



The Cancer Genome Atlas (TCGA) & International Cancer Genome Consortium (ICGC)

Session 4 – Rebecca Poulos

Never Stand Still

Medicine

Prince of Wales Clinical School

Introductory bioinformatics for human genomics workshop, UNSW

20th – 21st April 2017



Facts on cancer

- An estimated 134,000 new cases of cancer will be diagnosed in Australia this year, with that number set to rise to 150,000 by 2020
- Cancer is a leading cause of death in Australia. In 2014, > 44,000 people died from cancer, accounting for about 3 in every 10 deaths.

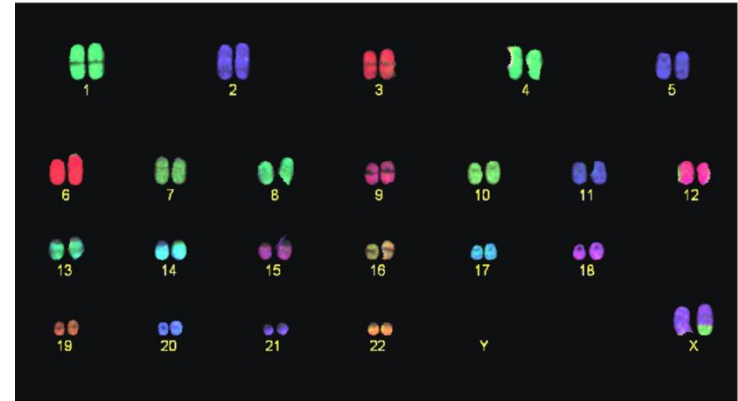
Source: Cancer Council Australia (2017)



Cancer is a disease of the genome

- Challenges in treating cancer:
 - Every patient is different
 - Every tumour is different, even in the same patient
 - Tumours can be highly heterogeneous
 - High rate of genomic abnormalities (few drivers, many passenger mutations)

Healthy 46 chromosomes



Example cancer 59 chromosomes

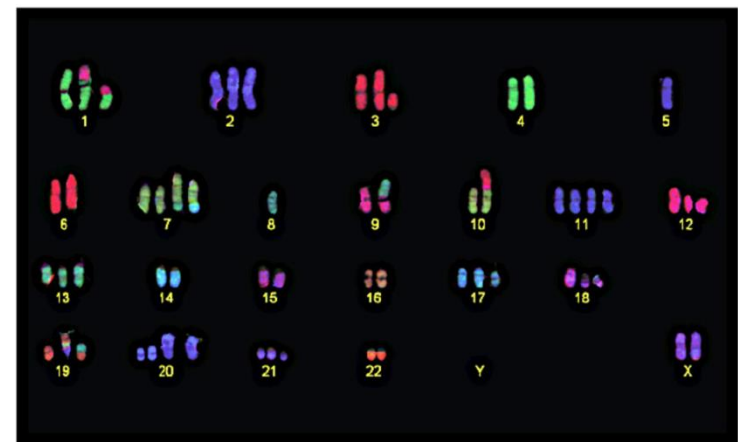


Image from Thompson & Compton *Chromosome Res* 2011.

What can go wrong in cancer genomes?

Types of changes	Some common technologies used to study these changes
DNA mutations - Point mutations - Insertions & deletions	WGS; WXS

WGS = whole genome sequencing, WXS = whole exome sequencing

RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing



What can go wrong in cancer genomes?

Types of changes	Some common technologies used to study these changes
DNA mutations <ul style="list-style-type: none">- Point mutations- Insertions & deletions	WGS; WXS
DNA structural variations	WGS

WGS = whole genome sequencing, WXS = whole exome sequencing

RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing



What can go wrong in cancer genomes?

Types of changes	Some common technologies used to study these changes
DNA mutations <ul style="list-style-type: none">- Point mutations- Insertions & deletions	WGS; WXS
DNA structural variations	WGS
Copy number variation (CNV)	CGH array; SNP array; WGS

WGS = whole genome sequencing, WXS = whole exome sequencing

RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing



What can go wrong in cancer genomes?

Types of changes	Some common technologies used to study these changes
DNA mutations <ul style="list-style-type: none">- Point mutations- Insertions & deletions	WGS; WXS
DNA structural variations	WGS
Copy number variation (CNV)	CGH array; SNP array; WGS
DNA methylation	Methylation array; RRBS; WGBS

WGS = whole genome sequencing, WXS = whole exome sequencing

RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing



What can go wrong in cancer genomes?

Types of changes	Some common technologies used to study these changes
DNA mutations <ul style="list-style-type: none">- Point mutations- Insertions & deletions	WGS; WXS
DNA structural variations	WGS
Copy number variation (CNV)	CGH array; SNP array; WGS
DNA methylation	Methylation array; RRBS; WGBS
mRNA expression changes	mRNA expression array; RNA-seq

WGS = whole genome sequencing, WXS = whole exome sequencing

RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing



What can go wrong in cancer genomes?

Types of changes	Some common technologies used to study these changes
DNA mutations <ul style="list-style-type: none">- Point mutations- Insertions & deletions	WGS; WXS
DNA structural variations	WGS
Copy number variation (CNV)	CGH array; SNP array; WGS
DNA methylation	Methylation array; RRBS; WGBS
mRNA expression changes	mRNA expression array; RNA-seq
miRNA expression changes	miRNA expression array; miRNA-seq

WGS = whole genome sequencing, WXS = whole exome sequencing

RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing



What can go wrong in cancer genomes?

Types of changes	Some common technologies used to study these changes
DNA mutations <ul style="list-style-type: none">- Point mutations- Insertions & deletions	WGS; WXS
DNA structural variations	WGS
Copy number variation (CNV)	CGH array; SNP array; WGS
DNA methylation	Methylation array; RRBS; WGBS
mRNA expression changes	mRNA expression array; RNA-seq
miRNA expression changes	miRNA expression array; miRNA-seq
Protein expression	Protein arrays; mass spectrometry

WGS = whole genome sequencing, WXS = whole exome sequencing

RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing



Goal of cancer genomics

- Identify changes in the genomes of tumors that drive cancer progression
- Understand how normal cells become cancerous
- Identify new targets for therapy
- Select drugs based on the genomics of the tumour – i.e. personalised therapy

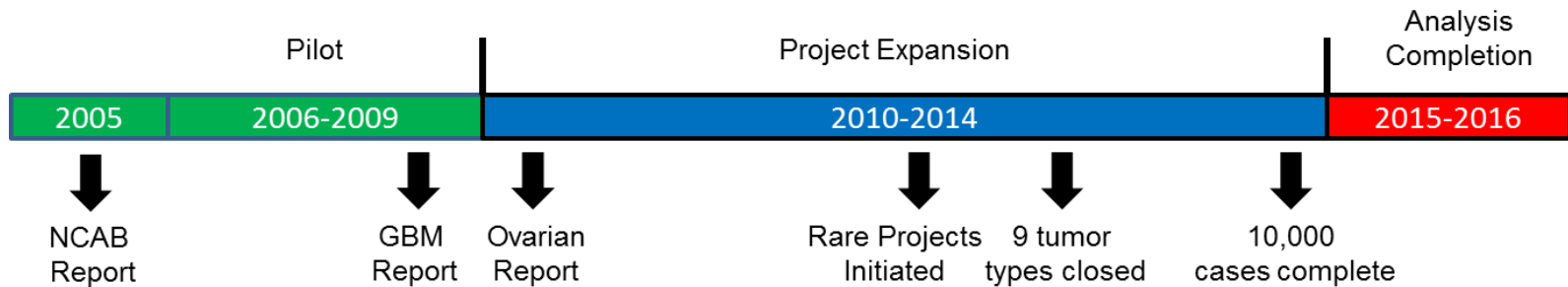


Cancer Sequencing Projects

The Cancer Genome Atlas (TCGA)

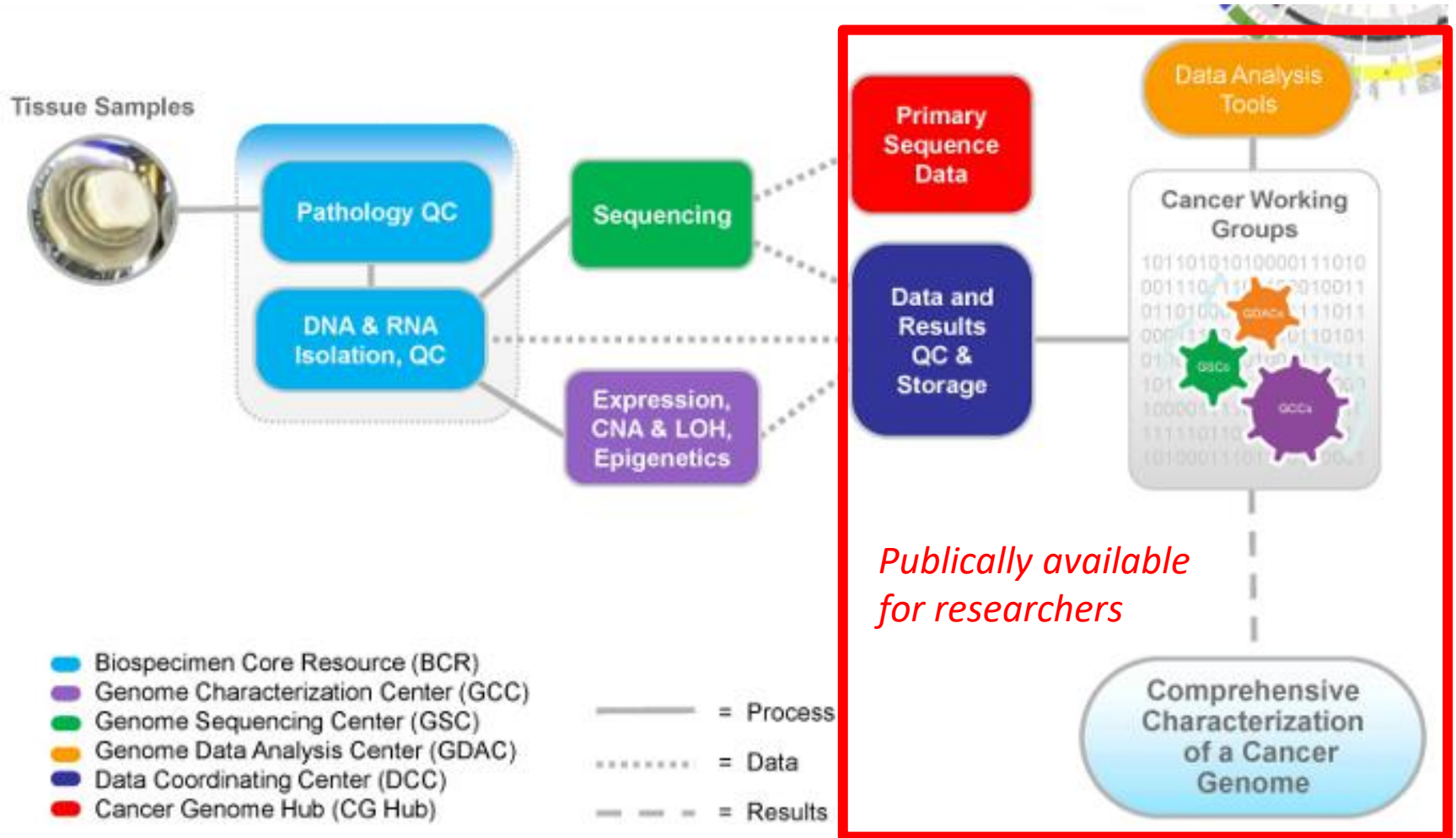


- Led by NIH
- Initiated in 2006 (as a pilot program) and expanded in 2009



- **Aim:**
To make the genomes of 20 cancers publically available
- **Update today:**
33 cancer types & subtypes analysed (11,000 samples)

