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`**

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 al 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.

Ensembl
http://www.ensembl.org/

Exercise

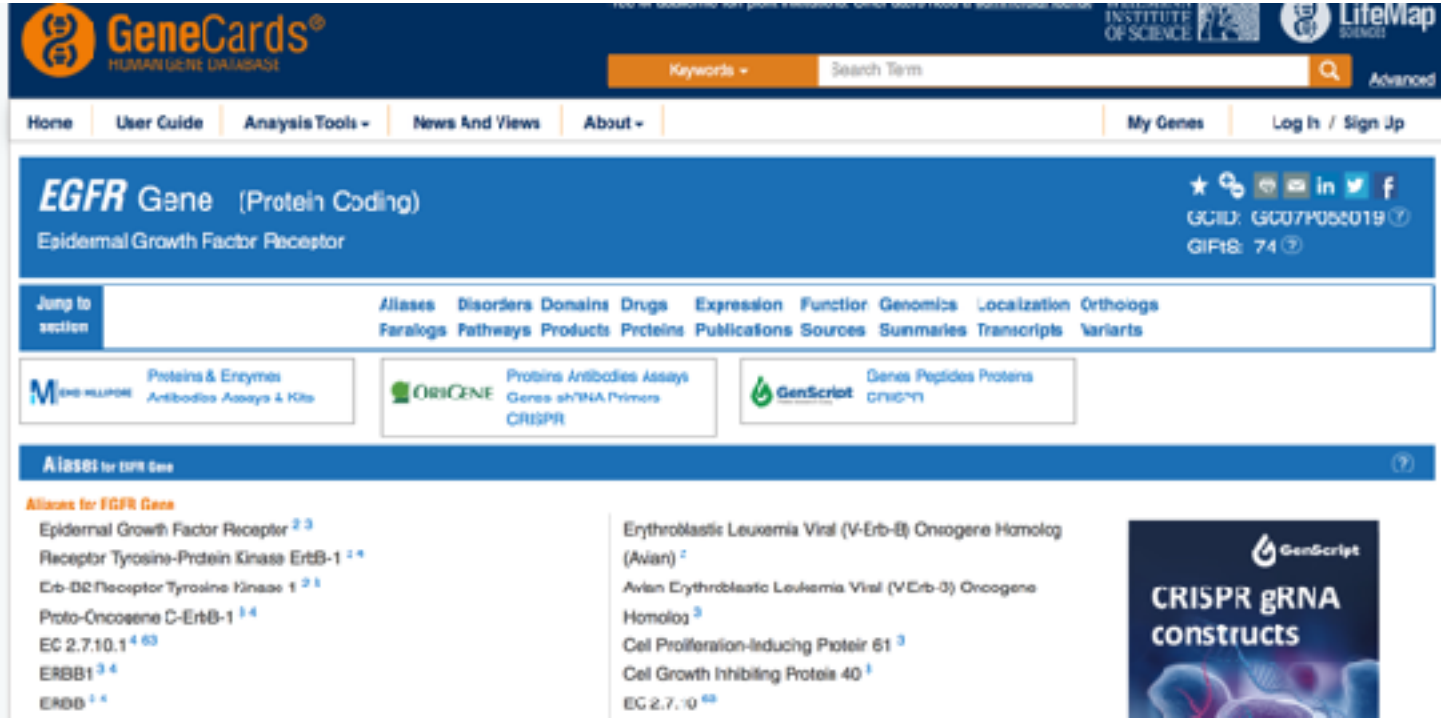
Gene & Protein DBs

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/



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

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/annotate.jsp

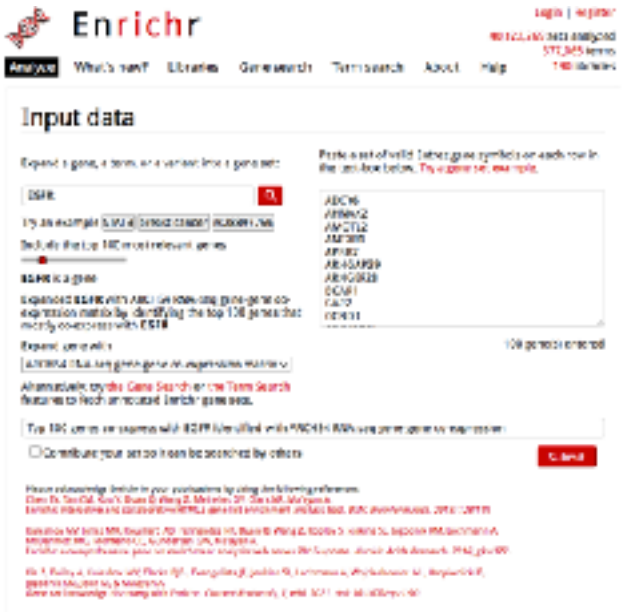
Enrichr
http://amp.pharm.mssm.edu/Enrichr/

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

GeneMania
http://genemania.org/

Exercise

Functional Enrichment



Exercise: Using EnrichR

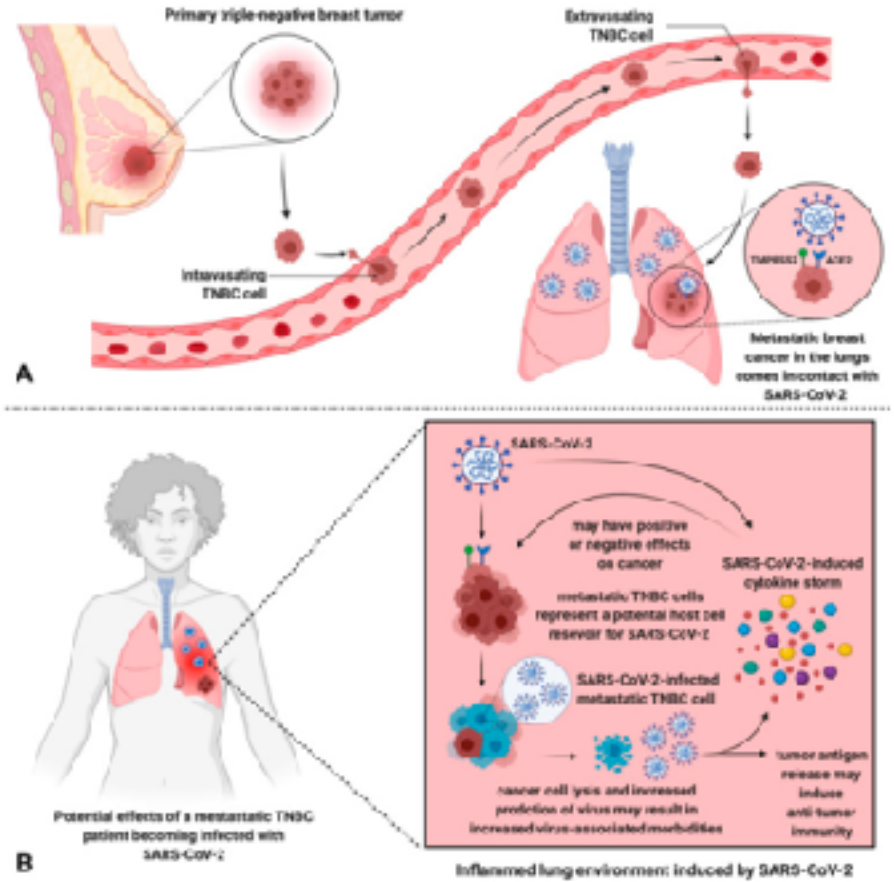
- Go to https://maayanlab.cloud/Enrichr/
- Expand into a gene set the term: **`triple negative breast cancer`**
- Submit the query
- Go to the Diseases/Drugs section
- Check the COVID-19 related gene sets 2021 collection

