox"/> TCGA-UY-A78N	<a href="#">TCGA-BLCA</a>	Bladder	Male	<a href="#">55</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">10</a>	<a href="#">14</a>	<a href="#">1</a>	<a href="#">1</a>	<a href="#">(2)</a>
<input type="checkbox"/> TCGA-FD-A3N5	<a href="#">TCGA-BLCA</a>	Bladder	Male	<a href="#">54</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">9</a>	<a href="#">14</a>	<a href="#">1</a>	<a href="#">1</a>	<a href="#">(2)</a>
<input type="checkbox"/> TCGA-XF-A9T3	<a href="#">TCGA-BLCA</a>	Bladder	Female	<a href="#">55</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">10</a>	<a href="#">14</a>	<a href="#">1</a>	<a href="#">1</a>	<a href="#">(2)</a>
<input type="checkbox"/> TCGA-KQ-A41R	<a href="#">TCGA-BLCA</a>	Bladder	Female	<a href="#">53</a>	<a href="#">4</a>	<a href="#">5</a>	<a href="#">16</a>	<a href="#">5</a>	<a href="#">1</a>	<a href="#">9</a>	<a href="#">13</a>	<a href="#">1</a>	<a href="#">1</a>	<a href="#">(1)</a>

# Genomic Data Commons – download WXS BAM files for all ERCC2 N238S mutants

FilesCases

Add a File Filter

File

Q e.g. 142682.bam, 4f6e2e7a-b...

Data Category

☐ sequencing reads18

Data Type

☐ Annotated Somatic Mutation36

☐ Raw Simple Somatic Mutation36

☒ Aligned Reads18

☐ Aggregated Somatic Mutation4

☐ Masked Somatic Mutation4

Experimental Strategy

☒ WXS18

☐ RNA-Seq9

☐ miRNA-Seq9

Workflow Type

☐ BWA with Mark Duplicates and Cocleaning18

ClearCaseINinput setANDData TypeISAligned ReadsANDAdvanced Search

Experimental StrategyISWXS

Add All Files to CartManifestView 9 Cases in ExplorationView ImagesBrowse Annotations

Files (18)Cases (9)184.38 GB

Primary SiteProjectData CategoryData TypeData Format

Show More

Showing 1 - 18 of 18 files

AccessFile NameCasesProjectData CategoryData FormatFile SizeAnnotations

controlled

C500.TCGA-E7-A85H-10B-01D-A34X-08.1\_gdc\_realn.bam1TCGA-BLCASequencing ReadsBAM9.82 GB0

controlled

C500.TCGA-E7-A677-10A-01D-A30H-08.1\_gdc\_realn.bam1TCGA-BLCASequencing ReadsBAM11.58 GB0

controlled

C500.TCGA-KQ-A41R-10G-01D-A34X-08.1\_gdc\_realn.bam1TCGA-BLCASequencing ReadsBAM8.95 GB0

controlled

C500.TCGA-FD-A3N5-01A-11D-A21A-08.2\_gdc\_realn.bam1TCGA-BLCASequencing ReadsBAM11.06 GB0

<https://portal.gdc.cancer.gov/legacy-archive/search/f>

**Has some extra datasets that have not been harmonised such as WGS and WBS data**

# UCSC Xena Browser

*Downloading tidy processed data*



## UCSC Xena

*See the bigger picture*

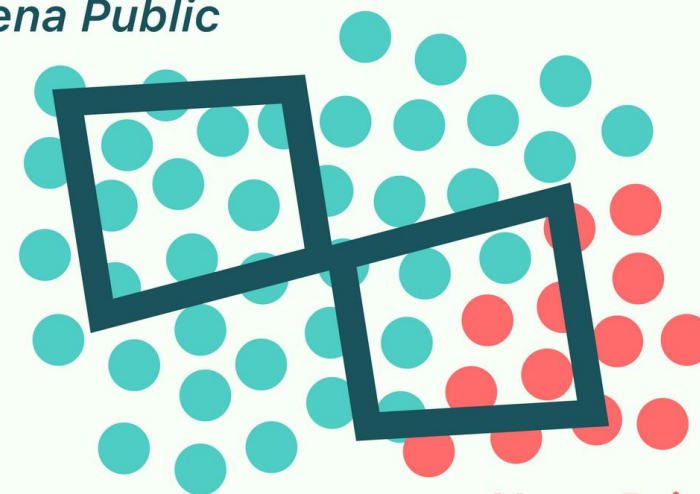
An online exploration tool for public and private,  
multi-omic and clinical/phenotype data

[Launch Xena](#)

### Tutorials and walkthroughs

Don't know where to start? Jump in with one of  
our tutorials or "How do I ..." walkthroughs

*Xena Public*



*Xena Private*



DATA SETS

VISUALIZATION

TRANSCRIPTS

DATA HUBS

VIEW MY DATA

HELP

MORE TOOLS

## 122 Cohorts, 1556 Datasets

Acute lymphoblastic leukemia (Mullighan 2008) (3 datasets)  
 Breast Cancer (Caldas 2007) (3 datasets)  
 Breast Cancer (Chin 2006) (3 datasets)  
 Breast Cancer (Haverty 2008) (2 datasets)  
 Breast Cancer (Hess 2006) (2 datasets)  
 Breast Cancer (Miller 2005) (2 datasets)  
 Breast Cancer (vantVeer 2002) (2 datasets)  
 Breast Cancer (Vijver 2002) (2 datasets)  
 Breast Cancer (Yau 2010) (2 datasets)  
 Breast Cancer Cell Lines (Heiser 2012) (4 datasets)  
 Breast Cancer Cell Lines (Neve 2006) (2 datasets)  
 Cancer Cell Line Encyclopedia (Breast) (4 datasets)  
 Cancer Cell Line Encyclopedia (CCLE) (9 datasets)  
 Connectivity Map (2 datasets)  
 GBM (Parsons 2008) (2 datasets)  
 GDC Pan-Cancer (PANCAN) (17 datasets)  
 GDC TARGET-ALL-P3 (6 datasets)  
 GDC TARGET-AML (6 datasets)  
 GDC TARGET-CCSK (6 datasets)  
 GDC TARGET-NBL (5 datasets)  
 GDC TARGET-OS (6 datasets)  
 GDC TARGET-RT (6 datasets)  
 GDC TARGET-WT (6 datasets)

