



# Medical and genomic tools and databases

Precision Oncology Course



CNIO BIOINFORMATICS UNIT

# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Genes

### NCBI Gene

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

### Ensembl

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

### GeneCards

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

### Expression Atlas

<https://www.ebi.ac.uk/gxa/home>

### Network of Cancer Genes

<http://ncg.kcl.ac.uk/index.php>

### GTEx

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

# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Protein

Tumor Protein Atlas

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

The Human Protein Atlas

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

# Exercise

## Gene & Protein DBs

### Exercise 1: Using NCBI gene.

- Go to NCBI's Gene database
- Do an Advanced search for
  - Organism: **homo sapiens**
  - Gene/Protein Name: **EGFR**

The screenshot shows the NCBI Gene search interface. The search term 'EGFR' is entered in the search bar. The results page for 'EGFR' in 'Homo sapiens (human)' is displayed, showing the gene's official symbol (EGFR), full name (epidermal growth factor receptor), primary source (HGNC:HGNC:3236), and various related entries like Ensembl, HPRD, MIM, and Vega. The 'Summary' section provides detailed information about the gene's function, being a transmembrane glycoprotein that binds to epidermal growth factor and leads to cell proliferation. It also lists orthologs in other species like mouse and all other species.

On which chromosome is the gene located?

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

# Exercise

## Gene & Protein DBs

### Exercise 1: Using NCBI gene.

Look at the **GeneRIFs** of this gene. GeneRIFs are short sentences that describe the function of a gene product that are extracted from scientific publications. NCBI encourages scientists to submit these GeneRIFs together with the Pubmed ID of the publication to the Gene database and curates the submissions. In this way you don't have to go through the scientific literature yourself to get idea about the function of a gene product.

**What is the function of the protein encoded by this gene ?**

*Check out all other info that you find here for EGFR and try to do a search on your favourite gene to see if you can find any new info on it.*

# Exercise

## Gene & Protein DBs

### Exercise 2: Using Ensembl

Search the human EGFR gene.

- Select the Human genome to search in
- Search for EGFR
- Click Go
- Click the EGFR (Human gene) link to go to the gene page of EGFR.

The screenshot shows the Ensembl gene page for EGFR. At the top, there's a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar at the top right contains the text "Search Human...". Below the navigation bar, the genome is set to "Human (GRCh38.p3)" and the location is "7:55,019,021-55,256,620". The gene identifier "Gene: EGFR" is shown. On the left, a sidebar titled "Gene-based displays" lists various options like Summary, Splice variants, Transcript comparison, Supporting evidence, Gene alleles, Sequence, Comparative Genomics, Genetic Variation, and External data. The "Summary" option is currently selected. The main content area on the right starts with a summary of EGFR, describing it as the "epidermal growth factor receptor [Source:HGNC Symbol;Acc:HGNC:3236]". It also mentions ERBB, ERBB1, Chromosome 7: 55,019,021-55,256,620 forward strand, and GRCh38.CM00669.2. Below this, there's information about transcripts, a summary section with details like Name (EGFR), CCDS (CCDS47587.1), UniProtKB (P00533), RefSeq (1956), LRG (LRG\_304), and Ensembl version (FNSGRNNNN14R&R 15). There's also a note about being a member of the Human CCDS set and having proteins corresponding to Uniprot identifiers.

**Ensembl**  
<http://www.ensembl.org/>

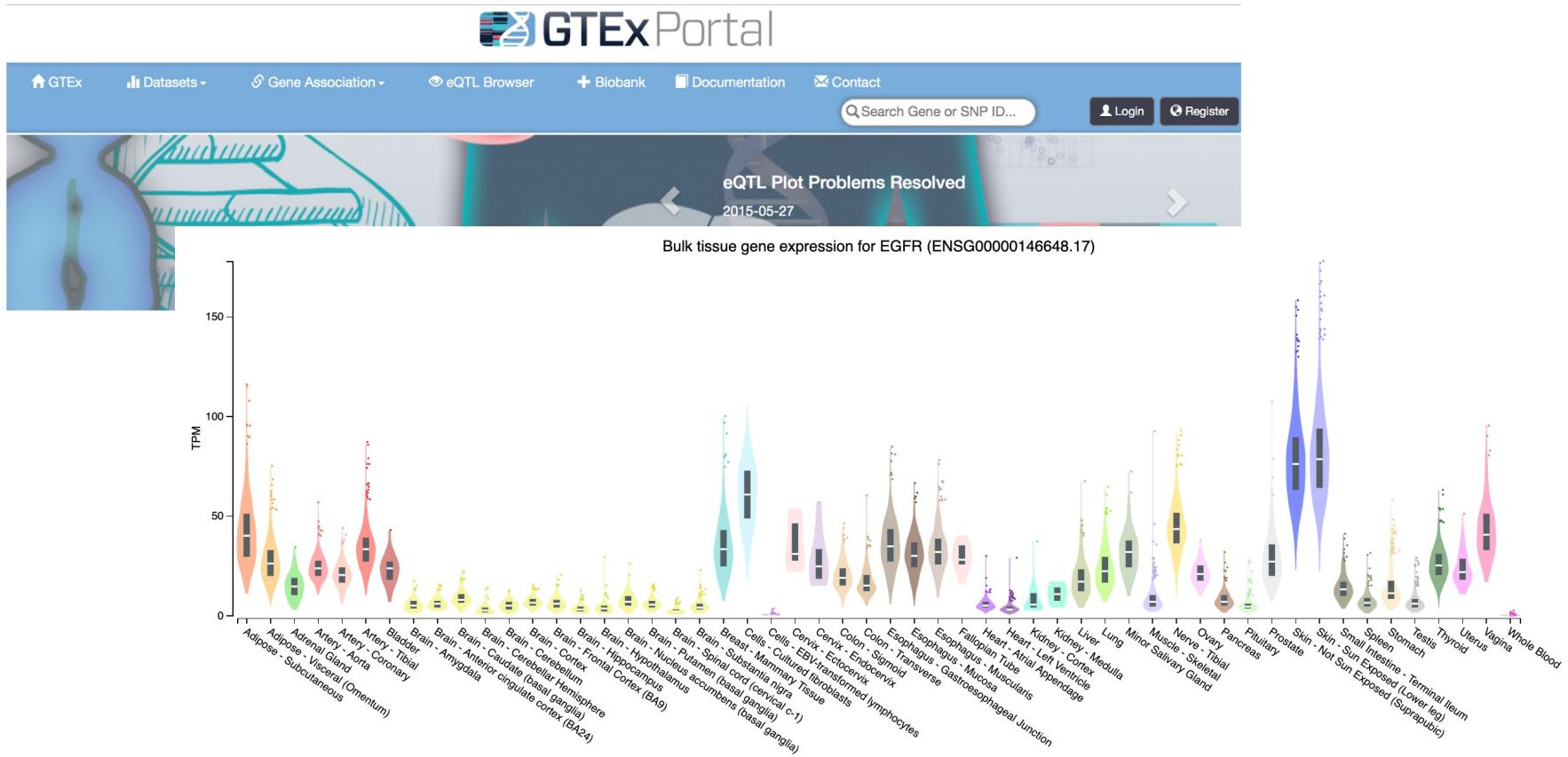
# Exercise

## Gene & Protein DBs

### Exercise 2: Using Ensembl

- On which chromosome and which strand of the genome is this gene located?
- How many transcripts (splice variants) does this gene have?
- How many CCDS are annotated for this gene?
- What's the name of the longest transcript?
- How long is the protein it encodes?
- Compare the sequence of the two longest protein-coding transcripts.

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



# GeneCards <http://www.genecards.org/>

Free for academic non-profit institutions. Other users need a commercial license

WEIZMANN INSTITUTE OF SCIENCE  LifeMap SCIENCES 

Keywords   Advanced

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

## EGFR Gene (Protein Coding)

Epidermal Growth Factor Receptor

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

 Proteins & Enzymes  
Antibodies Assays & Kits

 Proteins Antibodies Assays  
Genes shRNA Primers  
CRISPR

 Genes Peptides Proteins  
CRISPR

### Aliases for EGFR Gene

Epidermal Growth Factor Receptor <sup>2 3</sup>  
Receptor Tyrosine-Protein Kinase ErbB-1 <sup>3 4</sup>  
Erb-B2 Receptor Tyrosine Kinase 1 <sup>2 3</sup>  
Proto-Oncogene C-ErbB-1 <sup>3 4</sup>  
EC 2.7.10.1 <sup>4 63</sup>  
ERBB1 <sup>3 4</sup>  
ERBB <sup>3 4</sup>

Erythroblastic Leukemia Viral (V-Erb-B) Oncogene Homolog (Avian) <sup>2</sup>  
Avian Erythroblastic Leukemia Viral (V-Erb-B) Oncogene Homolog <sup>3</sup>  
Cell Proliferation-Inducing Protein 61 <sup>3</sup>  
Cell Growth Inhibiting Protein 40 <sup>3</sup>  
EC 2.7.10 <sup>63</sup>



