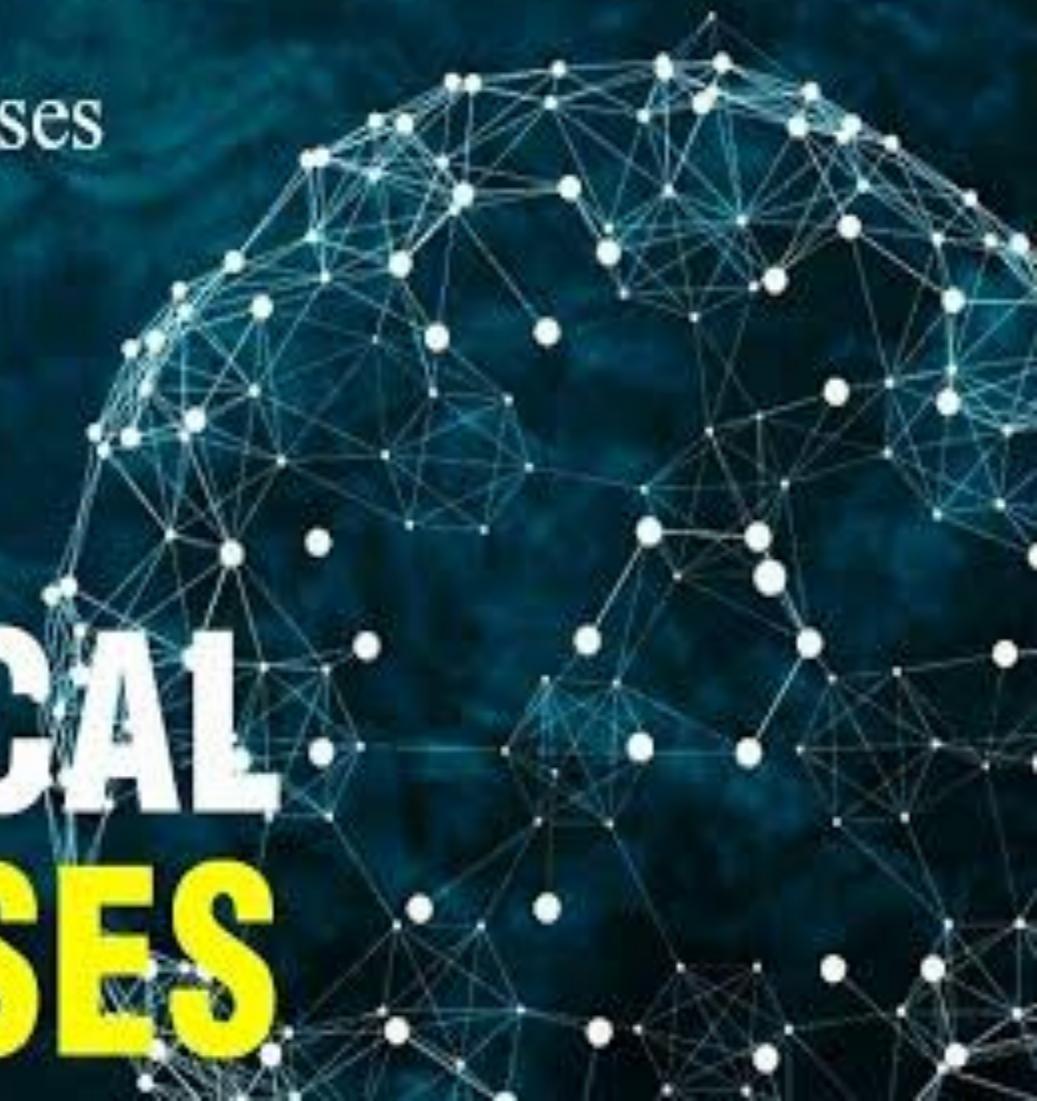


BIOINFORMATICA

Biological Databases

Overview
of
BIOLOGICAL
DATABASES



- Thesis :

- *Omics data integration using deep learning-based method for identification of survival associated subgroups in colon cancer .*



Siamak Salimy



Siamak-salimy



Siamak Salimy

Ph.D. Candidate Laboratory of Systems Biology and Bioinformatics
(LBB), Tehran University

Hamedan, Hamedan Province, Iran · [Contact info](#)

190 connections



Hamedan municipality

OMICS TECHNOLOGIES



Omics technologies represent **high-throughput** assays that are designed to **identify** and **quantify** all the biomolecules of a particular type – DNA, RNA, protein, and metabolite in a given biological sample.