TCGA pipeline



Types of Cancers

- **Breast**
 - Ductal carcinoma
 - Lobular carcinoma
- **Central nervous system**
 - Glioblastoma multiforme
 - Lower grade glioma
- **Endocrine**
 - Adrenocortical carcinoma
 - Papillary thyroid carcinoma
 - Paranglioma and pheochromocytoma
- **Gastrointestinal**
 - Cholangiocarcinoma
 - Colorectal Adenocarcinoma
 - Liver Hepatocellular Carcinoma
 - Pancreatic Ductal Adenocarcinoma
 - Stomach-Esophageal Cancer
- **Gynecological**
 - Cervical Cancer
 - Ovarian Serous Cystadenocarcinoma
 - Uterine Carcinosarcoma
 - Uterine Corpus Endometrial Carcinoma
- **Head and neck**
 - Squamous cell carcinoma
 - Uveal melanoma
- **Hematologic**
 - Acute myeloid leukemia
 - Thymoma
- **Skin**
 - Cutaneous melanoma
- **Soft tissue**
 - Sarcoma
- **Thoracic**
 - Lung Adenocarcinoma
 - Lung Squamous Cell Carcinoma
 - Mesothelioma
- **Urologic**
 - Chromophobe Renal Cell Carcinoma
 - Clear Cell Kidney Carcinoma
 - Papillary Kidney Carcinoma
 - Prostate Adenocarcinoma
 - Testicular Germ Cell Cancer
 - Urothelial Bladder Carcinoma



Datasets

Data types

- Clinical data
- Images
- Microsatellite instability
- DNA sequencing
- miRNA sequencing
- Protein expression
- mRNA & RNA sequencing
- Array-based expression
- DNA methylation
- Copy number

Data access tiers

- Open access
 - De-identified
 - Requires no certification
- Controlled access
 - No direct identifiers
 - Must complete Data Access Request (DAR) form



Genomic Data Commons (GDC)

- TCGA data is stored on the Genomic Data Commons (GDC) data portal: <https://portal.gdc.cancer.gov/>

The screenshot shows the GDC Data Portal homepage. At the top, there is a navigation bar with 'Home', 'Projects', 'Data', and 'Analysis' links. A search bar and utility icons for 'Quick Search', 'Login', 'Cart', and 'GDC Apps' are also present. The main content area features a blue header with 'Harmonized Cancer Datasets' and 'Genomic Data Commons Data Portal'. Below this, there are two prominent buttons: 'Projects' (highlighted with a red box) and 'Data'. A section titled 'Perform Advanced Search Queries, such as:' lists three search criteria with corresponding case and file counts: 'Cases of kidney cancer diagnosed at the age of 20 and below' (736 Cases, 1,519 Files), 'CNV data of female brain cancer cases' (459 Cases, 1,788 Files), and 'Gene expression quantification data in TCGA-GBM project' (166 Cases, 522 Files). To the right, a bar chart titled 'Cases by Primary Site' displays the number of cases for various cancer types, with 'Kidney' having the highest count at over 1,500. Below the chart, a summary row shows: 'DATA PORTAL SUMMARY' (Data Release 5.0 - March 16, 2017), 'PROJECTS 39', 'PRIMARY SITE 29', 'CASES 14,551', and 'FILES 274,724'. The bottom section is divided into three columns: 'Infrastructure' (12,800 Cores, 87.96 TB RAM, 4.98 PB Used, 5.42 PB Total), 'Documentation' (Browse Data using Facet Search, Search Data with Advanced Search Technology, Project Based Data Availability, Controlled Access Data, Visit the Documentation Website), and 'GDC Applications' (Data Portal, Website, Data Transfer Tool, API, Data Submission Portal, Documentation, Legacy Archive, GDC cBio Portal).

Exploring the “Data” option...

The dashboard is divided into several sections:

- Left Panel (Filters):** A sidebar with a red border containing search and filter options for Cases and Files. It includes sections for Case, Case Submitter ID Prefix, Primary Site (Kidney, Brain, Nervous System, Breast, Lung), Cancer Program (TCGA, TARGET), Project (TARGET-NBL, TCGA-BRCA, TARGET-AML, TARGET-WT, TCGA-GBM), and Disease Type.
- Top Bar:** Contains a search prompt, an 'Advanced' search button, and tabs for Summary, Cases (14,551), and Files (274,724). Buttons for 'Add all files to the Cart' and 'Download Manifest' are also present.
- Summary Cards:** Three cards at the top right showing 'FILES 274,724', 'CASES 14,551', and 'FILE SIZE 470.57 TB'.
- Charts:** Six pie charts arranged in a 2x3 grid, each representing a different facet: 'File Counts by Project' (39 Projects), 'File Counts by Access Level' (2 Access Levels), 'File Counts by Data Format' (7 Data Formats), 'File Counts by Primary Site', 'File Counts by Data Type', and 'File Counts by Experimental Strategy'.

Search and filter files using this utility

Let's find all processed RNA-seq data for colorectal cancer...

Cases Files << Hide Filters

[Add a Case/Biospecimen Filter](#)

Case

Search for Case Id

Case Submitter ID Prefix

Search for Submitter Id

Primary Site

- Colorectal 633
- Kidney 1,681
- 4,422

Cases Files << Hide Filters

[Add a File Filter](#)

File

Search for File Id

Data Category

- Transcriptome Profiling 698

Data Type

- Gene Expression Quantification 698

Experimental Strategy

- RNA-Seq 698

Workflow Type

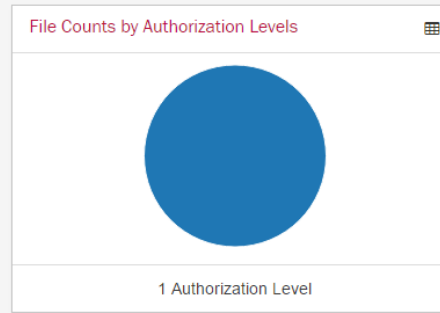
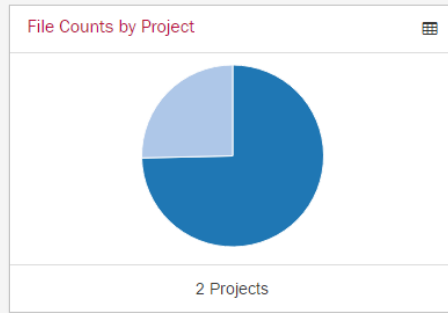
- HTSeq - FPKM-UQ 698
- HTSeq - Counts 698
- HTSeq - FPKM 698

Let's find all processed RNA-seq data for colorectal cancer...

FILES
698

CASES
623

FILE SIZE
350.96 MB



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.

Metadata Download Remove From Cart

Download

- Manifest
- Cart

Cart Items

Showing 1 - 20 of 698 cart items

Action	Access	File Name	Cases	Project	Data Type	Data Format	Size	Annotations
	Open	00106523-5b1d-44ad-a9f1-7d84db08722c.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	516 KB	1
	Open	00589871-e54f-492f-988f-502670edd606.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	490 KB	2
	Open	00cc9b4d-a847-464e-979a-7751e1a87ae3.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	449 KB	2
	Open	00f768f9-9e6c-4e84-bdba-c19368f7e522.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	518 KB	0
	Open	0134d0cc-e66b-4fda-804b-c4434ec00bd2.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	468 KB	0
	Open	020aa019-a3a4-4055-92ee-be824a597501.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	505 KB	0
	Open	02a81bc3-4672-4d2f-808d-d27f4b63bf85.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	492 KB	0
	Open	02dfa3d-13b3-4624-994d-62e740fa4a3d.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	502 KB	0
	Open	03109baa-936f-49ae-acb3-2d00ac03c7ab.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	474 KB	1
	Open	0319a4cc-ee3f-4c6f-83e1-7cdd5f865dc0.FPKM-UQ.txt.gz	1	TCGA-COAD	Gene Expression Quantification	TXT	508 KB	1
	Open	037bb811-58ff-45ba-9707-f4c11d05d333.FPKM-UQ.txt.gz	1	TCGA-READ	Gene Expression Quantification	TXT	512 KB	0

Genomic Data Commons (GDC)

- The GDC data portal is very user-friendly
- GDC is ideal for downloading data in large tab delimited format – perfect for a bioinformatician
- However, data portal files are difficult to use for the average biologist
- Fortunately there are some alternatives:
 - cBioPortal (www.cbioportal.org/)
 - ICGC data portal (<http://dcc.icgc.org/>)



cBioPortal (www.cbioportal.org/)

- A data analysis portal to TCGA data
- Provides functions for visualisation, analysis and download of data.
- Maintained by Memorial Sloan-Kettering Cancer Center



Features of cBioPortal

- Visualising frequency of mutations
- Correlation between occurrence of mutations
- Correlation of expression and CNV or methylation
- Visualisation of mutations
- Survival analysis
- Network analysis

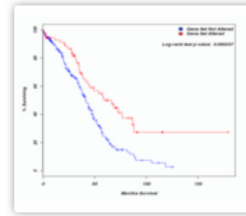
Gao et al (2013) Sci. Signal



The cBioPortal for Cancer Genomics provides **visualization, analysis** and **download** of large-scale **cancer genomics** data sets.

Please adhere to [the TCGA publication guidelines](#) when using TCGA data in your publications.

Please cite Gao et al. *Sci. Signal.* 2013 & Cerami et al. *Cancer Discov.* 2012 when publishing results based on cBioPortal.