## Active Data Hubs

- ☒ My computer hub
- ☒ UCSC Public Hub
- ☒ TCGA Hub
- ☒ Pan-Cancer Atlas Hub
- ☒ ICGC Hub
- ☒ UCSC Toil RNA-seq Recompute
- ☐ Treehouse Hub
- ☒ GDC Hub
- ☒ ATAC-seq Hub

cohort: GDC TCGA Lung Adenocarcinoma (LUAD)

VISUALIZE

copy number

[Copy Number Segment \(n=532\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/CNV\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/)

[Masked Copy Number Segment<sup>\\*</sup> \(n=532\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/CNV\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/)

copy number (gene-level)

[GISTIC - focal score by gene \(n=531\)](#) GDC Hub

The GDC GISTIC copy number dataset is derived from focal copy number estimates. Larger chromosomal-level deletions may not be captured in this dataset. More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/CNV\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/)

DNA methylation

[Illumina Human Methylation 27 \(n=150\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/Methylation\\_LO\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Methylation_LO_Pipeline/)

[Illumina Human Methylation 450 \(n=503\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/Methylation\\_LO\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Methylation_LO_Pipeline/)

gene expression RNAseq

[HTSeq - Counts \(n=585\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/Expression\\_mRNA\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/)

[HTSeq - FPKM \(n=585\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/Expression\\_mRNA\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/)

[HTSeq - FPKM-UQ<sup>\\*</sup> \(n=585\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/Expression\\_mRNA\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/)

phenotype

[Phenotype \(n=877\)](#) GDC Hub

[survival data \(n=738\)](#) GDC Hub

somatic mutation (SNPs and small INDELs)

[MuSE Variant Aggregation and Masking \(n=563\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/DNA\\_Seq\\_Variant\\_Calling\\_Pipeline/#somatic-variant-calling-workflow](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/DNA_Seq_Variant_Calling_Pipeline/#somatic-variant-calling-workflow)

[MuTect2 Variant Aggregation and Masking<sup>\\*</sup> \(n=567\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/DNA\\_Seq\\_Variant\\_Calling\\_Pipeline/#somatic-variant-calling-workflow](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/DNA_Seq_Variant_Calling_Pipeline/#somatic-variant-calling-workflow)

[SomaticSniper Variant Aggregation and Masking \(n=557\)](#) GDC Hub

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/DNA\\_Seq\\_Variant\\_Calling\\_Pipeline/#somatic-variant-calling-workflow](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/DNA_Seq_Variant_Calling_Pipeline/#somatic-variant-calling-workflow)

hub: <https://gdc.xenahubs.net>

More information on the GDC pipeline used to generate this data: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/DNA\\_Seq\\_Variant\\_Calling\\_Pipeline/#somatic-variant-calling-workflow](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/DNA_Seq_Variant_Calling_Pipeline/#somatic-variant-calling-workflow)

cohort	GDC TCGA Lung Adenocarcinoma (LUAD)
dataset ID	TCGA-LUAD.mutect2_snv.tsv
download	<a href="https://gdc.xenahubs.net/download/TCGA-LUAD.mutect2_snv.tsv.gz; Full metadata">https://gdc.xenahubs.net/download/TCGA-LUAD.mutect2_snv.tsv.gz; Full metadata</a>
samples	567
version	07-20-2019
type of data	somatic mutation (SNPs and small INDELs)
assembly	hg38
platform	Illumina
author	Genomic Data Commons
raw data	<a href="https://docs.gdc.cancer.gov/Data/Release_Notes/Data_Release_Notes/#data-release-180">https://docs.gdc.cancer.gov/Data/Release_Notes/Data_Release_Notes/#data-release-180</a>
raw data	<a href="https://api.gdc.cancer.gov/data/0458c57f-316c-4a7c-9294-ccd11c97c2f9">https://api.gdc.cancer.gov/data/0458c57f-316c-4a7c-9294-ccd11c97c2f9</a>
wrangling	

Data from different samples is combined into mutationVector; "Hugo\_Symbol", "Chromosome", "Start\_Position", "End\_Position", "Reference\_Allele", "Tumor\_Seq\_Allele2", "Tumor\_Sample\_Barcode", "HGVSp\_Short" and "Consequence" data are renamed accordingly and presented; "dna\_vaf" data is added and is calculated by "t\_alt\_count"/"t\_depth".

input data format Variant by Position (i.e. mutationVector)

[All Identifiers](#) [All Samples](#)

	alt	amino-acid	chrom	chromstart	chromend	effect	ref	sampleID
TCGA-17-Z056-01A	A	p.W142*	chr1	69516	69516	stop_gained	G	TCGA-17-Z056-01A
TCGA-53-A4EZ-01A	G	p.I218M	chr1	69744	69744	missense_variant	C	TCGA-53-A4EZ-01A
TCGA-55-8505-01A	CGA	p.H236_I237insR	chr1	69798	69799	inframe_insertion	-	TCGA-55-8505-01A
TCGA-L9-A444-01A	CGAATGG	p.I237Rfs*28	chr1	69798	69799	frameshift_variant	-	TCGA-L9-A444-01A

# GDC and Xena Browser

## Exercise...

*How many BAM files are present for TCGA-OR-A5J2?*

*2 x WXS, 1 x ATAC-seq, 1 x RNA-seq and 1 x miRNA-seq*

## Challenge...

*What is the Bcell\_receptors\_score for TCGA-02-2485-01?*

*(Hint: look for Pancancer analysis data in Xena browser. B-Cells are part of the immune system)*

