

Cancer Genomics Resources

Databases

Tools

NCBI Gene

<http://www.ncbi.nlm.nih.gov/gene/>

NCBI Resources ▾ How To ▾

Gene Gene Advanced

Full Report ▾ Send to: ▾

EGFR epidermal growth factor receptor [*Homo sapiens* (human)]

Gene ID: 1956, updated on 29-Nov-2015

Summary ↗ ?

Official Symbol EGFR provided by [HGNC](#)

Official Full Name epidermal growth factor receptor provided by [HGNC](#)

Primary source [HGNC:HGNC:3236](#)

See related [Ensembl:ENSG00000146648](#); [HPRD:00579](#); [MIM:131550](#); [Vega:OTTHUMG00000023661](#)

Gene type protein coding

RefSeq status REVIEWED

Organism [Homo sapiens](#)

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as ERBB; HER1; mENA; ERBB1; PIG61; NISBD2

Summary The protein encoded by this gene is a transmembrane glycoprotein that is a member of the protein kinase superfamily. This protein is a receptor for members of the epidermal growth factor family. EGFR is a cell surface protein that binds to epidermal growth factor. Binding of the protein to a ligand induces receptor dimerization and tyrosine autophosphorylation and leads to cell proliferation. Mutations in this gene are associated with lung cancer. Multiple alternatively spliced transcript variants that encode different protein isoforms have been found for this gene. [provided by RefSeq, Jul 2010]

Orthologs [mouse](#) [all](#)

Ensembl

http://www.ensembl.org/

Human (GRCh38.p3) ▾ Location: 7:55,019,021-55,256,620 Gene: EGFR

Login/Register

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Gene alleles
- Sequence
 - Secondary Structure
 - External references
 - Regulation
- Comparative Genomics
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
 - Ensembl protein families
- Phenotype
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- External data
 - Gene expression
 - Personal annotation

Gene: EGFR ENSG00000146648

Description epidermal growth factor receptor [Source:HGNC Symbol;Acc:[HGNC:3236](#)]

Synonyms ERBB, ERBB1

Location Chromosome 7: 55,019,021-55,256,620 forward strand.
GRCh38:CM000669.2

About this gene This gene has 11 transcripts ([splice variants](#)), [77 orthologues](#), [13 paralogues](#), is a member of [1 Ensembl protein family](#) and is associated with [14 phenotypes](#).

Transcripts [Show transcript table](#)

Summary [?](#)

Name [EGFR](#) (HGNC Symbol)

This gene is a member of the Human CCDS set: [CCDS47587.1](#), [CCDS5514.1](#), [CCDS5515.1](#), [CCDS5516.1](#)

This gene has proteins that correspond to the following Uniprot identifiers: [P00533](#)

Overlapping RefSeq Gene ID [1956](#) matches and has similar biotype of protein_coding

[LRG_304](#) provides a stable genomic reference framework for describing sequence variants for this gene
ENSG00000146648.15

GeneCards

http://www.genecards.org/

FREE for academic non-profit institutions. Other users need a Commercial license

WEIZMANN INSTITUTE OF SCIENCE

Lifemap SCIENCES

Keywords ▾ Search Term Advanced

Home User Guide Analysis Tools ▾ News And Views About ▾ My Genes Log In / Sign Up

EGFR Gene (Protein Coding)

Epidermal Growth Factor Receptor

★ + ⌂ ⌂ in Twitter Facebook
GCID: GC07P055019 ⓘ
GIFTs: 74 ⓘ

Jump to section Aliases Disorders Domains Drugs Expression Function Genomics Localization Orthologs Paralogs Pathways Products Proteins Publications Sources Summaries Transcripts Variants

EMD MILLIPORE Proteins & Enzymes Antibodies Assays & Kits

ORIGENE Proteins Antibodies Assays Genes shRNA Primers CRISPR

GenScript Make Research Easy Genes Peptides Proteins CRISPR

Aliases for EGFR Gene

Aliases for EGFR Gene

Epidermal Growth Factor Receptor ^{2 3}

Receptor Tyrosine-Protein Kinase ErbB-1 ^{3 4}

Erb-B2 Receptor Tyrosine Kinase 1 ^{2 3}

Proto-Oncogene C-ErbB-1 ^{3 4}

EC 2.7.10.1 ^{4 63}

ERBB1 ^{3 4}

ERBB ^{3 4}

Erythroblastic Leukemia Viral (V-Erb-B) Oncogene Homolog (Avian) ²

Avian Erythroblastic Leukemia Viral (V-Erb-B) Oncogene Homolog ³

Cell Proliferation-Inducing Protein 61 ³

Cell Growth Inhibiting Protein 40 ³

EC 2.7.10 ⁶³

GenScript CRISPR gRNA constructs

GTEX

<http://www.gtexportal.org/>

Correlations between genotype and tissue-specific gene expression levels will help identify regions of the genome that influence whether and how much a gene is expressed

CANCER BROWSERS

cBio Portal for Cancer genomics

<http://www.cbioportal.org/>

cBioPortal
for Cancer Genomics

Visualize, analyze, discover.

HOME DATA SETS WEB API R/MATLAB TUTORIALS FAQ NEWS TOOLS ABOUT

VISUALIZE YOUR DATA

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

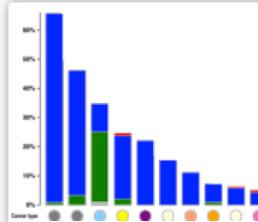
Please adhere to [the TCGA publication guidelines](#) when using TCGA data in your publications.

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 (105)
 - Adrenal Gland (1)
 - Adrenocortical Carcinoma (1)
 - Adrenocortical Carcinoma (TCGA, Provisional)** 92 samples
 - Biliary Tract (4)
 - Cholangiocarcinoma (3)
 - Intrahepatic Cholangiocarcinoma (Johns Hopkins University, Nature Genetics 2013)** 40 samples

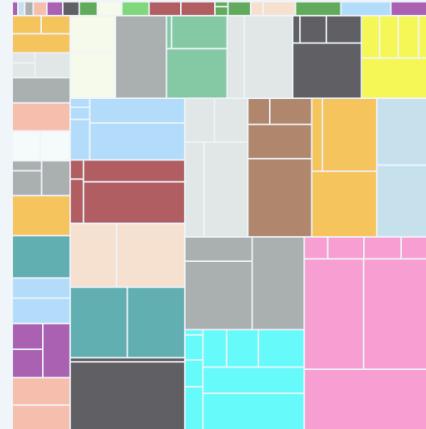
What's New

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 **105 cancer studies**.
[Details]



Example Queries

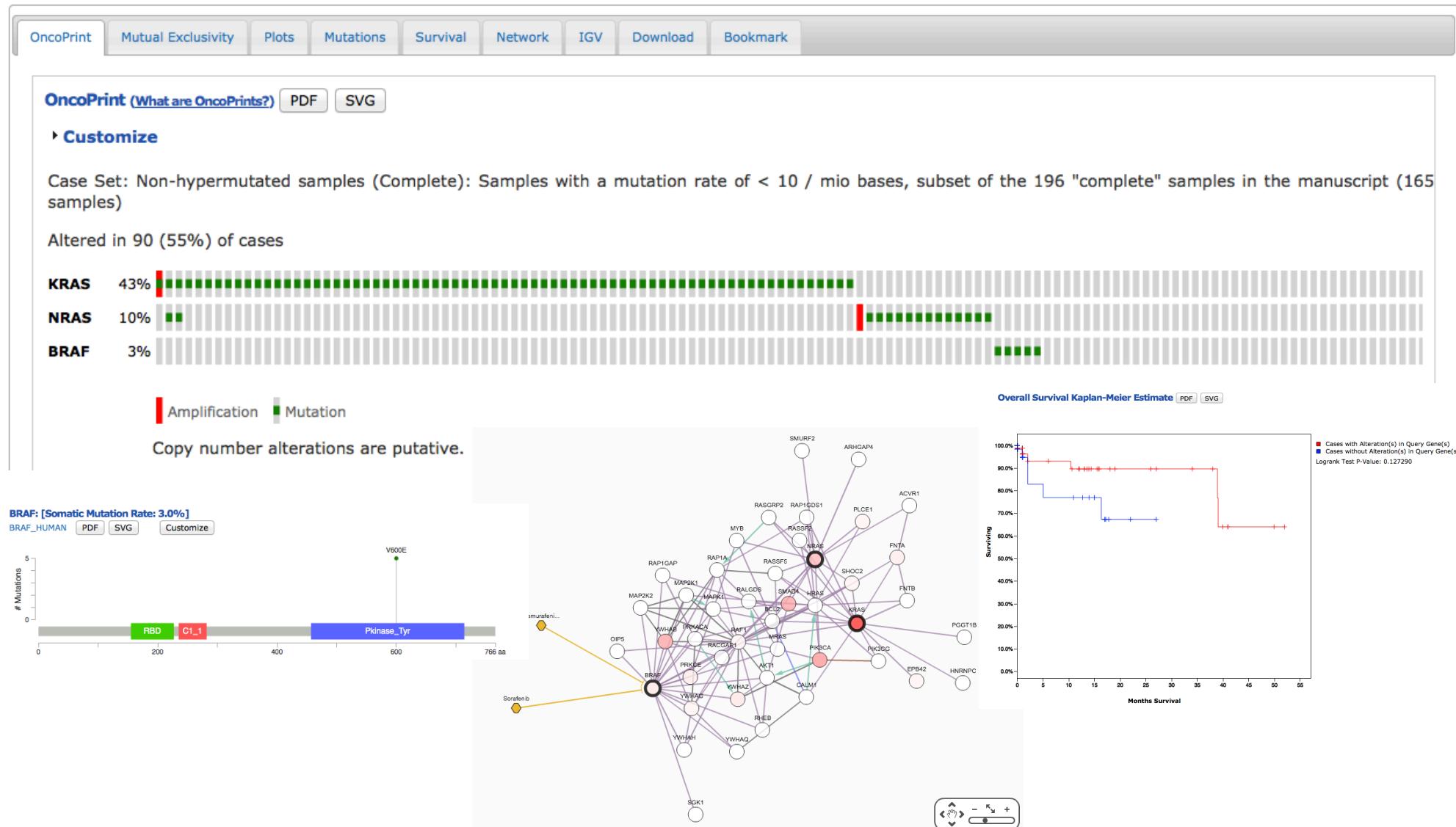
RAS/RAF alterations in colorectal cancer