Exercise

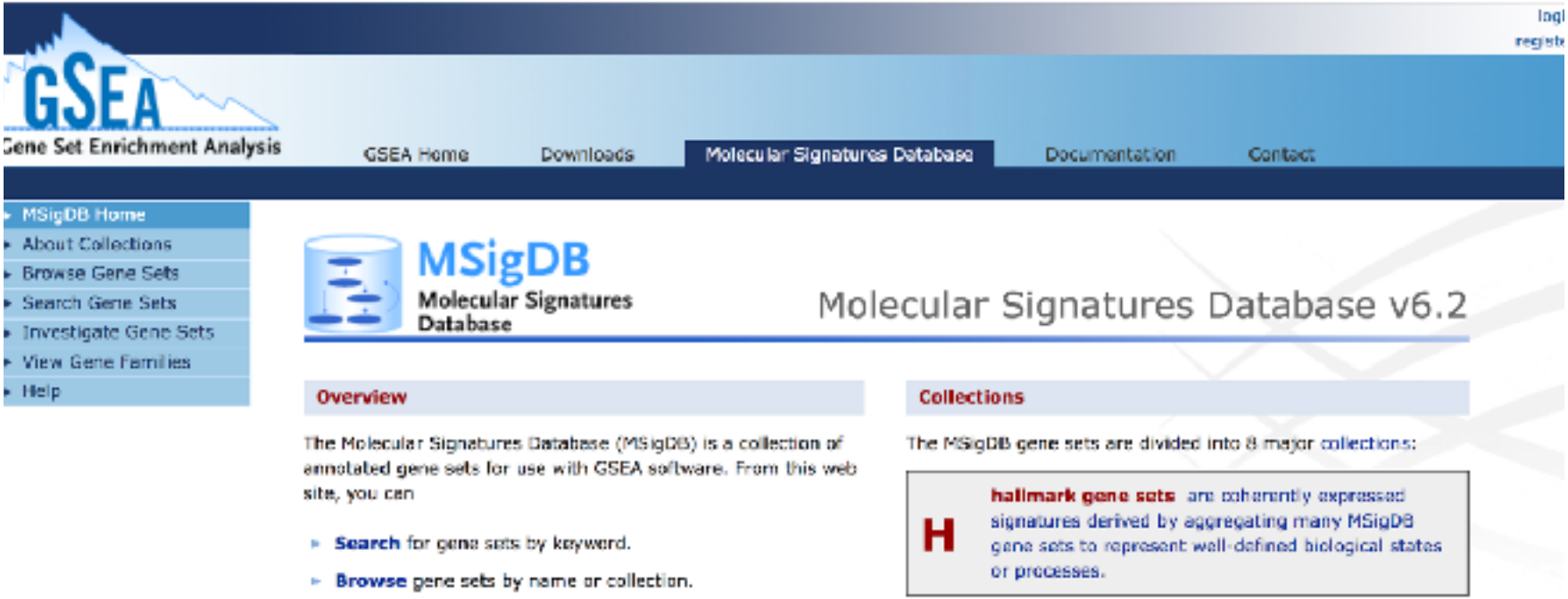
Functional Enrichment

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.



MSigDB http://www.software.broadinstitute.org/gsea/msigdb



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

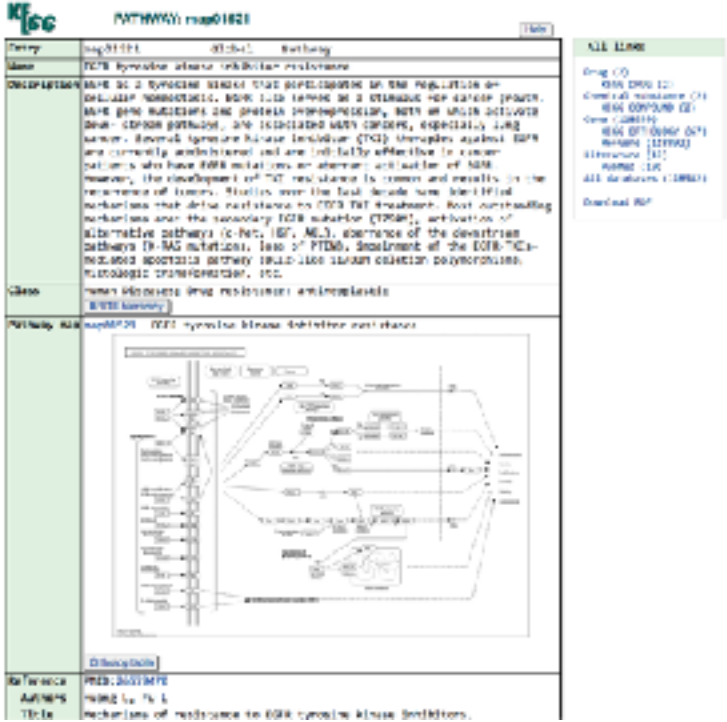
KEGG
https://www.genome.jp/kegg/

Exercise

Functional Enrichment

Exercise: Using KEGG

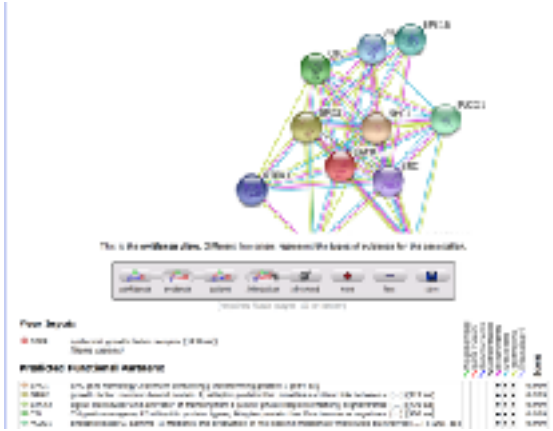
- Go to https://www.genome.jp/kegg/
- Load a map of the EGFR tyrosine kinase inhibitor resistance pathwaypathway



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

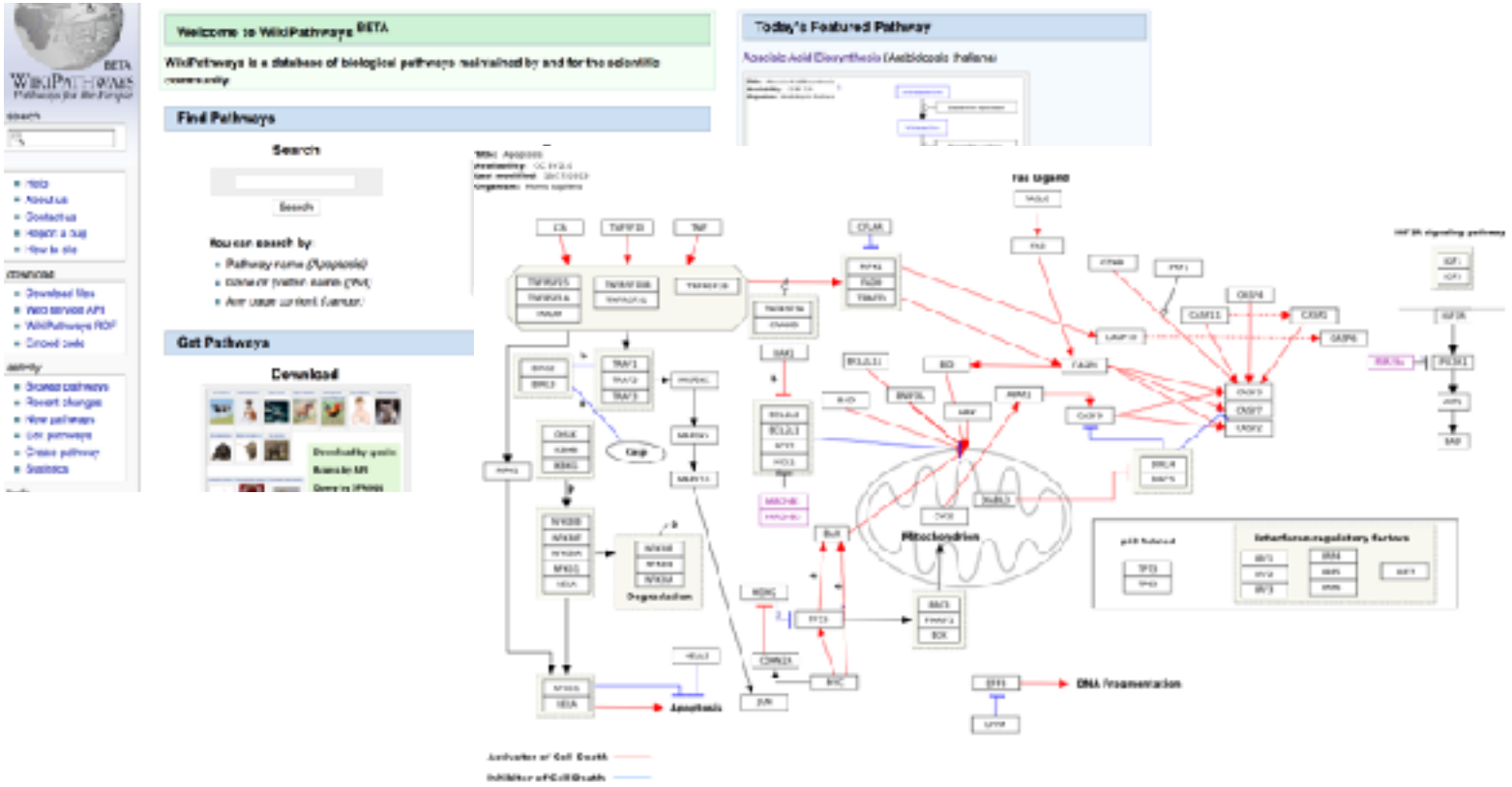
How to find the interaction network of a protein ?

