



Medical and genomic tools and databases

Precision Oncology Course



CNIO BIOINFORMATICS UNIT

Coral Fustero Torre
Bioinformatics Unit,
Structural Biology Programme.
cfustero@cnio.es | bioinformatics.cnio.es

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. A search for 'EGFR' has been performed. The results page displays detailed information about the EGFR gene, including its official symbol (EGFR), full name (epidermal growth factor receptor), primary source (HGNC:HGNC:3236), and various cross-references and orthologs. The 'Summary' section provides a brief description of the gene's function and its role in the epidermal growth factor family.

Gene: EGFR

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:ENSG0000146648; 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; Hominoidea; 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

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 is 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 name is "EGFR" with ID "ENSG00000146648".

The main content area is divided into sections:

- Gene-based displays:** A sidebar with links to 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:** This section includes a 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.CM00669.2), About this gene, Transcripts (link to "Show transcript table"), and a Summary section.
- Summary:** This section provides details about the gene's name (EGFR), CCDS sets (CCDS47587.1, CCDS5514.1, CCDS5515.1, CCDS5516.1), UniProtKB identifier (P00533), RefSeq, LRG, and Ensembl version (FNSG00000146648 15). It also notes that the gene has 11 transcripts, 77 orthologues, 13 paralogues, and is associated with 14 phenotypes.

Exercise

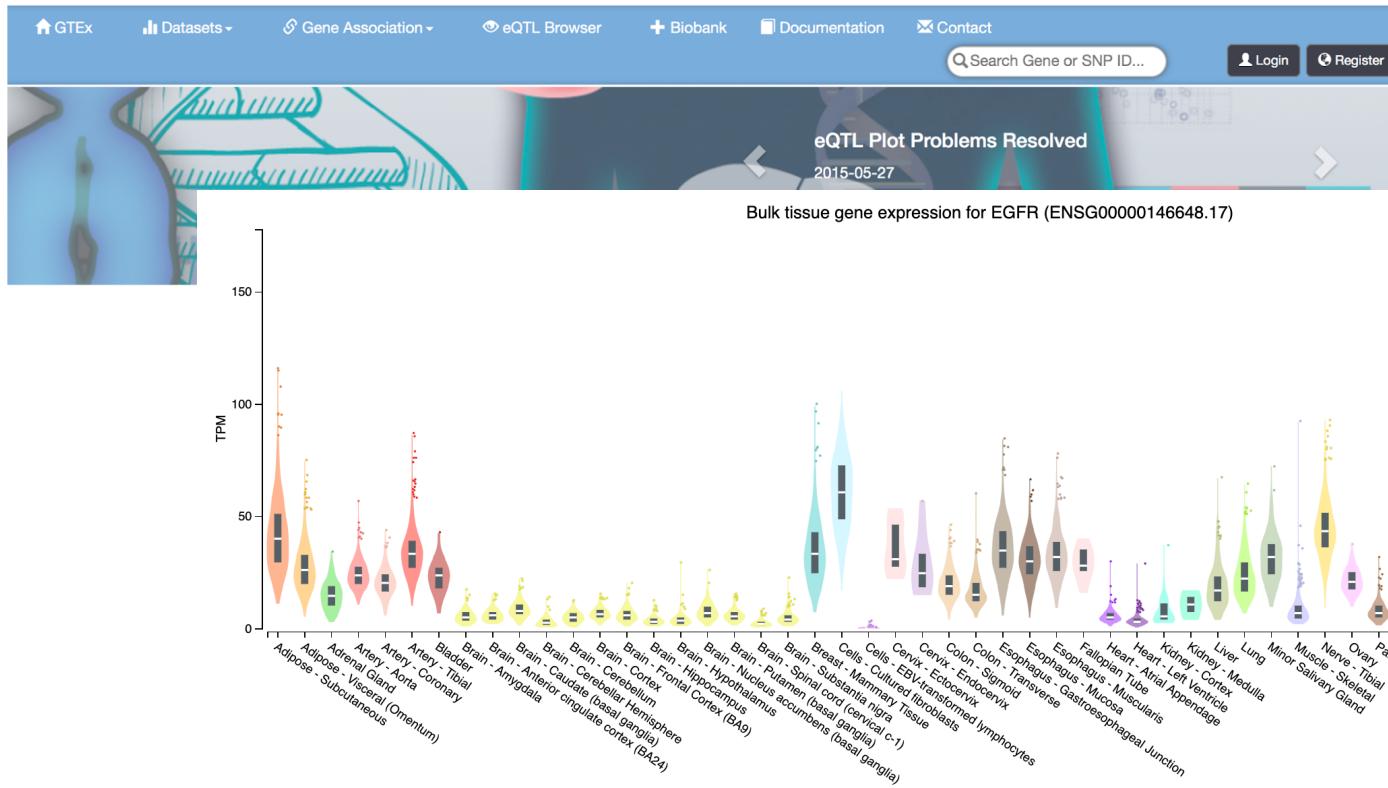
Gene & Protein DBs

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/>

GTEX Portal



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 ^{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