RAS/RAF alterations in colorectal cancer

Gene Set / Pathway is altered in 54.5% of all cases.

Colon and Rectum Adenocarcinoma (TCGA, Nature 2012)/Non-hypermutated samples (Complete): (165)/User-defined List/3genes

▶ **Modify Query**



International Cancer Genome Consortium

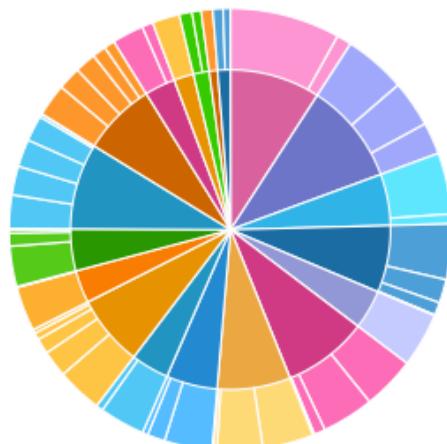
<http://dcc.icgc.org/>

[Cancer Projects](#)[Advanced Search](#)[Data Analysis](#)[Data Repository](#)

eg. BRAF, KRAS G12D, DO35100, MU7870, apoptosis, Cancer Gene Census, GO:0016049

Data Release 19
June 16th, 2015

Donor Distribution by Primary Site



Cancer projects	55
Cancer primary sites	21
Donors	12,979
Simple somatic mutations	16,459,160
Mutated genes	57,543

COSMIC

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

COSMIC
Catalogue of somatic mutations in cancer

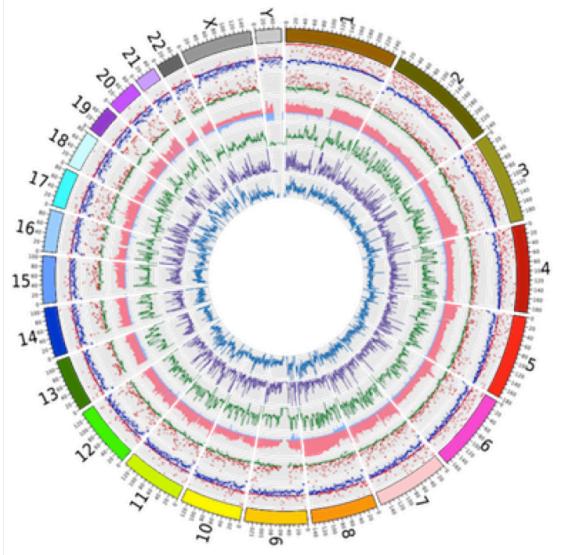
Home ▾ About ▾ Licensing ▾ Data Download ▾ News ▾ Help ▾ Enter search here... Login ▾

COSMIC v75

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

R Resources
Key COSMIC resources
[Cell Lines Project](#)
[COSMIC Whole Genomes](#)
[Cancer Gene Census](#)
[Drug Sensitivity](#)
[Mutational Signatures](#)
[GRCh37 Cancer Archive](#)

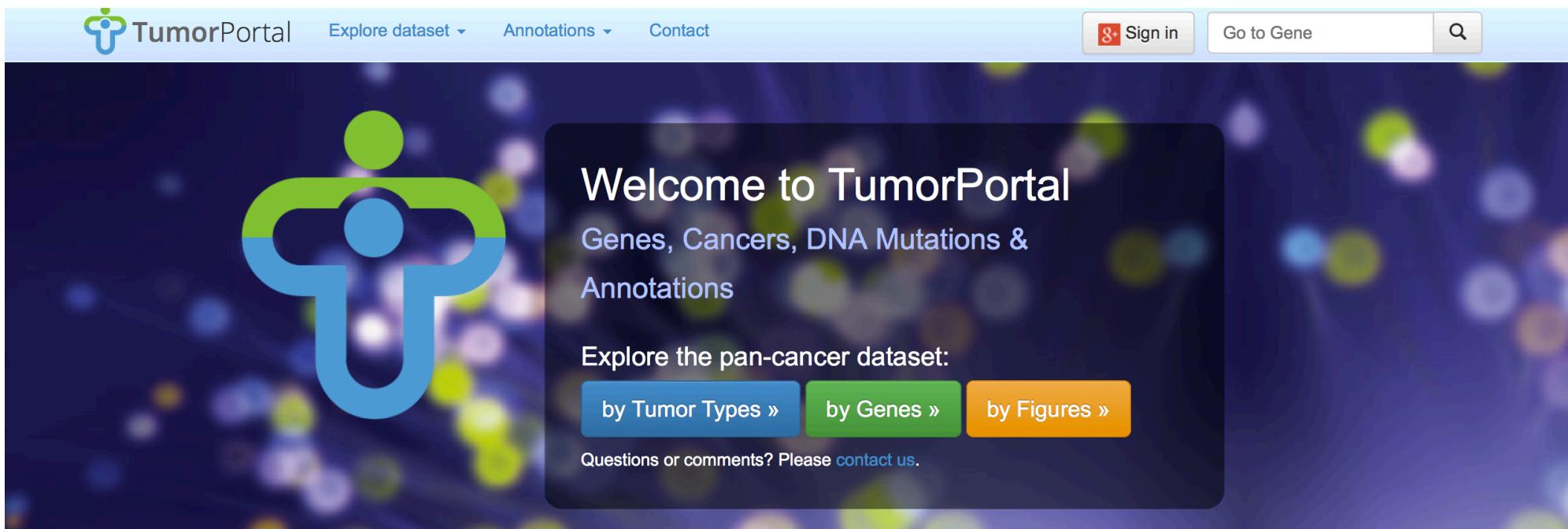
T Tools
Additional tools to explore COSMIC
[Cancer Browser](#)
[Genome Browser](#)
[CONAN](#)
[Beacon^{New}](#)
[COSMIC Mart](#)



Genomic Landscape of Cancer

TumorPortal

http://www.tumorportal.org/



The screenshot shows the TumorPortal homepage. At the top left is the TumorPortal logo. To its right are navigation links: "Explore dataset ▾", "Annotations ▾", and "Contact". On the far right are "Sign in" (with a Google+ icon), "Go to Gene", and a search icon. The main content area features a large, stylized green and blue "U" logo on the left. In the center, a dark rectangular box contains the text "Welcome to TumorPortal" and "Genes, Cancers, DNA Mutations & Annotations". Below this is the text "Explore the pan-cancer dataset:" followed by three buttons: "by Tumor Types ▷" (blue), "by Genes ▷" (green), and "by Figures ▷" (orange). At the bottom of the central box is the text "Questions or comments? Please contact us.".