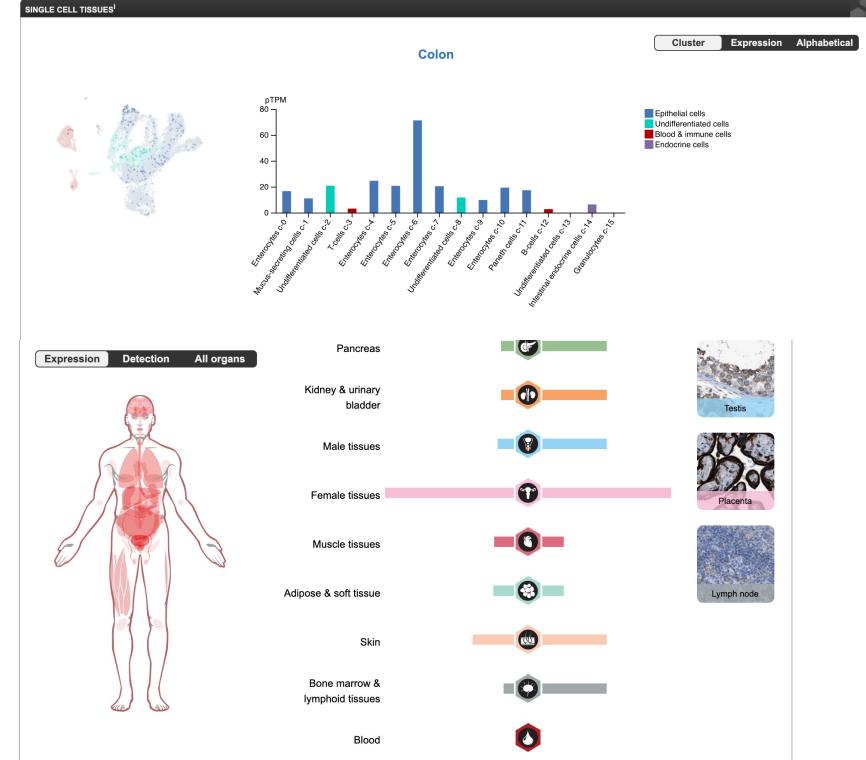
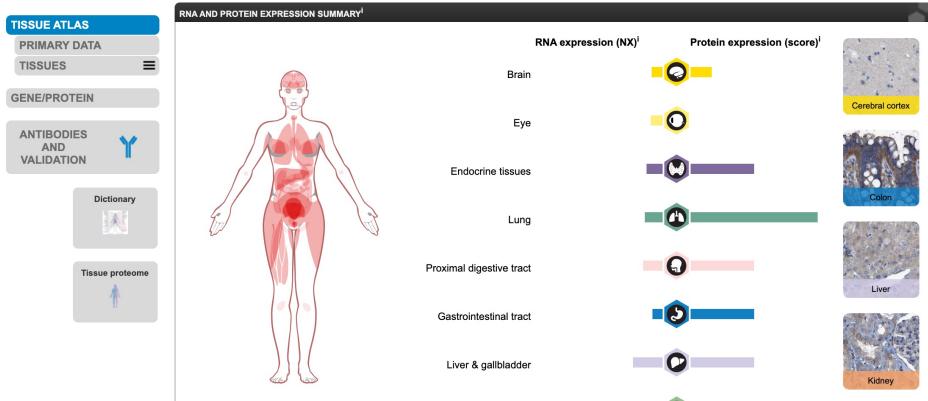
# Tumor protein atlas <http://www.proteinatlas.org/>

## EGFR



GENERAL INFORMATION <sup>1</sup>	
Gene name <sup>1</sup>	EGFR
Gene description <sup>1</sup>	Epidermal growth factor receptor
Protein class <sup>1</sup>	Cancer-related genes Disease related genes Enzymes FDA approved drug targets Plasma proteins RAS pathway related proteins
Predicted location <sup>1</sup>	Intracellular,Membrane,Secreted
Number of transcripts <sup>1</sup>	8

HUMAN PROTEIN ATLAS INFORMATION <sup>1</sup>	
RNA tissue specificity <sup>1</sup>	Tissue enhanced (placenta)
RNA tissue distribution <sup>1</sup>	Detected in many
Protein evidence <sup>1</sup>	Evidence at protein level
Protein expression <sup>1</sup>	Membranous and cytoplasmic expression in several tissues, most abundant in placenta.
IMMUNOHISTOCHEMISTRY DATA RELIABILITY	
Data reliability description <sup>1</sup>	Medium consistency between antibody staining and RNA expression data. At least one protein variant secreted, tissue location of RNA and protein might differ and correlation is complex.
Reliability score <sup>1</sup>	Enhanced
Antibodies <sup>1</sup>	<a href="#">HPA001200</a> , <a href="#">HPA018530</a> , <a href="#">CAB000035</a> , <a href="#">CAB068186</a> , <a href="#">CAB073534</a>
<a href="#">SHOW MORE</a>	



# Summary

Gene & Protein DBs

**Functional Enrichment**

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Gene Expression Signatures

**MSigDB**

<http://software.broadinstitute.org/gsea/msigdb/a>  
nnotate.jsp

**Enrichr**

<http://amp.pharm.mssm.edu/Enrichr/>

**Harmonizome**

<http://amp.pharm.mssm.edu/Harmonizome//>

**GeneMania**

<http://genemania.org/>

# Exercise

## Functional Enrichment

### Exercise: Using EnrichR

- Go to <https://maayanlab.cloud/Enrichr/>
- Expand into a gene set the term:  
**triple negative breast cancer**
- Submit the query
- Go to the Diseases/Drugs section
- Check the COVID-19 related gene sets 2021 collection

The screenshot shows the Enrichr web interface. At the top right, there are links for 'Login | Register' and statistics: '40,122,269 sets analyzed', '377,065 terms', and '190 libraries'. Below this is a navigation bar with tabs: 'Analyze' (which is active), 'What's new?', 'Libraries', 'Gene search', 'Term search', 'About', and 'Help'. The main area is titled 'Input data' and contains a text input field with 'EGFR' typed in. A red search icon is to the right of the input field. Below the input field, there is a placeholder text: 'Try an example [STAT3] [breast cancer] rs28897756'. A slider bar is present with the text 'Include the top 100 most relevant genes'. A note below the slider says 'EGFR is a gene'. Another note states 'Expanded EGFR with ARCHS4 RNA-seq gene-gene co-expression matrix by identifying the top 100 genes that mostly co-express with EGFR'. There is also a section for 'Expand gene with' containing 'ARCHS4 RNA-seq gene-gene co-expression matrix'. A note below it says 'Alternatively, try the Gene Search or the Term Search features to fetch annotated Enrichr gene sets.' At the bottom of the input area, there is a 'Submit' button. On the right side of the interface, there is a list of genes: ADCY6, AHNAK2, AMOTL2, ANTXR1, ARHGAP2, ARHGAP29, BCAR1, CAP21, CCND1. Below this list, it says '100 gene(s) entered'. At the very bottom, there is a note about acknowledging Enrichr in publications, listing references, and a copyright notice.

Input data

Expand a gene, a term, or a variant into a gene set:

EGFR

Try an example [\[STAT3\]](#) [\[breast cancer\]](#) [rs28897756](#)

Include the top 100 most relevant genes

EGFR is a gene

Expanded EGFR with ARCHS4 RNA-seq gene-gene co-expression matrix by identifying the top 100 genes that mostly co-express with EGFR

Expand gene with

ARCHS4 RNA-seq gene-gene co-expression matrix

Alternatively, try the Gene Search or the Term Search features to fetch annotated Enrichr gene sets.

Top 100 genes co-express with EGFR identified with ARCHS4 RNA-seq gene-gene co-expression

Contribute your set so it can be searched by others

Please acknowledge Enrichr in your publications by citing the following references:  
Chen BY, Tam CM, Kou Y, Duan Q, Wang Z, Mierelis GV, Clark NR, Maayan A.  
Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013; 12(14).  
Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A,  
McDermott MG, Monteiro CD, Gundersen GW, Maayan A.  
Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016; gkw377.  
Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E,  
Jagodnik KM, Jeon M, & Maayan A.  
Gene set knowledge discovery with Enrichr. *Current Protocols*, 1, e90. 2021. doi: 10.1002/cpz1.90

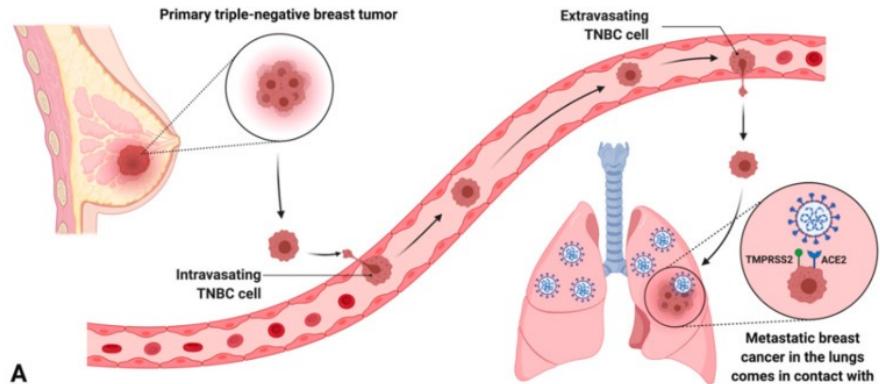
# Exercise

## Functional Enrichment

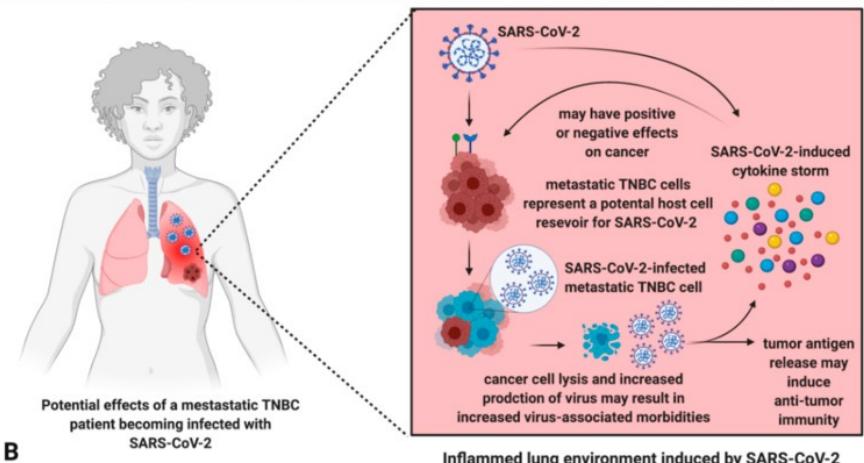
*Metastatic breast cancer cells may represent an additional host cell SARS-CoV-2 reservoir, leading to increased viral load and virus-associated morbidities, and unknown consequences on the progression of the cancer.*

Source:

Brown JM, Wasson MD, Marcato P. Triple-Negative Breast Cancer and the COVID-19 Pandemic: Clinical Management Perspectives and Potential Consequences of Infection. *Cancers*. 2021 Jan;13(2). DOI: 10.3390/cancers13020296. PMID: 33467411; PMCID: PMC7830590.



A



B

**MSigDB** <http://www.software.broadinstitute.org/gsea/msigdb>

The screenshot shows the MSigDB homepage. At the top, there's a navigation bar with links for "GSEA Home", "Downloads", "Molecular Signatures Database" (which is highlighted in blue), "Documentation", and "Contact". To the right of the navigation bar, there are "log in" and "register" buttons. On the far left, there's a sidebar with links to "MSigDB Home", "About Collections", "Browse Gene Sets", "Search Gene Sets", "Investigate Gene Sets", "View Gene Families", and "Help". The main content area features the GSEA logo at the top left and the MSigDB logo with the text "Molecular Signatures Database". Below the logos, there are two sections: "Overview" and "Collections". The "Overview" section contains text about the database and two bullet points: "Search for gene sets by keyword." and "Browse gene sets by name or collection.". The "Collections" section contains text about major collections and a callout box with a red letter "H" containing information about hallmark gene sets.