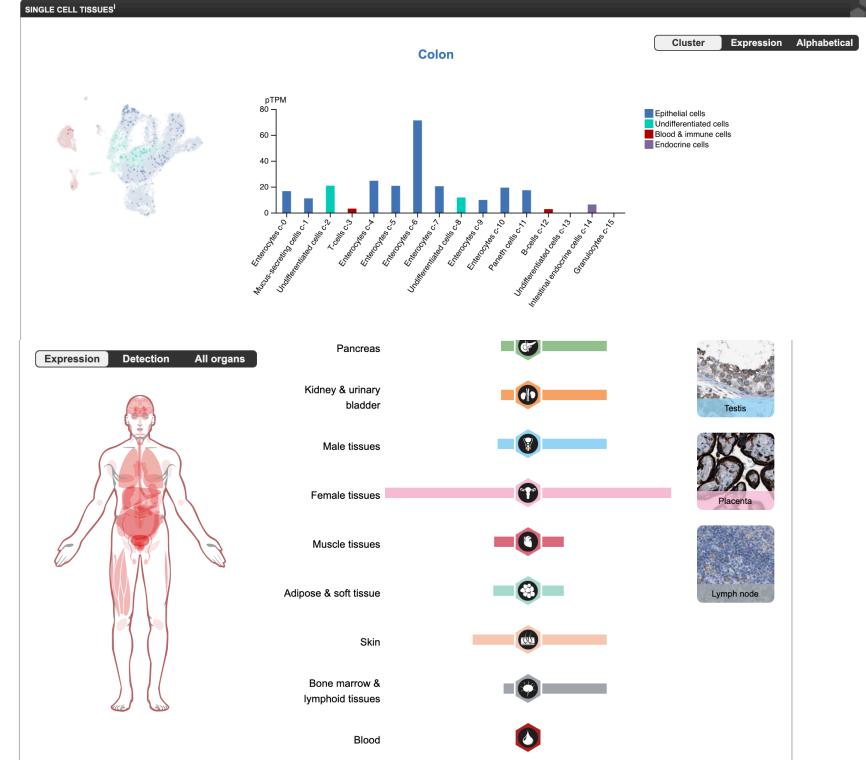
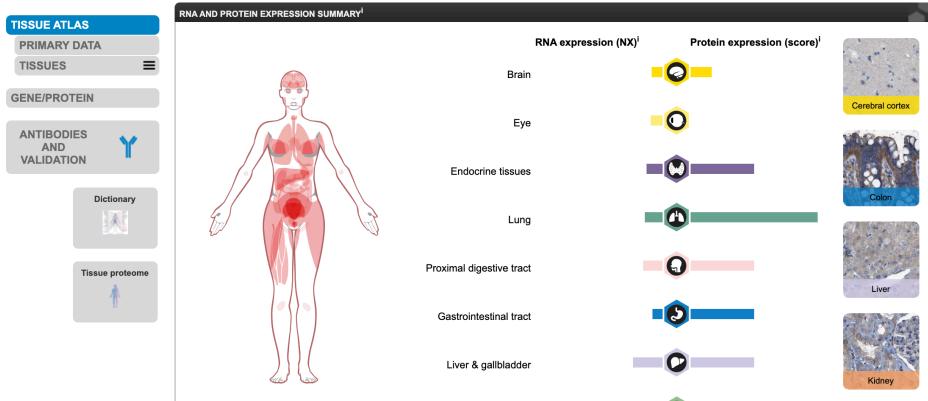
Tumor protein atlas <http://www.proteinatlas.org/>

EGFR



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

HUMAN PROTEIN ATLAS INFORMATION ¹	
RNA tissue specificity ¹	Tissue enhanced (placenta)
RNA tissue distribution ¹	Detected in many
Protein evidence ¹	Evidence at protein level
Protein expression ¹	Membranous and cytoplasmic expression in several tissues, most abundant in placenta.
IMMUNOHISTOCHEMISTRY DATA RELIABILITY	
Data reliability description ¹	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 ¹	Enhanced
Antibodies ¹	HPA001200 , HPA018530 , CAB000035 , CAB068186 , CAB073534
SHOW MORE	



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 is a placeholder text: 'Try an example [STAT3 |breast cancer| rs28897756]'. There is also a link to 'Include the top 100 most relevant genes'. A note below says 'EGFR is a gene'. To the right of the input field is a list of genes: ADCY6, AHNAK2, AMOTL2, ANTXR1, ARHGAP2, ARHGAP29, BCAR1, CAP21, CCND1. At the bottom right of this list is a note: '100 gene(s) entered'. Below the gene list is a button labeled 'Submit'. Further down, there is a note about acknowledging Enrichr in publications, followed by a list of references and a DOI: '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'.

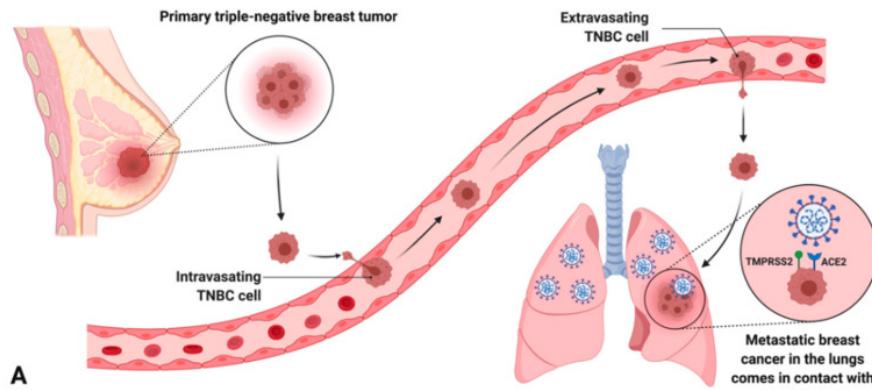
Exercise

Functional Enrichment

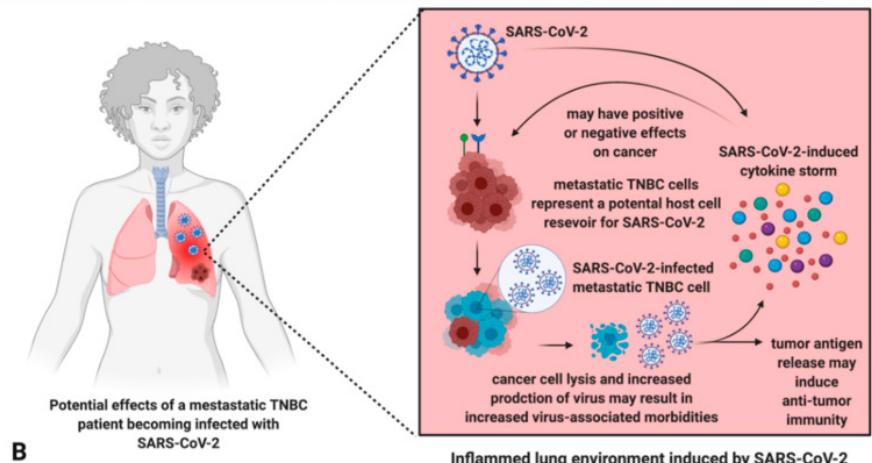
Metastatic breast cancer cells may represent an additional host cell SARS-CoV-2 reservoir, leading to increased viral load and virus-associated morbidities, and unknown consequences on the progression of the cancer.