Explore dataset by tumor types

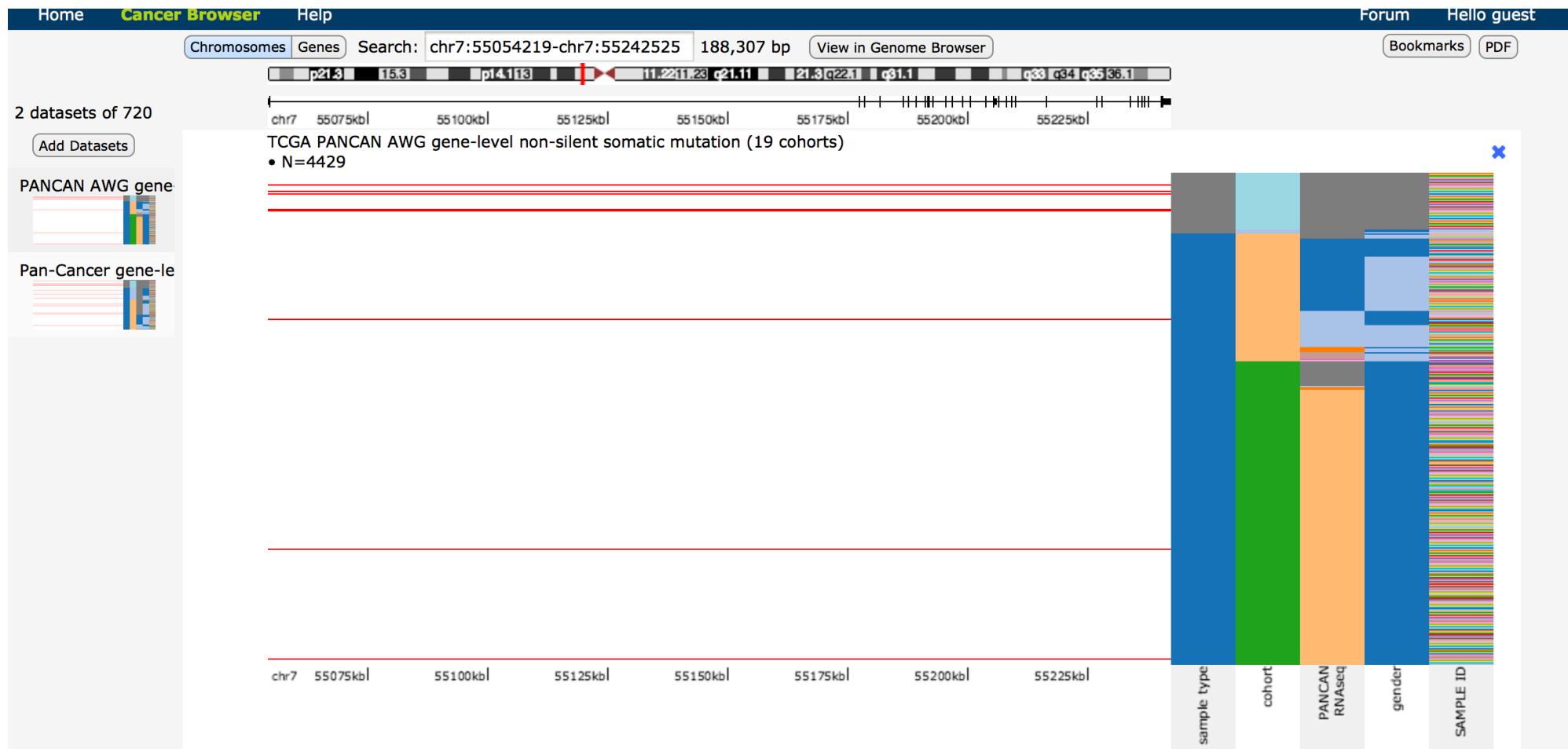
Click on a tumor type to see what genes are significantly mutated in it (and other details).

[◀ Show Annotation Activity](#)

Acute myeloid leukemia AML 196 patients	Bladder BLCA 99 patients	Breast BRCA 892 patients	Carcinoid CARC 54 patients	Chronic lymphocytic leukemia CLL 159 patients	Colorectal CRC 233 patients	Diffuse large B-cell lymphoma DLBCL 58 patients	Esophageal adenocarcinoma ESO 141 patients	Glioblastoma multiforme GBM 291 patients
--	---------------------------------------	---------------------------------------	---	--	--	--	---	---

Cancer Genome Browser UCSC

<https://genome-cancer.ucsc.edu/>



Integrative Onco Genomics

<https://www.intogen.org/>



e.g. Recurrence of TP53 mutations

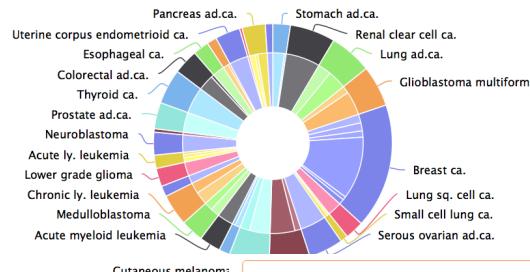
[Search example](#) | [Show more examples](#)

Release 2014.12

Plot Table

IntOGen Mutations 2014.12

Cancer types and projects chart



Cancer Types

28

Projects

48

Samples

6792

Somatic mutations

1341752

Coding sequence mutations (CSMs) ⓘ

21648

in driver genes

1341706

Protein affecting mutations (PAMs) ⓘ

18649

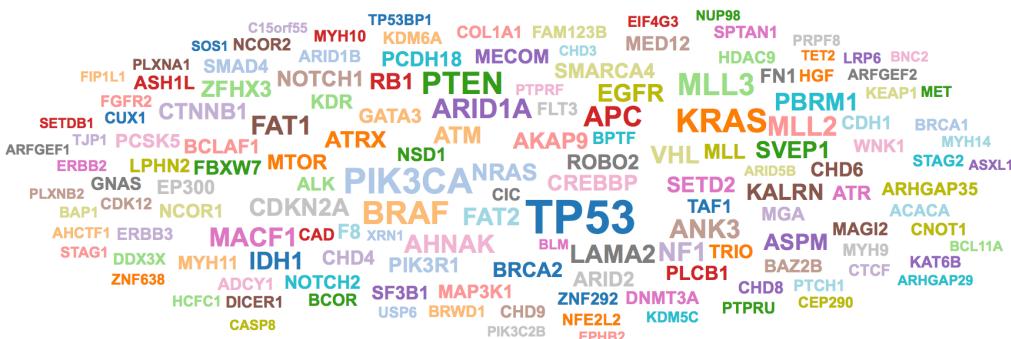
in driver genes

603770

Mutational cancer driver genes: 459

Cloud Plot Table

Downloads Print Copy



This driver cloud represents the most recurrently mutated cancer driver genes. The size of the gene symbol is relative to the count of samples with PAMs.

NETWORKS

Networks: STRING

<http://string-db.org/>

[Home](#) • [Download](#) • [Help](#) • [My Data](#)



STRING - Known and Predicted Protein-Protein Interactions

search by name search by protein sequence multiple names multiple sequences

protein name: (examples: #1 #2 #3)
|

(STRING understands a variety of protein names and accessions; you can also try a [random entry](#))

organism:
auto-detect ▾

interactors wanted:
COGs Proteins Reset GO!

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context



High-throughput Experiments



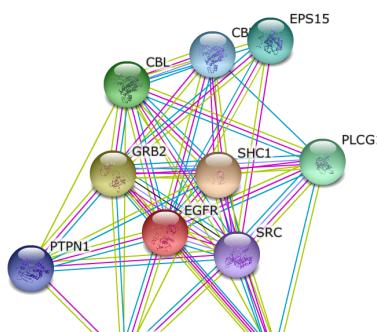
(Conserved) Coexpression



Previous Knowledge



STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 9'643'763 proteins from 2'031 organisms.



This is the **evidence view**. Different line colors represent the types of evidence for the association.



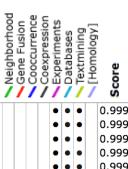
(requires Flash player 10 or better)

Your Input:

- EGFR epidermal growth factor receptor (1210 aa)
(*Homo sapiens*)

Predicted Functional Partners:

- | | |
|--|--|
| SHC1 | SHC (Src homology 2 domain containing) transforming protein 1 (584 aa)
growth factor receptor-bound protein 2; Adapter protein that provides a critical link between c [...] (217 aa) |
| GRB2 | signal transducer and activator of transcription 3 (acute-phase response factor); Signal transd [...] (770 aa) |
| STAT3 | Cbl proto-oncogene, E3 ubiquitin protein ligase; Adapter protein that functions as a negative r [...] (906 aa) |
| CBL | phosphoinositide-3-OH kinase, catalytic subunit gamma 1; Mediates the production of the second messenger molecules diacylglycerol [...] (11291 aa) |
| PI CG1 | |