Query **Download Data**

Select Cancer Study:

- Search... No studies selected.
- All (121)
 - Adrenal Gland (1)
 - Adrenocortical Carcinoma (1)
 - Adrenocortical Carcinoma (TCGA, Provisional) 92 samples
 - Biliary Tract (5)
 - Cholangiocarcinoma (4)
 - Intrahepatic Cholangiocarcinoma (Johns Hopkins University, Nat Genet 2013) 40 samples
 - Cholangiocarcinoma (National Cancer Centre of Singapore, Nat Genet 2013) 15 samples
 - Cholangiocarcinoma (National University of Singapore, Nat Genet 2012) 8 samples

Select Data Type Priority: Mutation and CNA Only Mutation Only CNA

Enter Gene Set: Advanced: Onco Query Language (OQL)

User-defined List

Enter HUGO Gene Symbols or Gene Aliases

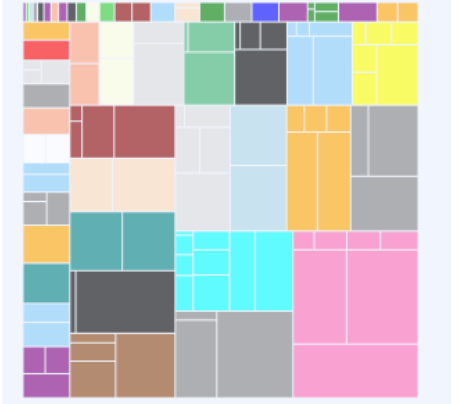
What's New

New Jobs available at Dana-Farber to work on cBioPortal
Sign up for low-volume email news alerts:

 Or follow us @cBioportal on Twitter

Data Sets

The Portal contains **121 cancer studies**. [Details]



Example Queries

- 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
- BRAF V600E mutations across cancer types
- Patient view of an endometrial cancer case

What People are Saying

"I want to thank you for the nice, useful and user-friendly interface you have generated and shared with the

In this query, we are telling cBioPortal to perform an analyse comparing all AML samples with ERG mutation or CNA and those without ERG mutation nor CNA.

Query [Download Data](#)

Select Cancer Study:

tcga

1 study selected. [Deselect all](#)

- Acute Lymphoid Leukemia (1)
 - Infant MLL-Rearranged Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015) 24 samples
- Acute Myeloid Leukemia (2)
 - Acute Myeloid Leukemia (TCGA, Provisional) 200 samples [Summary](#)
 - Acute Myeloid Leukemia (TCGA, NEJM 2013) 200 samples
- Multiple Myeloma (1)
 - Multiple Myeloma (Broad, Cancer Cell 2014) 205 samples
- Bone (2)
- Ewing Sarcoma (2)

Select cancer study
(AML, Provisional)

Select the type of
aberration you are
interested in
(Mutations & CNA)

Select Genomic Profiles:

- Mutations [?](#)
- Putative copy-number alterations from GISTIC [?](#)
- mRNA Expression data. Select one of the profiles below:
 - mRNA Expression z-Scores (RNA Seq V2 RSEM) [?](#)
 - mRNA Expression z-Scores (RNA Seq RPKM) [?](#)

Select the sample
set
(Tumour samples
with CAN data)

Select Patient/Case Set:

Tumor Samples with CNA data (191)

To build your own case set, try out our enhanced [Study View](#).

Enter Gene Set: [Advanced: Onco Query Language \(OQL\)](#)

User-defined List

Select Genes from Recurrent CNAs (Gistic)

ERG

Type in gene - can
accept any number.
(For this example,
we will look at ERG)

OncoPrint

OncoPrint Plots Mutations Co-Expression Enrichments Survival Network IGV Download Bookmark

Case Set: Tumor Samples with CNA data: All tumors with CNA data (191 samples)(191 patients / 191 samples)

Altered in 9 (5%) of 191 cases/patients

ERG

5%



Genetic Alteration

Amplification

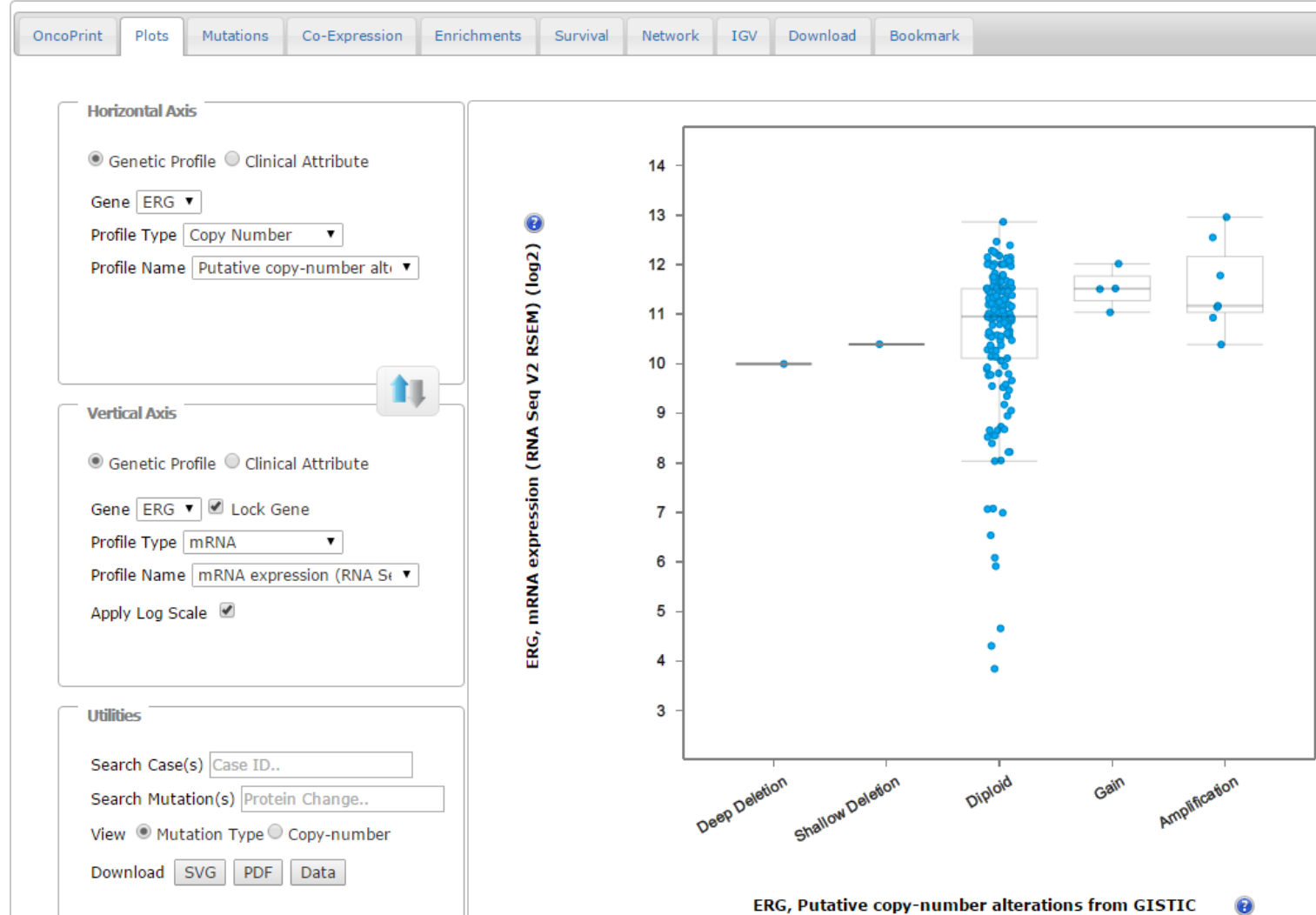
Deep Deletion

9 out of 191 samples have alteration in ERG:

- 8 samples have amplifications of ERG
- 1 sample has a deep deletion of ERG

Plots – correlation ERG expression with CNA

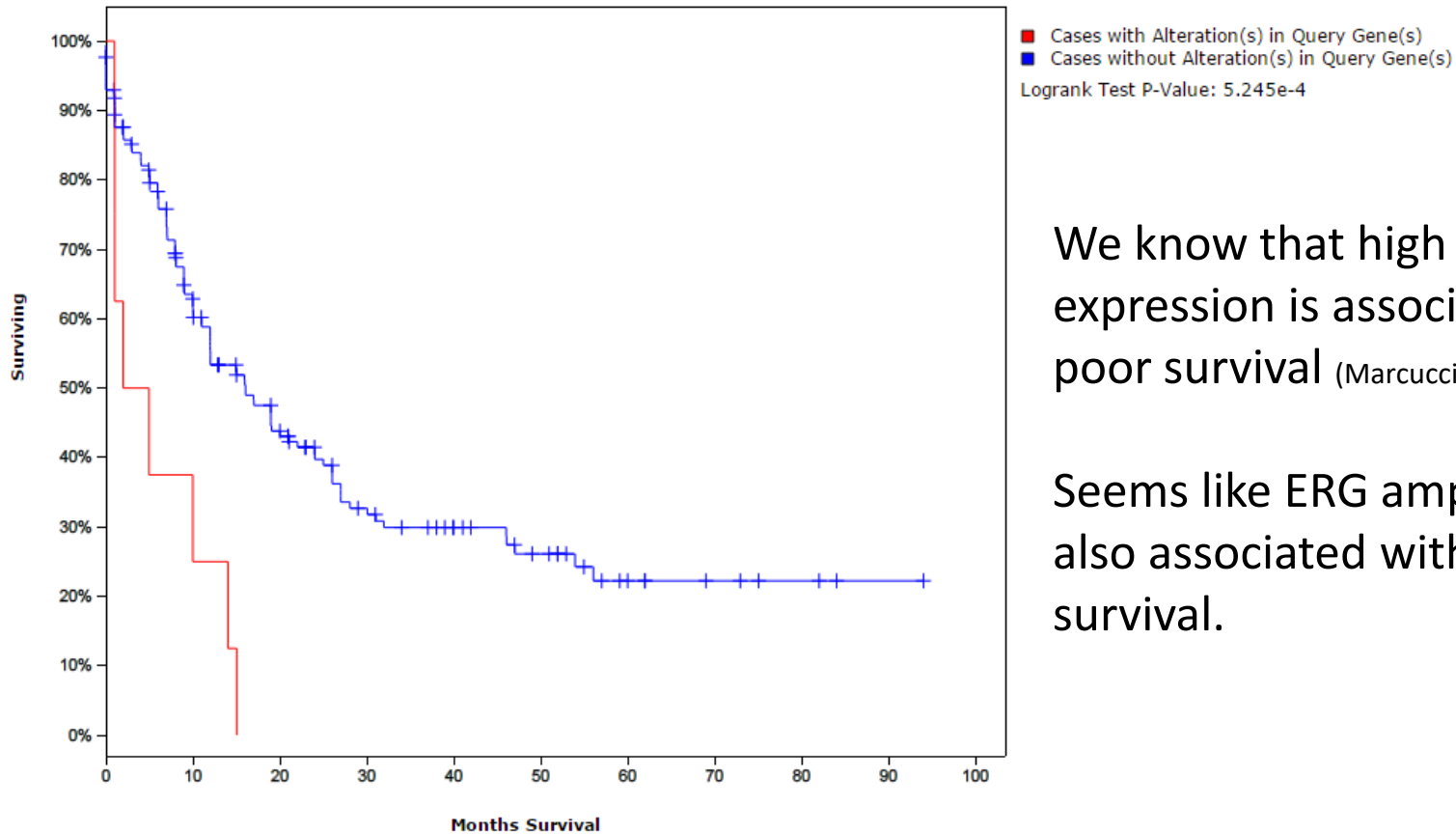
Samples with amplification possibly have higher expression



Survival analysis

OncoPrint Plots Mutations Co-Expression Enrichments **Survival** Network IGV Download Bookmark

Overall Survival Kaplan-Meier Estimate [SVG](#) [PDF](#) [Data](#)



We know that high ERG expression is associated with poor survival (Marcucci *et al* JCO 2005).