-0.4925

TCGA Pan-Cancer (PANCAN) → Signatures → Immune Signature Scores

# Cancer genomics cloud

*Accessing and analyzing data in the cloud*

# Cancer genomics cloud

<https://cgc.sbgenomics.com/>

**Projects**

Search

- CONTROLLED LAML-WGS**  
Created by [jasonwhwong](#) · Nov. 2, 2019 22:26
- CONTROLLED WGBS**  
Created by [jasonwhwong](#) · Sept. 24, 2019 15:37
- CONTROLLED POLE**  
Created by [jasonwhwong](#) · Sept. 24, 2019 13:17
- CONTROLLED TE**  
Created by [jasonwhwong](#) · June 3, 2019 12:36
- CONTROLLED Xiaoqiang**  
Created by [xiaoqiangzhu](#) · June 3, 2019 12:33

**+ Create a project** [View all projects](#)

**Public Data and Apps**

Analyze **1.142.997**

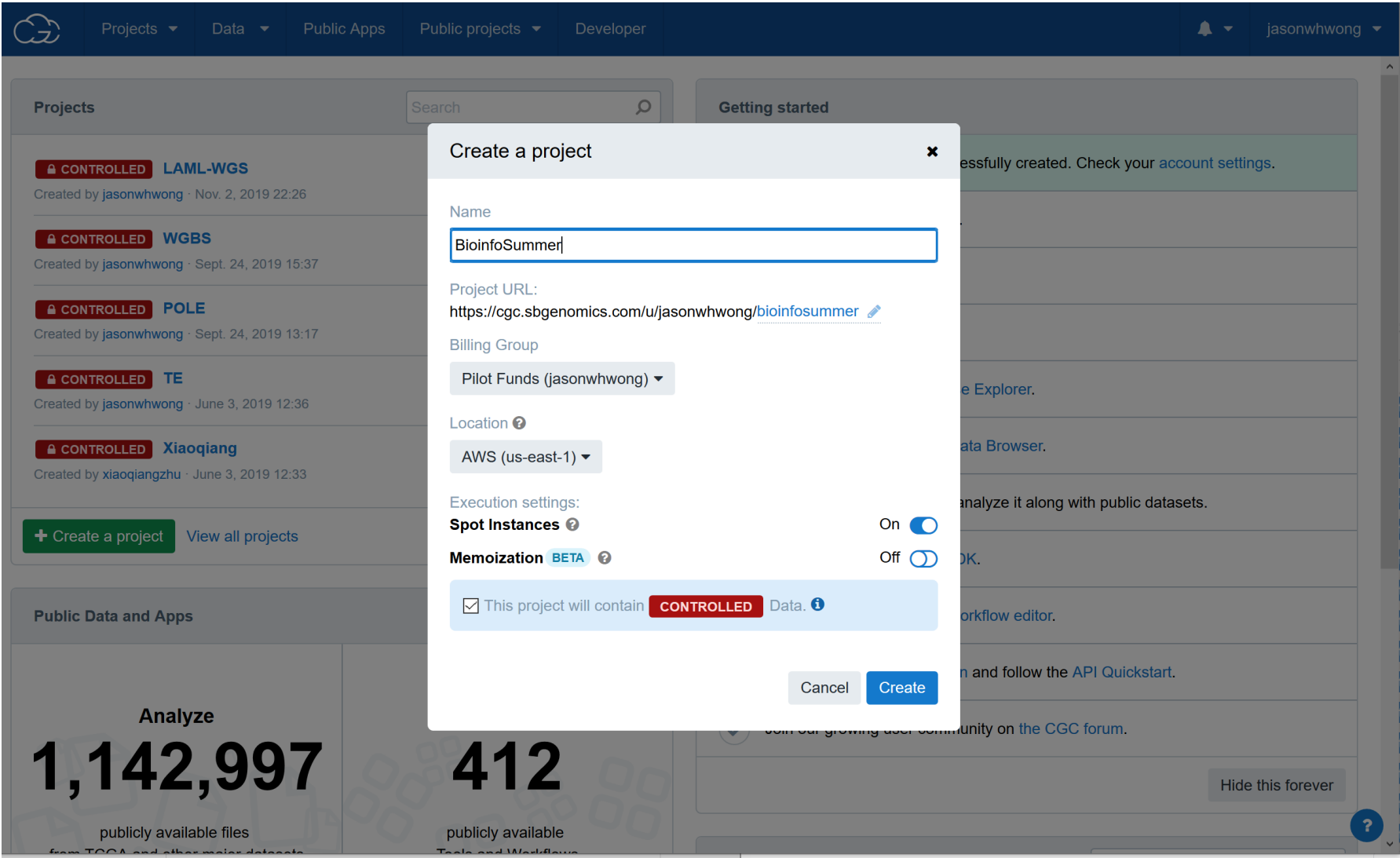
Use some of **412**

**Getting started**

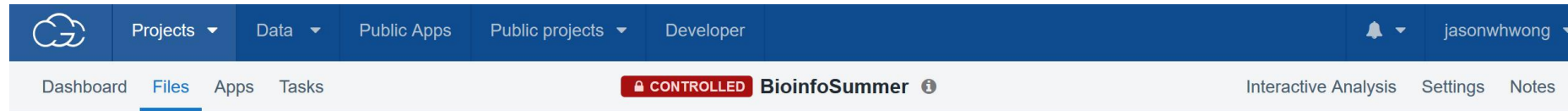
- ✓ Your account has been successfully created. Check your [account settings](#).
- ✓ Walk through the [QuickStart](#).
- ✓ Create a [project](#).
- ✓ Invite a [collaborator](#).
- ✓ Browse cases using [the Case Explorer](#).
- ✓ Search datasets using [the Data Browser](#).
- ✓ [Upload your private data](#) to analyze it along with public datasets.
- ✓ Bring your tools [using the SDK](#).
- ✓ Build a workflow using [the workflow editor](#).
- ✓ Get your [authentication token](#) and follow the [API Quickstart](#).
- ✓ Join our growing user community on [the CGC forum](#).