PATHiVAR

<http://pathivar.babelomics.org/>



PATHiVAR



PATHiVAR

Try it now

Overview

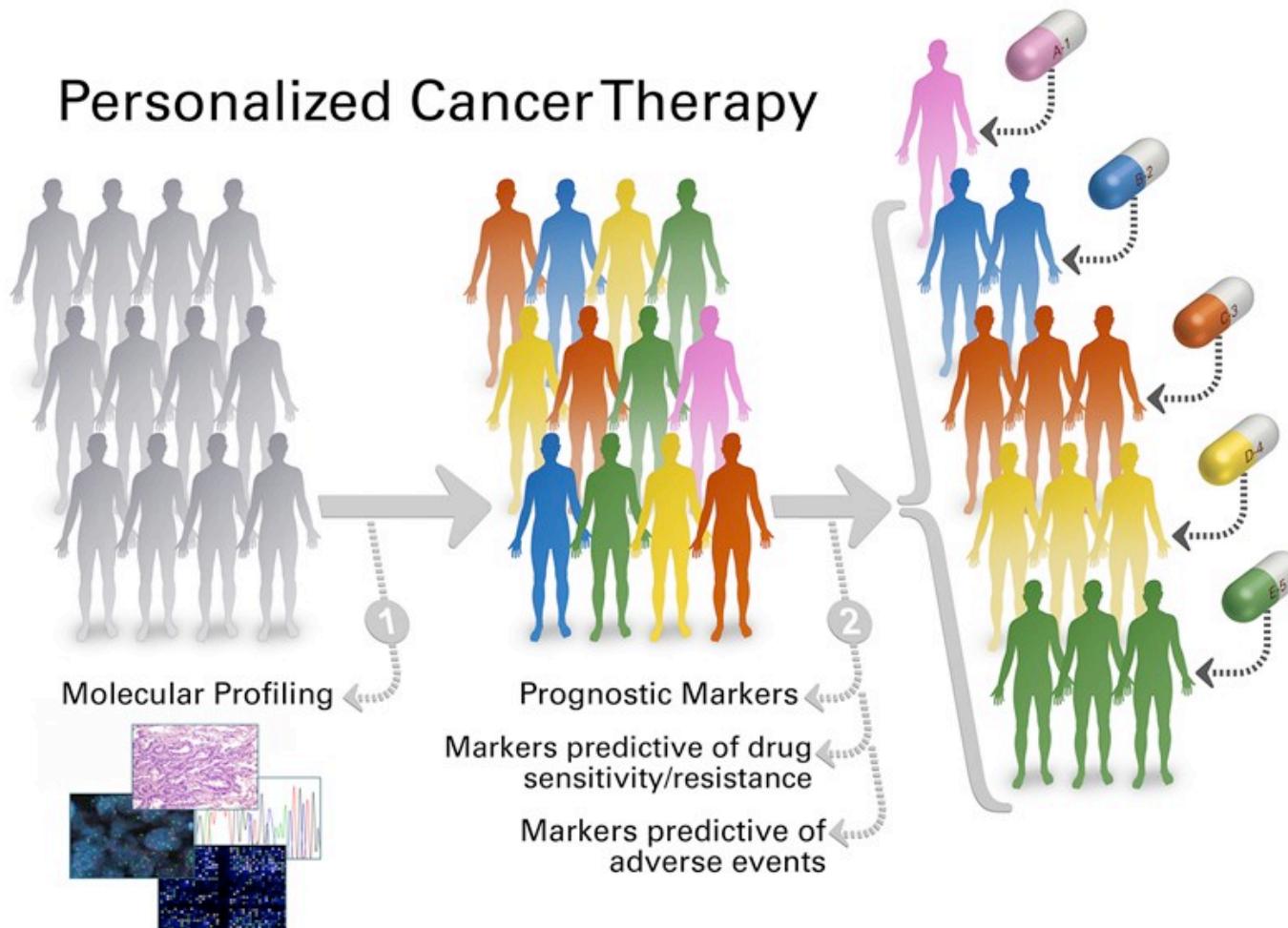
PATHiVAR estimates the functional impact that mutations have over the human signalling network.

PATHiVAR analyses VCF files, extract the deleterious mutations, locate them over the signalling pathways in the selected tissue (with the appropriate expression pattern) and provides a comprehensive, graphic and interactive view of the predicted signal transduction probabilities across the different signalling pathways.

DRUGS/THERAPIES

Personalized Cancer Therapy

<https://pct.mdanderson.org/>

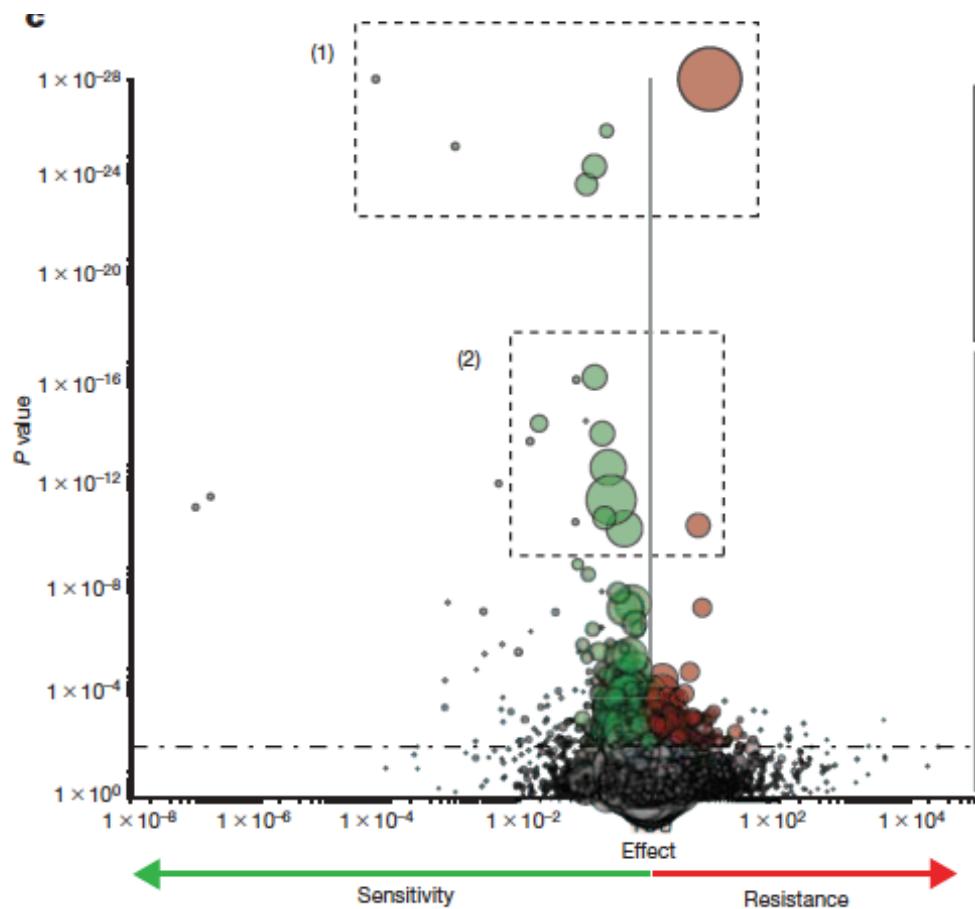


Genomics of Drug Sensitivity in Cancer

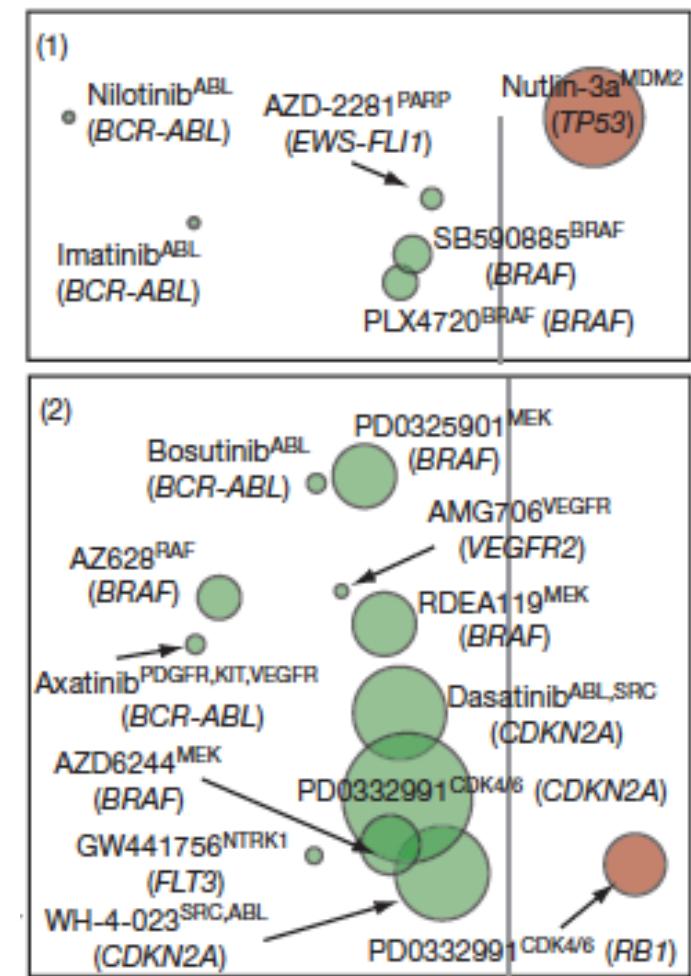
<http://www.cancerrxgene.org>

Systematic identification of genomic markers of drug sensitivity in cancer cells

Mathew J. Garnett^{1*}, Elena J. Edelman^{2*}, Sonja J. Heidorn^{1*}, Chris D. Greenman^{1†}, Anahita Dastur², King Wai Lau¹, Patricia Greninger², I. Richard Thompson¹, Xi Luo², Jorge Soares¹, Qingsong Liu^{3,4}, Francesco Iorio^{1,5}, Didier Surdez⁶, Li Chen², Randy J. Milano², Graham R. Bignell¹, Ah T. Tam², Helen Davies¹, Jesse A. Stevenson², Syd Barthorpe¹, Stephen R. Lutz², Fiona Kogera¹, Karl Lawrence¹, Anne McLaren-Douglas¹, Xeni Mitropoulos², Tatiana Mironenko¹, Helen Thi², Laura Richardson¹, Wenjun Zhou^{3,4}, Frances Jewitt¹, Tinghu Zhang^{3,4}, Patrick O'Brien¹, Jessica L. Boisvert², Stacey Price¹, Wooyoung Hur^{3,4}, Wanjuan Yang¹, Xiamming Deng^{3,4}, Adam Butler¹, Hwan Geun Choi^{3,4}, Jae Won Chang^{3,4}, Jose Baselga², Ivan Stamenkovic⁷, Jeffrey A. Engelmann², Sreenath V. Sharma^{3†}, Olivier Delattre⁶, Julio Saez-Rodriguez⁵, Nathanael S. Gray^{3,4}, Jeffrey Settleman², P. Andrew Futreal¹, Daniel A. Haber^{2,8}, Michael R. Stratton¹, Sridhar Ramaswamy², Ultan McDermott¹ & Cyril H. Benes²