Seems like ERG amplification is also associated with poor survival.

	#total cases	#cases deceased	median months survival
Cases with Alteration(s) in Query Gene(s)	8	8	2
Cases without Alteration(s) in Query Gene(s)	171	108	16.06

Network analysis

Modify Query

Acute Myeloid Leukemia (TCGA, Provisional)
Tumor Samples with QIA data (191 samples) / 1 Gene

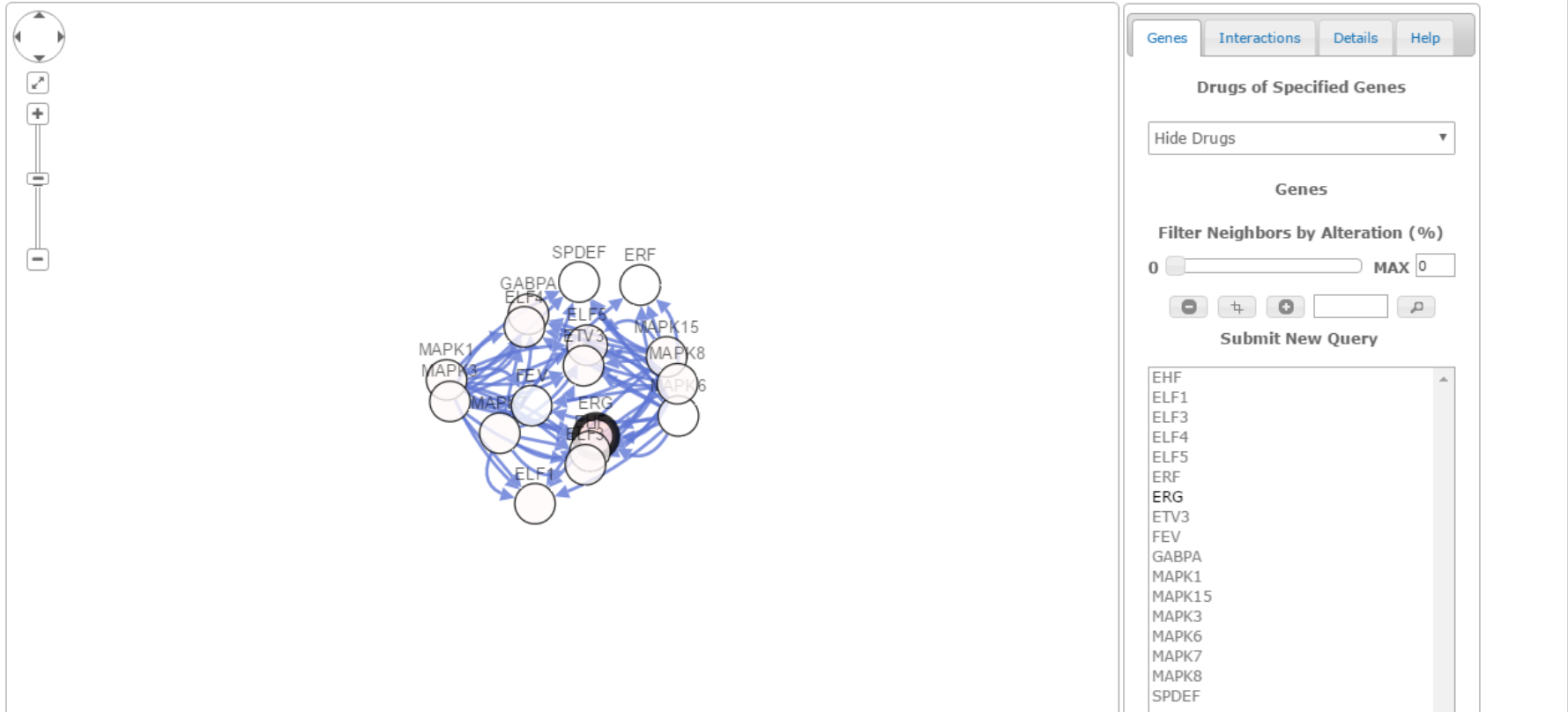
Gene Set / Pathway is altered in 9 (4.7%) of queried samples

OncoPrint Mutual Exclusivity Plots Mutations Co-Expression Enrichments Survival Network CN Segments Download Bookmark

The network below contains 17 nodes, including your 1 query gene and the 16 most frequently altered neighbor genes (out of a total of 17).
Download the complete network in [GraphML](#) or [SIF](#) for import into [Cytoscape](#) ([GraphMLReader plugin](#) is required for importing GraphML).

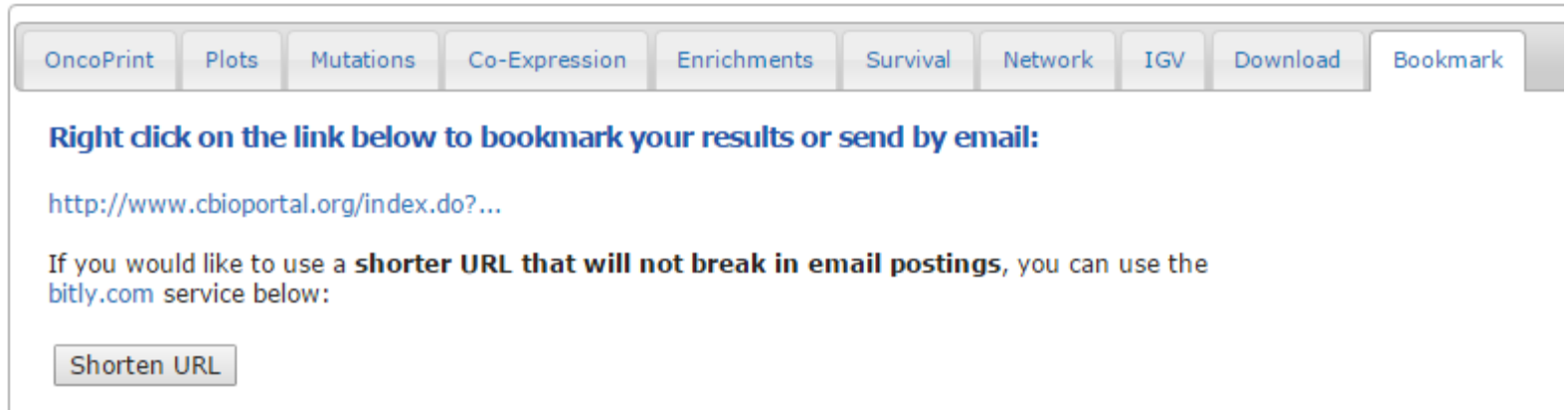
File Topology View Layout Legends

Double-click nodes/edges for details. Right-click edges for detailed process-level network view.



This network analysis is not that interesting, but it could be more useful with a larger input gene set.

Bookmark



OncoPrint Plots Mutations Co-Expression Enrichments Survival Network IGV Download Bookmark

Right click on the link below to bookmark your results or send by email:

<http://www.cbioportal.org/index.do?...>

If you would like to use a **shorter URL that will not break in email postings**, you can use the [bitly.com](#) service below:

Shorten URL

You can make a URL to immediately share analysis with collaborators

Gene summaries across cancer types

Query **Download Data**

Select Cancer Study:

Search... All studies selected. [Deselect all](#)

- All (121)
- Adrenal Gland (1)
 - Adrenocortical Carcinoma (1)
 - Adrenocortical Carcinoma (TCGA, Provisional) 92 samples
- Biliary Tract (5)
 - Cholangiocarcinoma (4)
 - Intrahepatic Cholangiocarcinoma (Johns Hopkins University, Nat Genet 2013) 40 samples
 - Cholangiocarcinoma (National Cancer Centre of Singapore, Nat Genet 2013) 15 samples
 - Cholangiocarcinoma (National University of Singapore, Nat Genet 2012) 8 samples

Select all cancers

Select the type of aberration you are interested in (Mutations & CNA)

Select Genomic Profiles:

Select Data Type Priority: Mutation and CNA Only Mutation Only CNA

Select Patient/Case Set:

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

Enter Gene Set: Advanced: Onco Query Language (OQL)

User-defined List

ERG

Type in gene - can accept any number. (For this example, we will look at ERG)

Cancer Sequencing Projects

International Cancer Genome Consortium (ICGC)



International Cancer Genome Consortium

- Collaboration between 22 countries
- Initiated in 2007
- **Aim:**
To catalogue genomic abnormalities in tumours from 50 different cancer types & subtypes
- **Update today:**
70 projects, 21 primary sites, >16,246 tumour DNA data
- Uses data from TCGA and the Sanger Cancer Genome Project

Working groups

INTERNATIONAL CANCER GENOME CONSORTIUM (ICGC) WORKING GROUPS						
Clinical and Pathology Issues	Quality Standards of Samples	Genome Analyses	Informed Consent and Privacy Protections	Sample Size/Study Design	Data Management/ Databases and Coordination	Data Release, Data Tiers, Intellectual Property, and Publications
Lynda Chin Jean-Yves Blay William Dalton Tony Green Stan Hamilton Timothy Ley Ed Liu Paul Mischel Kenneth Pienta Rajiv Sarin Daniel Tan	Peter Lichter Carolyn Compton Andy Futreal Youyong Lu Miguel Angel Piris	Mike Stratton Olli Kallioniemi Ed Liu Marco Marra John McPherson Brad Ozenberger Henk Stunnenberg Daniel Tan Brandon Wainwright Rick Wilson	Bartha Knoppers Martin Bobrow Wylie Burke Kazuto Kato Karen Kennedy Brad Ozenberger Daniel Tan Susan Wallace Henry Yang	Eric Lander Ron DePinho Doug Easton Gaddy Getz Partha P. Majumder	Lincoln Stein Cameron Brennan Arul Chinnaiyan Peter Good Joe Gray J Gowrishankar David Haussler David Housman Tim Hubbard Subha Madhavan Paul Spellman	Mark Guyer Daniela Gerhard Karen Kennedy Brad Ozenberger



ICGC Samples

Biliary Tract Cancer Japan 🇯🇵	Biliary Tract Cancer Singapore 🇸🇬	Bladder Cancer China 🇨🇳
Bladder Cancer United States 🇺🇸	Blood Cancer China 🇨🇳	Blood Cancer Singapore 🇸🇬
Blood Cancer South Korea 🇰🇷	Blood Cancer United States 🇺🇸	Blood Cancer United States 🇺🇸
Bone Cancer France 🇫🇷	Bone Cancer United Kingdom 🇬🇧	Brain Cancer Canada 🇨🇦
Brain Cancer China 🇨🇳	Brain Cancer United States 🇺🇸	Brain Cancer United States 🇺🇸
Breast Cancer China 🇨🇳	Breast Cancer European Union / United Kingdom 🇪🇺 🇬🇧	Breast Cancer France 🇫🇷
Breast Cancer Mexico 🇲🇽	Breast Cancer South Korea 🇰🇷	Breast Cancer South Korea 🇰🇷
Breast Cancer United Kingdom 🇬🇧	Breast Cancer United States 🇺🇸	Cervical Cancer United States 🇺🇸
Chronic Lymphocytic Leukemia Spain 🇪🇸	Chronic Myeloid Disorders United Kingdom 🇬🇧	Colon Cancer United States 🇺🇸
Colorectal Cancer China 🇨🇳	Endometrial Cancer United States 🇺🇸	Esophageal Cancer China 🇨🇳