[Hide this forever](#)

# Cancer genomics cloud



# Cancer genomics cloud – Get WXS BAM files for ERCC2 mutant bladder cancers



Files

Files are the basis  
of every analysis.

New folder

+ Add files

[learn more about different ways to add files.](#)

# Cancer genomics cloud

Add files to "BioinfoSummer"

Public Files

[Case Explorer and Data Browser](#)

Projects

FTP / HTTP

Data Tools

Volumes

Import from PDC

## Case Explorer

Explore processed TCGA data.

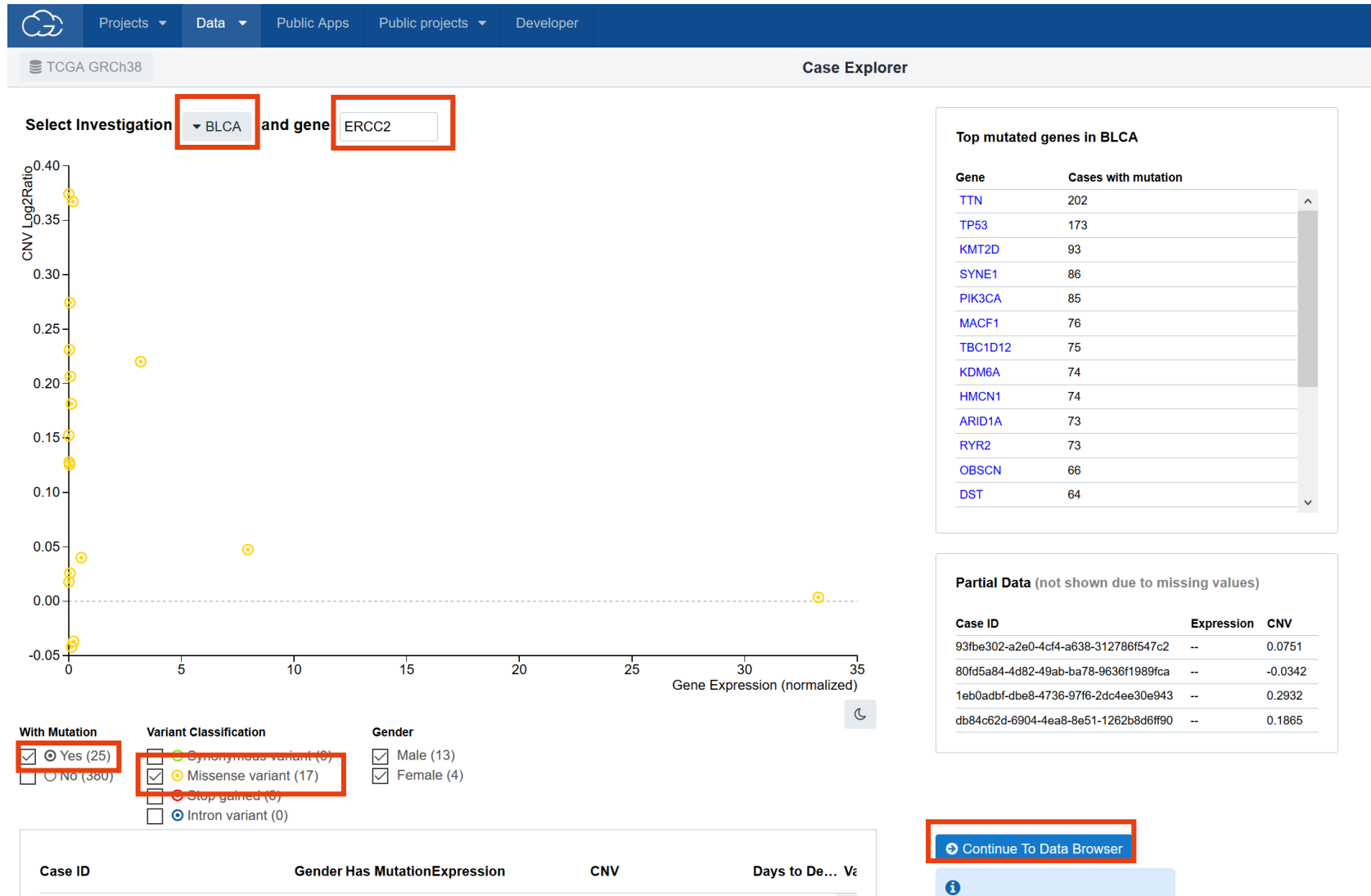
[Go to Case Explorer](#)

## Data Browser

Query TCGA metadata and add files to your projects.

[Go to Data Browser](#)

# Cancer genomics cloud



# Cancer genomics cloud

TCGA GRCh38 New query Edited

68 out of 68 file(s) copied to BioinfoSummer

Create new query Queries Search by ID Copy files to project

Choose the target project

CONTROLLED BioinfoSummer

CONTROLLED LAML-WGS

CONTROLLED WGBS

CONTROLLED POLE

CONTROLLED TE

CONTROLLED Xiaoqiang

Files marked with a lock can be used inside CONTROLLED projects Learn More

Case 17 File 34

Export Details Analytics

File

Details for C500.TCGA-GC-A3RC-01A-11D-A22Z-08.7\_gdc\_realn.bam

Connections

Inbound: Case 1

Outbound:

No outbound connections

TCGA GRCh38

Access level Controlled

Created datetime 2016-05-04T04:46:55.814Z

Data category Sequencing Reads

Data format BAM

Data type Aligned Reads

File

C500.TCGA-GC-A3RC-01A-11D-A22Z-08.7\_gdc\_realn.bam

C500.TCGA-DK-AA77-10A-01D-A394-08.1\_gdc\_realn.bam

C500.TCGA-XF-A9SM-01A-11D-A42E-08.2\_gdc\_realn.bam

C500.TCGA-MV-A51V-01A-11D-A26M-08.5\_gdc\_realn.bam

C500.TCGA-XF-AAMZ-01A-11D-A42E-08.2\_gdc\_realn.bam

# Cancer genomics cloud

Cloud icon

Projects

Data

Public Apps

Public projects

Developer

Dashboard

Files

Apps

Tasks

CONTROLLED

BioinfoSummer

Interactive Analysis

Settings

Notes

Jasonwhwong

Files

New folder

Add files

Search

Type: All

Sample ID: All

Task ID: All

Tags: All

Clear filters

Name	Experimental strategy	Type	Size	Sample ID
<div>TCGA GRCH38</div> C500.TCGA-YF-AA3L-10A-01D-A38J-08.2_gdc_realn.bam	WXS	BAM	7.4 GiB	TCGA-YF-AA3L-10A
<div>TCGA GRCH38</div> C500.TCGA-YF-AA3L-10A-01D-A38J-08.2_gdc_realn.bai	WXS	BAI	5.9 MiB	TCGA-YF-AA3L-10A
<div>TCGA GRCH38</div> C500.TCGA-YF-AA3L-01A-11D-A38G-08.2_gdc_realn.bam	WXS	BAM	8.6 GiB	TCGA-YF-AA3L-01A
<div>TCGA GRCH38</div> C500.TCGA-YF-AA3L-01A-11D-A38G-08.2_gdc_realn.bai	WXS	BAI	5.9 MiB	TCGA-YF-AA3L-01A
<div>TCGA GRCH38</div> C500.TCGA-XF-AAMZ-10A-01D-A42H-08.2_gdc_realn.bam	WXS	BAM	5.8 GiB	TCGA-XF-AAMZ-10A
<div>TCGA GRCH38</div> C500.TCGA-XF-AAMZ-10A-01D-A42H-08.2_gdc_realn.bai	WXS	BAI	5.9 MiB	TCGA-XF-AAMZ-10A
<div>TCGA GRCH38</div> C500.TCGA-XF-AAMZ-01A-11D-A42E-08.2_gdc_realn.bam	WXS	BAM	7.7 GiB	TCGA-XF-AAMZ-01A
<div>TCGA GRCH38</div> C500.TCGA-XF-AAMZ-01A-11D-A42E-08.2_gdc_realn.bai	WXS	BAI	5.9 MiB	TCGA-XF-AAMZ-01A
<div>TCGA GRCH38</div> C500.TCGA-XF-AAMX-10A-01D-A42H-08.2_gdc_realn.bam	WXS	BAM	7.5 GiB	TCGA-XF-AAMX-10A
<div>TCGA GRCH38</div> C500.TCGA-XF-AAMX-10A-01D-A42H-08.2_gdc_realn.bai	WXS	BAI	5.9 MiB	TCGA-XF-AAMX-10A
<div>TCGA GRCH38</div> C500.TCGA-XF-AAMX-01A-11D-A42E-08.2_gdc_realn.bam	WXS	BAM	7.9 GiB	TCGA-XF-AAMX-01A
<div>TCGA GRCH38</div> C500.TCGA-XF-AAMX-01A-11D-A42E-08.2_gdc_realn.bai	WXS	BAI	5.9 MiB	TCGA-XF-AAMX-01A
<div>TCGA GRCH38</div> C500.TCGA-XF-A9SM-10A-01D-A42H-08.2_gdc_realn.bam	WXS	BAM	6.1 GiB	TCGA-XF-A9SM-10A
<div>TCGA GRCH38</div> C500.TCGA-XF-A9SM-10A-01D-A42H-08.2_gdc_realn.bai	WXS	BAI	5.9 MiB	TCGA-XF-A9SM-10A

Refresh

Showing 1-68 of 68

# Cancer genomics cloud

Dashboard Files Apps Tasks

 **CONTROLLED** BioinfoSummer 

[Interactive Analysis](#)

## Explore genomics data

Understand complex genomics data with interactive analysis tools.

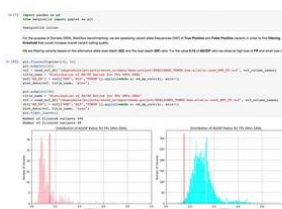


### Genome Browser

Visualize alignments, SNV/Indels, annotation tracks, check coverage and mismatch, assess alignments and variants

34 files

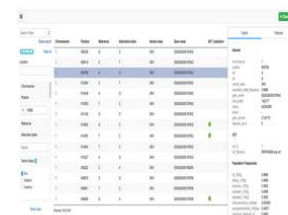
[Open](#)



### Data Cruncher

Analyze and explore data using JupyterLab or RStudio

[Open](#)