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 v6.2 homepage. At the top, there's a dark blue header bar with the GSEA logo on the left and navigation links: GSEA Home, Downloads, Molecular Signatures Database (which is highlighted in white), Documentation, and Contact. To the right of the header, there are partially visible links for login and registration. Below the header is a light blue main content area. On the left side, there's a sidebar with a light blue background containing a list of links: MSigDB Home, About Collections, Browse Gene Sets, Search Gene Sets, Investigate Gene Sets, View Gene Families, and Help. In the center, there's a logo for "Molecular Signatures Database" featuring a blue cylinder icon with three dots inside. To the right of the logo, the text "Molecular Signatures Database v6.2" is displayed. The page is divided into two main sections: "Overview" on the left and "Collections" on the right. 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 border and a large red letter "H" containing information about hallmark gene sets.

GSEA
Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact

logi
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

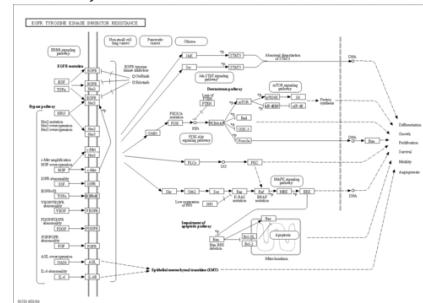
Functional Enrichment

Exercise: Using KEGG

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

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

Kegg PATHWAY: map01521 [Help](#)

Entry	map01521	Global Pathway
Name	EGFR tyrosine kinase inhibitor resistance	
Description	EGFR is a tyrosine kinase that participates in the regulation of cellular homeostasis. EGFR also serves as a stimulus for cancer growth. EGFR gene mutations and protein overexpression, both of which activate down-stream pathways, are associated with cancers, especially lung cancer. Several tyrosine kinase inhibitor (TKI) therapies against EGFR are currently administered and are initially effective in cancer patients who have EGFR mutations on aberrant activation of EGFR. However, the development of TKI resistance is common and results in the recurrence of tumors. Studies over the last decade have identified mechanisms that drive resistance to EGFR TKI treatment. Most outstanding mechanisms are: the secondary EGFR mutation (T790M), activation of alternative pathways (c-Met, HGF, AXL), aberrance of the downstream pathways (K-RAS mutations, loss of PTEN), impairment of the EGFR-TKIs-mediated apoptosis pathway (BCL2-like 11/BIM deletion polymorphism), histologic transformation, etc.	
Class	Human Diseases; Drug resistance: antineoplastic BRITE hierarchy	
Pathway map	map01521 EGFR tyrosine kinase inhibitor resistance	
		
Ortholog table		
Reference	PMID: 26579470	
Authors	Huang L, Fu L	
Title	Mechanisms of resistance to EGFR tyrosine kinase inhibitors.	

All links

- Drug (2)
- KEGG DRUG (2)
- Chemical substance (2)
- KEGG COMPOUND (2)
- Gene (118959)
- KEGG ORTHOLOGY (67)
- RefGene (118892)
- Literature (19)
- PubMed (19)
- All databases (118982)

[Download RDF](#)

Exercise

Functional Enrichment

Exercise: Using STRING

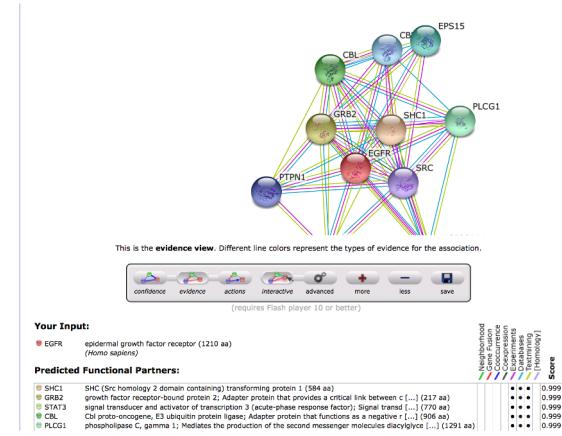
Go to the STRING [website](#)

How to find the interaction network of a protein ?

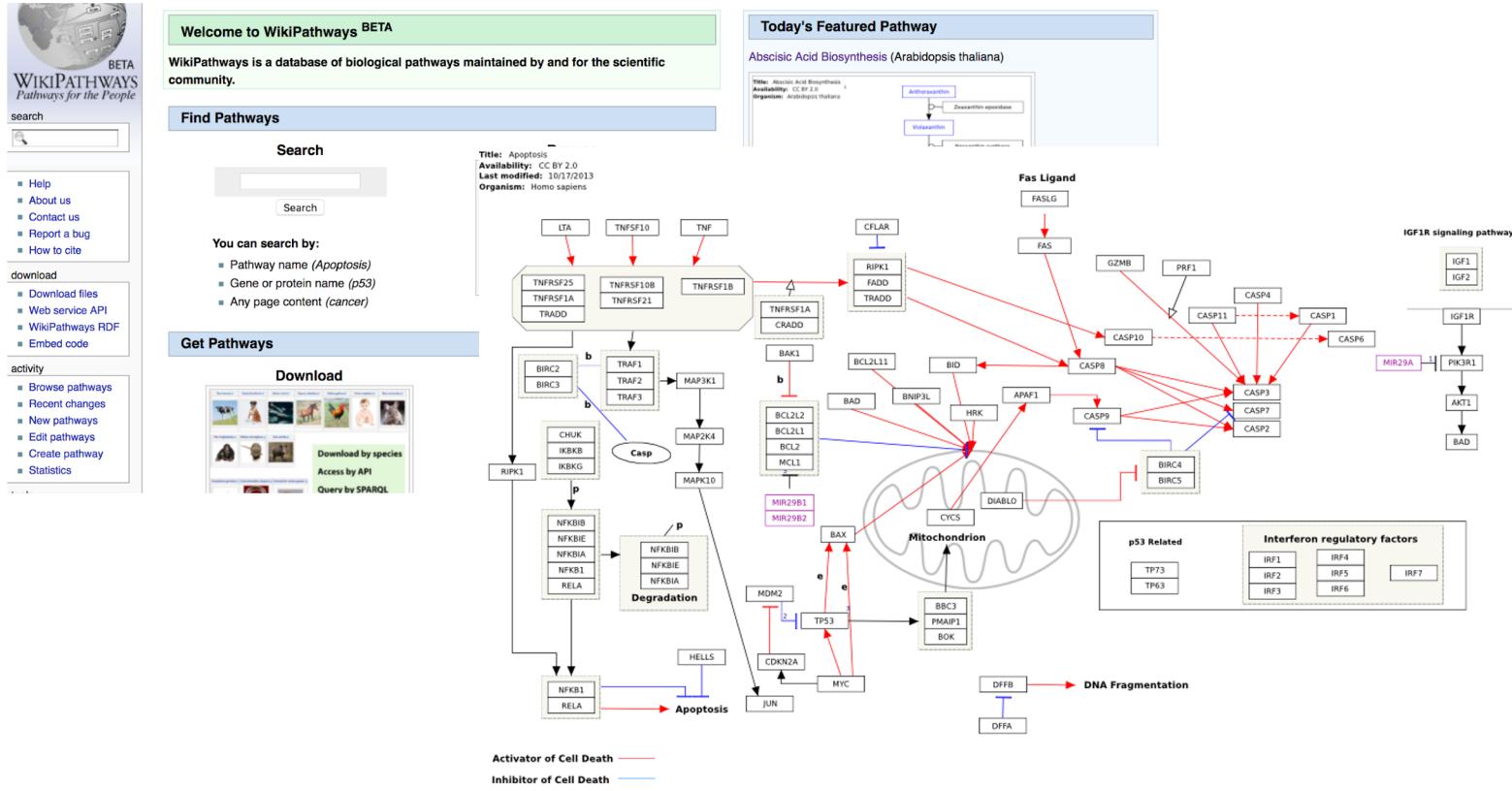
On the top of the results page, the interaction network is visualized.