Garnett et al. Nature 2012



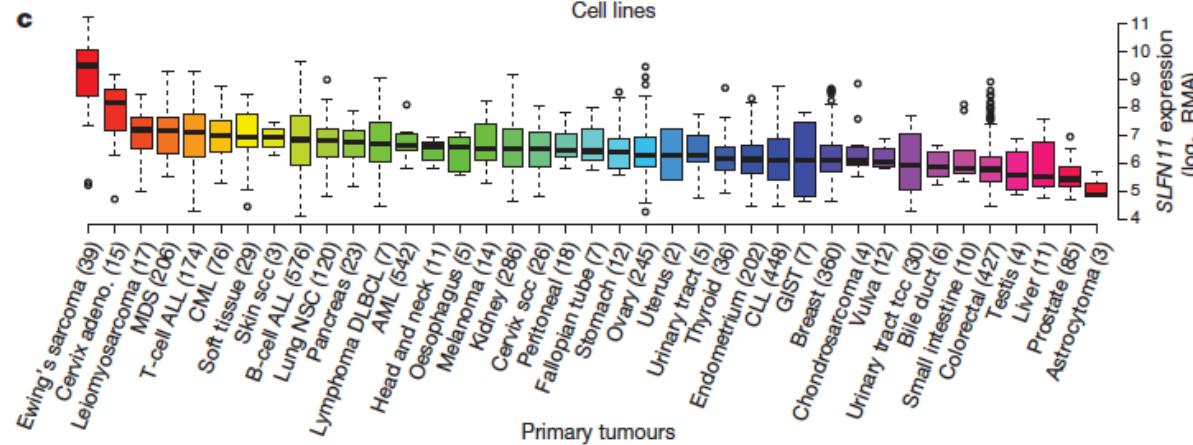
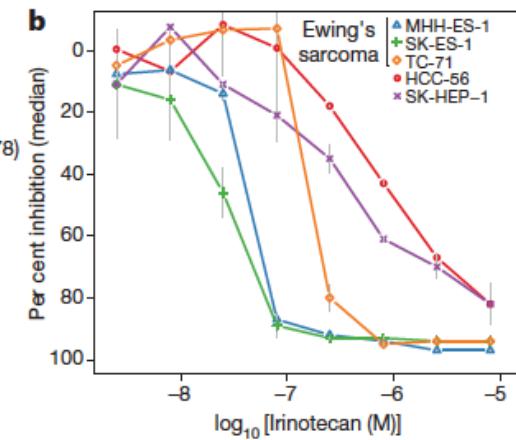
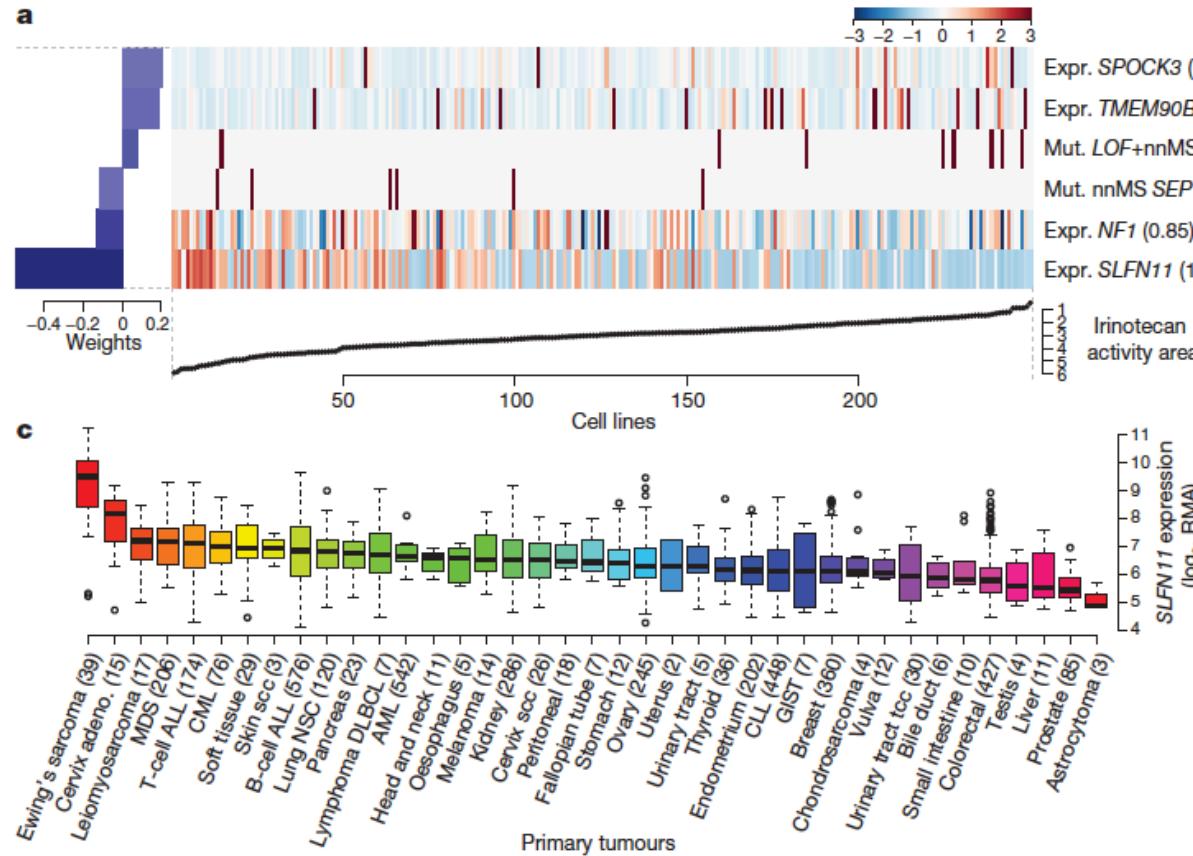
Cancer Cell Line Encyclopedia

<http://www.broadinstitute.org/ccle/>

The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity

Barretina et al. Nature 2012

Jordi Barretina^{1,2,3†*}, Giordano Caponigro^{4*}, Nicolas Stransky^{1*}, Kavitha Venkatesan^{4*}, Adam A. Margolin^{1†*}, Sungjoon Kim⁵, Christopher J. Wilson⁴, Joseph Lehár⁴, Gregory V. Kryukov⁴, Dmitry Sonkin⁴, Anupama Reddy⁴, Manway Liu⁴, Lauren Murray¹, Michael F. Berger^{1†}, John E. Monahan⁴, Paula Morais¹, Jodi Meltzer⁴, Adam Korejwa¹, Judit Jané-Valbuena^{1,2}, Felipa A. Mapa^{4*}, Joseph Thibault², Eva Bric-Purlong⁴, Pichai Raman⁴, Aaron Shipway⁶, Ingo H. Engels⁶, Jill Cheng⁶, Guoying K. Yu⁶, Jianjun Yu⁶, Peter Aspesi Jr⁴, Melanie de Silva⁴, Kalpana Jagtap⁴, Michael D. Jones⁴, Li Wang⁴, Charles Hatton³, Emanuele Palestro³, Supriya Gupta¹, Scott Mahan¹, Carrie Sougnez¹, Robert C. Onofrio¹, Ted Liefeld¹, Laura MacConaill³, Wendy Winckler¹, Michael Reich¹, Nanxin Li³, Jill P. Mesirov¹, Stacey B. Gabriel¹, Gad Getz¹, Kristin Ardlie¹, Vivien Chan⁶, Vic E. Myer¹, Barbara L. Weber⁴, Jeff Porter⁴, Markus Wermuth⁴, Peter Finan⁴, Jennifer L. Harris⁵, Matthew Meyerson^{1,2,3}, Todd R. Golub^{1,3,7,8}, Michael P. Morrissey^{4*}, William R. Sellers^{4*}, Robert Schlegel^{4*} & Levi A. Garraway^{1,2,3*}



Cancer Therapeutics Response Portal

<http://www.broadinstitute.org/ctrp/>

Cancer Therapeutics Response Portal v2

COMPOUNDS FEATURES TARGETS CLUSTER

inhibitor of SHH pathway
J Am Chem Soc (2013)

The Cancer Therapeutics Response Portal (CTR) links genetic, lineage, and other cellular features of cancer cell lines to small-molecule sensitivity with the goal of accelerating discovery of patient-matched cancer therapeutics. We generated an 'Informer Set' of 481 small-molecule probes and drugs that selectively target distinct nodes in cell circuitry and that collectively modulate a broad array of cell processes. We quantitatively measured the sensitivity of 860 deeply characterized cancer-cell lines to Informer Set compounds, and have undertaken analyses connecting sensitivity to cancer features, including mutations, gene expression, copy-number variation, and lineage. These analyses, and links to the underlying data, are provided openly on the CTRP.