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

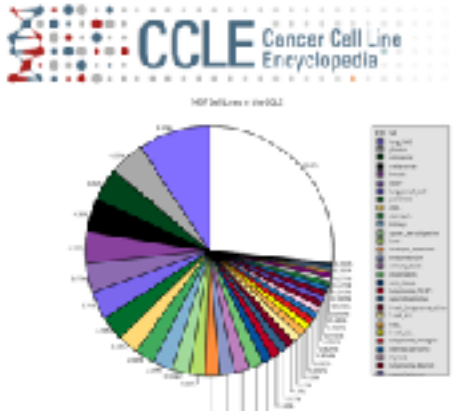
Survival, Clinical & other useful info

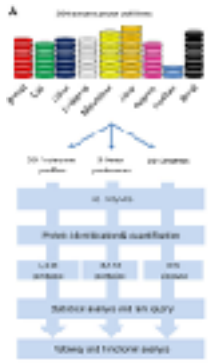
Collections

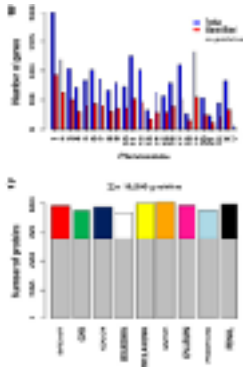
Cell lines

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 2211-1247 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

Tumor samples

NCI60
https://discover.nci.nih.gov/cellr

Cancer Cell Line Encycloped
https://portals.broadinstitute.org

The Cancer Genome Atlas
https://cancergenome.nih.gov/

NIH 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

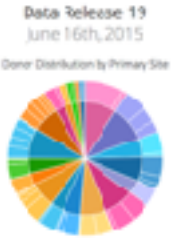
NCI60
https://discover.nci.nih.gov/cellminer/

Cancer Cell Line Encyclopedia
https://portals.broadinstitute.org/ccle

Tumor samples

ICGC Data Portal

International Cancer Genome Consortium
http://dcc.icgc.org/



Cancer projects	55
Cancer primary sites	41
Donors	12,979
Simple somatic mutations	16,159,160
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/

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/

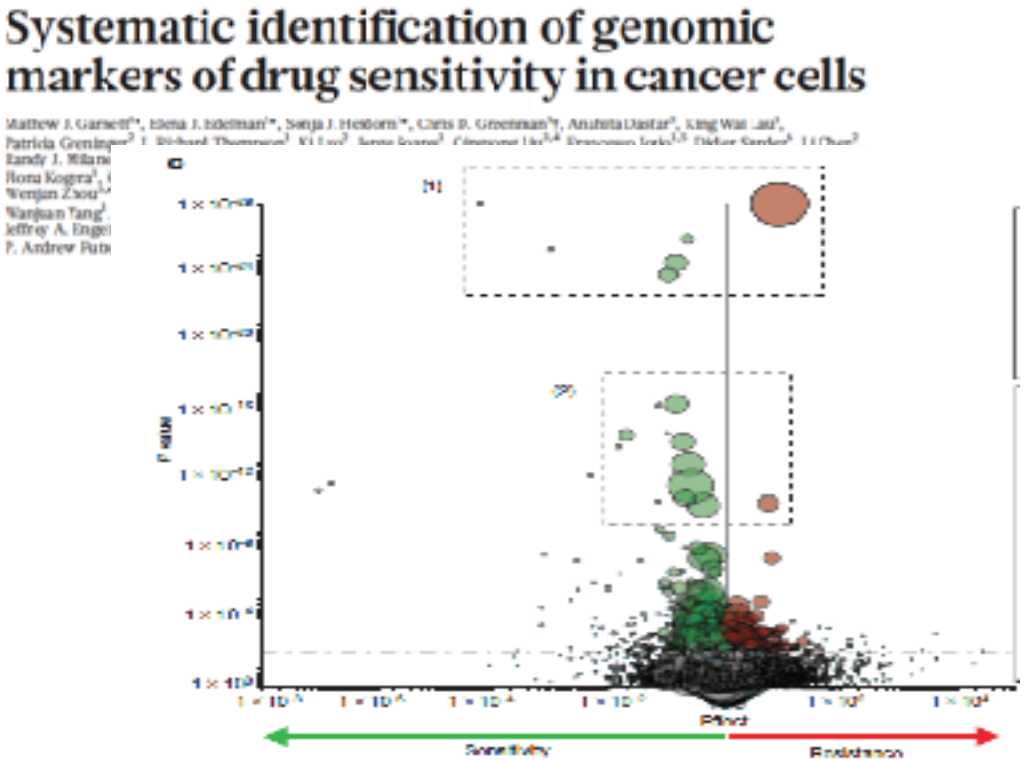
Cancer Therapeutics Response Portal
http://www.broadinstitute.org/ctrp/

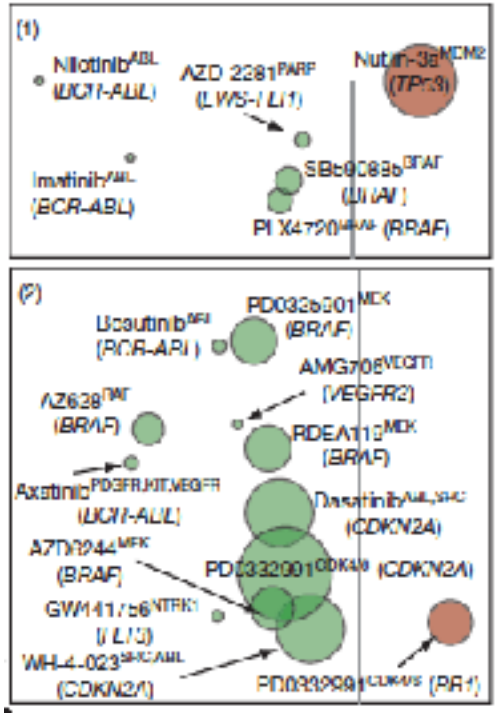
Open Targets
https://www.opentargets.org/