GSEA  
Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact

log in register

► MSigDB Home

► About Collections

► Browse Gene Sets

► Search Gene Sets

► Investigate Gene Sets

► View Gene Families

► Help

  **MSigDB**  
Molecular Signatures Database

## Molecular Signatures Database v6.2

**Overview**

The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets for use with GSEA software. From this web site, you can

- **Search** for gene sets by keyword.
- **Browse** gene sets by name or collection.

**Collections**

The MSigDB gene sets are divided into 8 major collections:

**H** **hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.

# Summary

Gene & Protein DBs

**Functional Enrichment**

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Pathways and Networks

**Networks: STRING**  
<http://string-db.org>

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

**WikiPathways**  
<https://www.wikipathways.org/index.php/WikiPathways>

**KEGG**  
<https://www.genome.jp/kegg/>

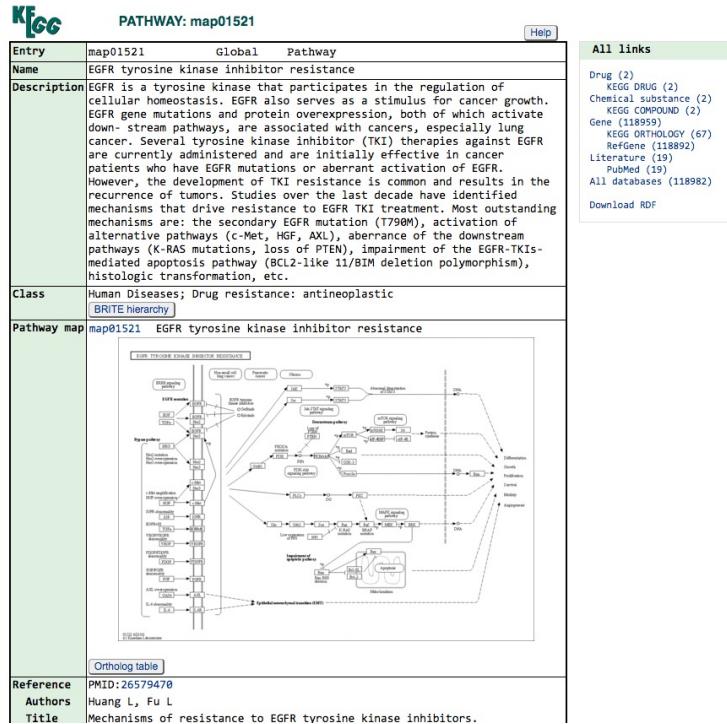
# Exercise

## Functional Enrichment

### Exercise: Using KEGG

- Go to <https://www.genome.jp/kegg/>
- Load a map of the EGFR tyrosine kinase inhibitor resistance pathway

What are the names of the chemical compounds that are related to the EGFR pathway according to KEGG ?



# Exercise

## Functional Enrichment

### Exercise: Using STRING

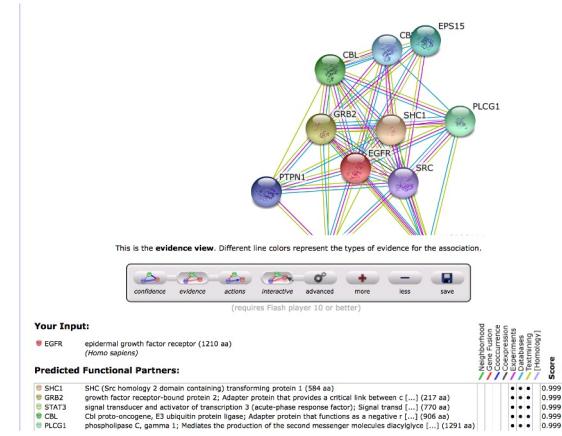
Go to the STRING [website](http://string-db.org) <http://string-db.org>

### How to find the interaction network of a protein ?

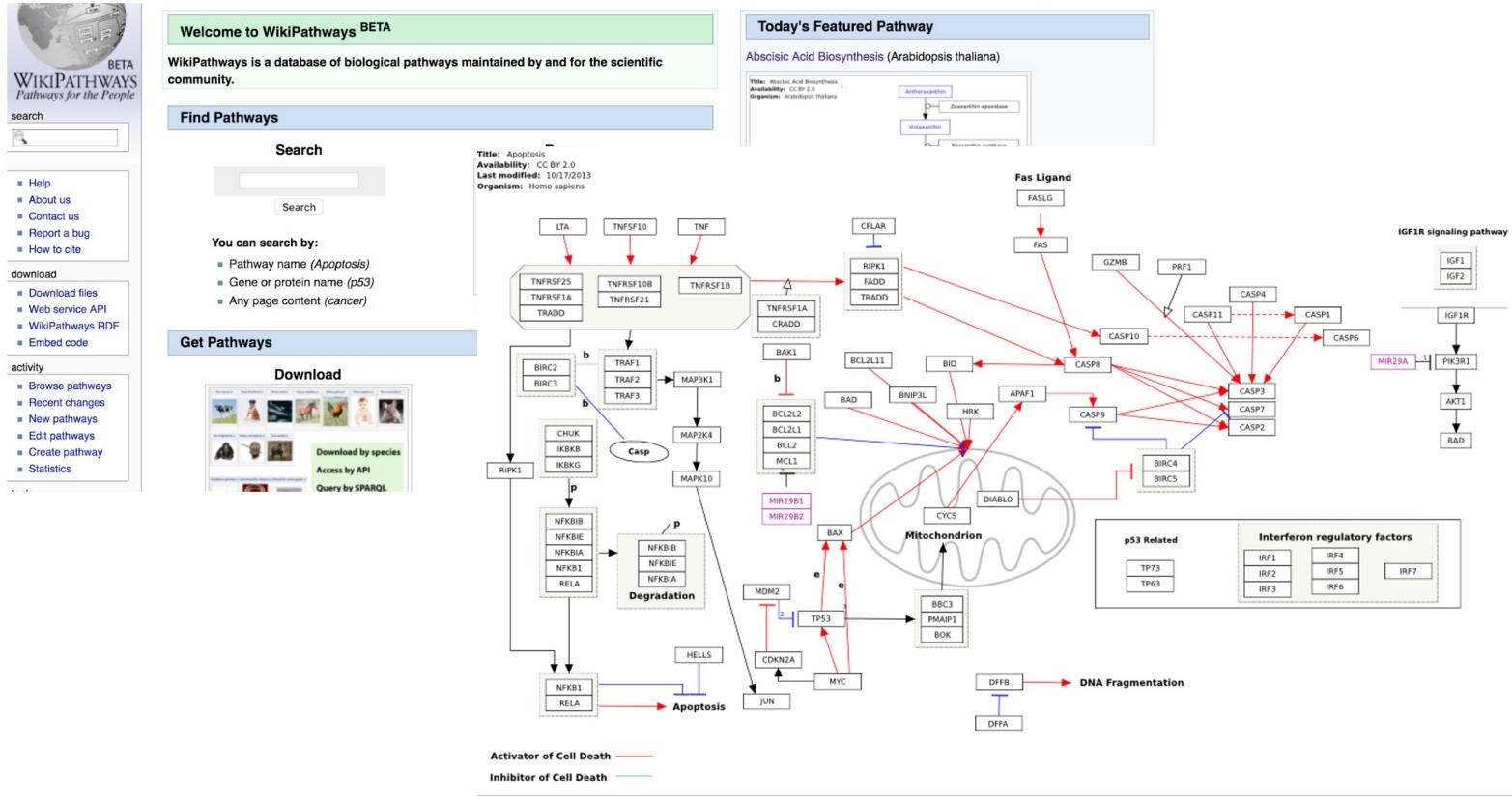
On the top of the results page, the interaction network is visualized.

- The network nodes are proteins.
- The edges represent the predicted functional associations.
- The color of the edges reflects the evidence:

- Red line - indicates the presence of fusion evidence
- Green line - neighborhood evidence
- Blue line - cooccurrence evidence
- Purple line - experimental evidence
- Yellow line - textmining evidence
- Light blue line - database evidence
- Black line - coexpression evidence.



# WikiPathways <https://www.wikipathways.org/index.php/WikiPathways>



# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Collections

NCI60

<https://discover.nci.nih.gov/cellminer/>

Cancer Cell Line Encyclopedia

<https://portals.broadinstitute.org/ccle>

International Cancer Genome Consortium

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

The Cancer Genome Atlas

<https://cancergenome.nih.gov/>

COSMIC

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

# Gene & Protein DBs

# Functional Enrichment

# Cancer Collections & Browsers

## Therapy response

# Cancer dependencies

# Survival, Clinical & other useful info

# Collections

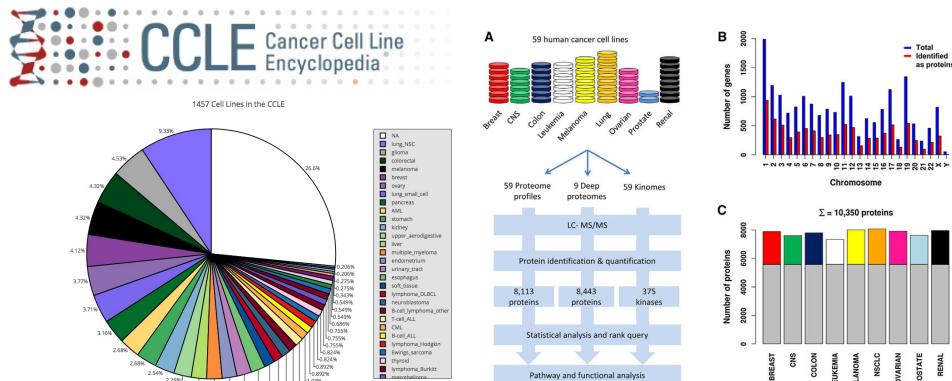
## Cell lines

NCI60

<https://discover.nci.nih.gov/cellminer/>

# Cancer Cell Line Encyclopedia

<https://portals.broadinstitute.org/ccle>



Amin Moghaddas Gholami, Hannes Hahne, Zhixiang Wu, Florian Johann Auer, Chen Meng, Mathias Wilhelm, Bernhard Kuster, Global Proteome Analysis of the NCI-60 Cell Line Panel, *Cell Reports*, Volume 4, Issue 3, 2013, Pages 609–620, ISSN 2211-2124 <https://doi.org/10.1016/j.celrep.2013.07.018>.