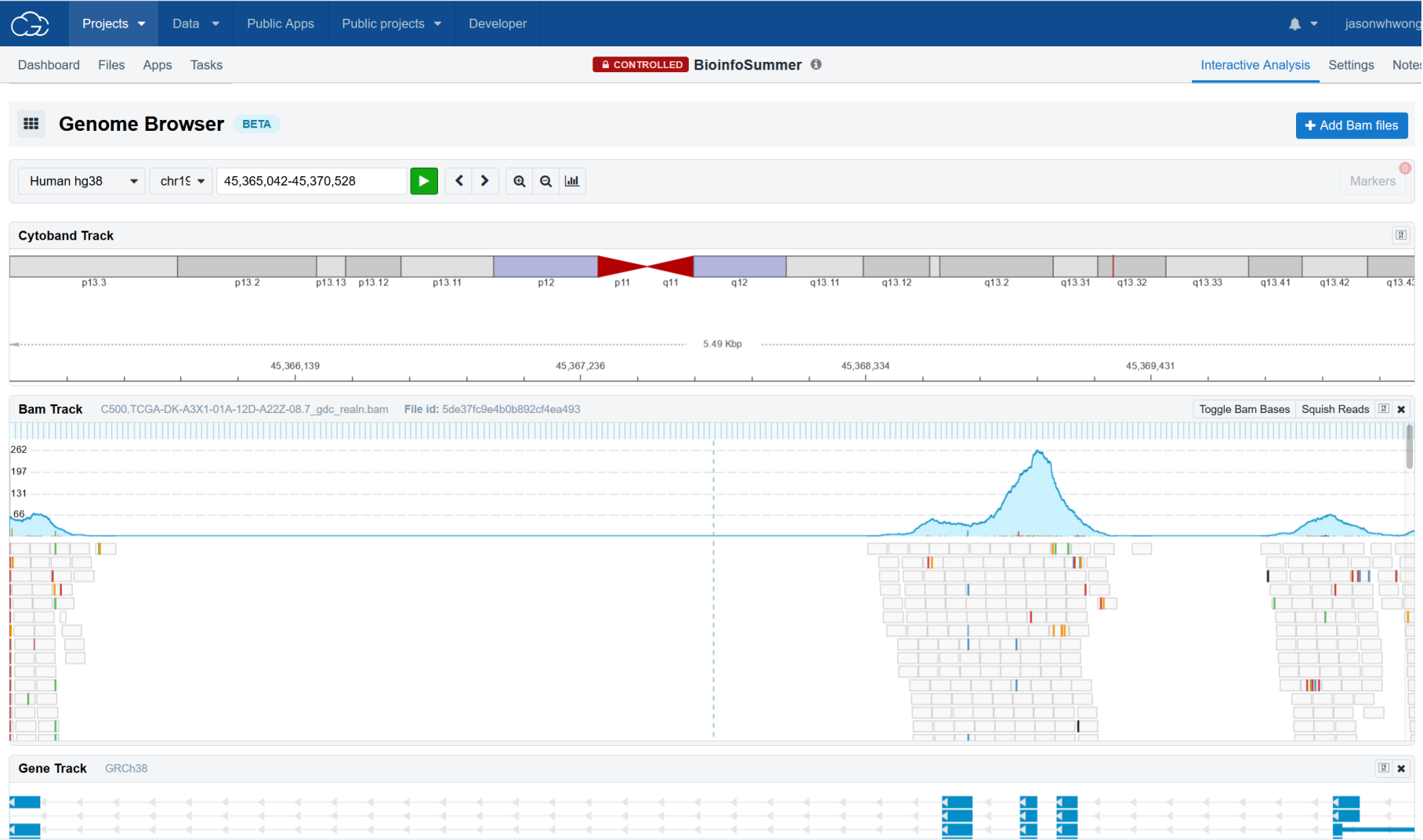
### Variant Browser BETA

Filter and interpret your annotated data

0 files

[Open](#)

# Cancer genomics cloud



# Cancer genomics cloud

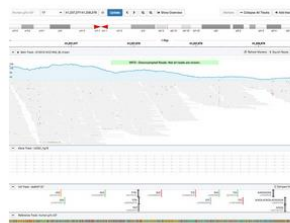
Dashboard Files Apps Tasks

 **CONTROLLED** BioinfoSummer 

[Interactive Analysis](#)

## Explore genomics data

Understand complex genomics data with interactive analysis tools.

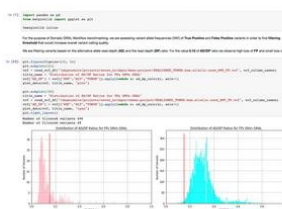


### Genome Browser

Visualize alignments, SNV/Indels, annotation tracks, check coverage and mismatch, assess alignments and variants

34 files

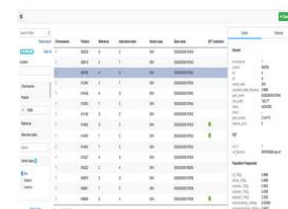
Open



### Data Cruncher

Analyze and explore data using JupyterLab or RStudio

Open



### Variant Browser BETA

Filter and interpret your annotated data

0 files

Open

# Cancer genomics cloud

Create new analysis

Basic information

Compute requirements

Analysis name

Test

Environment

JupyterLab

Web-based UI for Project Jupyter

RStudio

BETA

IDE for R

Skip wizard

Previous

Next

Interfacing across multiple projects

Your analyses will appear here

Create your first analysis

Create new analysis

Basic information

Compute requirements

Select an instance type with adequate CPU, memory and storage allocation for your analysis. This can be changed between analysis runs, but not while the analysis is running.

Instance type

m4.large (1024GB EBS, 2vCPUs, 8GB RA...

Suspend time

30

Minutes

On

Instance types

m4.16xlarge (1024GB EBS, 64vCPUs, 256GB RAM)

m4.2xlarge (1024GB EBS, 8vCPUs, 32GB RAM)

m4.4xlarge (1024GB EBS, 16vCPUs, 64GB RAM)

m4.large (1024GB EBS, 2vCPUs, 8GB RAM)