Connectivity Map
https://clue.io/

PanDrugs
http://pandrugs.bioinfo.cnio.es/

Genomics of drug sensitivity in cancer https://www.cancerrxgene.org/





Garnett et al. Nature 2012

Exercise

Therapy response

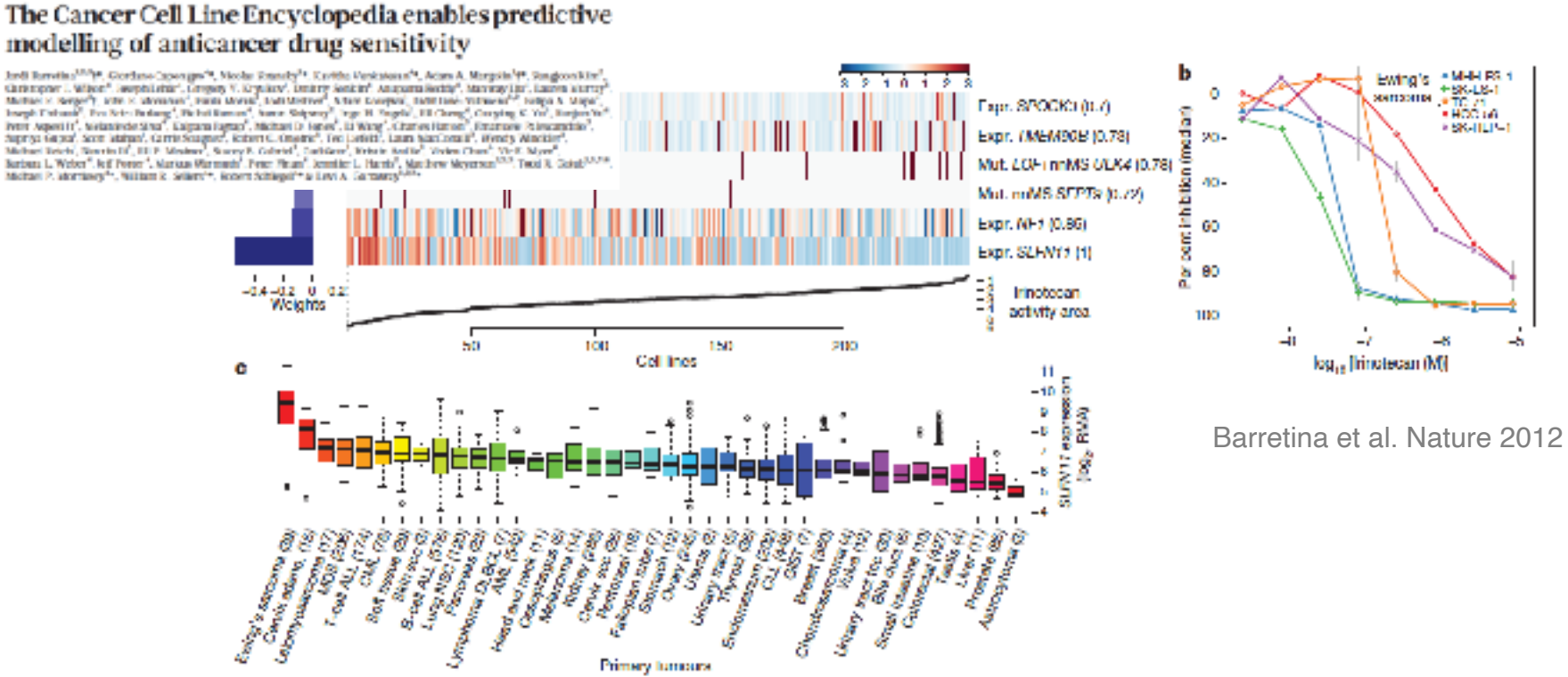
Exercise: Using GDSC

- 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



Barretina et al. Nature 2012

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

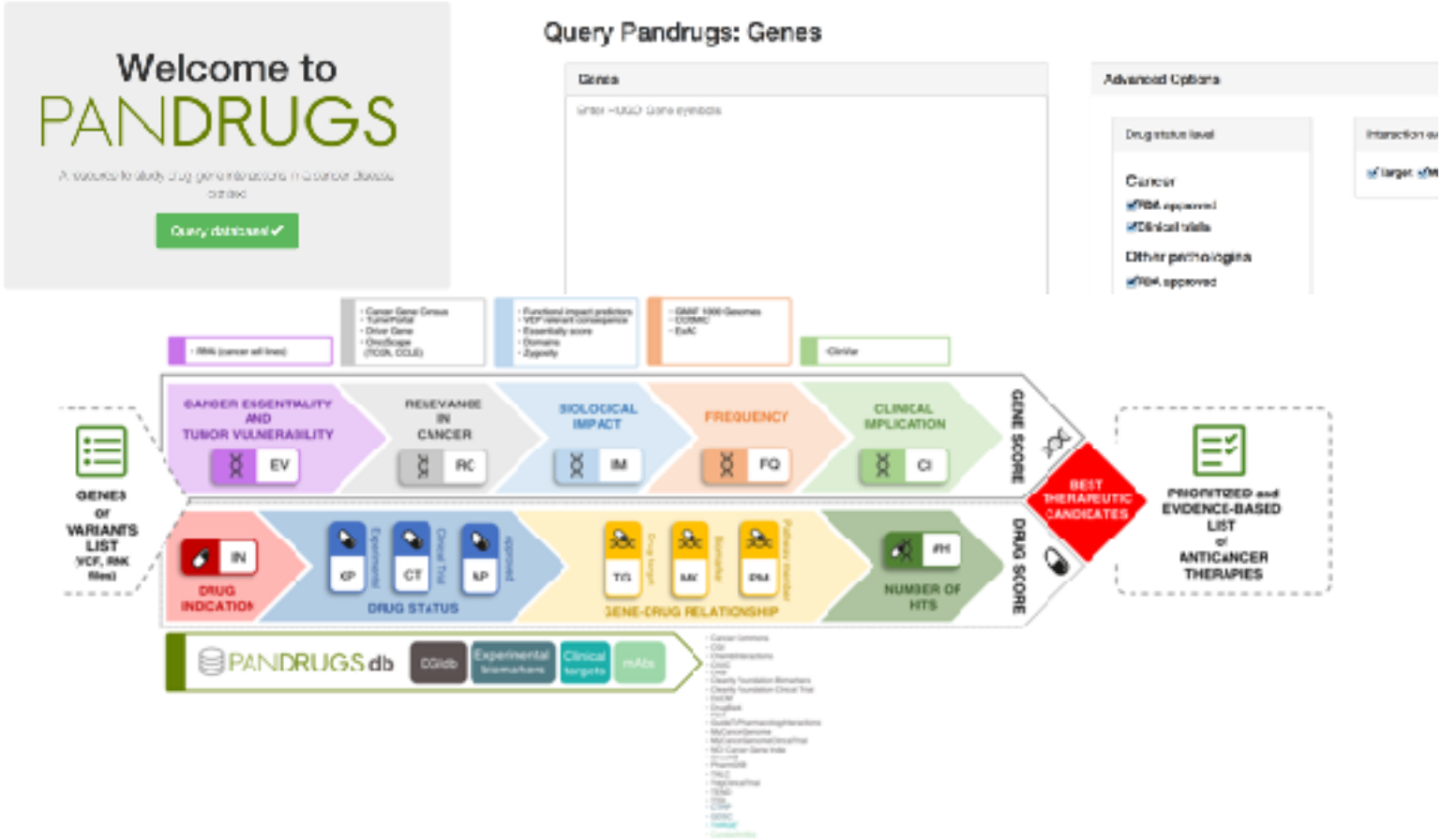
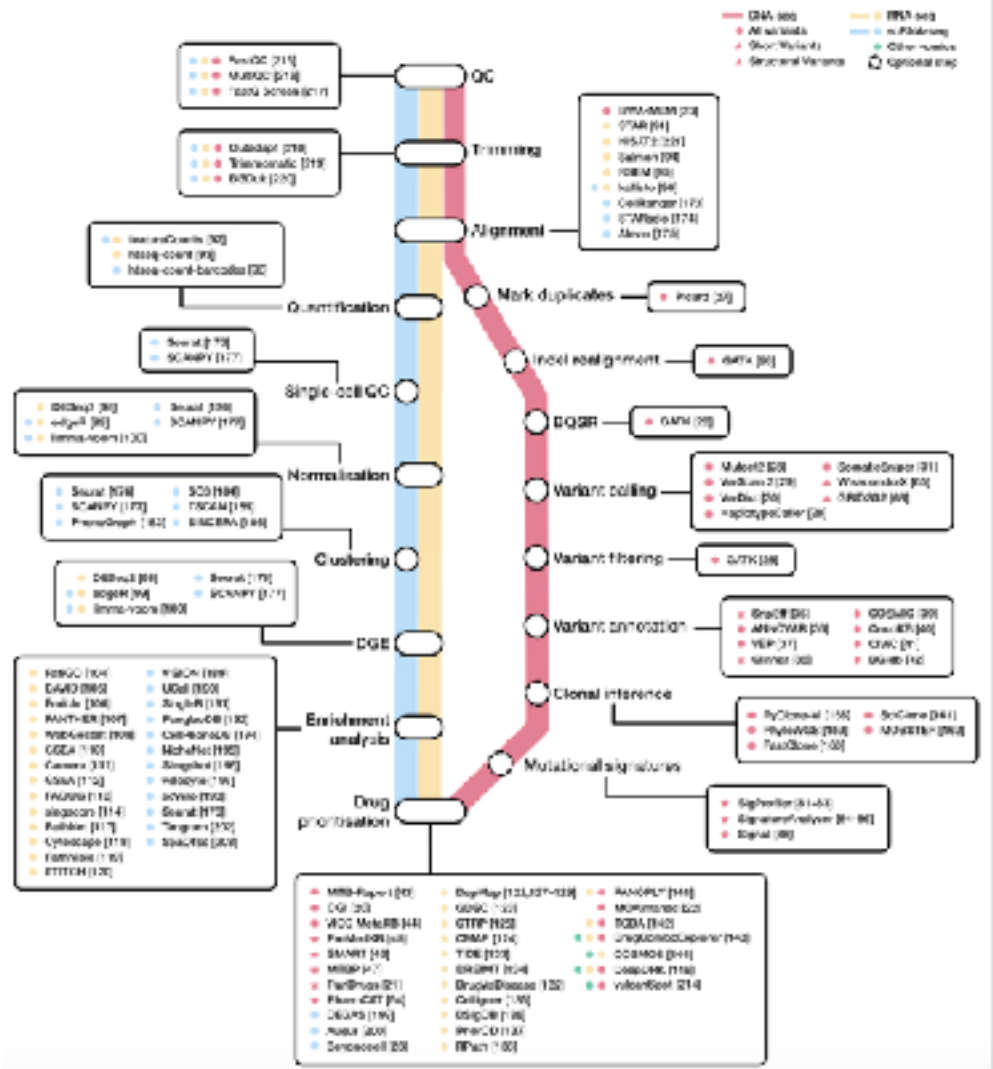


Table 1: Bioinformatics tools for genomics-based drug prioritisation.