- 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

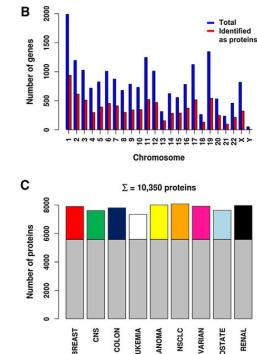
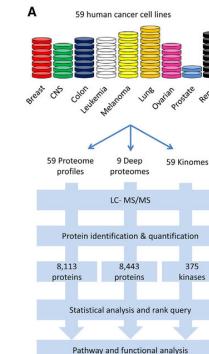
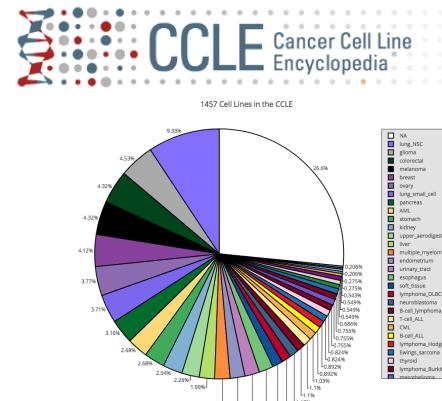
Cell lines

NCI60

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

Cancer Cell Line Encyclopedia

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



Amin Moghaddas Ghafari, Hannes Hahne, Zhixiang Wu, Florian Johann Auer, Chen Meng, Matthias Wilhelm, Bernhard Kuster, Global Proteome Analysis of the NCI-60 Cell Line Panel, *Cell Reports*, Volume 4, Issue 3, 2013, Pages 609-620, ISSN 2211-1247 <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/cell/>

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



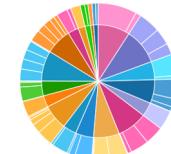
ICGC Data Portal

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', 'Set Operations', 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 'Select' and 'Demo' button.

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

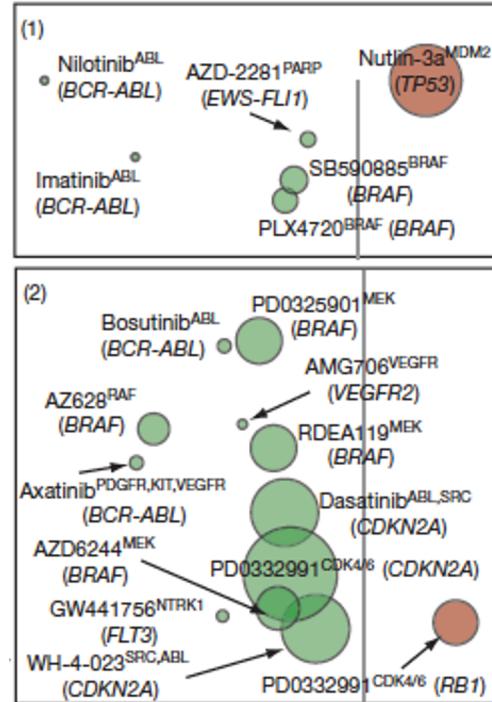
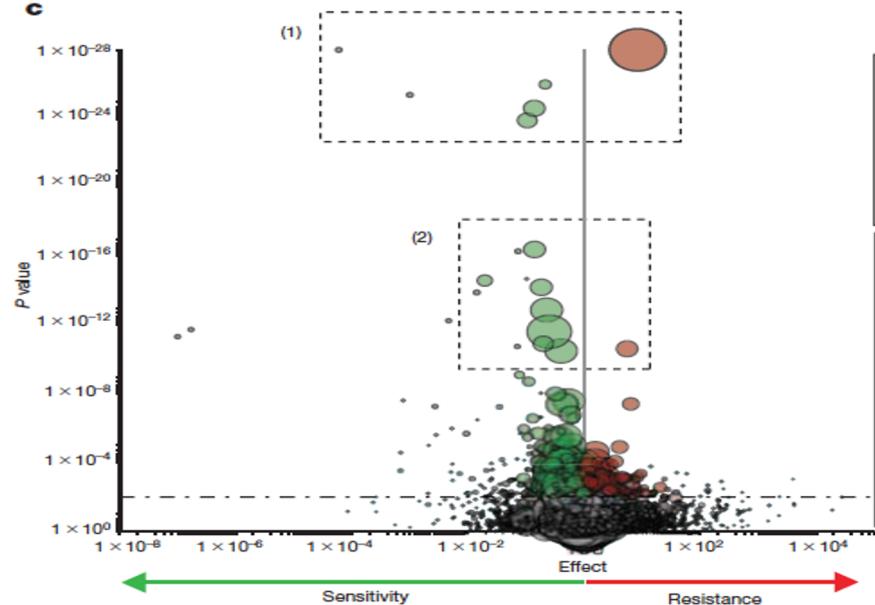
<https://clue.io/>