Esophageal Cancer United Kingdom 🇬🇧	Eye Cancer France 🇫🇷	Gastric Cancer China 🇨🇳
Gastric Cancer Japan 🇯🇵	Gastric Cancer United States 🇺🇸	Head And Neck Cancer United States 🇺🇸
Head and Neck Cancer Mexico 🇲🇽	Head and Neck Cancer United States 🇺🇸	Liver Cancer China 🇨🇳
Liver Cancer France 🇫🇷	Liver Cancer Japan 🇯🇵	Liver Cancer United States 🇺🇸
Lung Cancer China 🇨🇳	Lung Cancer South Korea 🇰🇷	Lung Cancer United States 🇺🇸
Lung Cancer United States 🇺🇸	Malignant Lymphoma Germany 🇩🇪	Melanoma Brazil 🇧🇷
Nasopharyngeal Cancer China 🇨🇳	Non-Hodgkin Lymphoma Mexico 🇲🇽	Oral Cancer India 🇮🇳
Ovarian Cancer Australia 🇦🇺	Ovarian Cancer China 🇨🇳	Ovarian Cancer United States 🇺🇸
Pancreatic Cancer Australia 🇦🇺	Pancreatic Cancer Australia 🇦🇺	Pancreatic Cancer Canada 🇨🇦
Pancreatic Cancer China 🇨🇳	Pancreatic Cancer United States 🇺🇸	Pediatric Brain Tumors Germany 🇩🇪

Prostate Cancer France 🇫🇷	Prostate Cancer Germany 🇩🇪	Prostate Cancer United Kingdom 🇬🇧
Prostate Cancer United States 🇺🇸	Rare Pancreatic Tumors Italy 🇮🇹	Rectal Cancer United States 🇺🇸
Renal Cancer China 🇨🇳	Renal Cancer European Union / France 🇪🇺 🇫🇷	Renal Cancer United States 🇺🇸
Renal Cancer United States 🇺🇸	Skin Cancer United States 🇺🇸	Soft tissue cancer France 🇫🇷
Thyroid Cancer China 🇨🇳	Thyroid Cancer Saudi Arabia 🇸🇦	Uterine Cancer France 🇫🇷

Data Release 23

Dec 7, 2016

Donor Distribution by Primary Site



Cancer projects	70
Cancer primary sites	21
Donors with molecular data in DCC	16,246
Total Donors	19,305
Simple somatic mutations	46,693,172
Mutated Genes	57,658

Data types

- Mandatory:
 - Genomic DNA analyses of tumors (and matching control DNA) are core elements of the project.
- Complementary (Recommended):
 - Additional studies of DNA methylation and RNA expression are recommended **on the same samples that are used to find somatic mutations.**
- Optional:
 - Proteomic analyses
 - Metabolomic analyses
 - Immunohistochemical analyses



Data access policy

ICGC Open Access Datasets	ICGC Controlled Access Datasets
<ul style="list-style-type: none">➤ Cancer Pathology<ul style="list-style-type: none">Histologic type or subtypeHistologic nuclear grade➤ Patient/Person<ul style="list-style-type: none">GenderAge range➤ Gene Expression (normalized)➤ DNA methylation➤ Genotype frequencies➤ Computed Copy Number and Loss of Heterozygosity➤ Newly discovered somatic variants	<ul style="list-style-type: none">➤ Detailed Phenotype and Outcome Data<ul style="list-style-type: none">Patient demographyRisk factorsExaminationSurgery/Drugs/RadiationSample/SlideSpecific histological featuresProtocolAnalyte/Aliquot➤ Gene Expression (probe-level data)➤ Raw genotype calls➤ Gene-sample identifier links➤ Genome sequence files

ICGC data portal (<http://dcc.icgc.org/>)

Click on cancer projects



ICGC Data Portal

☰ Cancer Projects

🔍 Advanced Search

🧪 Data Analysis

☰ DCC Data Releases

☁ Data Repositories

🔍 e.g. BRAF, KRAS G12D, DO35100, MU7870, FI998, apoptosis, Cancer Gene Census, GO:0016049

About Us

The [ICGC Data Portal](#) provides tools for visualizing, querying and downloading the data released quarterly by the consortium's member projects.

To access ICGC controlled tier data, please read these [instructions](#).

New features will be regularly added by the DCC development team. [Feedback](#) is welcome.

Data Release 20 November 27th, 2015

Donor Distribution by Primary Site



Tutorial

EXAMPLE QUERIES

1. BRAF missense mutations in colorectal cancer
2. Most frequently mutated genes by high impact mutation in stage III malignant lymphoma
3. Brain cancer donors with frameshift mutations and having methylation data available



Cancer project view

▼ Project

e.g. PACA-CA, Brain, TCGA

▼ Primary Site

- Blood 8
- Kidney 7
- Liver 6
- Pancreas 5
- Head and neck 4
- 16 more

▼ Country

- United States 29
- China 6
- United Kingdom 6
- France 5
- Australia 4
- 11 more

▼ Available Data Type

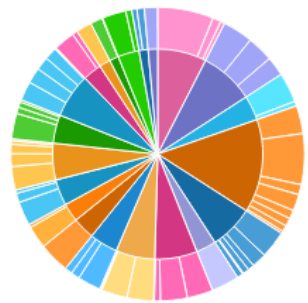
- SSM 55
- CNSM 41
- EXP-S 28
- METH-A 26
- miRNA-S 22
- 6 more

▼ Tumour Type

- Renal cancer 7
- Liver cancer 6
- Blood cancer 5
- Lung cancer 4
- Pancreatic cancer 4

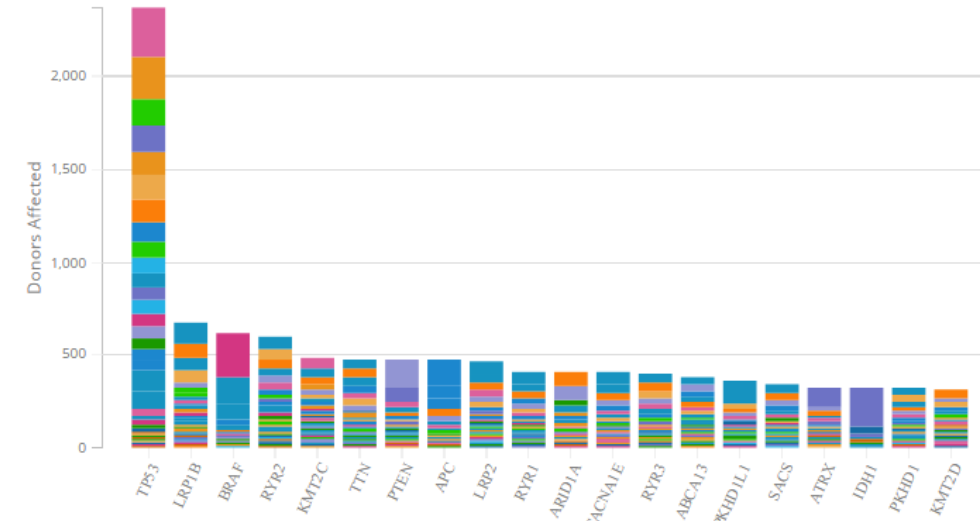
Summary Details History

Donor Distribution
17,867 Donors across 66 Projects

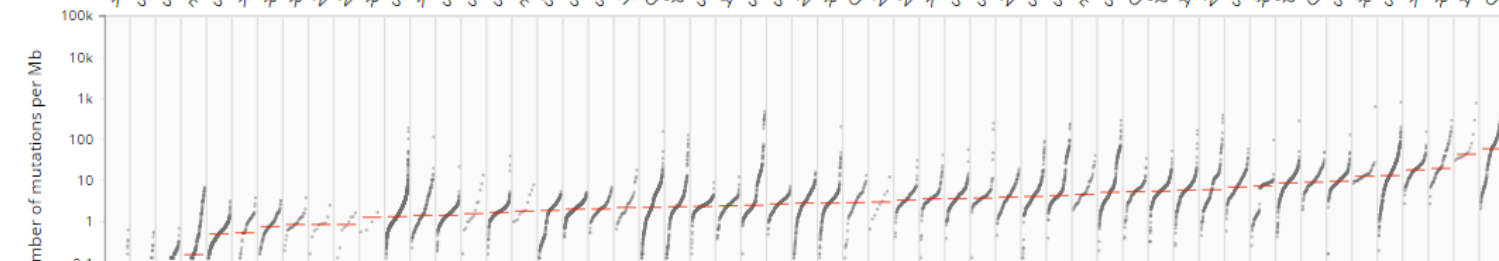


click on BRCA-US

Top 20 Mutated Genes with High Functional Impact SSMs
9,155 Unique SSM-Tested Donors



Number of Somatic Mutations in Donor's Exomes Across Cancer Projects





Summary

Mutated Genes

Mutations

Donors

Publications

Page Filters

▼ Mutation Impact

High

Low

Unknown

Summary

Code	BRCA-US
Name	Breast Cancer - TCGA, US
Primary Site	Breast
Tumour Type	Breast cancer
Tumour Subtype	Ductal & lobular
Countries	United States
Number of donors in PCAWG	92
Number of donors with molecular data in DCC	1,045
Total number of donors	1,099

Experimental Analyses

non-NGS	2107 samples from 1045 donors
WGS	19 samples from 19 donors
WXS	970 samples from 955 donors
RNA-Seq	1177 samples from 1041 donors
miRNA-Seq	1153 samples from 1026 donors

[Download Sample Sheet](#)

Raw data is available at

- [The Cancer Genome Atlas](#)
- [Cancer Genomics Hub](#)

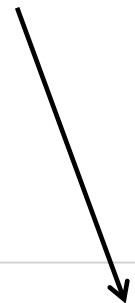
An approved [data access request](#) is required.