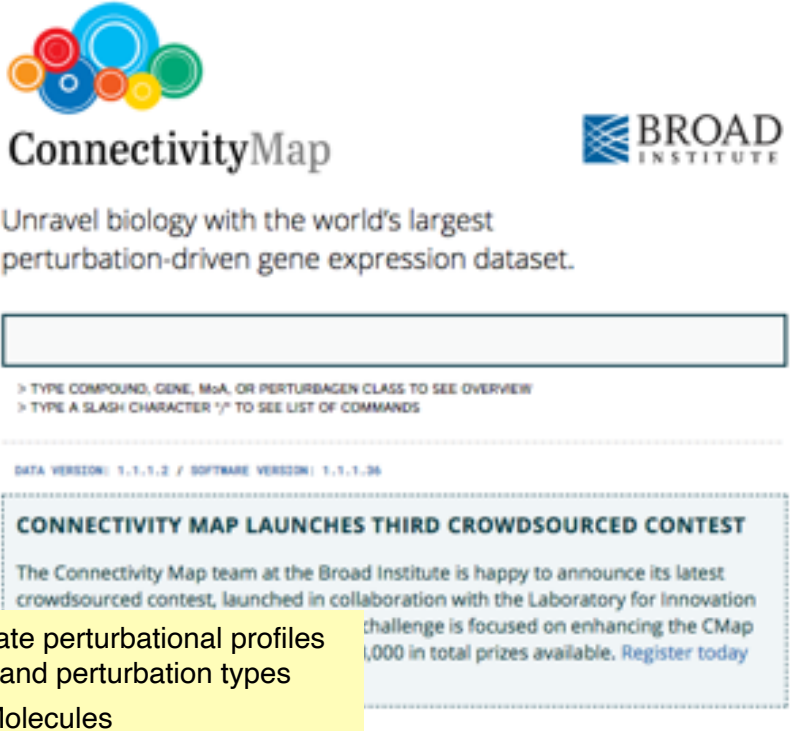
Name	Description	Input	Output	URL
MTB-Report [43]	R script that filters and classifies cancer variants into levels of evidence using gene-drug databases.	Tables with SNVs, CNVs and gene fusions (somatic).	Molecular Tumour Board (MTB) report with actionable variants in PDF.	https://github.com/jperera-bel/MTB-Report
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).	Downloadable tables of a) Annotated variants, including information about the oncogenicity and biological consequence, and b) drug-variant associations with evidence level and response prediction.	https://www.cancergenomeinterpreter.org/home
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.	https://search.cancervariants.org/#*
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.	http://www.fudan-pgx.org/premedkb/index.html#/home
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.	https://smart-cancer-navigator.github.io/home
PanDrugs [21]	Web tool to prioritise anticancer drug treatments according to individual genomics data. PanDrugs computes two scores, the Gene Score (GScore) and the Drug Score (DScore). The GScore ranges from 0 to 1 and is estimated according to gene essentiality and tumoural 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.	https://www.pandrugs.org/#!/
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.	https://mtbp.org/
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 germinal variants.	https://pharmcat.org/

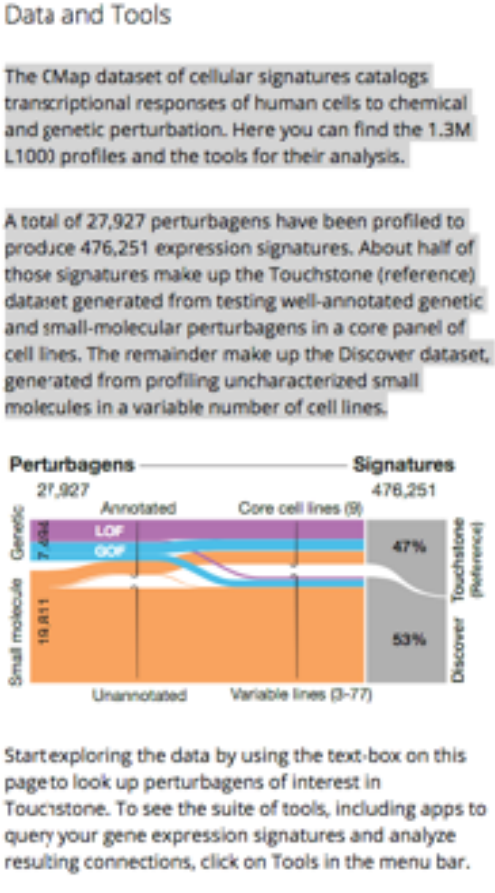


The bioinformatics roadmap for therapy selection in cancer genomics

Jiménez-Santos et al. 2022 coming soon

Connectivity Map https://clue.io

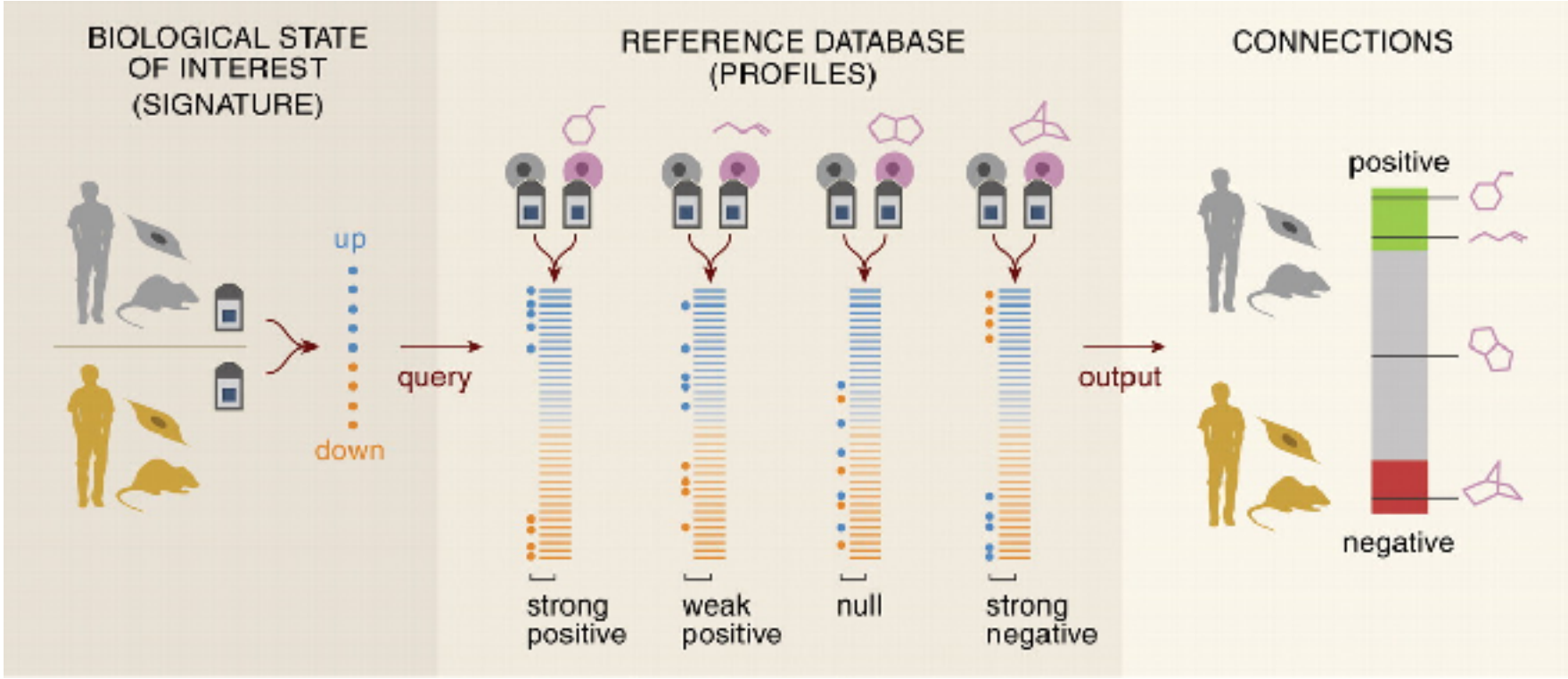




The aim is to generate perturbational profiles across multiple cell and perturbation types