The CTRP is a living resource for the biomedical research community that can be mined to develop insights into small-molecule mechanisms of action and novel therapeutic hypotheses, and to support future discovery of drugs matched to patients based on predictive biomarkers.

CTR P v2

- 481 compounds X 860 CCLs
- interactive interface to explore clustering by small molecule and CCL
- cluster enrichments for small molecule and CCL annotations
- annotations for small molecules by protein target
- annotations for CCLs by mutation and lineage

<< You are here

Complementary Resources

The National Cancer Institute's CTD2 Network maintains an Open-Access Data Portal that makes available raw data downloads from member Centers, including all raw sensitivity and enrichment data and other supporting information from the Broad related to this portal.

The Cancer Cell Line Encyclopedia provides public access to genomic data, analysis and visualization for about 1000 cell lines.

Publications



Please cite our cancer cell-line profiling Resource by referencing:
"Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset" Seashore-Ludlow *et al.*, *Cancer Discovery*, **5**, 1210-1223 (2015), and
"An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules" Basu, Bodycombe, Cheah, *et al.*, *Cell*, **154**, 1151-1161 (2013).

Acknowledgements

Complementary Resources



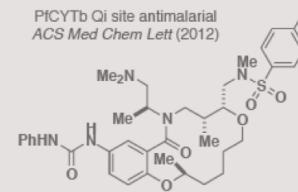
The National Cancer Institute's CTD2 Network maintains an Open-Access Data Portal that makes available raw data downloads from member Centers, including all raw sensitivity and enrichment data and other supporting information from the Broad related to this portal.

The Cancer Cell Line Encyclopedia provides public access to genomic data, analysis and visualization for about 1000 cell lines.

CTR P v1

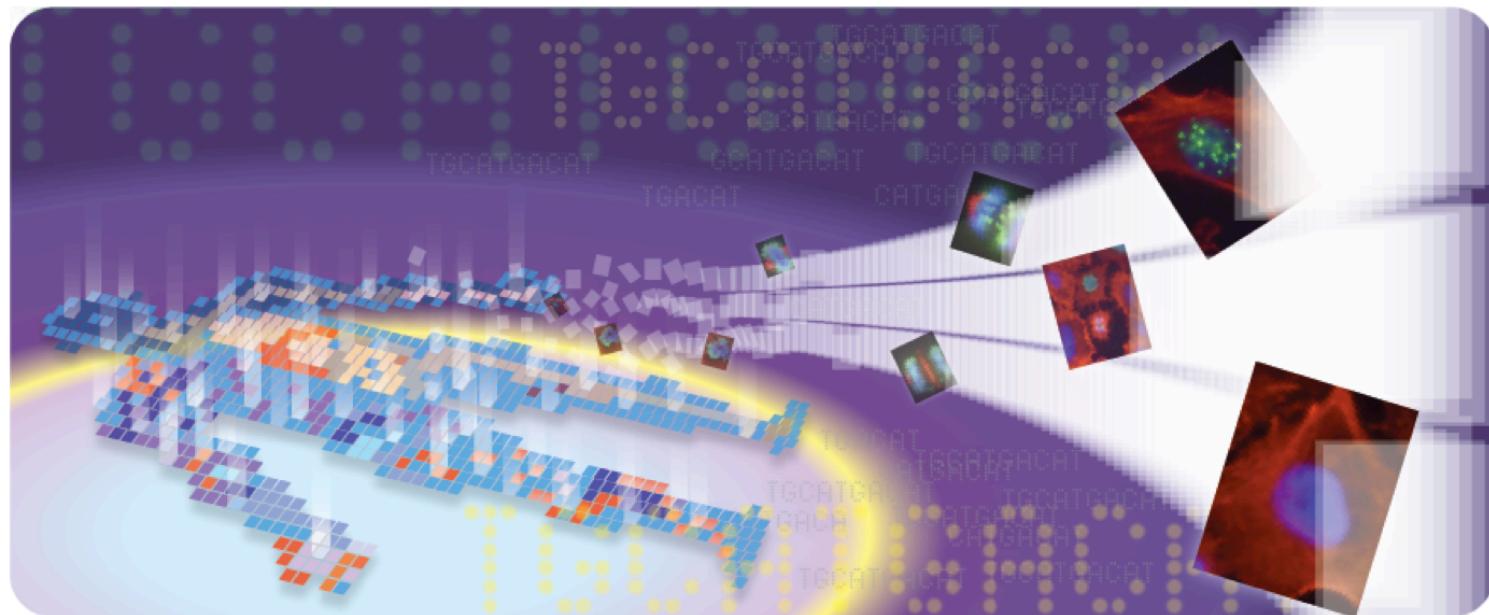
- 185 compounds X 242 CCLs
- pre-computed enrichment analysis and visualizations
- filter by lineage, CCLE mutation source, confounding factors
- 76,703 significant connections ($q < 0.01$)

[Visit CTRPv1 >>](#)



Project Achilles

<http://www.broadinstitute.org/achilles>



The Project Achilles website features a purple header bar with the project name and the Broad Institute logo. A navigation menu includes links for Data, Resources, About, and Contact, along with a Login button. Below the header is a large, central graphic illustrating the project's methodology. It shows a 3D ribbon of colored squares representing genetic data, with several small square images of microscopy slides floating around it. The background is a purple gradient with a grid pattern. The main text area describes the project's goal of identifying cancer genetic dependencies by perturbing individual genes and screening their effects on cell survival across many cancer cell lines, linking these findings to molecular characteristics for therapeutic development.

Project Achilles is a systematic effort aimed at identifying and cataloging genetic vulnerabilities across hundreds of genetically characterized cancer cell lines. The project uses genome-wide genetic perturbation reagents (shRNAs or Cas9/sgRNAs) to silence or knock-out individual genes and identify those genes that affect cell survival. Large-scale functional screening of cancer cell lines provides a complementary approach to those studies that aim to characterize the molecular alterations (e.g. mutations, copy number alterations) of primary tumors, such as The Cancer Genome Atlas (TCGA). The overall goal of the project is to identify cancer genetic dependencies and link them to molecular characteristics in order to prioritize targets for therapeutic development and identify the patient population that might benefit from such targets.

Networks: STITCH

<http://stitch.embl.de/>

STITCH 4.0

[Input Page](#) | [Downloads](#) | [Help/Info](#) | [My Data](#)

search by name chemical structure(s) protein sequence(s) multiple names batch import

name: (examples: #1 #2 #3)

(STITCH understands a variety of chemical/protein names, accessions and InChIKeys; you can also try a [random entry](#))

organism:

auto-detect (in case of chemicals, the organism with the most likely interaction is chosen)

[Reset](#) [GO !](#)

please enter your protein or chemical of interest...

STITCH: Chemical-Protein Interactions

Interaction network around tryptophan:

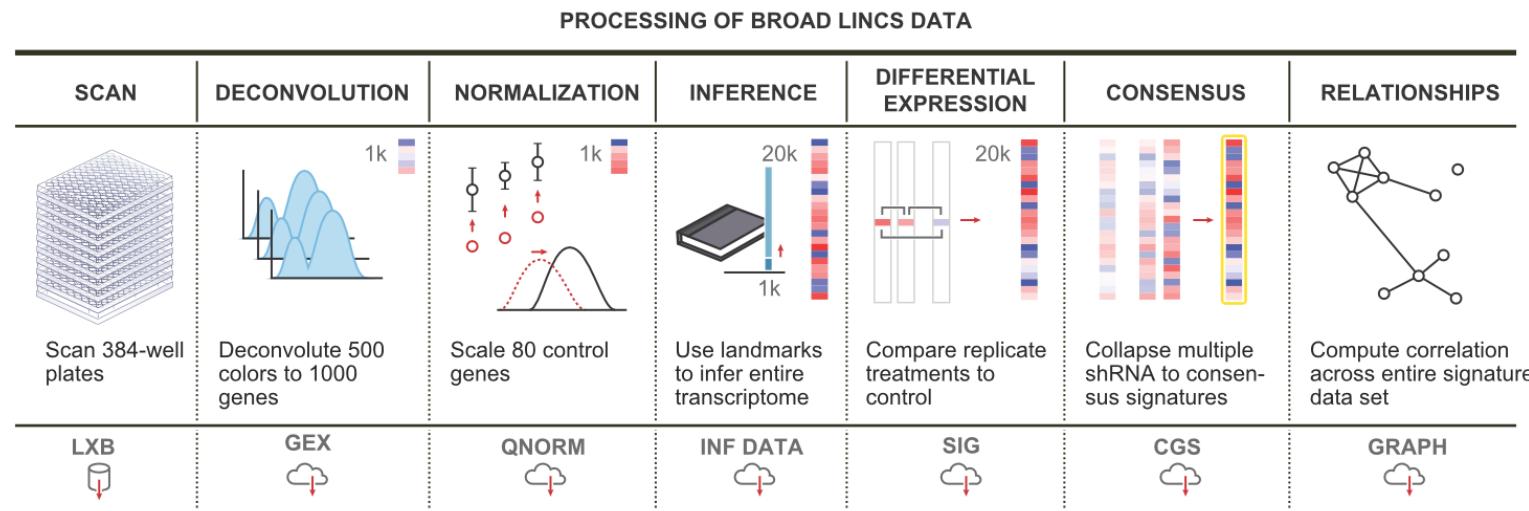
The diagram illustrates the interaction network around tryptophan. Tryptophan is a tryptamine derivative (C[C@H](C[C@H]1=CC=C(C=C1)N)C(=O)O). It interacts with several proteins: trpR, trpB, trpS, IGPS, trpD, and trpE. trpR is associated with trpB. trpS is associated with trpB and trpD. IGPS is associated with trpB and trpE. trpD is associated with trpE.