Available Data Types

Clinical Data	1,099 donors
Simple Somatic Mutations (SSM)	955 donors
Copy Number Somatic Mutations (CNSM)	1,045 donors
Structural Somatic Mutations (StSM)	--
Simple Germline Variants (SGV)	--
Array-based DNA Methylation (METH-A)	1,013 donors
Sequence-based DNA Methylation (METH-S)	--
Array-based Gene Expression (EXP-A)	529 donors
Sequence-based Gene Expression (EXP-S)	1,041 donors
Protein Expression (PEXP)	298 donors
Sequence-based miRNA Expression (miRNA)	1,026 donors
Exon junction (JCN)	--

OPEN IN [DCC Data Releases](#)

Click on Genome Viewer



Most Frequently Mutated Genes

[OPEN IN ADVANCED SEARCH](#) | [GENOME VIEWER](#)





GENOME VIEWER

View top mutated genes

Genes Mutations

Filter genes

Query: Project IS BRCA-US

Homo sapiens GRCh37

Karyotype Chromosome Region Overview

Region overview Window size: 809415 nts

178507184 178911891 179316598

phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha > [protein_coding]

chr3:178865902-178957881

TP53

tumor protein p53

chr17:7565097-7590856

TTN

titin

chr2:179390716-179695529

TTN-AS1

TTN antisense RNA 1

chr2:179385910-179639402

RP11-245C23.3

RP11-245C23.3

chr3:178954479-178955163

PCDHGA1

protocadherin gamma subfamily 1

chr5:140710252-140892546

PCDHGA2

protocadherin gamma subfamily 2

chr5:140718539-140892546

PCDHGA3

protocadherin gamma subfamily 3

chr5:140723601-140892546

Detailed information Window size: 101177 nts

178861303 178911891 178962479

Sequence

ICGC Genes

phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha > [protein_coding]

PIK3CA-001 > [protein_coding]

PIK3CA-003 > [protein_coding]

PIK3CA-002 > [protein_coding]

Small nucleolar RNA SNORA25 [snoRNA]

ICGC Mutations

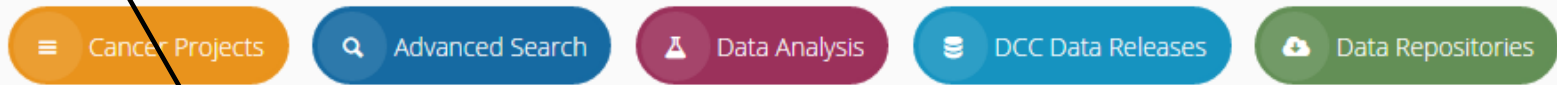
Powered by Genome Maps

178911891



Looking at mutations in specific genes

Type in BRAF here



Search bar containing the text: "e.g. BRAF, KRAS G12D, DO35100, MU7870, FI998, apoptosis, Cancer Gene Census, GO:0016049". An arrow points from the text "Type in BRAF here" to the search bar.

About Us

The [ICGC Data Portal](#) provides tools for visualizing, querying and downloading the data released quarterly by the consortium's member projects.

To access ICGC controlled tier data, please read these [instructions](#).

New features will be regularly added by the DCC development team. [Feedback](#) is welcome.

Data Release 20 November 27th, 2015

Donor Distribution by Primary Site



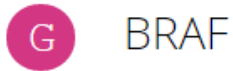
Tutorial

EXAMPLE QUERIES

1. BRAF missense mutations in colorectal cancer
2. Most frequently mutated genes by high impact mutations in stage III malignant lymphoma
3. Brain cancer donors with frameshift mutations and having methylation data available



Gene centric view



General information

- Summary
- Cancer Distribution
- Protein
- Genomic Context
- Mutations
- Compounds

Page Filters

▼ Mutation Impact

- High
- Low
- Unknown

Summary

Symbol	BRAF
Name	v-raf murine sarcoma viral oncogene homolog B
Synonyms	BRAF1
Type	Protein coding
Location	chr7:140419127-140624564 (GRCh37)
Strand	-
Description	This gene encodes a protein belonging to the raf/mil family of serine/threonine protein kinases. This protein plays a role in regulating the MAP kinase/ERKs signaling pathway, which affects cell division, differentiation, and secretion. Mutations in ... more

Annotation

Reactome Pathways	ARMS-mediated activation CREB phosphorylation through the activation of Ras Frs2-mediated activation MAP2K and MAPK activation Negative feedback regulation of MAPK pathway 4 more
GO Terms	ATP binding CD4-positive, alpha-beta T cell differentiation Fc-epsilon receptor signaling pathway MAP kinase kinase kinase activity MAPK cascade 51 more
Curated Gene Set	Cancer Gene Census

External References

HGNC Gene	1097
Ensembl (release 75)	ENSG00000157764
COSMIC	BRAF
Entrez Gene	673
OMIM	164757
UniProtKB/Swiss-Prot	P15056

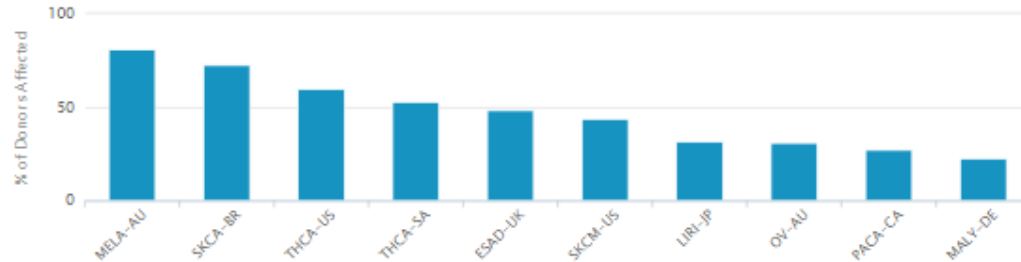
Then scroll down...

Gene centric view

Cancer Distribution

[OPEN IN ADVANCED SEARCH](#)

1,144 DONORS AFFECTED BY 1,244 MUTATIONS ACROSS 43 PROJECTS



Types of cancers with the mutation

Showing 43 projects



Project	Site	Tumour Type	Tumour Subtype	# Donors affected	# Mutations
MELA-AU	Skin	Skin cancer	Melanoma	149 / 183 (81.42%)	501
SKCA-BR	Skin	Melanoma		48 / 66 (72.73%)	81
THCA-US	Head and neck	Head and Neck cancer	Thyroid carcinoma	238 / 400 (59.50%)	5
THCA-SA	Head and neck	Thyroid cancer	Papillary thyroid carcinoma	8 / 15 (53.33%)	2
ESAD-UK	Esophagus	Esophageal cancer	Esophageal adenocarcinoma	58 / 119 (48.74%)	93
SKCM-US	Skin	Skin cancer	Cutaneous melanoma	148 / 335 (44.18%)	22
LIRI-JP	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)	83 / 260 (31.92%)	131
OV-AU	Ovary	Ovarian cancer	Serous cystadenocarcinoma	29 / 93 (31.18%)	46
PACA-CA	Pancreas	Pancreatic cancer	Ductal adenocarcinoma	55 / 204 (26.96%)	75
MALY-DE	Blood	Malignant Lymphoma	Germinal center B-cell derived lymphomas	10 / 44 (22.73%)	19
PRAD-UK	Prostate	Prostate cancer	Adenocarcinoma	21 / 108 (19.44%)	21
RECA-EU	Kidney	Renal cancer	Renal cell carcinoma (Focus on but not limited to clear cell subtype)	18 / 95 (18.95%)	23
COAD-US	Colorectal	Colon cancer	Adenocarcinoma	33 / 216 (15.28%)	11
LINC-JP	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)	30 / 244 (12.30%)	34

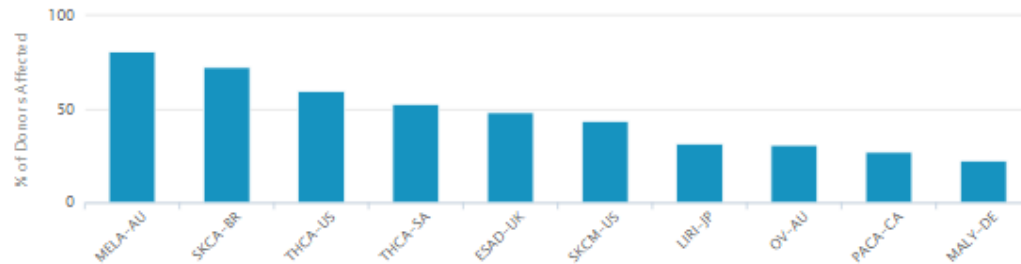


Gene centric view

Cancer Distribution

[OPEN IN ADVANCED SEARCH](#)

1,144 DONORS AFFECTED BY 1,244 MUTATIONS ACROSS 43 PROJECTS



More detailed information



Showing 43 projects



Project	Site	Tumour Type	Tumour Subtype	# Donors affected ▾	# Mutations
MELA-AU	Skin	Skin cancer	Melanoma	149 / 183 (81.42%)	501
SKCA-BR	Skin	Melanoma		48 / 66 (72.73%)	81
THCA-US	Head and neck	Head and Neck cancer	Thyroid carcinoma	238 / 400 (59.50%)	5
THCA-SA	Head and neck	Thyroid cancer	Papillary thyroid carcinoma	8 / 15 (53.33%)	2
ESAD-UK	Esophagus	Esophageal cancer	Esophageal adenocarcinoma	58 / 119 (48.74%)	93
SKCM-US	Skin	Skin cancer	Cutaneous melanoma	148 / 335 (44.18%)	22
LIRI-JP	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)	83 / 260 (31.92%)	131
OV-AU	Ovary	Ovarian cancer	Serous cystadenocarcinoma	29 / 93 (31.18%)	46
PACA-CA	Pancreas	Pancreatic cancer	Ductal adenocarcinoma	55 / 204 (26.96%)	75
MALY-DE	Blood	Malignant Lymphoma	Germinal center B-cell derived lymphomas	10 / 44 (22.73%)	19
PRAD-UK	Prostate	Prostate cancer	Adenocarcinoma	21 / 108 (19.44%)	21
RECA-EU	Kidney	Renal cancer	Renal cell carcinoma (Focus on but not limited to clear cell subtype)	18 / 95 (18.95%)	23
COAD-US	Colorectal	Colon cancer	Adenocarcinoma	33 / 216 (15.28%)	11
LINC-JP	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)	30 / 244 (12.30%)	34

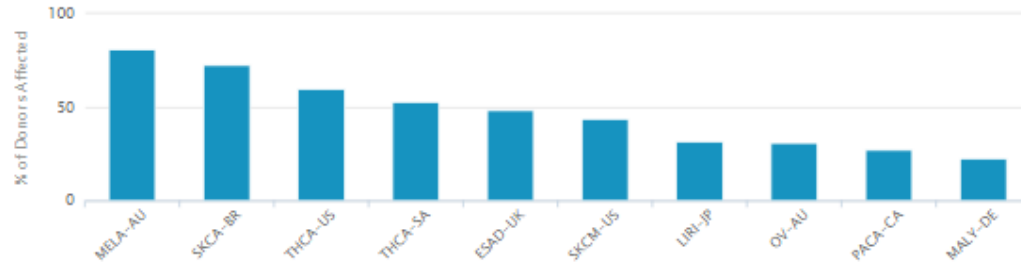


Gene centric view

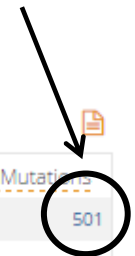
Cancer Distribution

1,144 DONORS AFFECTED BY 1,244 MUTATIONS ACROSS 43 PROJECTS

[OPEN IN ADVANCED SEARCH](#)