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

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

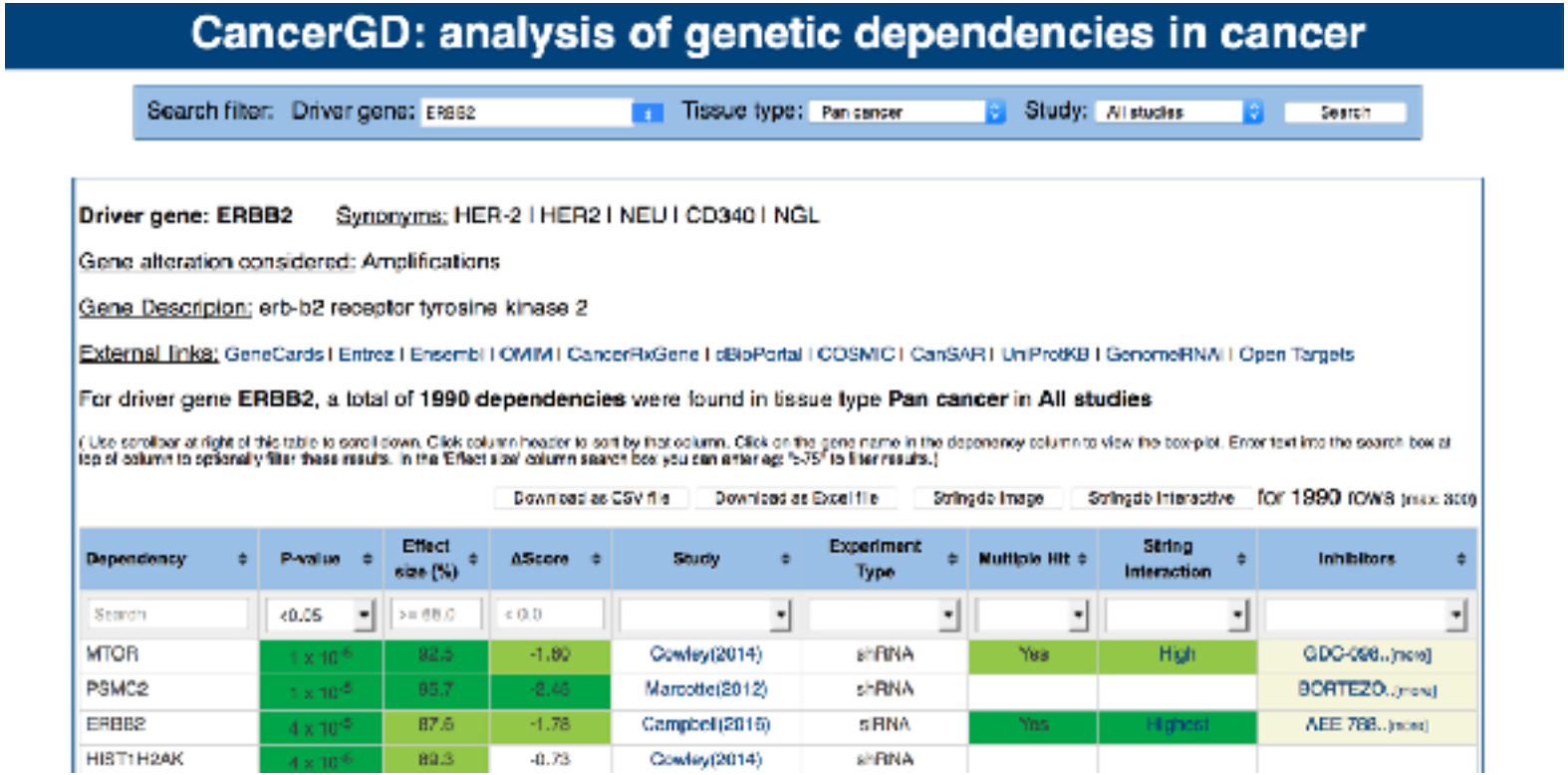
DepMap
https://depmap.org/portal/

CancerGD
http://www.cancergd.org/

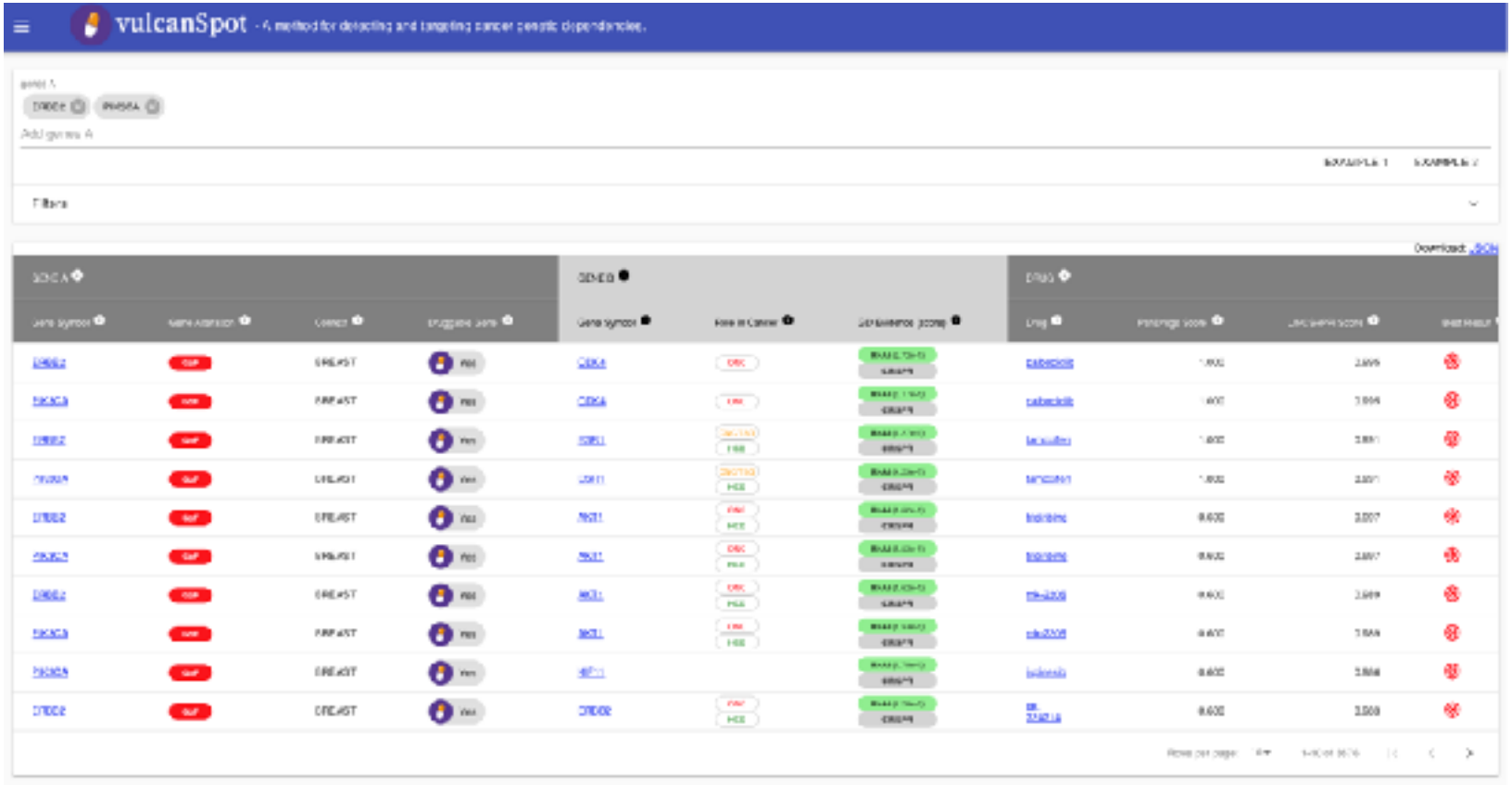
vulcanSpot
http://www.vulcanspot.org/

Cancer dependency map https://www.depmap.org/portal/