# Summary

Gene & Protein DBs

Functional Enrichment

**Cancer Collections & Browsers**

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Collections

NCI60

<https://discover.nci.nih.gov/cellminer/>

Cancer Cell Line Encyclopedia

<https://portals.broadinstitute.org/ccle>

Tumor samples

The Cancer Genome Atlas

<https://cancergenome.nih.gov/>

International Cancer Genome Consortium

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

# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Collections

NCI60

<https://discover.nci.nih.gov/cellline>

Cancer Cell Line Encyclopedia

<https://portals.broadinstitute.org>

Tumor samples

The Cancer Genome Atlas

<https://cancergenome.nih.gov/>



**THE CANCER GENOME ATLAS**  
National Cancer Institute  
National Human Genome Research Institute



# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Collections

NCI60

<https://discover.nci.nih.gov/cellminer/>

Cancer Cell Line Encyclopedia

<https://portals.broadinstitute.org/ccle>

Tumor samples



**ICGC** Data Portal

International Cancer Genome Consortium

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

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

# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Collections

NCI60

<https://discover.nci.nih.gov/cellminer/>

Cancer Cell Line Encyclopedia

<https://portals.broadinstitute.org/ccle>

Tumor samples



**ICGC** Data Portal

International Cancer Genome Consortium

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

A screenshot of the ICGC Data Portal's user interface. The top navigation bar includes 'DATA ANALYSIS' (selected), 'Cohort Comparison', 'Enrichment Analysis', and 'OncoGrid'. Below the navigation are four main sections: 'Enrichment Analysis' (with a 'DNA' icon), 'Cohort Comparison' (with a 'people' icon), 'Set Operations' (with a 'Venn diagram' icon), and 'OncoGrid' (with a 'grid' icon). Each section has a brief description and 'Select' or 'Demo' buttons.

# Summary

Gene & Protein DBs

Functional Enrichment

**Cancer Collections & Browsers**

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Collections

NCI60

<https://discover.nci.nih.gov/cellminer/>

Cancer Cell Line Encyclopedia

<https://portals.broadinstitute.org/ccle>

International Cancer Genome Consortium

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

The Cancer Genome Atlas

<https://cancergenome.nih.gov/>

Both

COSMIC

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

# Summary

Gene & Protein DBs

Functional Enrichment

**Cancer Collections & Browsers**

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Browsers

**cBio Portal for Cancer genomics**

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

**Integrative Onco Genomics**

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

**Oncomine**

<https://www.oncomine.org/resource/login.html>

**TumorPortal**

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

**XenaBrowser**

<https://xena.ucsc.edu/>

# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Therapy response

GDSC

<http://www.cancerrxgene.org>

Cancer Cell Line Encyclopedia

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

Cancer Therapeutics Response Portal

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

Open Targets

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

Connectivity Map

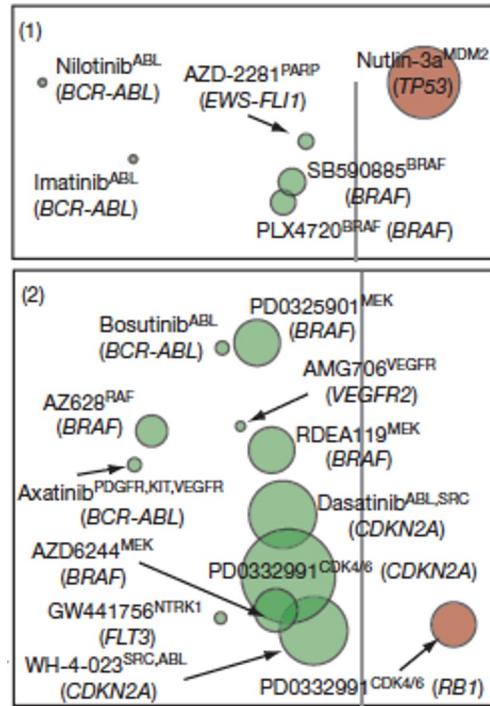
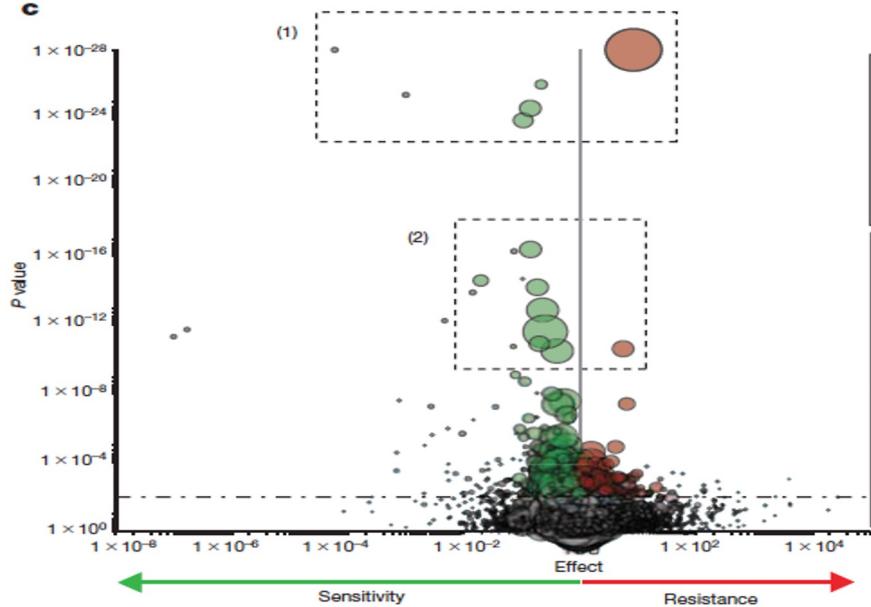
<https://clue.io/>

PanDrugs

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

## Systematic identification of genomic markers of drug sensitivity in cancer cells

Mathew J. Garnett<sup>1\*</sup>, Elena J. Edelman<sup>2\*</sup>, Sonja J. Heidorn<sup>1\*</sup>, Chris D. Greenman<sup>1†</sup>, Anahita Dastur<sup>2</sup>, King Wai Lau<sup>1</sup>, Patricia Gréninger<sup>2</sup>, Richard Thompson<sup>1</sup>, Yi Huo<sup>2</sup>, Jorge Soares<sup>1</sup>, Giuseppe Llin<sup>3,4</sup>, Francesco Iorio<sup>1,5</sup>, Didier Strelak<sup>6</sup>, Li Chen<sup>2</sup>, Randy J. Milano<sup>1</sup>, Fiona Kogera<sup>1,8</sup>, Wenjun Zhou<sup>3,4</sup>, Wanjuan Yang<sup>1</sup>, Jeffrey A. Engel<sup>1</sup>, P. Andrew Futreal<sup>1</sup>



Garnett et al. Nature 2012

# Exercise

## Therapy response

### Exercise: Using GDSC

- Go to the GDSC [website](#)
- Look for the EGFR gene

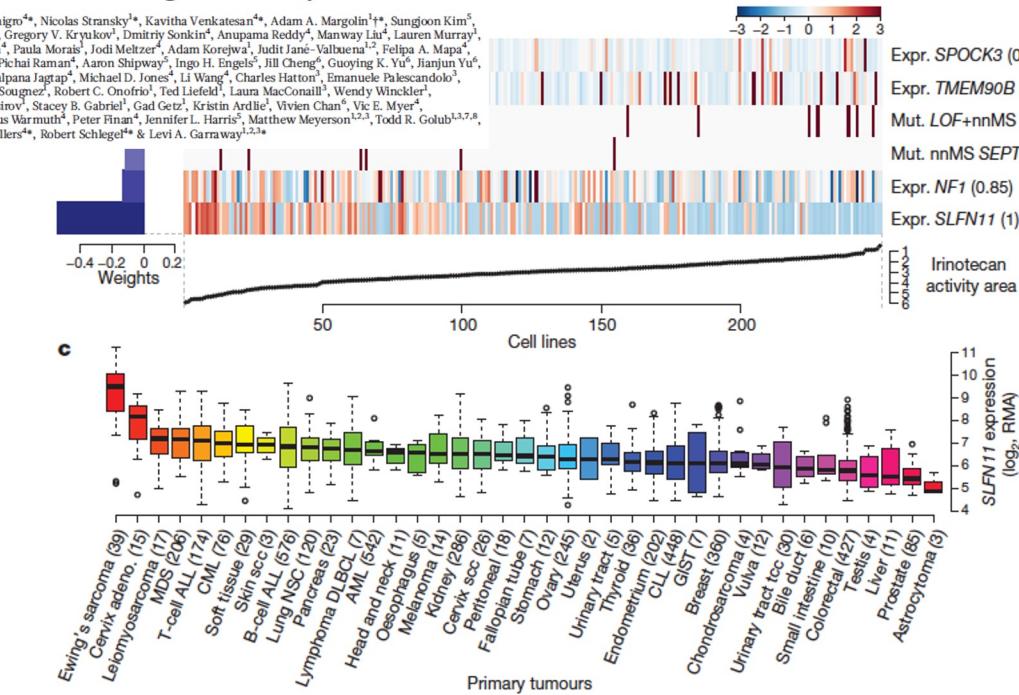
**To which drugs are sensitive the EGFR mutated cell lines?**

- Choose a compound of interest
- Look for other sensitivity/resistance associations related to that same compound