Click for more detail on the mutations



Showing 43 projects

Project	Site	Tumour Type	Tumour Subtype	# Donors affected	# Mutations
MELA-AU	Skin	Skin cancer	Melanoma	149 / 183 (81.42%)	501
SKCA-BR	Skin	Melanoma		48 / 66 (72.73%)	81
THCA-US	Head and neck	Head and Neck cancer	Thyroid carcinoma	238 / 400 (59.50%)	5
THCA-SA	Head and neck	Thyroid cancer	Papillary thyroid carcinoma	8 / 15 (53.33%)	2
ESAD-UK	Esophagus	Esophageal cancer	Esophageal adenocarcinoma	58 / 119 (48.74%)	93
SKCM-US	Skin	Skin cancer	Cutaneous melanoma	148 / 335 (44.18%)	22
LIRI-JP	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)	83 / 260 (31.92%)	131
OV-AU	Ovary	Ovarian cancer	Serous cystadenocarcinoma	29 / 93 (31.18%)	46
PACA-CA	Pancreas	Pancreatic cancer	Ductal adenocarcinoma	55 / 204 (26.96%)	75
MALY-DE	Blood	Malignant Lymphoma	Germinal center B-cell derived lymphomas	10 / 44 (22.73%)	19
PRAD-UK	Prostate	Prostate cancer	Adenocarcinoma	21 / 108 (19.44%)	21
RECA-EU	Kidney	Renal cancer	Renal cell carcinoma (Focus on but not limited to clear cell subtype)	18 / 95 (18.95%)	23
COAD-US	Colorectal	Colon cancer	Adenocarcinoma	33 / 216 (15.28%)	11
LINC-JP	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)	30 / 244 (12.30%)	34



Hover over the section of the graph to see what region it represents

▼ Mutation
e.g. MU123
Upload Mutation Set

▼ Consequence Type
Missense 20
Splice Region 1
5 UTR 1
Upstream 39
Synonymous 2
4 more

▼ Functional Impact
High 16
Low 4
Unknown 501

▼ Type
Substitution 469
MSub 25
Deletion 7

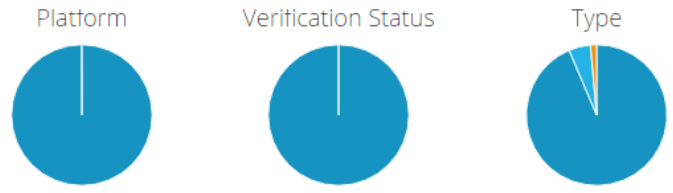
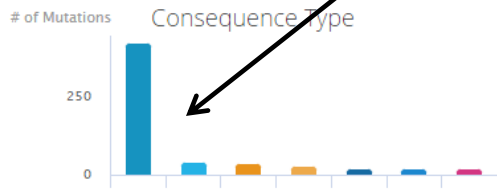
▶ Platform

▶ Analysis Type

▶ Verification Status

▼ Location
e.g. chr12:43566-3457633
Search

Donors 149
Genes 1
Mutations 501



Mutations 501
Mutation Occurrences 572

Each mutation has a unique ID (click for more info)

Mutations

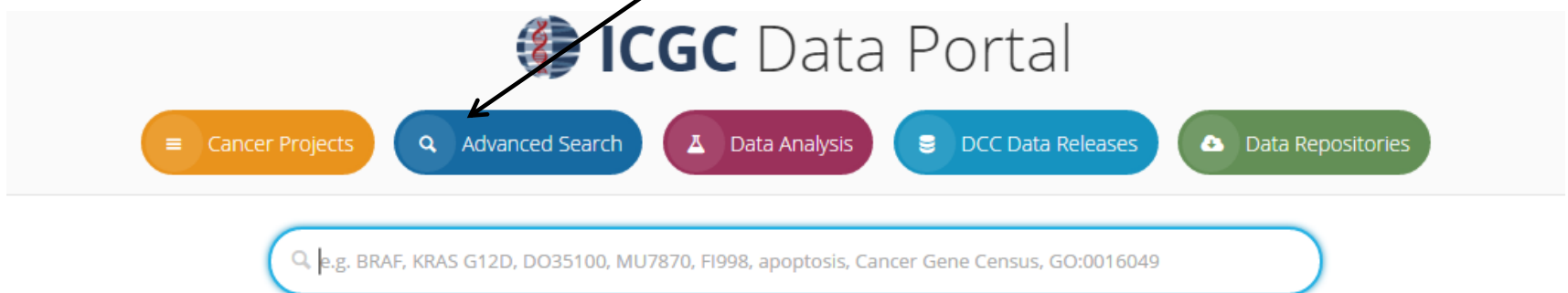
Showing 1 - 10 of 501 mutations

ID	DNA change	Type	Consequences	# Donors affected		Chart
				Total	Across all Projects	
<input type="checkbox"/> MU62030	chr7:g.140453136A>T	single base substitution	Missense: BRAF V207E, V600E 3 UTR: BRAF Exon: BRAF	47 / 546 (8.61%)	546 / 10,648 (5.13%)	
<input type="checkbox"/> MU40909253	chr7:g.140453136AC>TT	multiple base substitution (>=2bp and <=200bp)	Missense: BRAF V600K, V207K 3 UTR: BRAF Exon: BRAF	17 / 17 (100.00%)	17 / 10,648 (0.16%)	
<input type="checkbox"/> MU48018461	chr7:g.1404565962G>A	single base substitution	Intron: BRAF	2 / 2 (100.00%)	2 / 10,648 (0.02%)	
<input type="checkbox"/> MU42617524	chr7:g.140472618G>A	single base substitution	Intron: BRAF	2 / 2 (100.00%)	2 / 10,648 (0.02%)	
<input type="checkbox"/> MU4440100	chr7:g.140481411C>T	single base substitution	Missense: BRAF G466E, G73E 3 UTR: BRAF	2 / 6 (33.33%)	6 / 10,648 (0.06%)	

Advanced Search

Find out which cancers commonly have BRAF missense mutations

Go to the home page and click
“advanced search”



Advanced Search

Find out which cancers commonly have BRAF missense mutations

Search for "BRAF"

The screenshot displays an advanced search interface for BRAF mutations. On the left, there are filters for Gene (with a search box containing "e.g. KRAS, ENSG0000013703" and an "Upload Gene Set" button) and Type (with checkboxes for Protein coding (20,327), Pseudogene (13,920), lincRNA (7,109), Antisense (5,273), and miRNA (3,049), plus a "24 more" link). Below these are filters for Pathway (with a search box containing "e.g. R-HSA-983168, TNF Signaling").

At the top, summary statistics are shown: Donors (17,867), Genes (57,773), and Mutations (36,985,985). Below this are five pie charts representing distributions: Project, Primary Site, Gender, Tumour Stage, and Vital Status. A "Show More" button is located below the charts.

Below the charts, there are options to "SAVE DONOR SET", "DOWNLOAD DONOR DATA", and "VIEW IN DATA REPOSITORIES".

The main section is titled "Donors" and shows "Showing 1 - 17,867 of 17,867 donors". Below this is a table with the following columns: ID, Project, Site, Gender, Age, Stage, Survival (days), and Available Data Types (SSM, CNSM, SISM, SGV, METH-A, METH-S, EXP-A, EXP-S, PDX, miRNA-S, JCN). The table also includes "# Mutations" and "# Genes" columns.

Advanced Search

Find out which cancers commonly have BRAF missense mutations

Go to the mutations tab

The screenshot shows a search interface for BRAF missense mutations. The search criteria are: Gene IS BRAF AND Consequence Type IS Missense. The results show 619 donors, 1 gene, and 93 mutations. The interface includes several filters and charts:

- Mutation:** Search box with "e.g. MU123".
- Consequence Type:** Missense (93), Frameshift (5), Stop Gained (10), Splice Acceptor (2), Splice Donor (3), Splice Region (10), and 11 more.
- Functional Impact:** High (66), Low (27), Unknown (62).
- Type:** Substitution (87), MSub (6).
- Platform:** (Expanded).
- Charts:** Five pie charts for Project, Primary Site, Gender, Tumour Stage, and Vital Status. An arrow points to the Project chart with the text "Select 'missense'".
- Donors:** A table showing 1 - 619 of 619 donors. An arrow points to the table with the text "Hover mouse to see details.".

ID	Project	Site	Gender	Age	Stage	Survival (days)	Available Data Types										# Mutations	# Genes
							SSM	CNSM	SISM	SGV	METH-A	METH-S	EXP-A	EXP-S	PDP	mRNA-S		
DO218900	SKCA-BR	Skin	Male	66	4	512	✓	✓	✓	--	--	--	--	--	--	--	1	1
DO218859	SKCA-BR	Skin	Male	29	2	2,262	✓	✓	✓	--	--	--	--	--	--	--	1	1
DO38907	THCA-US	Head and neck	Female	41			✓	✓	--	✓	--	--	✓	✓	✓	--	1	1

Most common cancers with BRAF missense mutations are thyroid cancer and melanoma.

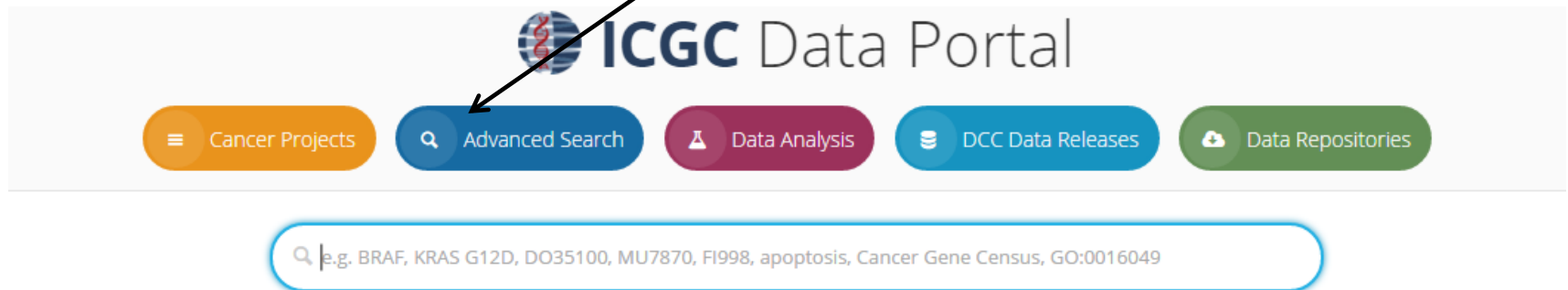
Limitations of data portal

- The data portal is mutation centric
 - i.e. All queries are related to retrieving tumours/samples with particular mutations in a particular gene
- If you just want expression/methylation data for a particular gene you still have to download the data



Downloading data from ICGC

Go to the home page and click
“advanced search”