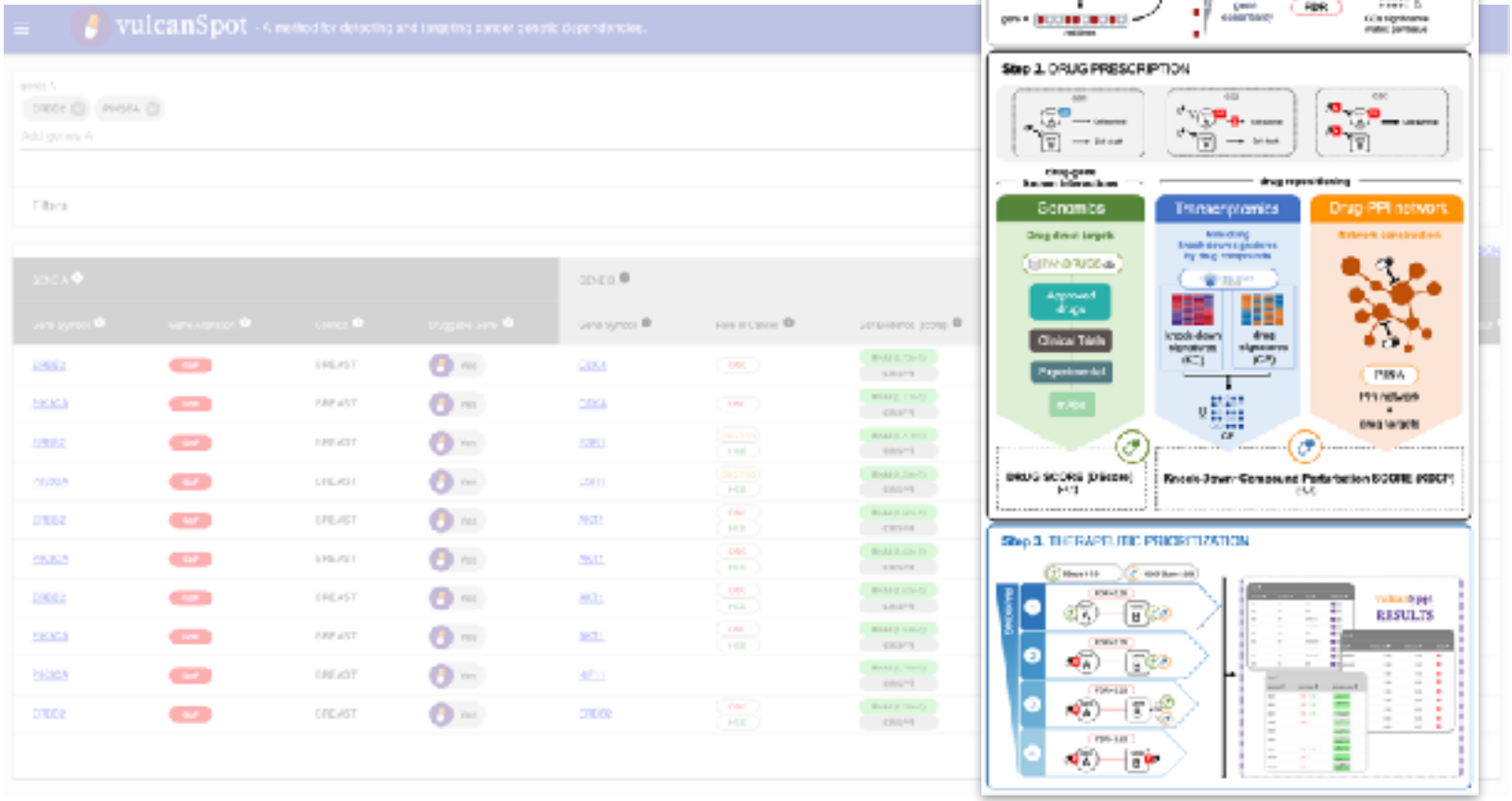
CancerGD https://www.cancergd.org



vulcanspot https://www.vulcanspot.org



vulcanspot https://www.vulcanspot.org



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

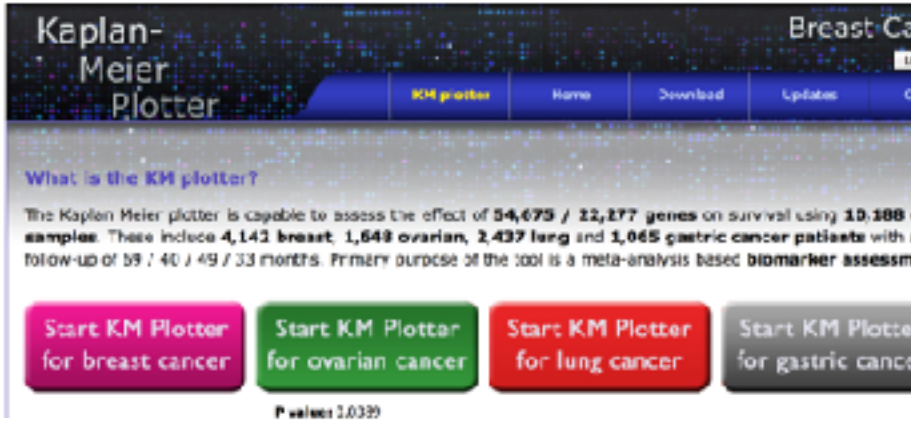
Clinical Variants Annotators

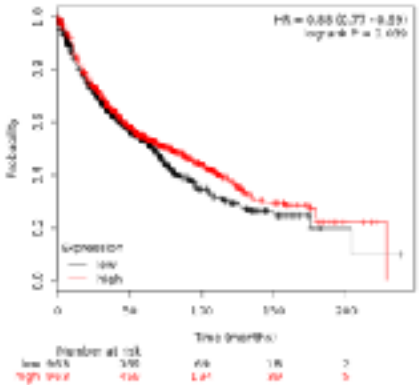
ClinVAR
http://www.ncbi.nlm.nih.gov/clinvar/