# Cancer cell line encyclopedia <https://www.broadinstitute.org/ccle>

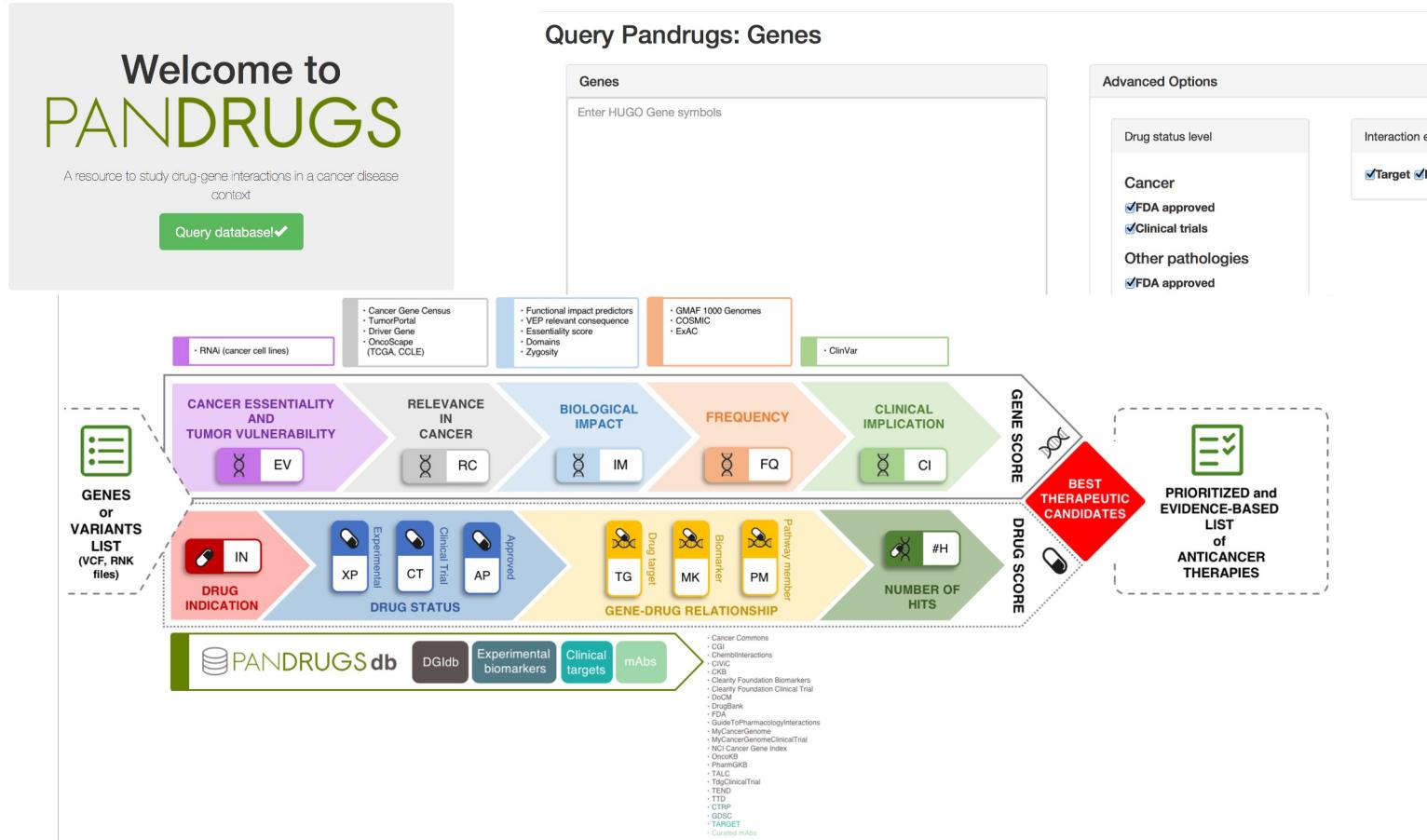
## The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity

Iordi Barretina<sup>1,2,3,4\*</sup>, Giordano Caponigro<sup>4\*</sup>, Nicolas Stransky<sup>4\*</sup>, Kavitha Venkatesan<sup>4\*</sup>, Adam A. Margolin<sup>1+\*</sup>, Sungjoon Kim<sup>5</sup>, Christopher J. Wilson<sup>6</sup>, Joseph Lehár<sup>4</sup>, Gregory V. Kryukov<sup>6</sup>, Dmitry Sorkin<sup>6</sup>, Anupama Reddy<sup>4</sup>, Manway Liu<sup>4</sup>, Lauren Murray<sup>4</sup>, Michael F. Berger<sup>4</sup>, John E. Monahan<sup>4</sup>, Paula Morais<sup>4</sup>, Jodi Meltzer<sup>4</sup>, Adam Korejwa<sup>4</sup>, Judit Jane-Valbuena<sup>1,2</sup>, Felipa A. Mapa<sup>4</sup>, Joseph Thibault<sup>2</sup>, Eva Bric-Purlong<sup>4</sup>, Pichai Raman<sup>4</sup>, Aaron Shipway<sup>2</sup>, Ingo H. Engels<sup>2</sup>, Jill Cheng<sup>4</sup>, Guoying K. Yu<sup>6</sup>, Jianjun Yu<sup>6</sup>, Peter Aspesi<sup>1,7</sup>, Melanie de Silva<sup>1</sup>, Kalpana Jagtap<sup>4</sup>, Michael D. Jones<sup>4</sup>, Li Wang<sup>4</sup>, Charles Hatton<sup>4</sup>, Emanuele Palascandolo<sup>3</sup>, Supriya Gupta<sup>1</sup>, Scott Mahan<sup>4</sup>, Carrie Sougner<sup>4</sup>, Robert C. Onofrio<sup>4</sup>, Ted Liefeld<sup>4</sup>, Laura MacConell<sup>3</sup>, Wendy Winkler<sup>4</sup>, Michael Reich<sup>1</sup>, Nannxin Li<sup>8</sup>, Jill P. Mesirov<sup>4</sup>, Stacey B. Gabriel<sup>4</sup>, Gad Getz<sup>4</sup>, Kristin Ardile<sup>4</sup>, Vivien Chan<sup>6</sup>, Vick E. Myer<sup>4</sup>, Barbara L. Weber<sup>4</sup>, Jeff Porter<sup>4</sup>, Markus Warmuth<sup>4</sup>, Peter Finan<sup>4</sup>, Jennifer L. Harris<sup>4</sup>, Matthew Meyerson<sup>1,2,4</sup>, Todd R. Golub<sup>1,3,7,8</sup>, Michael P. Morrissey<sup>4</sup>, William R. Sellers<sup>4</sup>, Robert Schlegel<sup>4</sup> & Levi A. Garraway<sup>1,2,3</sup>