LINCS Cloud

<http://www.lincscloud.org/>



- Library of Integrated Cellular Signatures
- The aim is to generate perturbational profiles across multiple cell and perturbation types
 - Small Molecules
 - 1300 FDA approved drugs
 - 5500 bioactive compounds
 - Gene Knock-Outs and Over Expression



VCF ANNOTATORS

Variant Effect Predictor

<http://www.ensembl.org/info/docs/tools/vep/index.html/>

The screenshot shows the Ensembl web interface with the VEP tool selected. The top navigation bar includes links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar at the top right allows searching across all species. The main content area features a sidebar with various web tools, including Web Tools, BLAST/BLAT, Variant Effect Predictor (which is highlighted), and Assembly Converter. Below this is a 'New VEP job:' section with a button for 'VEP for Human GRCh37'. A note indicates that users looking for VEP for Human GRCh37 should go to the [GRCh37 website](#). The main form area is titled 'Input' and contains fields for 'Species' (set to Human (Homo sapiens)), 'Assembly' (set to GRCh38.p3), 'Name for this data (optional)', and a large text area for 'Either paste data:'.

StructurePPI

http://structureppi.bioinfo.cnio.es/

Structure PPi

Structure – Annotate

Annotates genomic mutations based on the protein features that are overlapping amino-acid changes

database (default: UniProt)

UniProt

principal (default: true)

mutations

Browse... No file selected. or use the text area below

Structure PPi

organism (default: Hsa/feb2014)

Hsa/feb2014

Tasks

annotate
Annotates genomic mutations based on the protein features that are overlapping amino-acid changes

annotate_mi
Annotates mutated isoforms based on the protein features that are overlapping amino-acid changes

annotate_mi_neighbours
Annotates mutated isoforms based on the protein features that are in close physical proximity to amino-acid changes

annotate_neighbours
Annotates genomic mutations based on the protein features that are in close physical proximity to amino-acid changes

interfaces

This workflow offers several functionalities to explore the consequence of protein mutations. It reports features that overlap the mutations, or that are in close physical proximity.

The features reported include protein domains, variants, helices, ligand binding residues, catalytic sites, transmembrane domains, InterPro domains, and known somatic mutations in different types of cancer. This information is extracted from resources such as UniProt, COSMIC, InterPro and Appris. It can also identify mutations affecting the interfaces of protein complexes.

This workflow makes use of PDB files to calculate residues in close proximity. This information is used to find features close to the mutations, at a distance of 5 angstroms, or mutations in residues close to residues in a complex partner, at a distance of up to 8 angstroms.

PDBs are extracted from Interactome3d, which organizes thousands of PDBs, for both experimental structures and structure models, of individual proteins and protein complexes.

Pairwise (Smith-Waterman) alignment is used to fix all inconsistencies between protein sequences in PDBs, Uniprot and Ensembl Protein ID.

Reference:

Vazquez M, Valencia A, Pons T. (2015) Structure-PPi: a module for the annotation of cancer-related single-nucleotide variants at protein-protein interfaces. *Bioinformatics* (2015); 31(14):2397-2399 (doi: 10.1093/bioinformatics/btv142)

contributes has been selected based on expert opinion and guided by empirical results on the COSMIC and 1000 Genomes data. The scoring scheme is as follows:

- Appris features: we add 2 if at least one ligand binding or catalytic site annotated in Firestar is affected; if none of the affected features meets this condition we add only 1
- COSMIC mutations: 3 if more than ten COSMIC samples have

PanDrugs

<http://pandrugs.bioinfo.cnio.es/>

Welcome to PANDRUGS

A resource to study drug-gene interactions in a cancer disease context

Query database! ✓

Query Pandrugs: Genes

Genes

Enter HUGO Gene symbols

Advanced Options

Drug status level

Cancer

FDA approved

Clinical trials

Other pathologies

FDA approved

Clinical trials Experimental

Interaction evidence level

Target Marker

Select Cancer Types

Select all Clear all

SURVIVAL, CLINICAL INFORMATION

Survival: Prognoscan

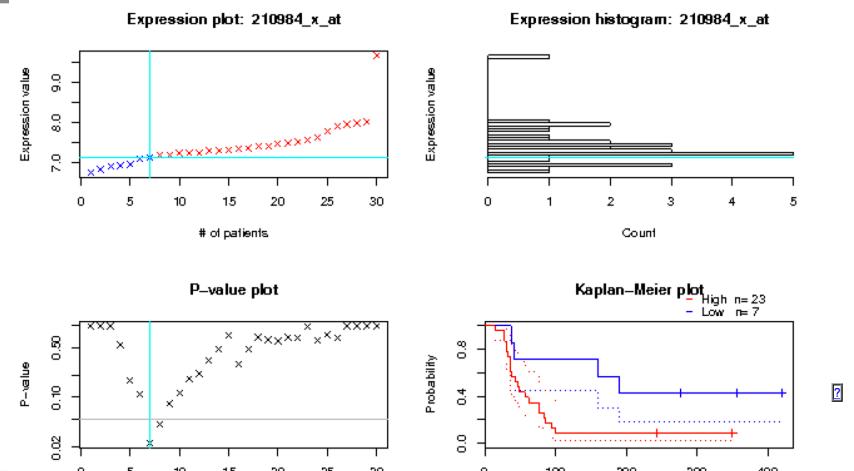
<http://www.abren.net/Prognoscan/>

Prognoscan: A new database for meta-analysis of the prognostic value of genes.

Enter gene identifier(s) [[Find gene at Entrez](#)]

GENE_SYMBOL EGFR
GENE_DESCRIPTION epidermal growth factor receptor

DATASET	CANCER_TYPE	SUBTYPE	ENDPOINT	COHORT	CONTRIBUTOR	ARRAY TYPE	PROBE	# of patients	P-value	Probability
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	210984_x_at	30	0.23	0.407818
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	201983_s_at	30	0.43	0.553565
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	211551_at	30	0.43	0.689576
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	201984_s_at	30	0.87	0.530147
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	211607_x_at	30	0.33	0.352078
GSE5287	Bladder cancer		Overall Survival	Aarhus (1995-2004)	Als	HG-U133A	211550_at	30	0.73	0.304392
GSE13507	Bladder cancer		Overall Survival	CNUH	Kim	Human-6 v2	ILMN_1696521	165	0.90	0.021788
GSE13507	Bladder cancer	Transitional cell carcinoma	Disease Specific Survival	CNUH	Kim	Human-6 v2	ILMN_1696521	165	0.90	0.088225
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCG (1999-2003)	Metzeler	HG-U133A	201984_s_at	163	0.15	0.381582
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCG (1999-2003)	Metzeler	HG-U133A	211550_at	163	0.10	0.406127
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCG (1999-2003)	Metzeler	HG-U133A	211607_x_at	163	0.89	0.646211
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCG (1999-2003)	Metzeler	HG-U133A	210984_x_at	163	0.11	0.760847
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCG (1999-2003)	Metzeler	HG-U133A	201983_s_at	163	0.29	0.636596
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCG (1999-2003)	Metzeler	HG-U133A	211551_at	163	0.11	0.468405
GSE12417-GPL96	Blood cancer	AML	Overall Survival	AMLCG (2004)	Metzeler	HG-U133_Plus_2	210984_x_at	79	0.11	0.520371



Survival: Kaplan Meier Plotter

<http://kmplot.com/analysis/index.php>

Kaplan-Meier Plotter

Breast Cancer

Lung

KM plotter Home Download Updates Contact

What is the KM plotter?

The Kaplan Meier plotter is capable to assess the effect of **54,675 / 22,277 genes** on survival using **10,188 cancer samples**. These include **4,142 breast, 1,648 ovarian, 2,437 lung and 1,065 gastric cancer patients** with a mean follow-up of 69 / 40 / 49 / 33 months. Primary purpose of the tool is a meta-analysis based **biomarker assessment**.

Start KM Plotter for breast cancer Start KM Plotter for ovarian cancer Start KM Plotter for lung cancer Start KM Plotter for gastric cancer

Kaplan-Meier Plotter Lung Cancer Lung

Affy id/Gene symbol: 204009_z_at Use multigene classifier
Survival: OS (n=1926)
Split patients by: median Auto select best cutoff:
Follow up threshold: all Censor at threshold:

Compute median over entire dataset:

Probe set options:
 user selected probe set
 all probe sets per gene
 only JetSet best probe set

Plot beeswarm graph of probe distribution:

Using the selected parameters, the analysis will run on 1926 patients.