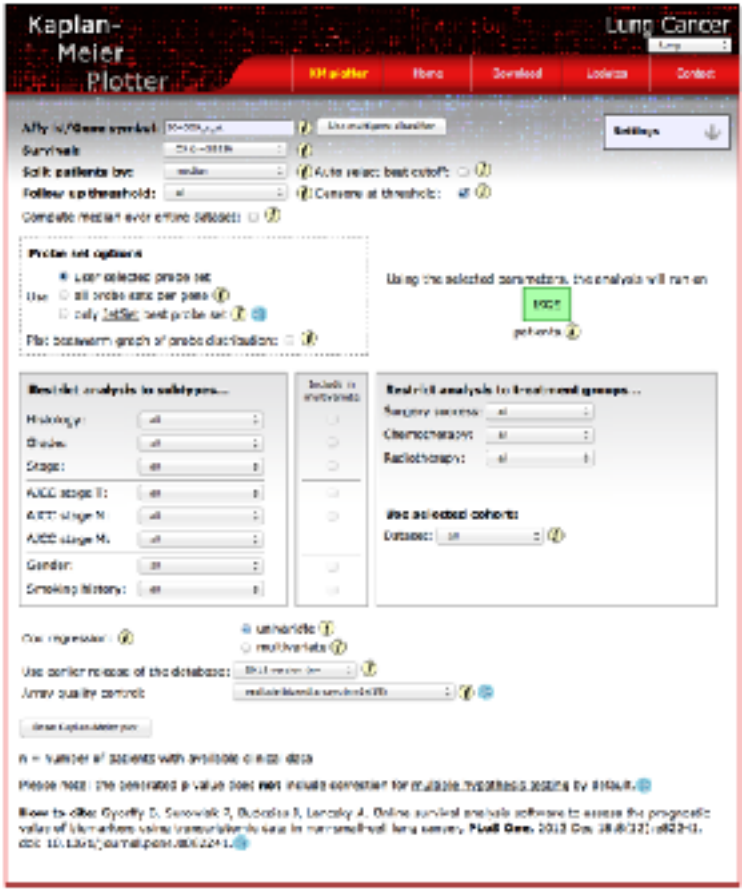
ExAC
http://exac.broadinstitute.org

gnomAD
http://gnomAD.broadinstitute.org

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







Summary

Gene & Protein DBs

Functional Enrichment

Cancer Collections & Browsers

Therapy response

Cancer dependencies

Survival, Clinical & other useful info

Survival Analysis

https://www.tcga-survival.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

Mutational signatures

https://signal.mutationalsignatures.com/

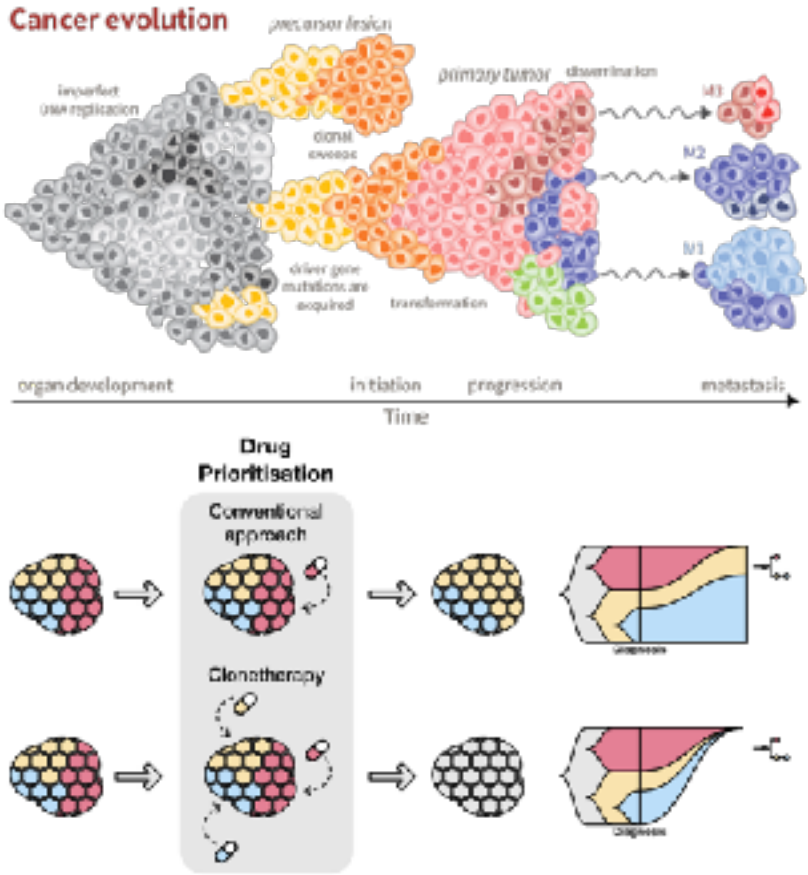
April 2022 Update — Signal presents newly published single and double base substitution signature data derived from more than 12,000 cancer samples recruited via the UK National Health Service. Click here to view all of the cancer signature data on Signal.

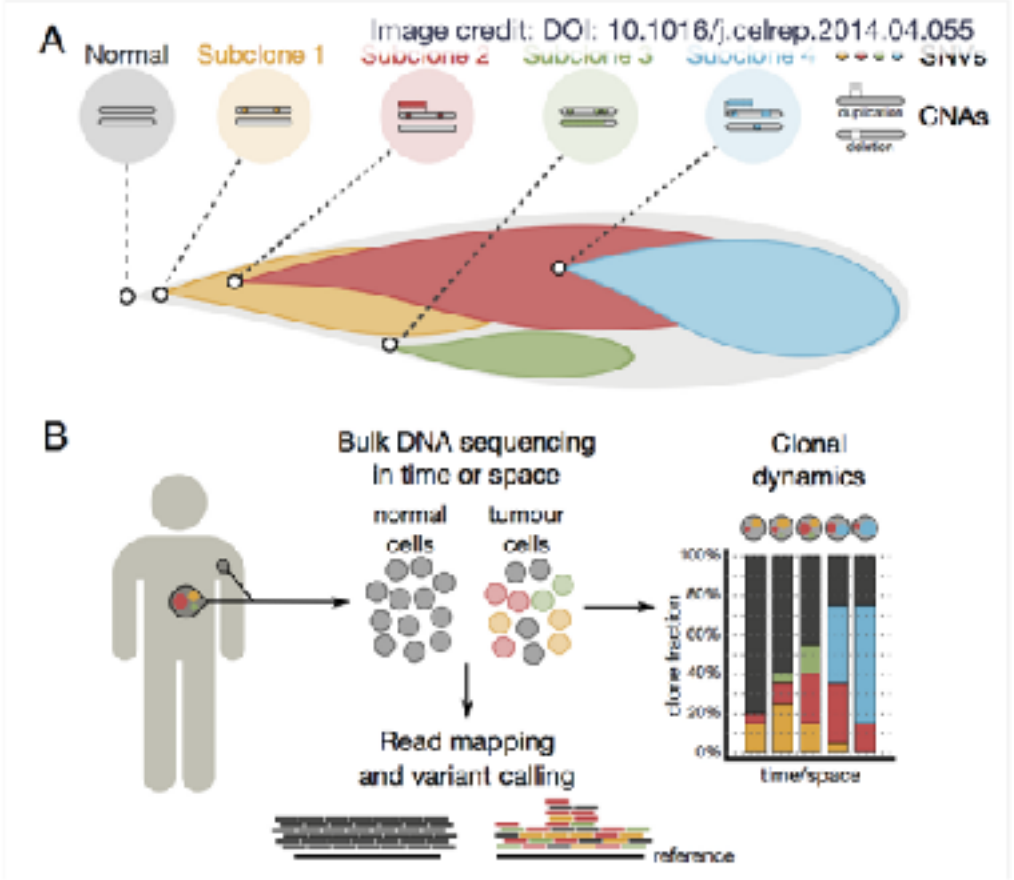
Discover Mutational Signatures

Somatic mutations arise in the DNA of all the cells of the
are the consequence of mutational processes that may b
cellular activities, and are a key component in understand
Different mutational processes generate observable patt
patterns are called mutational signatures.

Tumour clonality

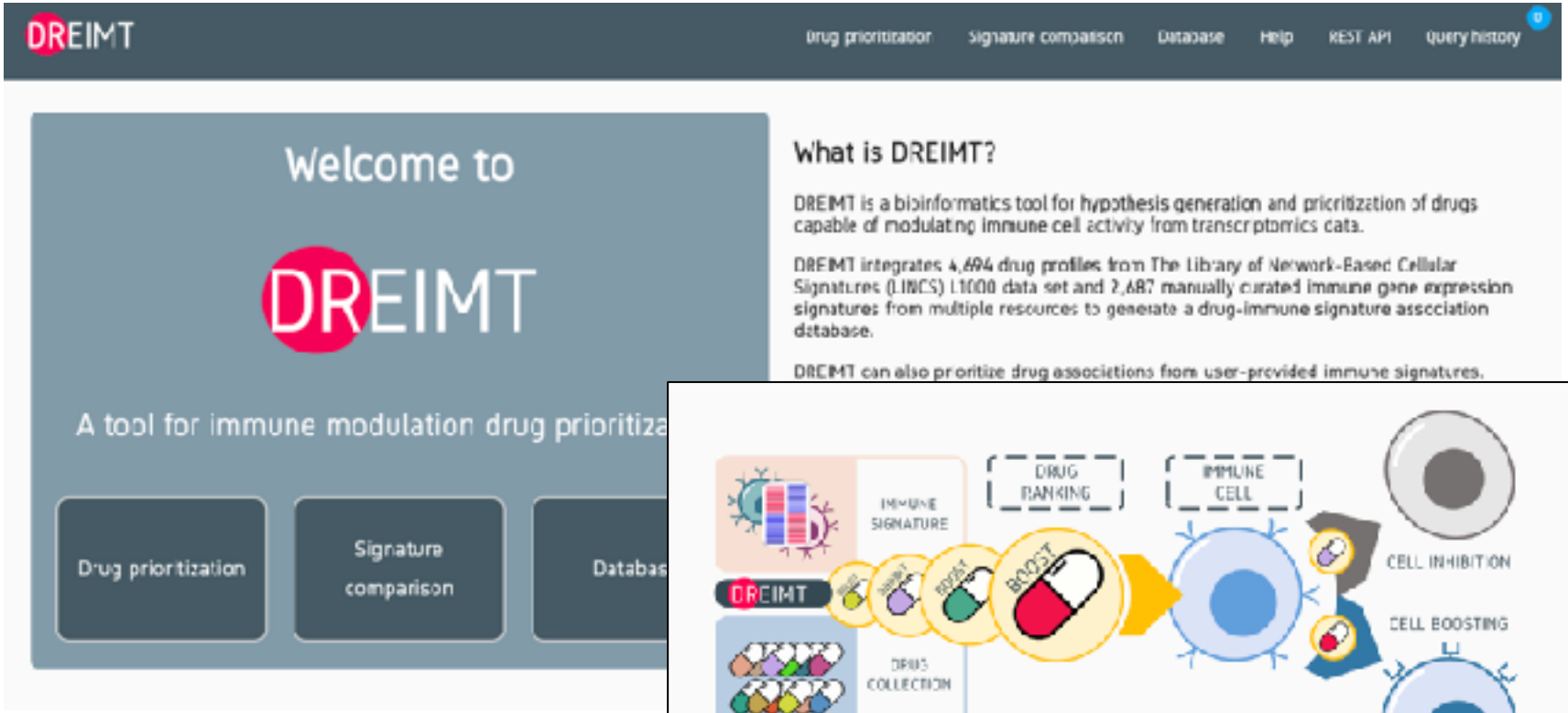
http://tracerx.co.uk/





PyClone2, FastClone, Subclone, etc

Dreimt https://www.dreimt.org/



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!

CNIO **BIOINFORMATICS** UNIT