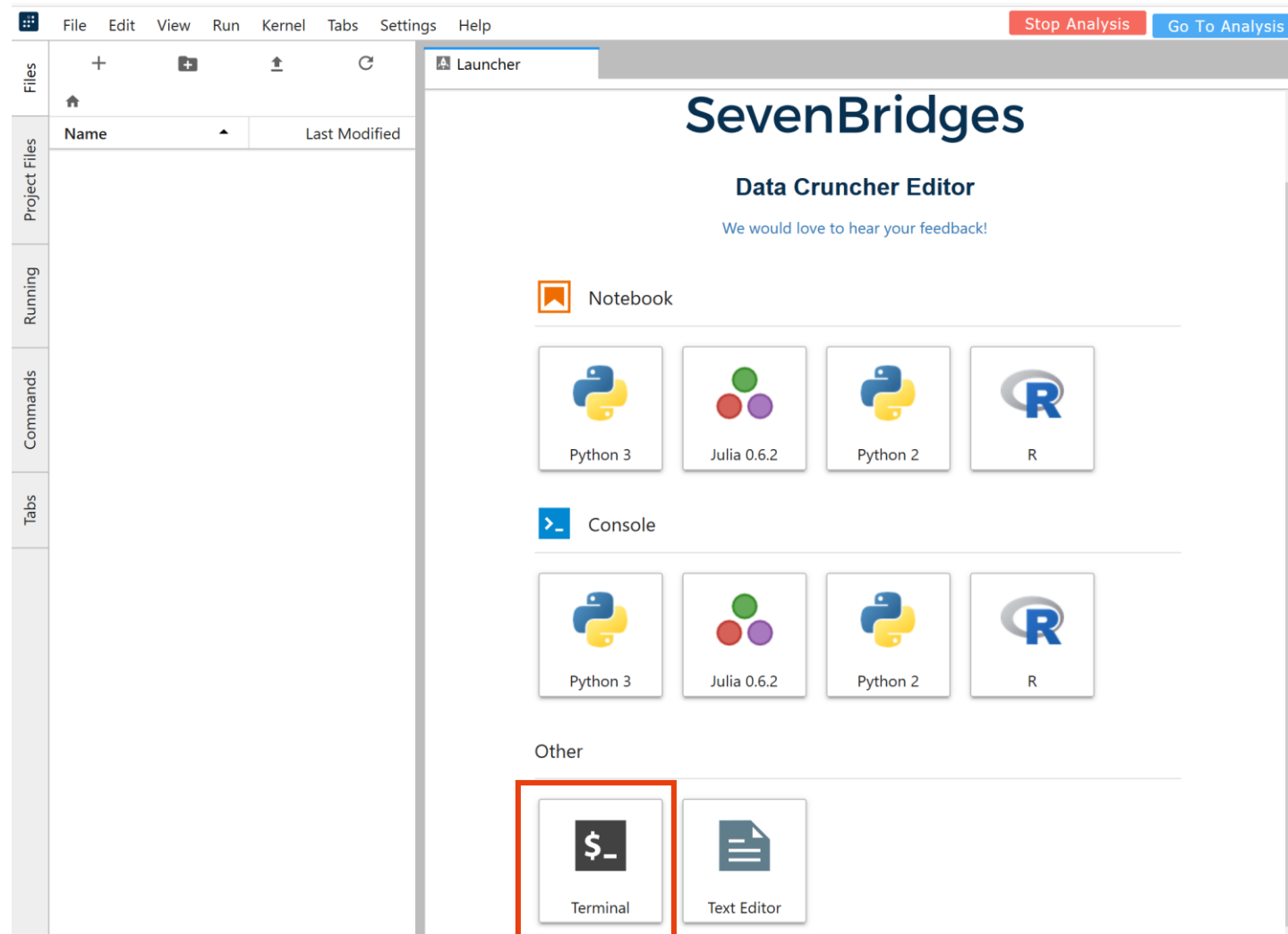
m4.xlarge (1024GB EBS, 4vCPUs, 16GB RAM)

Start the analysis

# Cancer genomics cloud

The screenshot displays the Cancer Genomics Cloud web interface. At the top, a dark blue navigation bar contains a cloud icon, dropdown menus for 'Projects', 'Data', and 'Public A...', a notification bell, and the user name 'jasonwhwong'. Below this is a secondary navigation bar with 'Dashboard', 'Files', 'Apps', and 'Tasks'. A green notification box states: 'Your instance is up and running! Open the editor to start working on your analysis.' The main content area shows a 'RUNNING Test' session. It includes a timestamp 't...', session details 'Session started on Dec. 1, 2019 20:02 by jasonwhwong', environment info 'Environment: JupyterLab | Price: \$0.02 | Duration: 7 minutes', and a 'View Sessions' button. Action buttons for 'Copy', 'Stop', and 'Open in editor' (highlighted with a red box) are present. A light blue information banner states: 'Your instance is up and running. After 30 minutes of inactivity we will stop the instance and save your analysis. Learn more'. Below this, tabs for 'Files' and 'Settings' are shown. The 'Files' tab is active, displaying 'Analysis files:' with 'No notebooks' and 'Produced by this analysis' with 'No files'. A large watermark graphic of overlapping document icons is centered over the main workspace area, with the text 'Select analysis file to preview.'

# Cancer genomics cloud



# Cancer genomics cloud

Project files are in “/sbgenomics/project-files”  
Write files to “/sbgenomics/workspace”