PanDrugs

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

Systematic identification of genomic markers of drug sensitivity in cancer cells

Mathew J. Garnett^{1*}, Elena J. Edelman^{2*}, Sonja J. Heidorn^{1*}, Chris D. Greenman^{1†}, Anahita Dastur², King Wai Lau¹, Patricia Gréninger², Richard Thompson¹, Yi Lu², Jorge Sáez¹, Omasana Lin^{3,4}, Francesco Iorio^{1,5}, Didier Strelak⁶, Li Chen²
 Randy J. Milano
 Fiona Kogera^{1,8}
 Wenjun Zhou^{3,4}
 Wanjuan Yang¹
 Jeffrey A. Engel¹
 P. Andrew Futreal¹



Garnett et al. Nature 2012

Exercise

Therapy response

Exercise: Using GDSC

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

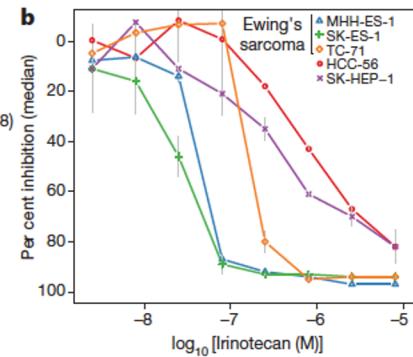
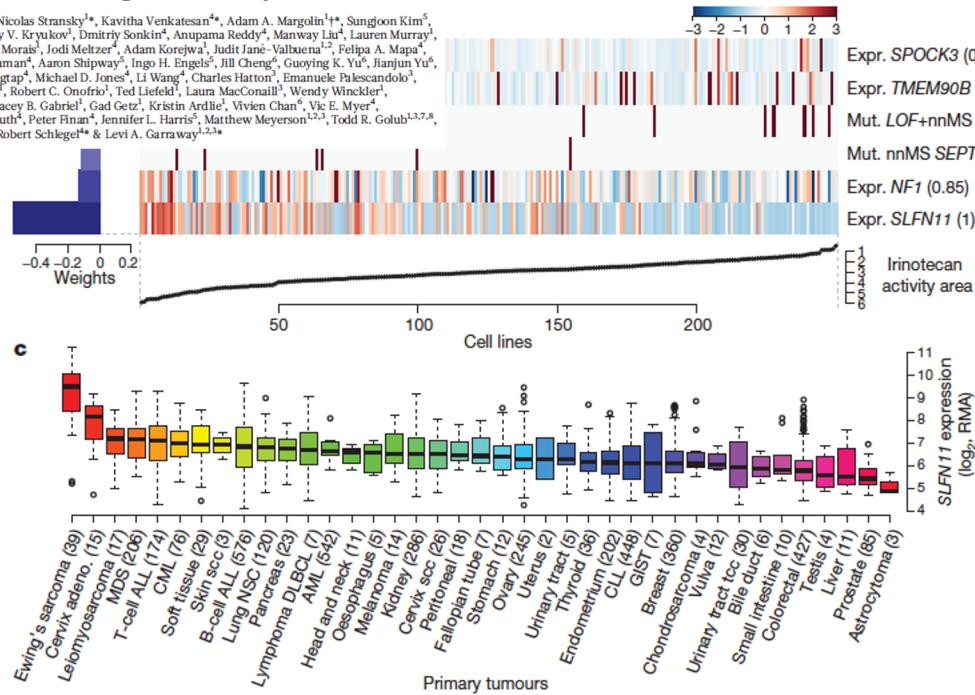
To which drugs are sensitive the EGFR mutated cell lines?

- 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>

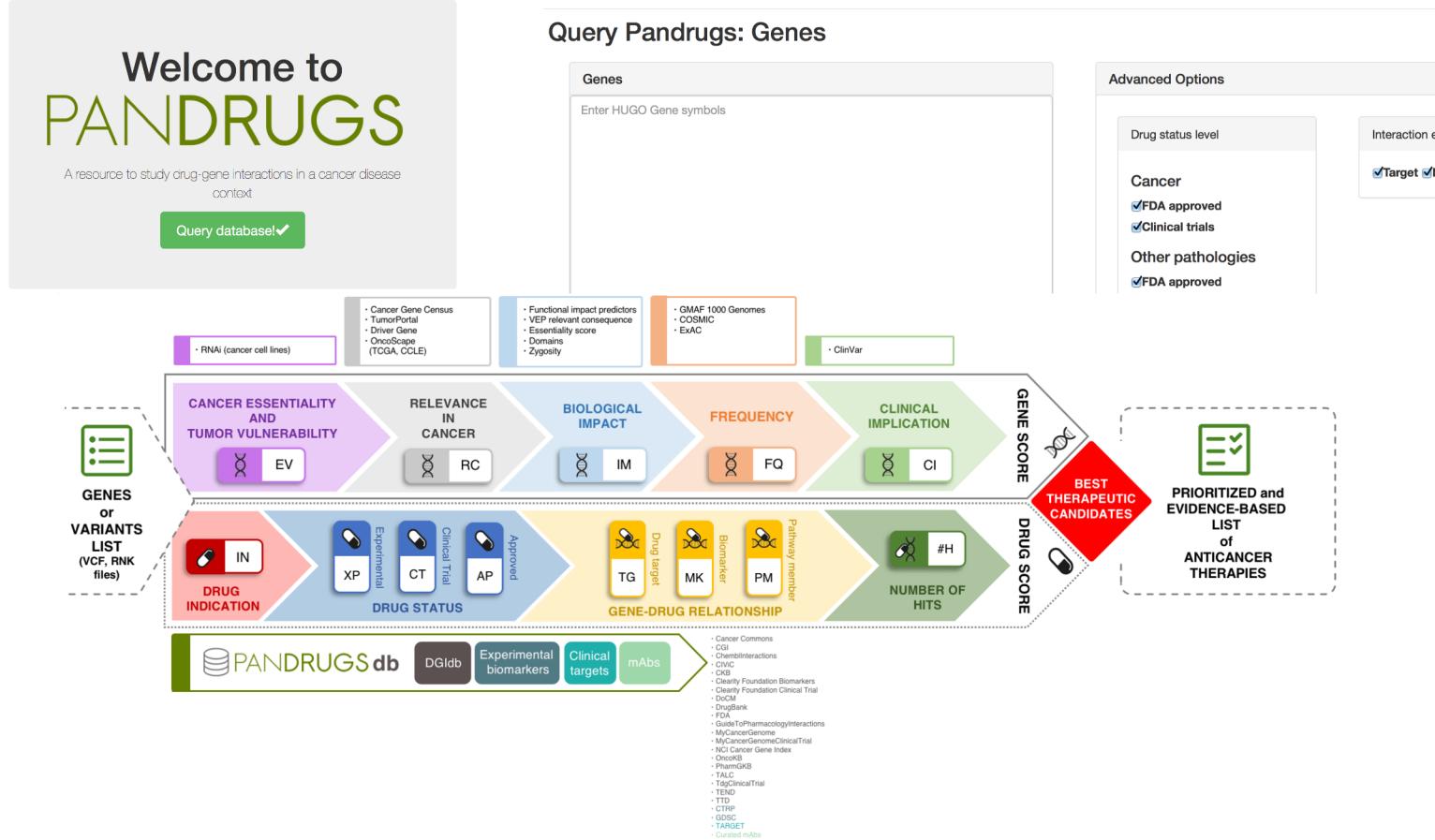
The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity

Jordi Barretina^{1,2,3,4*}, Giordano Caponigro^{4*}, Nicolas Strancky^{4*}, Kavitha Venkateswaran^{4*}, Adam A. Margolin^{1+*}, Sungjoon Kim⁵, Christopher J. Wilson⁶, Joseph Lehár⁴, Gregory V. Kryukov⁶, Dmitry Sankin⁶, Anupama Reddy⁴, Manway Liu⁴, Lauren Murray⁴, Michael F. Berger⁴, John E. Monahan⁷, Paula Moran⁸, Jodi Meltzer⁹, Adam Korejwa¹⁰, Judit Jane-Valbuena^{1,2}, Felipa A. Mapa⁴, Joseph Thibault¹⁰, Eva Bric-Purlong¹⁰, Pichai Raman¹⁰, Aaron Shipway², Ingo H. Engels², Jill Cheng², Guoying K. Yu², Jianjun Yu⁶, Peter Aspesi¹¹, Melanie de Silva¹², Kalpana Jagtap⁴, Michael D. Jones¹³, Li Wang¹⁴, Charles Hatton¹⁵, Emanuele Palascandolo³, Supriya Gupta¹⁶, Scott Mahan¹⁷, Carrie Sougner¹⁸, Robert C. Onofrio¹⁹, Ted Liefeld¹⁹, Laura MacConell²⁰, Wendy Winkler⁴, Michael Reich²¹, Nannxin Li²², Jill P. Mesinly²³, Stacey B. Gabriel²⁴, Gad Getz²⁵, Kristin Ardlie²⁶, Vivien Chan²⁶, Vick E. Myer⁴, Barbara L. Weber²⁷, Jeff Porter²⁸, Markus Warmuth²⁹, Peter Finan⁴, Jennifer L. Harris²⁶, Matthew Meyerson^{1,2,3}, Todd R. Golub^{1,3,7,8}, Michael P. Morrissey⁴, William R. Sellers⁴, Robert Schlegel⁴ & Levi A. Garraway^{1,2,3,4}



Barretina et al. Nature 2012

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



Connectivity Map <https://clue.io>



ConnectivityMap

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

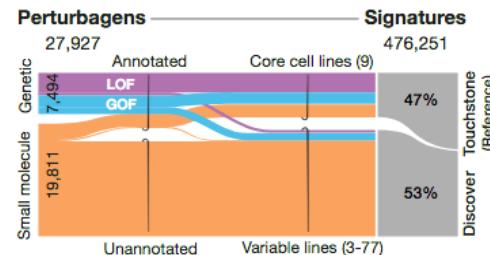
The aim is to generate perturbational profiles across multiple cell and perturbation types

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

Data and Tools

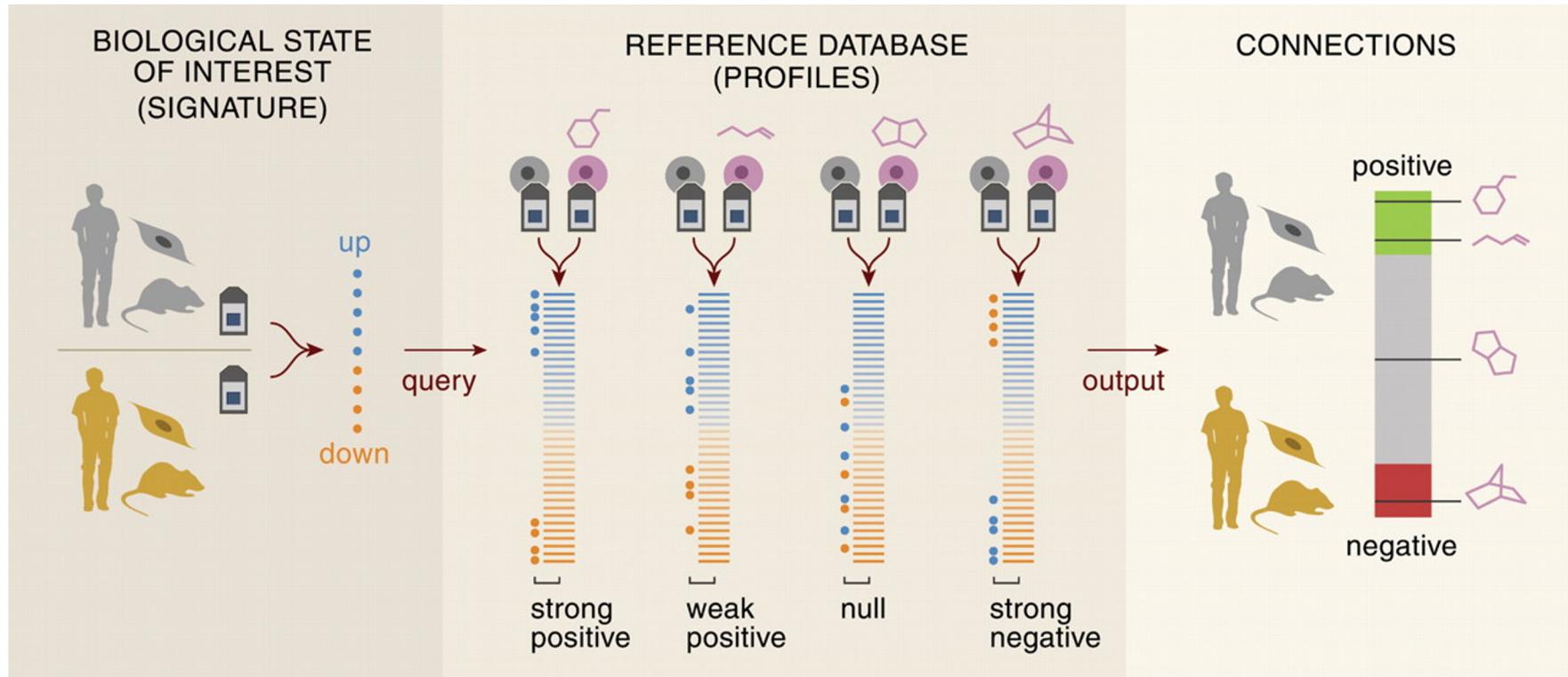
The CMap dataset of cellular signatures catalogs transcriptional responses of human cells to chemical and genetic perturbation. Here you can find the 1.3M L1000 profiles and the tools for their analysis.