P value: 0.0389

HR = 0.88 (0.77 - 0.99) logrank P = 0.039

Probability

Time (months)

Number at risk

Expression low high

0 50 100 150 200

963 369 69 18 2 5

963 459 134 39

Draw Kaplan-Meier plot

n = number of patients with available clinical data

Please note: the generated p value does **not** include correction for multiple hypothesis testing by default.

How to cite: Györfi B, Surowiak P, Budczies J, Lanczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One*, 2013 Dec 18;8(12):e82241. doi: 10.1371/journal.pone.0082241.

ClinVAR

<http://www.ncbi.nlm.nih.gov/clinvar/>

ClinVar ClinVar EGFR[gene] Search Help

Home About Access Using the website How to submit Statistics FTP site

Gene Tabular 100 per page Sort by Location Download: Customize this list...

Clinical significance Showing for results for variants in the EGFR gene. [Search instead for all ClinVar records that mention EGFR](#)

Uncertain significance (6) Pathogenic (6)

Likely pathogenic (1) Review status

Pathogenic (6) Single submitter (50)

At least one star (50) Allele origin

Single submitter (50) Germline (22)

At least one star (50) Somatic (149)

Allele origin Method type

Germline (22) Literature only (17)

Somatic (149) Clinical testing (156)

Method type Molecular consequence

Literature only (17) Frameshift (6)

Clinical testing (156) ...

Search results

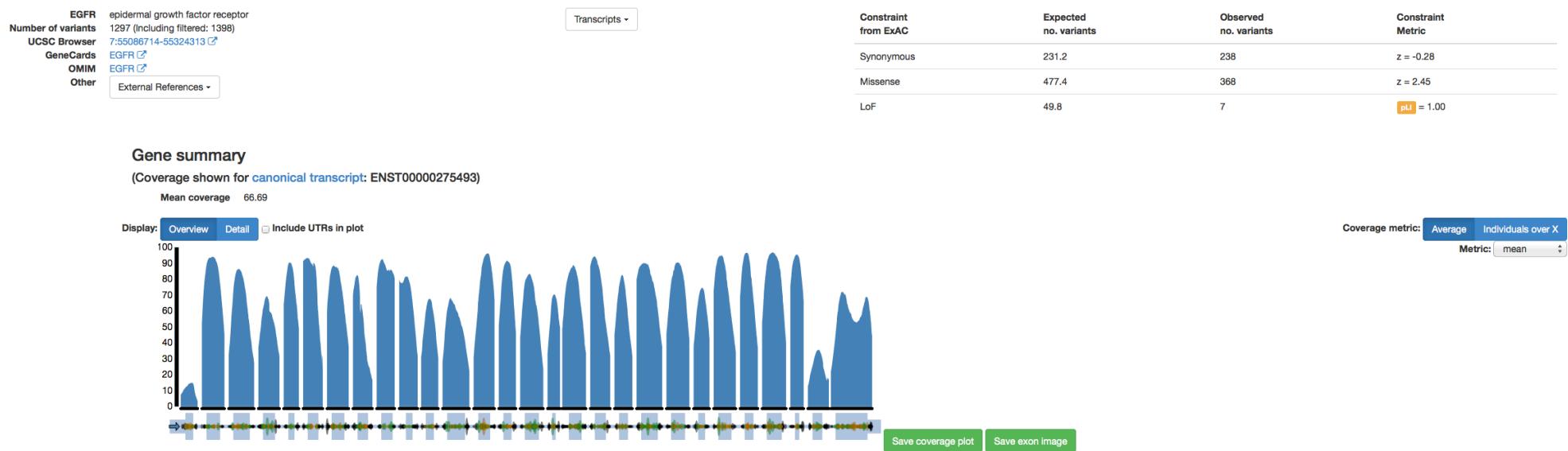
Items: 1 to 100 of 185 << First < Prev Page 1 of 2 Next > Last >>

	Variation Location	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)	Review status
1.	NM_005228.3(EGFR):c.493C>T (p.Arq165Trp) GRCh37: Chr7:55214367 GRCh38: Chr7:55146674	EGFR	not specified		not provided (Sep 19, 2013)	no assertion provided
2.	NM_005228.3(EGFR):c.1177G>C (p.Asp393His) GRCh37: Chr7:55224495 GRCh38: Chr7:55156802	EGFR	not specified		not provided (Sep 19, 2013)	no assertion provided

ExAC

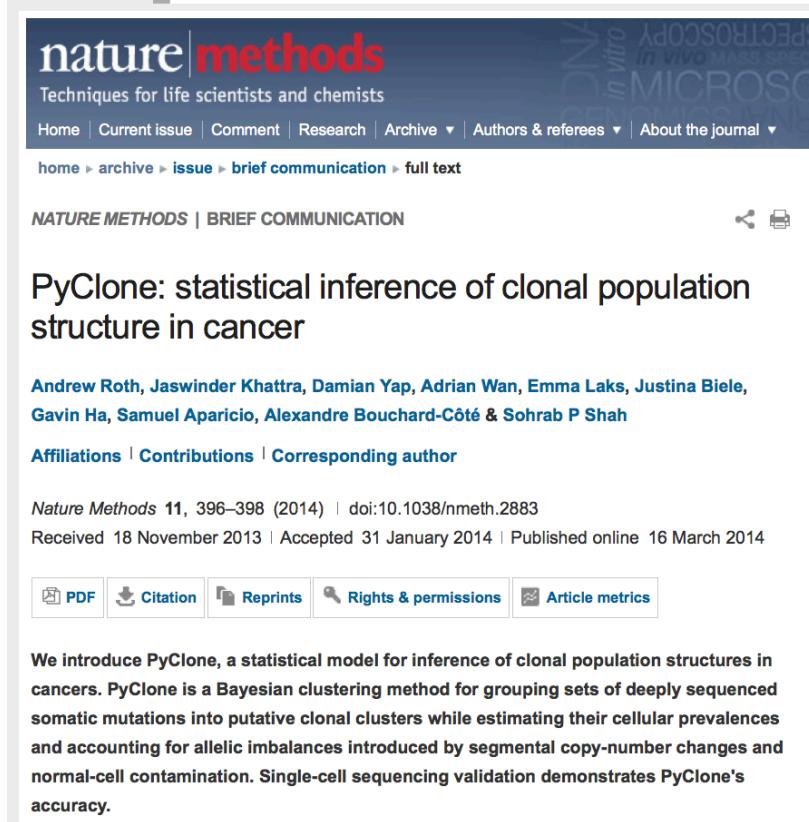
<http://exac.broadinstitute.org>

Gene: EGFR



Other Variant Callers

- **MuTect:** MuTect is a method for the reliable and accurate identification of somatic point mutations in next generation sequencing data of cancer genomes. [Check MuTect2]
<https://www.broadinstitute.org/cancer/cga/mutect>
- **PyClone:**



The screenshot shows a journal article from **Nature Methods**. The title of the article is **PyClone: statistical inference of clonal population structure in cancer**. The authors listed are Andrew Roth, Jaswinder Khatra, Damian Yap, Adrian Wan, Emma Laks, Justina Biele, Gavin Ha, Samuel Aparicio, Alexandre Bouchard-Côté & Sohrab P Shah. The article was published in **Nature Methods** 11, 396–398 (2014) | doi:10.1038/nmeth.2883. It was received on 18 November 2013, accepted on 31 January 2014, and published online on 16 March 2014. Below the article summary, there are links for PDF, Citation, Reprints, Rights & permissions, and Article metrics. The abstract text reads: "We introduce PyClone, a statistical model for inference of clonal population structures in cancers. PyClone is a Bayesian clustering method for grouping sets of deeply sequenced somatic mutations into putative clonal clusters while estimating their cellular prevalences and accounting for allelic imbalances introduced by segmental copy-number changes and normal-cell contamination. Single-cell sequencing validation demonstrates PyClone's accuracy."

<http://bcb.io/tags/>

Variant Callers Benchmarking

ABOUT RECENT TOPICS SEARCH

Blue Collar Bioinformatics

Community built tools for biological data analysis

CANCER • SMALL-VARIANTS • STRUCTURAL-VARIATION • VALIDATION

Validating multiple cancer variant callers and prioritization in tumor-only samples

- **OncoCNV:** Detection of copy number changes in Deep Sequencing data
- **CNVkit:** Genome-wide copy number from targeted DNA sequencing
- **THetA:**

Method

Open Access

THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data

Layla Oesper¹*, Ahmad Mahmoody¹ and Benjamin J Raphael^{1,2}*

* Corresponding authors: Layla Oesper layla@cs.brown.edu - Benjamin J Raphael raphael@cs.brown.edu

▼ Author Affiliations

¹ Department of Computer Science, Brown University, 115 Waterman Street, Providence, RI 02912, USA

² Center for Computational Molecular Biology, Brown University, Box 1910, Providence, RI 02912, USA

For all author emails, please [log on](#).

Genome Biology 2013, **14**:R80 doi:10.1186/gb-2013-14-7-r80

The electronic version of this article is the complete one and can be found online at:
<http://genomebiology.com/2013/14/7/R80>

Received: 26 April 2013

Revisions received: 31 May 2013

Accepted: 29 July 2013

Published: 29 July 2013