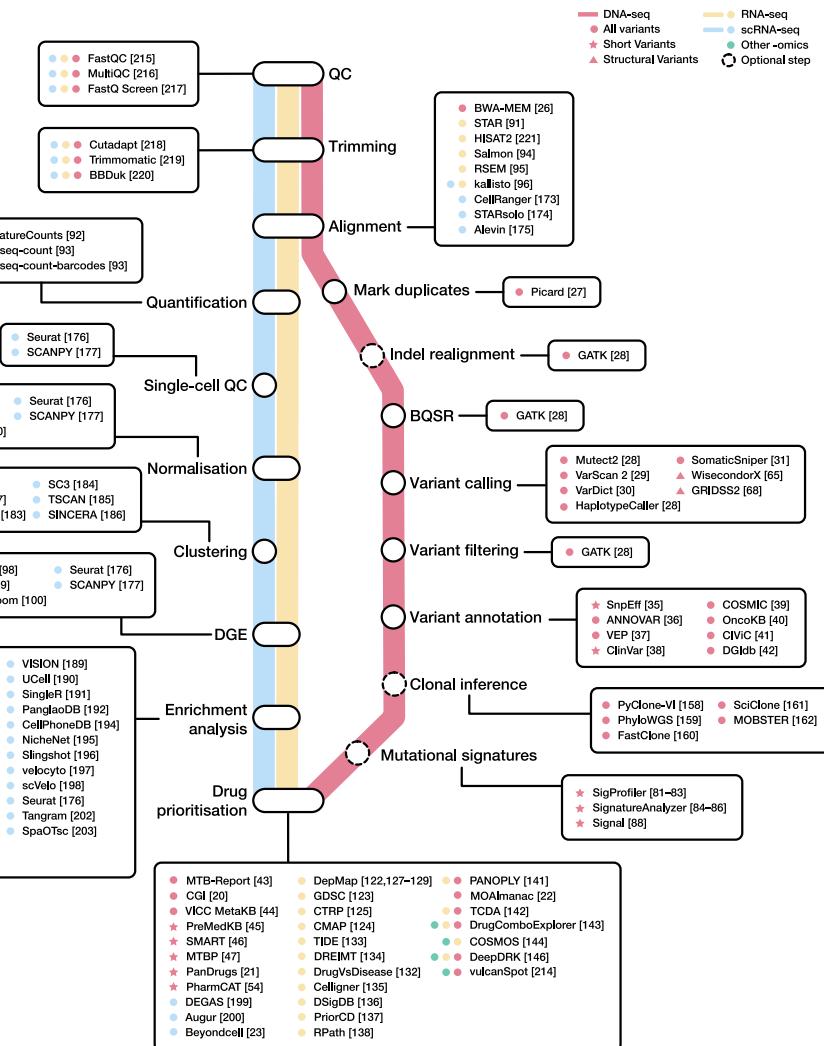
Barretina et al. Nature 2012

# PanDrugs <https://pandrugs.bioinfo.cnio.es>



**Table 1: Bioinformatics tools for genomics-based drug prioritisation.**

Name	Description	Input	Output	URL
MTB-Report [43]	R script that filters and classifies cancer variants into levels of evidence using gene-drug databases.	Tables with SNVs, CNVs and gene fusions (somatic).	Molecular Tumour Board (MTB) report with actionable variants in PDF. Downloadable tables of a) Annotated variants, including information about the oncogenicity and biological consequence, and b) drug-target associations with evidence level and response prediction.	<a href="https://github.com/iperera-hel/MTB-Report">https://github.com/iperera-hel/MTB-Report</a>
Cancer Genome Interpreter (CGI) [20]	Web tool that annotates cancer variants and identifies potential oncogenic alterations and genomic biomarkers of drug response.	List of SNVs, indels, CNVs and/or gene fusions (somatic).	Annotated variants, including information about the oncogenicity and biological consequence, and b) drug-target associations with evidence level and response prediction.	<a href="https://www.cancergenomeinterpreter.org/">https://www.cancergenomeinterpreter.org/</a>
VICC MetaKB [44]	Web tool for cancer variant interpretation that harmonises 6 different variant annotation knowledgebases with information about variant, gene, disease and drug associations and their corresponding evidence levels.	List of variants (somatic), including gene fusions, genes, diseases and/or drugs.	Interactive report with variant-gene-disease-drug associations, each one with its evidence label and supporting links.	<a href="https://search.cancervariants.page">https://search.cancervariants.page</a>
PreMedKB [45]	Web tool for integrating information on diseases, genes, variants, drugs, and the relationships between any two or more of these four components.	List of short variants (somatic), genes, drugs and/or diseases.	Interactive semantic network displaying components as nodes and their relationships as edges. Results can be downloaded in either JSON or PNG format.	<a href="http://www.fudap.gx.cn/premedkb/index.html#home">http://www.fudap.gx.cn/premedkb/index.html#home</a>
SMART Cancer Navigator [46]	Web application for variant interpretation that associates the corresponding genes to diseases, known drugs and relevant clinical trials.	List of short variants (somatic and germline).	Interactive report with variant, gene, disease and drug information.	<a href="https://smart-cancer.navigator.github.io/home">https://smart-cancer.navigator.github.io/home</a>
PanDrugs [21]	Web tool to prioritise anticancer drug treatments according to individual genomic data. PanDrugs computes two scores, the GScore (GScore) and the DScore (DScore). The GScore ranges from 0 to 1 and is estimated according to gene essentiality and tumour vulnerability, gene relevance in cancer, the biological impact of mutations, the frequency of gene alterations and their clinical implications. The DScore ranges from -1 to 1, considers drug indication and status, gene-drug associations and number of hits and estimates resistance (negative values) or sensitivity (positive values).	VCF, a list or a ranking of genes or a drug query (somatic).	Report with a prioritised list of anticancer therapies. PanDrugs resolves the Best Therapeutic Candidates based on the accumulated and weighted scoring of the GScore and the DScore.	<a href="https://www.pandrugs.org/#/">https://www.pandrugs.org/#/</a>
MTBP [47]	Web tool that annotates somatic and germline short variants (SNVs and indels) functionally and clinically, categorising the cancer biomarkers (diagnosis, prognosis and drug response) found in the tumour.	VCF or a list of short variants (somatic and germline).	HTML report with annotated variants, the evidence supporting the variants' functional classification and their associated actionability.	<a href="https://mtbp.org/">https://mtbp.org/</a>
PharmCAT [54]	A tool for identifying germinal variants, inferring patient's haplotypes and diplotypes and suggesting treatments following the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines.	VCF (germline).	HTML/JSON report with drug suggestions based on germline variants.	<a href="https://pharmcat.org/">https://pharmcat.org/</a>



## The bioinformatics roadmap for therapy selection in cancer genomics

Jiménez-Santos et al. 2022 coming soon

# Connectivity Map <https://clue.io>



Unravel biology with the world's largest perturbation-driven gene expression dataset.

> TYPE COMPOUND, GENE, MoA, OR PERTURBAGEN CLASS TO SEE OVERVIEW  
> TYPE A SLASH CHARACTER "/" TO SEE LIST OF COMMANDS

DATA VERSION: 1.1.1.2 / SOFTWARE VERSION: 1.1.1.36

## CONNECTIVITY MAP LAUNCHES THIRD CROWDSOURCED CONTEST

The Connectivity Map team at the Broad Institute is happy to announce its latest crowdsourced contest, launched in collaboration with the Laboratory for Innovation challenge is focused on enhancing the CMap 1,000 in total prizes available. Register today

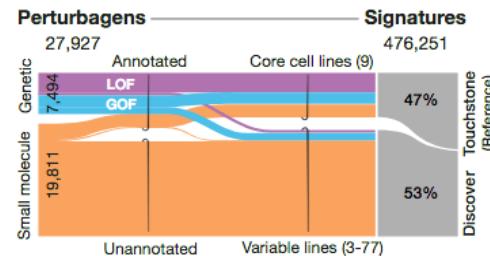
The aim is to generate perturbational profiles across multiple cell and perturbation types

- Small Molecules
  - 1300 FDA approved drugs
  - 5500 bioactive compounds
- Gene Knock-Outs and Over Expression

## Data and Tools

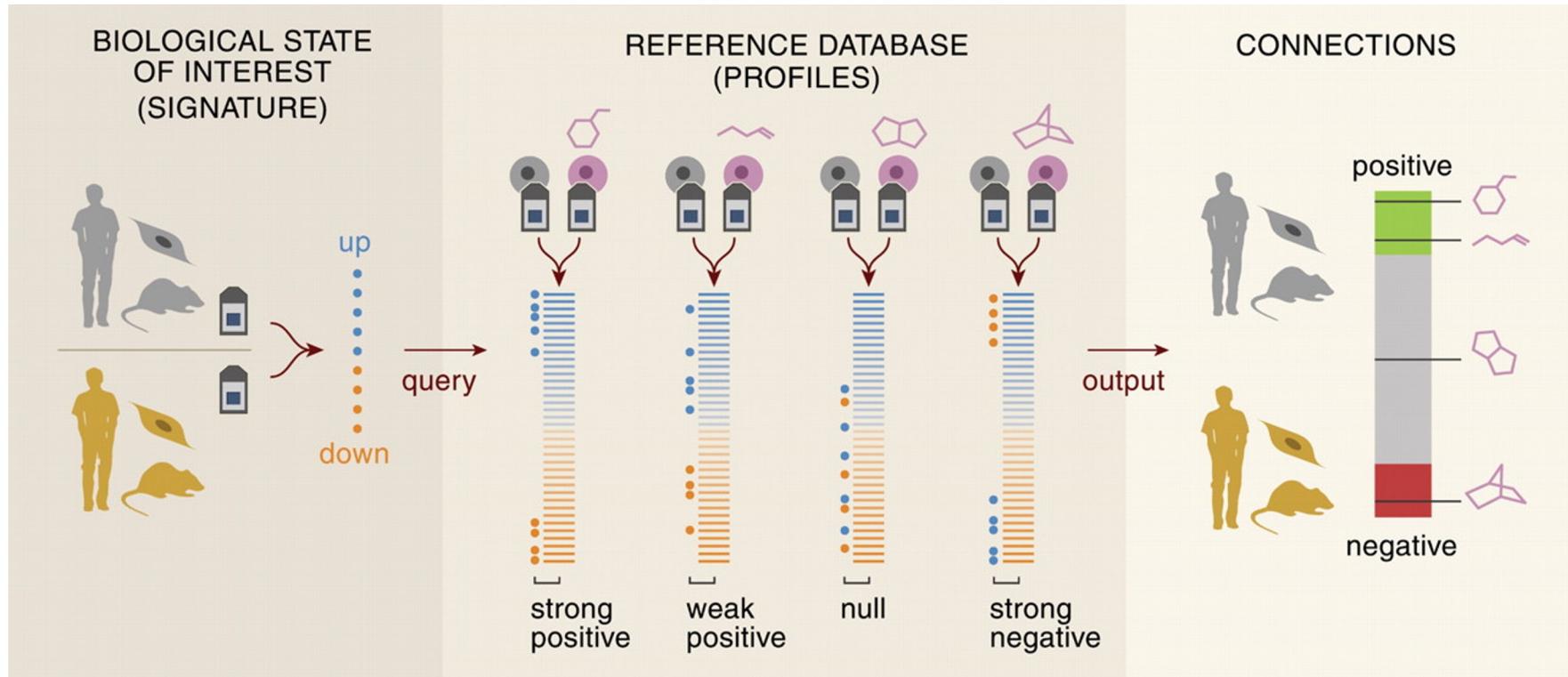
The CMap dataset of cellular signatures catalogs transcriptional responses of human cells to chemical and genetic perturbation. Here you can find the 1.3M L1000 profiles and the tools for their analysis.

A total of 27,927 perturbagens have been profiled to produce 476,251 expression signatures. About half of those signatures make up the Touchstone (reference) dataset generated from testing well-annotated genetic and small-molecular perturbagens in a core panel of cell lines. The remainder make up the Discover dataset, generated from profiling uncharacterized small molecules in a variable number of cell lines.



Start exploring the data by using the text-box on this page to look up perturbagens of interest in Touchstone. To see the suite of tools, including apps to query your gene expression signatures and analyze resulting connections, click on Tools in the menu bar.

# Connectivity Map <https://clue.io>



# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

**Cancer dependencies**

Survival, Clinical & other useful info

## Cancer Dependencies

**DepMap**

<https://depmap.org/portal/>

**CancerGD**

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

**vulcanSpot**

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

# Cancer dependency map <https://www.depmap.org/portal/>

The goal of the Cancer Dependency Map is to create a comprehensive preclinical reference map connecting tumor features with tumor dependencies to accelerate the development of precision treatments. Our strategy is to systematically characterize cellular models of cancers and to test those models for sensitivity to genetic and small-molecule perturbations. By integrating data beyond those collected at the Broad, DepMap hopes to develop a complete understanding of the vulnerabilities of cancer, identify targets for therapeutic development, and design strategies to optimize patient responses to those therapies.