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**

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 ⁻⁵	92.5	-1.60	Cowley(2014)	shRNA	Yes	High	GDC-098..[more]
PSMC2	1 x 10 ⁻⁵	95.7	-2.45	Marcotte(2012)	shRNA			BORTEZO..[more]
ERBB2	4 x 10 ⁻⁵	87.6	-1.78	Campbell(2016)	siRNA	Yes	Highest	AEE 788..[more]
HIST1H2AK	4 x 10 ⁻⁵	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 ⓘ
ERBB2		BREAST	Yes	CDK4		CRISPR	palbociclib	1.000	0.995	
PIK3CA		BREAST	Yes	CDK4		CRISPR	palbociclib	1.000	0.995	
ERBB2		BREAST	Yes	ESR1		CRISPR	tamoxifen	1.000	0.991	
PIK3CA		BREAST	Yes	ESR1		CRISPR	tamoxifen	1.000	0.991	
ERBB2		BREAST	Yes	AKT1		CRISPR	triciribine	0.600	0.997	
PIK3CA		BREAST	Yes	AKT1		CRISPR	triciribine	0.600	0.997	
ERBB2		BREAST	Yes	AKT1		CRISPR	mk-2206	0.600	0.989	
PIK3CA		BREAST	Yes	AKT1		CRISPR	mk-2206	0.600	0.989	
PIK3CA		BREAST	Yes	KIF11		CRISPR	ispinesib	0.600	0.986	
ERBB2		BREAST	Yes	ERBB2		CRISPR	cr-724714	0.600	0.983	

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

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

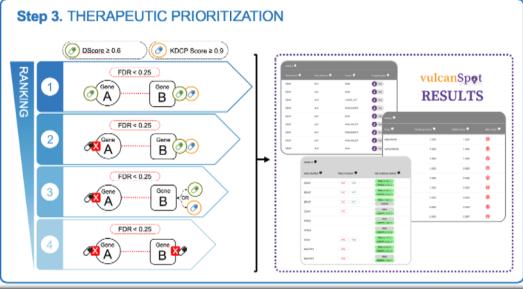
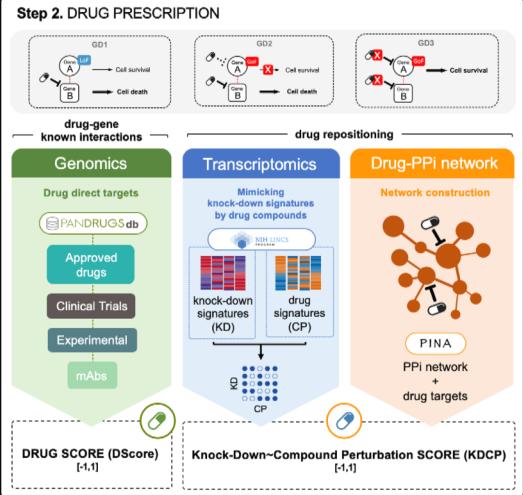
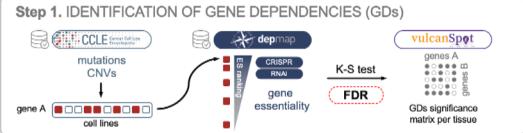
vulcanSpot - A method for detecting and targeting cancer genetic dependencies.

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)
PIK3CA	GoP	BREAST		Yes	CDK4	ONC		CRISPR RNAi (3.17e-1)
ERBB2	GoP	BREAST		Yes	ESR1	ONC/TSG	HCD	CRISPR RNAi (1.27e-1)
PIK3CA	GoP	BREAST		Yes	ESR1	ONC/TSG	HCD	CRISPR RNAi (4.25e-1)
ERBB2	GoP	BREAST		Yes	AKT1	ONC	HCD	CRISPR RNAi (1.62e-1)
PIK3CA	GoP	BREAST		Yes	AKT1	ONC	HCD	CRISPR RNAi (1.53e-1)
ERBB2	GoP	BREAST		Yes	AKT1	ONC	HCD	CRISPR RNAi (1.62e-1)
PIK3CA	GoP	BREAST		Yes	KIF11	ONC	HCD	CRISPR RNAi (1.73e-1)
PIK3CA	GoP	BREAST		Yes	ERBB2	ONC	HCD	CRISPR RNAi (7.29e-1)