▶

Downloading data from ICGC

ADVANCED SEARCH

Donors: 17,867 | Genes: 57,773 | Mutations: 36,985,985

Click download donor data

Project | Primary Site | Gender | Tumour Stage | Vital Status

Donors

Showing 1 - 17,867 of 17,867 donors

ID	Project	Site	Gender	Age	Stage	Survival (days)	Available Data Types										# Mutations	# Genes	
							SGM	CNSM	SSM	SGV	METHA	METHS	DP-A	DP-S	FPD	mRNA-S			JCN
DO222843	MELA-AU	Skin	Male	76	IIC	907	✓	--	✓	--	--	--	--	--	--	--	--	964,360	51,565
DO222837	MELA-AU	Skin	Male	82	IIB	1,110	✓	--	✓	--	--	--	--	--	--	--	--	786,166	49,858
DO222363	MELA-AU	Skin	Male	81	IIC	154	✓	--	✓	--	--	--	--	--	--	--	--	775,848	47,839
DO222875	MELA-AU	Skin	Male	79	IIA	1,192	✓	--	✓	--	--	--	--	--	--	--	--	819,954	45,993
DO222702	MELA-AU	Skin	Male	80	IIC	900	✓	--	✓	--	--	--	--	--	--	--	--	696,598	45,379
DO220886	MELA-AU	Skin	Male	56	IA/IB	7,730	✓	--	✓	--	--	--	--	--	--	--	--	419,022	44,032
DO220906	MELA-AU	Skin	Female	70	IB	842	✓	--	✓	--	--	--	--	--	--	--	--	471,943	43,340
			Male	47	III	1,527	✓	--	✓	--	--	--	--	--	--	--	--	548,756	42,820
			Male	78	IA/IB	1,324	✓	--	✓	--	--	--	--	--	--	--	--	453,462	42,244
			Male	83	IB	1,160	✓	--	✓	--	--	--	--	--	--	--	--	385,914	41,431

Showing 10 rows

SELECT DONOR SET | **DOWNLOAD DONOR DATA** | VIEW IN DATA REPOSITORIES

Donor filters: Blood (2,571), Brain (1,491), Breast (1,346), Liver (1,222), Kidney (1,212), 16 more

Project filters: ALL-US (1,002), AML-US (322), BLCA-CN (103), BLCA-US (412), BOCA-FR (100), BOCA-UK (145), BRCA-EU (81), BRCA-UK (166), BRCA-US (1,099), BTCA-JP (239), CLLE-ES (549), CMDI-UK (161), COAD-US (459)

→

↗

Select cancer type of interest

Note: go back to Advance search on home page

Downloading data from ICGC

Download Donor Data

To browse and download unfiltered data, please visit the [DCC Data Releases](#) section.

Search criteria results in donors. Select data types to download

Data Type	File Size (Estimated)	Select
Clinical Data	1.78 MB	<input type="checkbox"/>
Simple Somatic Mutation	256.18 MB	<input type="checkbox"/>
Copy Number Somatic Mutation	275.04 MB	<input type="checkbox"/>
Structural Somatic Mutations	--	--
Array-based Gene Expression	1.76 GB	<input type="checkbox"/>
Sequencing-based Gene Expression	5.96 GB	<input type="checkbox"/>
Protein Expression	16.67 MB	<input type="checkbox"/>
Sequence-based miRNA Expression	343.85 MB	<input type="checkbox"/>
Exon Junctions	--	--
Array-based DNA Methylation	50.95 GB	<input type="checkbox"/>
Sequencing-based DNA Methylation	--	--

Total file size 59.54 GB **0 files selected (0 B)**

Select the data types of interest

Click "Submit"

Downloading data from ICGC

Or download from the data repository



ICGC Data Portal

Cancer Projects Advanced Search Data Analysis DCC Data Releases Data Repositories

e.g. BRAF, KRAS G12D, DO35100, MU7870, FI998, apoptosis, Cancer Gene Census, GO:0016049

↳



DATA REPOSITORIES

Click through filters to choose what data you want

The ICGC DCC is periodically aggregating and indexing ICGC data hosted in multiple archives and cloud repositories worldwide. [More information here.](#)

Files Donors

File **Download manifests** 121,167 Files 11,371 Donors 1.26 PB

Enter File ID, Name or Object ID

Showing 1 - 25 of 121,167 files

Then download the data you selected

SAVE DONOR SET

	File ID	Donor	Repository	Project	Study	Data Type	Strategy	Format	Size	
<input type="checkbox"/>	FI99998	DO15366	CGHub	HNSC-US		Aligned Reads	RNA-Seq	BAM	6.89 GB	
<input type="checkbox"/>	FI99996	DO14274	CGHub	HNSC-US		Aligned Reads	RNA-Seq	BAM	3.52 GB	
<input type="checkbox"/>	FI99994	DO14462	CGHub	HNSC-US		Aligned Reads	miRNA-Seq	BAM	97.27 MB	
<input type="checkbox"/>	FI99992	DO14169	CGHub	HNSC-US		Aligned Reads	miRNA-Seq	BAM	221.17 MB	
<input type="checkbox"/>	FI99990	DO14158	CGHub	HNSC-US		Aligned Reads	miRNA-Seq	BAM	435.70 MB	
<input type="checkbox"/>	FI99988	DO16061	CGHub	HNSC-US		Aligned Reads	miRNA-Seq	BAM	192.24 MB	
<input type="checkbox"/>	FI99986	DO14686	CGHub	HNSC-US		Aligned Reads	WGS	BAM	183.91 GB	
<input type="checkbox"/>	FI99984	DO14408	CGHub	HNSC-US		Aligned Reads	miRNA-Seq	BAM	187.34 MB	
<input type="checkbox"/>	FI99982	DO14286	CGHub	HNSC-US		Aligned Reads	RNA-Seq	BAM	7.27 GB	
<input type="checkbox"/>	FI99980	DO14246	CGHub	HNSC-US		Aligned Reads	miRNA-Seq	BAM	385.00 MB	
<input type="checkbox"/>	FI9998	DO217962	PCAWG - Heidelberg	BRCA-EU	PCAWG	CNSM	WGS	VCF	1.83 KB	
<input type="checkbox"/>	FI99978	DO14886	CGHub	HNSC-US		Aligned Reads	RNA-Seq	BAM	5.41 GB	
<input type="checkbox"/>	FI99976	DO14214	CGHub	HNSC-US		Aligned Reads	RNA-Seq	BAM	6.69 GB	
<input type="checkbox"/>	FI99974	DO14218	CGHub	HNSC-US		Aligned Reads	miRNA-Seq	BAM	192.05 MB	

The advantage of ICGC is that data for all samples is in a single file so it is easier to work with in Excel (if file is small) or Galaxy (if file is big).



COSMIC database

<http://cancer.sanger.ac.uk/cosmic>



← Select "Cancer Gene Census"

Home ▾ Resources ▾ **Curation ▾** Tools ▾ Data ▾ News ▾ Help ▾ About ▾ Search COSMIC... Login ▾

COSMIC v80, released 13 FEB 17

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below, or by browsing a region of the human genome using the map to the right.

eg: *Braf, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell* **SEARCH**

R Resources

Key COSMIC resources

- Cell Lines Project
- COSMIC
- Cancer Gene Census
- Drug Sensitivity [↗](#)
- Mutational Signatures
- GRCh37 Cancer Archive [↗](#)

C Expert Curation

High quality curation by expert postdoctoral

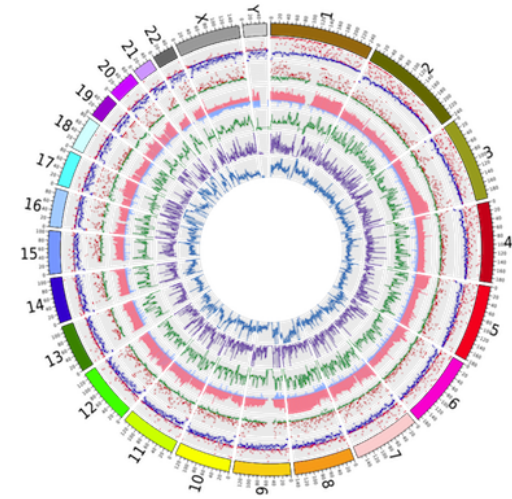
T Tools

Additional tools to explore COSMIC

- Cancer Browser
- Genome Browser
- GA4GH Beacon
- COSMIC-3D **Beta**
- COSMIC In BigQuery [↗](#) **New**
- CONAN

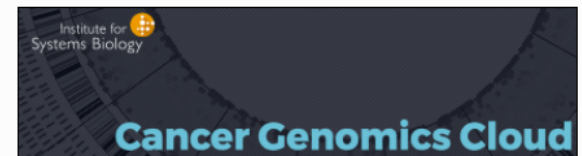
D Data

Further details on using COSMIC's content



Browse the **genomic landscape** of cancer

COSMIC in BigQuery hosted by ISB-CGC



Cancer Gene Census



Census Breakdown Abbreviations

The cancer Gene Census is an ongoing effort to catalogue those genes for which mutations have been causally implicated in cancer. The original census and analysis was published in [Nature Reviews Cancer](#) and [supplemental analysis information](#) related to the paper is also available.

The census is not static but rather is updated regularly/as needed. In particular we are grateful to Felix Mitelman and his colleagues in providing information on more genes involved in uncommon translocations in leukaemias and lymphomas. Currently, more than 1% of all human genes are implicated via mutation in cancer. Of these, approximately 90% have somatic mutations in cancer, 20% bear germline mutations that predispose to cancer and 10% show both somatic and germline mutations.

Show entries

Export: Search:

Gene Symbol ▲	Name ⇅	Entrez GeneId ⇅	Genome Location ⇅	Chr Band ⇅	Somatic ⇅	Germline ⇅	Tumour Types(Somatic) ⇅	Tumour Types(Germline) ⇅	Cancer Syndrome ⇅	Tissue Type ⇅	Molecular Genetics ⇅	Role in Cancer ⇅	M
ABI1	abl-interactor 1	10006	10:26748570-26860863	10p11.2	yes		AML			L	Dom	TSG	T
ABL1	v-abl Abelson murine leukemia viral oncogene homolog 1	25	9:130835447-130885683	9q34.1	yes		CML; ALL; T-ALL			L	Dom	oncogene	T;
ABL2	c-abl oncogene 2; non-receptor tyrosine kinase	27	1:179107718-179143044	1q24-q25	yes		AML			L	Dom	oncogene	T
ACKR3	atypical chemokine receptor 3	57007		2q37.3	yes		lipoma			M	Dom	oncogene	T
ACSL3	acyl-CoA synthetase long-chain family member 3	2181	2:222908773-222941654	2q36	yes		prostate			E	Dom		T

Summary

- There are global cancer genome sequencing projects with publically available data
- TCGA data can downloaded or easily viewed through cBioPortal
- ICGC data can be downloaded or viewed from the user interface
- COSMIC database allows you to easily select cancer-associated genes



Exercises

1. Download patient clinical annotations for AML (TCGA dataset) using GDC data portal and then using the ICGC data portal.

2. Using the ICGC data portal:
 - a. What is the cancer with most frequent RUNX1 mutations?
 - b. Which cancer has the most RUNX1 frameshift mutations?

3. Using cBioPortal and COSMIC:
 - a. Do kidney renal papillary cell carcinoma patients with BAP1 mutations have worse survival than those without?
 - b. Is this gene listed in the Cancer Gene Census and, if so, what is its role in cancer?