Genetic screens      Cellular models      Drug sensitivity  
Predictive modeling

**CANCER DEPENDENCY MAP**

Genetic targets   Therapeutic leads   Patient stratification

Enter a gene, cell line, lineage or compound

Tools   Data   About

**Look up dependencies**

Enter a gene, cell line, lineage or compound

**Use this portal to:**

- UNDERSTAND** Dependency profiles at genome-scale across more than 500 human cell lines
- FIND** Detailed genetic and pharmacologic characterization of over 1000 cell lines
- IDENTIFY** Genetic drivers that have functional importance as potential drug targets
- SEARCH** For cell line models that best represent your research interests
- EXPORT** Presentation-quality figures

**CELL LINES PROFILED**

Screen Type	Completed	1 year goal	5 year goal
CRISPR screen	~500	~2,000	~10,000
Drug screen	~500	~1,500	~10,000

Data Release: **EVERY 90 DAYS**

Completed   1 year goal   5 year goal

## CancerGD: analysis of genetic dependencies in cancer

Search filter: Driver gene: ERBB2 Tissue type: Pan cancer Study: All studies Search

**Driver gene:** ERBB2    Synonyms: HER-2 | HER2 | NEU | CD340 | NGL

Gene alteration considered: Amplifications

Gene Description: erb-b2 receptor tyrosine kinase 2

External links: GeneCards | Entrez | Ensembl | OMIM | CancerRxGene | cBioPortal | COSMIC | CanSAR | UniProtKB | GenomeRNAi | Open Targets

For driver gene **ERBB2**, a total of **1990 dependencies** were found in tissue type **Pan cancer** in **All studies**

( Use scrollbar at right of this table to scroll down. Click column header to sort by that column. Click on the gene name in the dependency column to view the box-plot. Enter text into the search box at top of column to optionally filter these results. In the 'Effect size' column search box you can enter eg: ">75" to filter results.)

for 1990 rows (max: 300)

Dependency	P-value	Effect size (%)	ΔScore	Study	Experiment Type	Multiple Hit	String Interaction	Inhibitors
Search	<0.05	>= 65.0	< 0.0					
MTOR	1 x 10 <sup>-5</sup>	92.5	-1.60	Cowley(2014)	shRNA	Yes	High	GDC-098..[more]
PSMC2	1 x 10 <sup>-5</sup>	95.7	-2.45	Marcotte(2012)	shRNA			BORTEZO..[more]
ERBB2	4 x 10 <sup>-5</sup>	87.6	-1.78	Campbell(2016)	siRNA	Yes	Highest	AEE 788..[more]
HIST1H2AK	4 x 10 <sup>-5</sup>	89.3	-0.73	Cowley(2014)	shRNA			

# vulcanspot <https://www.vulcanspot.org>

 vulcanSpot - A method for detecting and targeting cancer genetic dependencies.

genes A  
 ERBB2  PIK3CA 

Add genes A

EXAMPLE 1   EXAMPLE 2

Filters

Download: [jJSON](#)

GENE A ⓘ				GENE B ⓘ			DRUG ⓘ			
Gene Symbol ⓘ	Gene Alteration ⓘ	Context ⓘ	Druggable Gene ⓘ	Gene Symbol ⓘ	Role in Cancer ⓘ	GD Evidence (score) ⓘ	Drug ⓘ	PanDrugs Score ⓘ	LINCS+PPI Score ⓘ	Best Result ⓘ
<a href="#">ERBB2</a>		BREAST	 Yes	<a href="#">CDK4</a>		 CRISPR	<a href="#">palbociclib</a>	1.000	0.995	
<a href="#">PIK3CA</a>		BREAST	 Yes	<a href="#">CDK4</a>		 CRISPR	<a href="#">palbociclib</a>	1.000	0.995	
<a href="#">ERBB2</a>		BREAST	 Yes	<a href="#">ESR1</a>	 	 CRISPR	<a href="#">tamoxifen</a>	1.000	0.991	
<a href="#">PIK3CA</a>		BREAST	 Yes	<a href="#">ESR1</a>	 	 CRISPR	<a href="#">tamoxifen</a>	1.000	0.991	
<a href="#">ERBB2</a>		BREAST	 Yes	<a href="#">AKT1</a>	 	 CRISPR	<a href="#">triciribine</a>	0.600	0.997	
<a href="#">PIK3CA</a>		BREAST	 Yes	<a href="#">AKT1</a>	 	 CRISPR	<a href="#">triciribine</a>	0.600	0.997	
<a href="#">ERBB2</a>		BREAST	 Yes	<a href="#">AKT1</a>	 	 CRISPR	<a href="#">mk-2206</a>	0.600	0.989	
<a href="#">PIK3CA</a>		BREAST	 Yes	<a href="#">AKT1</a>	 	 CRISPR	<a href="#">mk-2206</a>	0.600	0.989	
<a href="#">PIK3CA</a>		BREAST	 Yes	<a href="#">KIF11</a>		 CRISPR	<a href="#">ispinesib</a>	0.600	0.986	
<a href="#">ERBB2</a>		BREAST	 Yes	<a href="#">ERBB2</a>	 	 CRISPR	<a href="#">cr-724714</a>	0.600	0.983	

Rows per page: 10 ▾ 1-10 of 3676 < < >

# vulcanspot <https://www.vulcanspot.org>

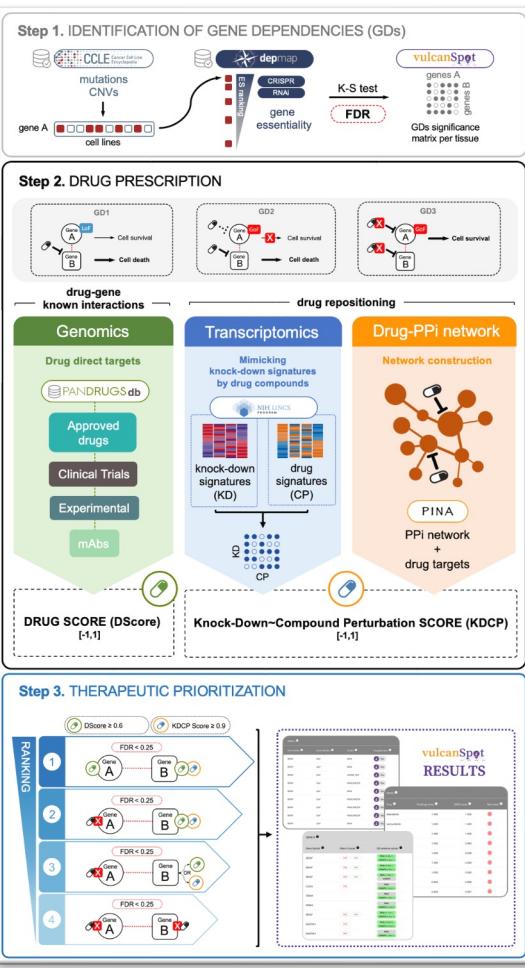
## vulcanSpot - A method for detecting and targeting cancer genetic dependencies.

genes A  
ERBB2 X PIK3CA X

Add genes A

Filters

GENE A ⓘ	Gene Symbol ⓘ	Gene Alteration ⓘ	Context ⓘ	Druggable Gene ⓘ	GENE B ⓘ	Gene Symbol ⓘ	Role in Cancer ⓘ	GD Evidence (score) ⓘ	
ERBB2	GoP	BREAST	?	Yes	CDK4	ONC	RNAi (2.72e-1)	CRISPR	
PIK3CA	GoP	BREAST	?	Yes	CDK4	ONC	RNAi (3.17e-1)	CRISPR	
ERBB2	GoP	BREAST	?	Yes	ESR1	ONC/TSG	HCD	RNAi (1.27e-1)	CRISPR
PIK3CA	GoP	BREAST	?	Yes	ESR1	ONC/TSG	HCD	RNAi (4.25e-1)	CRISPR
ERBB2	GoP	BREAST	?	Yes	AKT1	ONC	HCD	RNAi (1.62e-1)	CRISPR
PIK3CA	GoP	BREAST	?	Yes	AKT1	ONC	HCD	RNAi (1.53e-1)	CRISPR
ERBB2	GoP	BREAST	?	Yes	AKT1	ONC	HCD	RNAi (1.62e-1)	CRISPR
PIK3CA	GoP	BREAST	?	Yes	KIF11	ONC	HCD	RNAi (1.73e-1)	CRISPR
PIK3CA	GoP	BREAST	?	Yes	ERBB2	ONC	HCD	RNAi (7.29e-1)	CRISPR



# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## VCF annotators

### Variant Effect Predictor

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

### SnpEff

<http://snpeff.sourceforge.net/>

# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Clinical Variants Annotators

**ClinVAR**

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

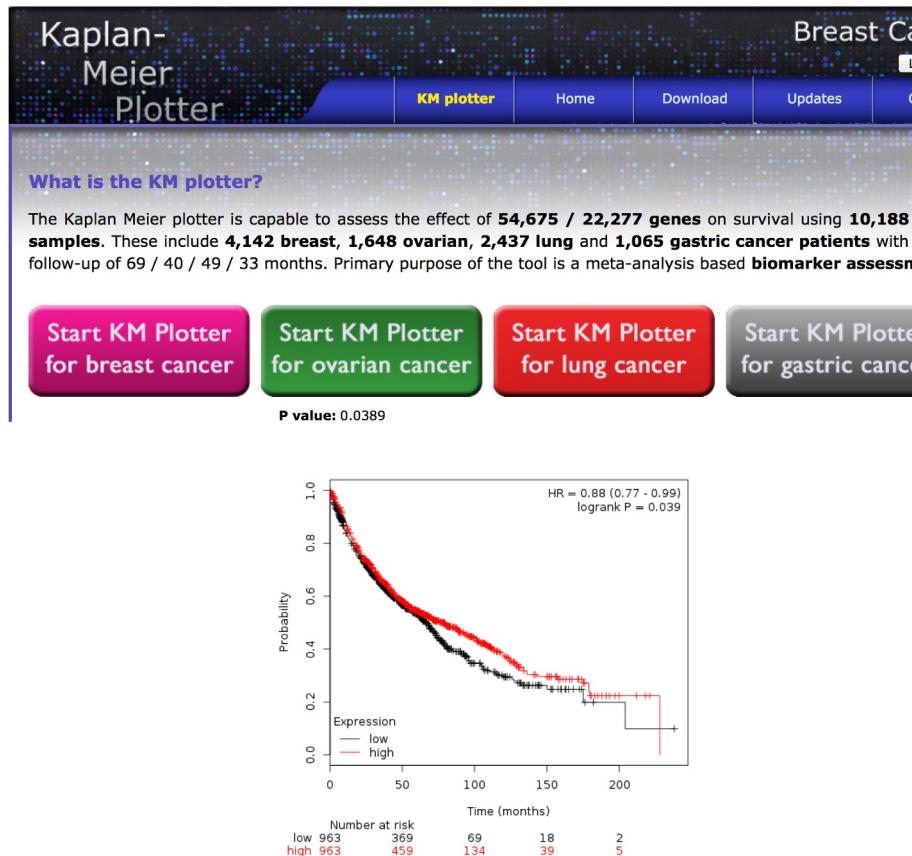
**ExAC**

<http://exac.broadinstitute.org>

**gnomAD**

<http://gnomAD.broadinstitute.org>

# Survival Kaplan meier plotter <http://kmplot.com/analysis/index.php>



Kaplan-Meier Plotter Lung Cancer Lung

Affy id/Gene symbol: 204009\_s\_at Use multigene classifier

Survival: OS (n=1928) Auto select best cutoff:  Censor at threshold:

Split patients by: median Auto select best cutoff:  Censor at threshold:

Follow up threshold: all Censore at threshold:

Compute median over entire dataset:

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

Probe set options

user selected probe set  
all probe sets per gene   
only JetSet best probe set

Plot beeswarm graph of probe distribution:

Restrict analysis to subtypes...