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/>

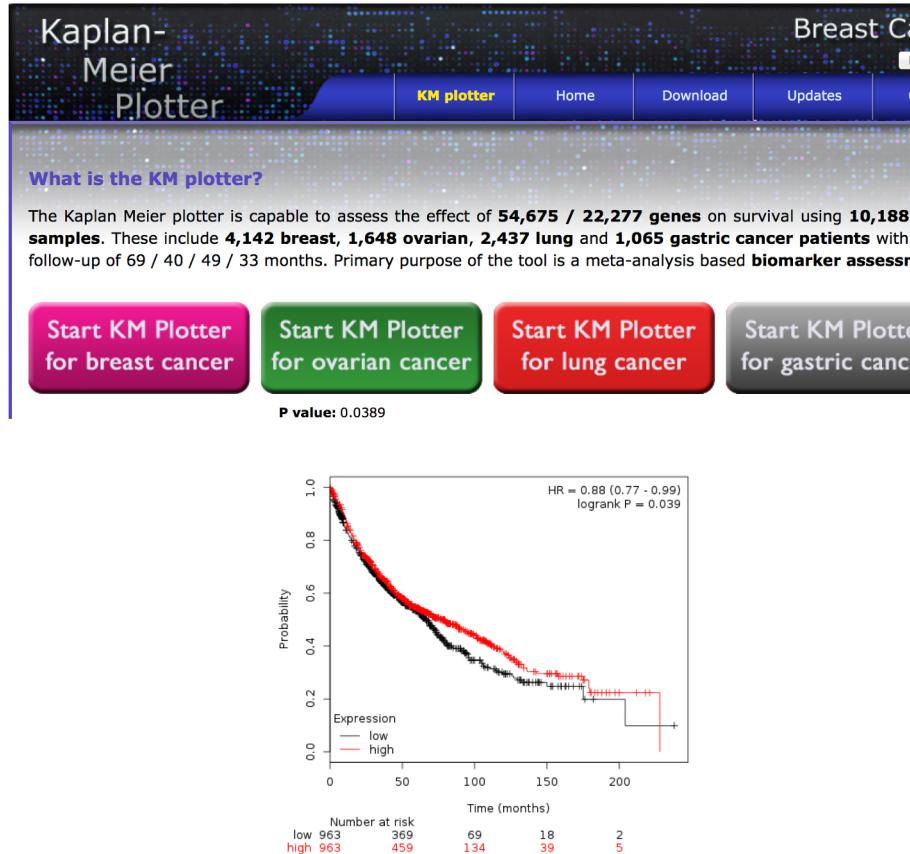
ExAC

<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:

Split patients by: median Auto select best cutoff:

Follow up threshold: all Censor 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 letSet 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

Survival: Kaplan Meier Plotter

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

Survival: Prognoscan

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

Other

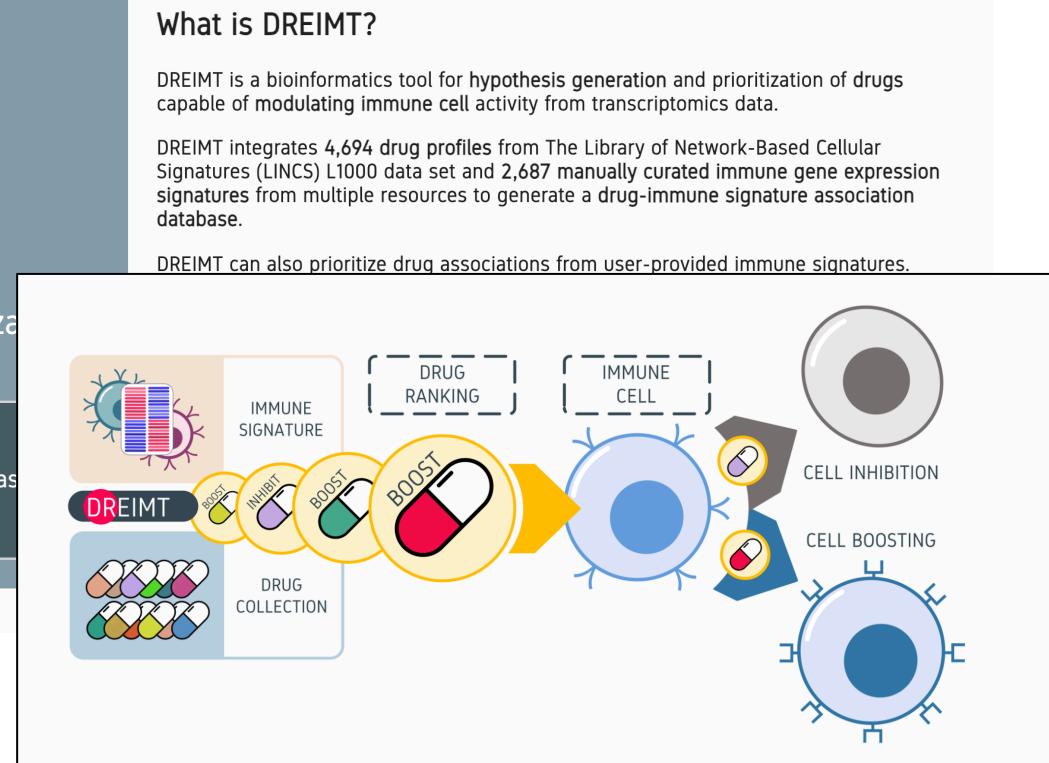
DREIMT

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

Welcome to DREIMT

A tool for immune modulation drug prioritization

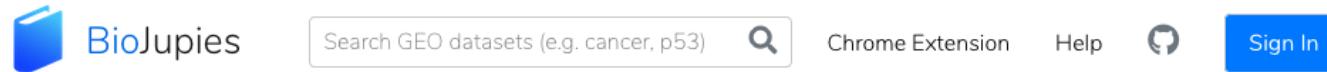
Drug prioritization Signature comparison Database





Find the best of bioinformatics

Ask biological questions, get computational answers

A vertical flow diagram showing three steps: Step 1 (Upload or Fetch RNA-seq Data), Step 2 (Select Data Analysis Tools), and Step 3 (Generate Your Notebook). Each step has a corresponding icon (database, wrench, book) and a list of actions.

- Step 1.** Upload or Fetch RNA-seq Data
 - Upload your raw or processed RNA-seq data
 - Fetch >8,000 public RNA-seq datasets published in the Gene Expression Omnibus
- Step 2.** Select Data Analysis Tools
 - Select from multiple state-of-the-art RNA-seq data analysis tools
 - Contribute your computational tool as a plugin
- Step 3.** Generate Your Notebook
 - Access and share your results through a permanent URL
 - Download, rerun and customize your notebook using Docker

BioJupies Automatically Generates RNA-seq Data Analysis Notebooks

With BioJupies you can produce in seconds a customized, reusable, and interactive report from your own raw or processed RNA-seq data through a simple user interface

[Get Started](#)

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



Thanks!