File Edit View Run Kernel Tabs Settings Help

Stop Analysis Go To Analysis

Files

+

+

+

↺

Home

Name

▲

Last Modified

Project Files

Running

Commands

Tabs

joyvan@6b036: X

/sbgenomics/workspace\$ ll /sbgenomics/project-files/

total 325891758

-rw-r--r-- 1 1001 1002 6185888 Dec 1 08:54 C500.TCGA-DK-A3X1-01A-12D-A22Z-08.7\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 10705997580 Dec 1 08:54 C500.TCGA-DK-A3X1-01A-12D-A22Z-08.7\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6243376 Dec 1 08:54 C500.TCGA-DK-A3X1-10A-01D-A22Z-08.9\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 12245607442 Dec 1 08:54 C500.TCGA-DK-A3X1-10A-01D-A22Z-08.9\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6234200 Dec 1 08:54 C500.TCGA-DK-A6AW-01A-11D-A30E-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 10259162456 Dec 1 08:54 C500.TCGA-DK-A6AW-01A-11D-A30E-08.1\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6261384 Dec 1 08:54 C500.TCGA-DK-A6AW-10A-01D-A30H-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 12132432621 Dec 1 08:54 C500.TCGA-DK-A6AW-10A-01D-A30H-08.1\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6191664 Dec 1 08:54 C500.TCGA-DK-AA77-01A-11D-A391-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 9210282051 Dec 1 08:54 C500.TCGA-DK-AA77-01A-11D-A391-08.1\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6226784 Dec 1 08:54 C500.TCGA-DK-AA77-10A-01D-A394-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 10058581085 Dec 1 08:54 C500.TCGA-DK-AA77-10A-01D-A394-08.1\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6295424 Dec 1 08:54 C500.TCGA-E7-A7XN-01A-11D-A34U-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 12784067246 Dec 1 08:54 C500.TCGA-E7-A7XN-01A-11D-A34U-08.1\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6237632 Dec 1 08:54 C500.TCGA-E7-A7XN-10A-01D-A34X-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 10036280024 Dec 1 08:54 C500.TCGA-E7-A7XN-10A-01D-A34X-08.1\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6281200 Dec 1 08:54 C500.TCGA-FD-A43X-01A-11D-A23U-08.7\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 12530900104 Dec 1 08:54 C500.TCGA-FD-A43X-01A-11D-A23U-08.7\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6195112 Dec 1 08:54 C500.TCGA-FD-A43X-10A-01D-A23U-08.7\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 11891368515 Dec 1 08:54 C500.TCGA-FD-A43X-10A-01D-A23U-08.7\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 5959712 Dec 1 08:54 C500.TCGA-FD-A5BX-01A-11D-A26M-08.5\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 5508436311 Dec 1 08:54 C500.TCGA-FD-A5BX-01A-11D-A26M-08.5\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6099080 Dec 1 08:54 C500.TCGA-FD-A5BX-10A-01D-A26K-08.5\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 6569476911 Dec 1 08:54 C500.TCGA-FD-A5BX-10A-01D-A26K-08.5\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6283792 Dec 1 08:54 C500.TCGA-FJ-A3Z7-01A-12D-A23M-08.3\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 13982022480 Dec 1 08:54 C500.TCGA-FJ-A3Z7-01A-12D-A23M-08.3\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6204624 Dec 1 08:54 C500.TCGA-FJ-A3Z7-10A-01D-A23K-08.3\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 11157407292 Dec 1 08:54 C500.TCGA-FJ-A3Z7-10A-01D-A23K-08.3\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6103856 Dec 1 08:54 C500.TCGA-GC-A3RC-01A-11D-A22Z-08.7\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 9457140584 Dec 1 08:54 C500.TCGA-GC-A3RC-01A-11D-A22Z-08.7\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6159448 Dec 1 08:54 C500.TCGA-GC-A3RC-10B-01D-A22Z-08.9\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 9651371746 Dec 1 08:54 C500.TCGA-GC-A3RC-10B-01D-A22Z-08.9\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6215448 Dec 1 08:54 C500.TCGA-GU-AATO-01A-11D-A391-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 10145211195 Dec 1 08:54 C500.TCGA-GU-AATO-01A-11D-A391-08.1\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6273616 Dec 1 08:54 C500.TCGA-GU-AATO-10A-01D-A394-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 7439992923 Dec 1 08:54 C500.TCGA-GU-AATO-10A-01D-A394-08.1\_gdc\_realn.bam

-rw-r--r-- 1 1001 1002 6199584 Dec 1 08:54 C500.TCGA-GU-AATP-01A-11D-A391-08.1\_gdc\_realn.bai

-rw-r--r-- 1 1001 1002 9805654032 Dec 1 08:54 C500.TCGA-GU-AATP-01A-11D-A391-08.1\_gdc\_realn.bam

# Cancer genomics cloud

Upload your precompiled Ubuntu tools

File Edit View Run Kernel Tabs Settings Help

Stop Analysis Go To Analysis

Files

+

+

↑

↺

Home

Project Files

Running

Commands

Tabs

Name	Last Modified
Tools	a year ago
Tools_for_linux.tar.gz	seconds ago

joyan@6b036: X Launcher X

```
/sbgenomics/workspace$ ll
total 2980
-rw-r--r-- 1 joyan users 3050748 Dec  1 09:19 Tools_for_linux.tar.gz
/sbgenomics/workspace$ tar -zxvf Tools_for_linux.tar.gz
Tools/
Tools/annotateBed
Tools/bedtools
Tools/shiftBed
Tools/samtools
Tools/coverageBed
Tools/windowBed
Tools/genomeCoverageBed
Tools/maskFastaFromBed
Tools/bamToBed
Tools/linksBed
Tools/nucBed
Tools/subtractBed
Tools/fastFromBed
Tools/pairToBed
Tools/bwa
Tools/flankBed
Tools/mapBed
Tools/clusterBed
Tools/lofreq2_call_pparallel.py
Tools/slopBed
Tools/intersectBed
Tools/randomBed
Tools/complementBed
Tools/shuffleBed
Tools/mergeBed
Tools/closestBed
Tools/sortBed
Tools/multiIntersectBed
Tools/lofreq
Tools/lofreq2_somatic.py
/sbgenomics/workspace$
```

<https://cloudstor.aarnet.edu.au/plus/s/O7gw6CRJU3H0rIk/download>

# Cancer genomics cloud

[illegible]

# Cancer genomics cloud

The screenshot displays the Cancer Genomics Cloud interface. On the left, a sidebar contains a file explorer with a table of files:

Name	Last Modified
Tools	
header.txt	
Tools_for_linux.t	

A context menu is open over the 'header.txt' file, listing actions such as 'Save To Project', 'Open', 'Rename', 'Delete', 'Cut', 'Copy', 'Paste', 'Duplicate', 'Download' (highlighted with a red box), 'Shutdown Kernel', and 'Copy Path'. The main panel on the right is a terminal window titled 'jovyan@6b036: X Launcher X' showing the command prompt and the execution of 'samtools view -H /sbgenomics/project-files/C500.TCGA-MV-A51V-01A-11D-A26M-08.5\_gdc\_realn.bam > header.txt'. At the top right of the interface, there are two buttons: 'Stop Analysis' (highlighted with a red box) and 'Go To Analysis'.

# Cancer genomics cloud

Public Apps

## Public apps for your data analysis

Browse 430 publicly available Common Workflow Language workflows and tools to enable reproducible bioinformatics.

Search workflows and tools or Explore all apps

### Featured Apps

#### RNA-seq Alignment - STAR 2.5.4b

Toolkit: STAR 2.5.4b

This workflow performs the first step of RNA-seq analysis - alignment to a reference genome and transcriptome.

ALIGNMENT RNA

Copy Run