Histology: all Grade: all Stage: all AJCC stage T: all AJCC stage N: all AJCC stage M: all Gender: all Smoking history: all

Include in multivariate:

Restrict analysis to treatment groups...

Surgery success: all Chemotherapy: all Radiotherapy: all

Use selected cohort:

Dataset: all

Cox regression:  univariate  multivariate

Use earlier release of the database: 2015 version (n= exclude biased arrays (n=2435))

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: Gyorffy 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.

# Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

## Survival Analysis

<https://www.tcgasurvival.com/> Cell Reports 2022

<https://bbisr.shinyapps.winship.emory.edu/SurvivalGenie/>

Survival: Kaplan Meier Plotter  
<http://kmplot.com/analysis/index.php>

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

Other



# signal

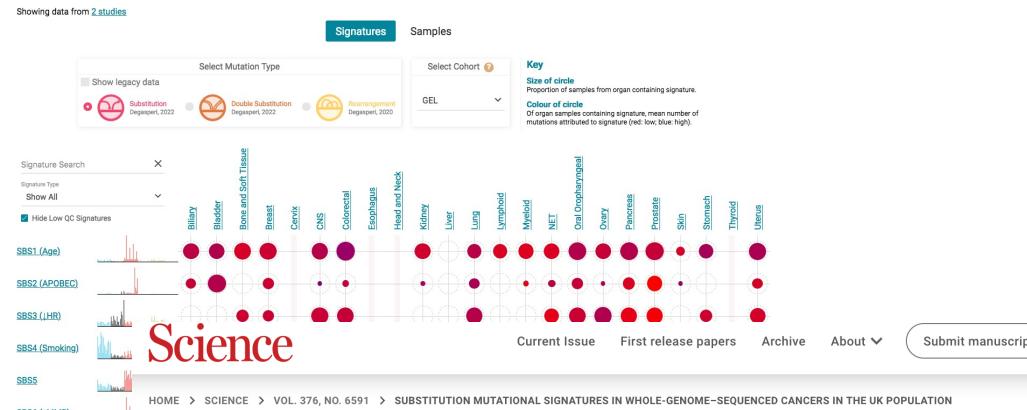
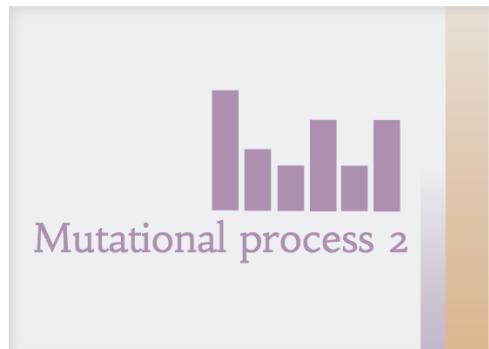
## Mutational signatures

<https://signal.mutationalsignatures.com/>

April 2022 Update — Signal presents newly published single and double base substitution signature data derived from more than 12,000 cancer samples recruited via the UK National Health Service. [Click here to view all of the cancer signature data on Signal.](#)

## Discover Mutational Signatures

Somatic mutations arise in the DNA of all the cells of the body. They are the consequence of mutational processes that may be induced by cellular activities, and are a key component in understanding cancer. Different mutational processes generate observable patterns. These patterns are called mutational signatures.



Science

HOME > SCIENCE > VOL. 376, NO. 6591 > SUBSTITUTION MUTATIONAL SIGNATURES IN WHOLE-GENOME-SEQUENCED CANCERS IN THE UK POPULATION

RESEARCH ARTICLE | CANCER GENOMICS

## Substitution mutational signatures in whole-genome-sequenced cancers in the UK population

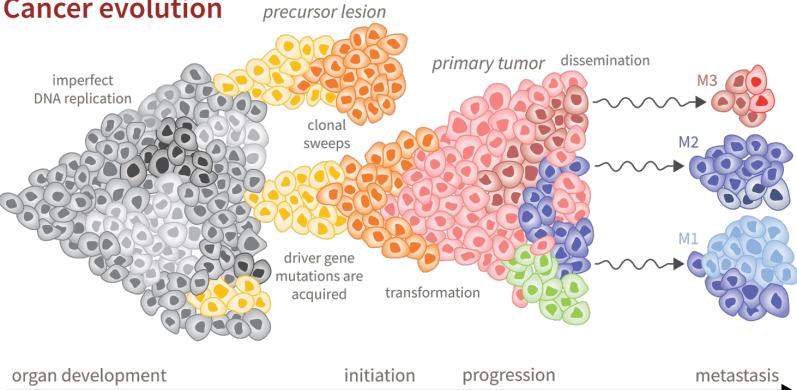
ANDREA DEGASPERI , XUEQING ZOU , TAUJANNE DIAS AMARANTE , ANDREA MARTINEZ-MARTINEZ , GENE CHING CHIEK KOH , JOÃO M. L. DIAS ,

Laura Heskin , Lucia Chmelenova , Giuseppe Rinaldi , ... SERENA NIK-ZAINAL , +14 authors [Authors Info & Affiliations](#)

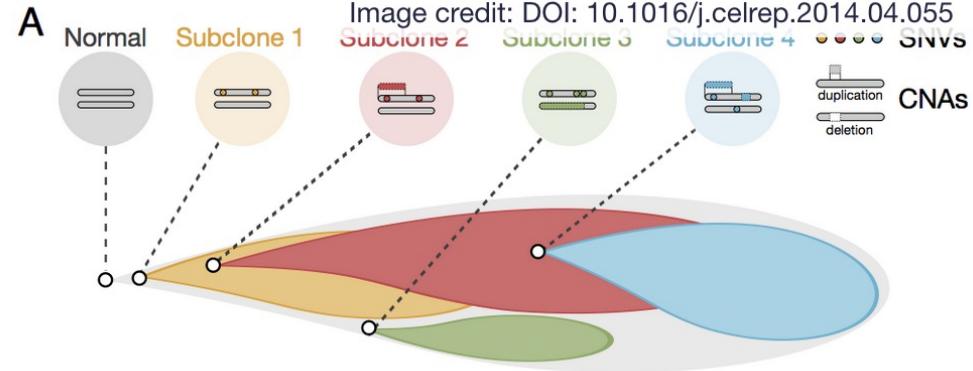
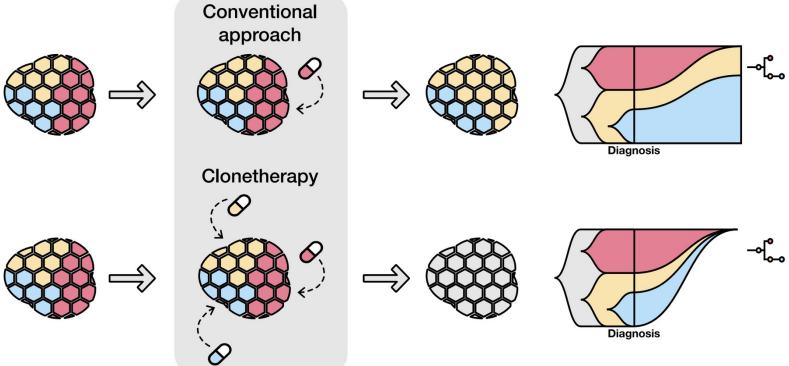
# Tumour clonality

<http://tracerx.co.uk/>

## Cancer evolution

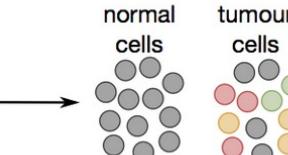


## Drug Prioritisation

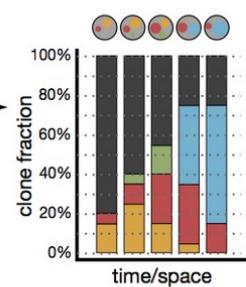


**B**

## Bulk DNA sequencing in time or space



## Clonal dynamics



Read mapping  
and variant calling



PyClone2, FastClone, Subclone, etc

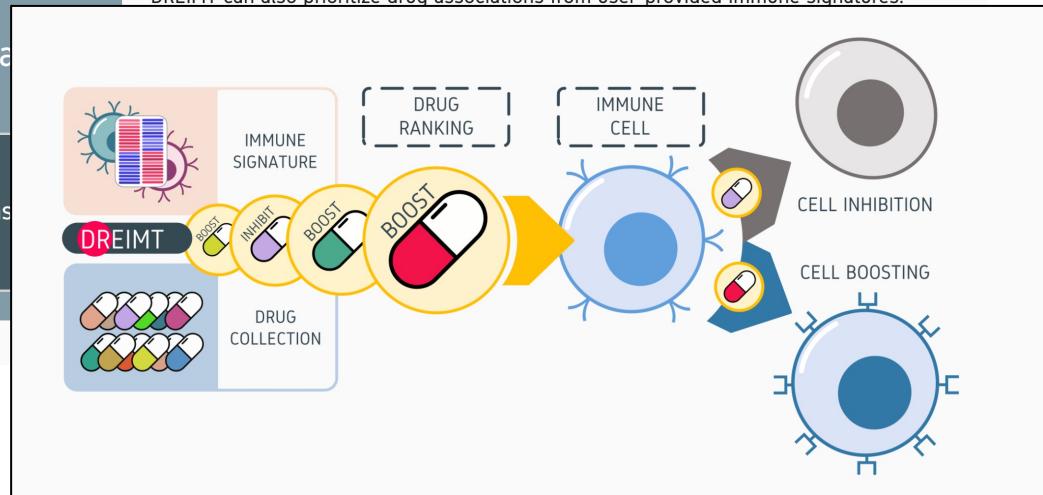
DREIMT

Drug prioritization   Signature comparison   Database   Help   REST API   Query history (0)

## Welcome to DREIMT

A tool for immune modulation drug prioritization

Drug prioritization   Signature comparison   Database



# Extra

## Downloading the reference genome

- **Gencode:** <https://www.gencodegenes.org/>
- **UCSC:** <https://hgdownload.soe.ucsc.edu/downloads.html>

## Downloading published experiments

- **GEO:** <https://www.ncbi.nlm.nih.gov/geo/>
- **SRA:** <https://www.ncbi.nlm.nih.gov/sra>
- **Single Cell Portal:** [https://portals.broadinstitute.org/single\\_cell](https://portals.broadinstitute.org/single_cell)



# Thanks!