Review Article

Onco-Multi-OMICS Approach: A New Frontier in Cancer Research

Sajib Chakraborty ,¹ Md. Ismail Hosen ,¹ Musaddeque Ahmed,² and Hossain Uddin Shekhar ,¹

NGS based techniques :
Genomics, Epigenomics, and transcriptomics

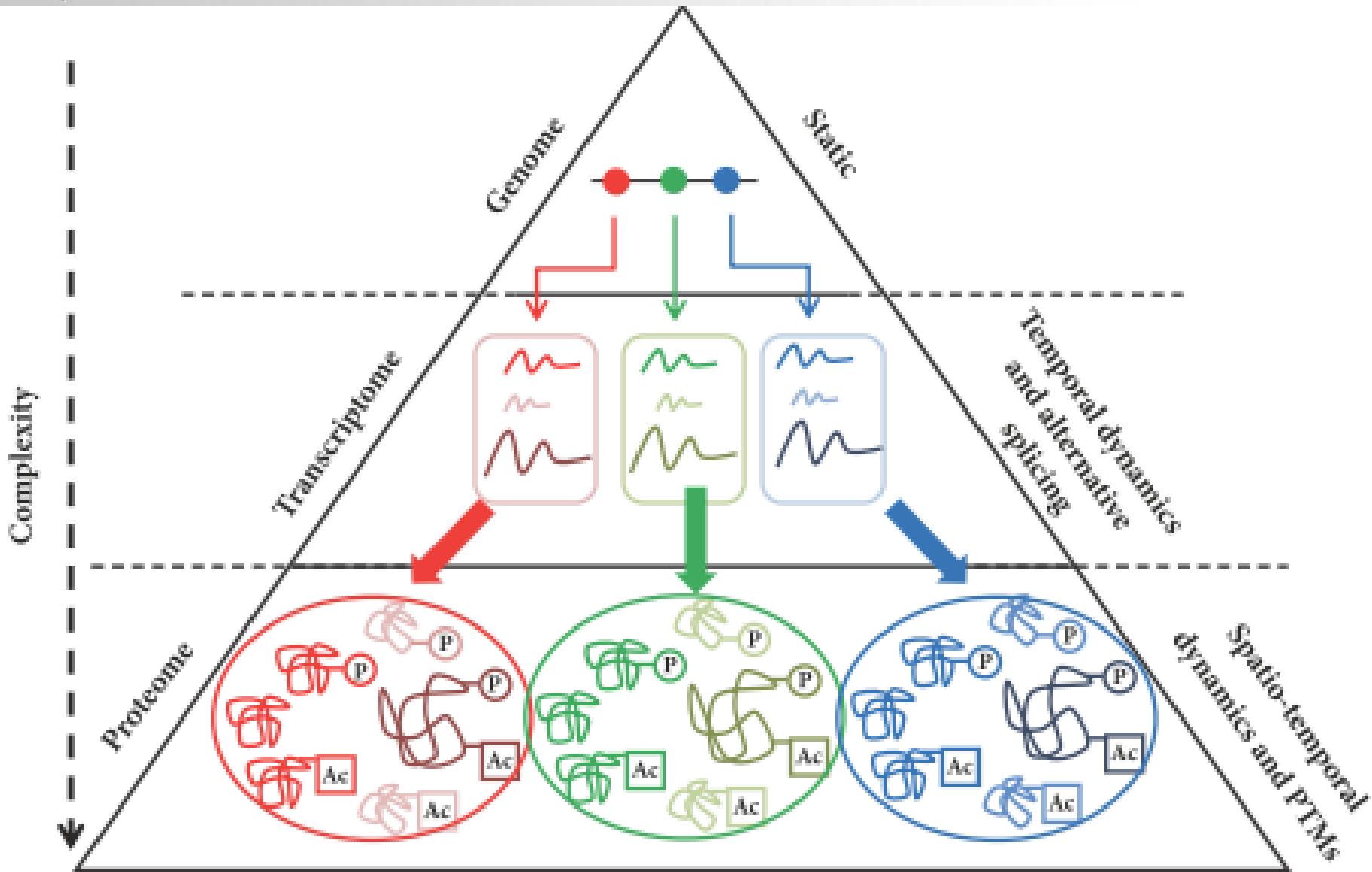
Mass-spectrometry based techniques :
Proteomics and metabolomics .

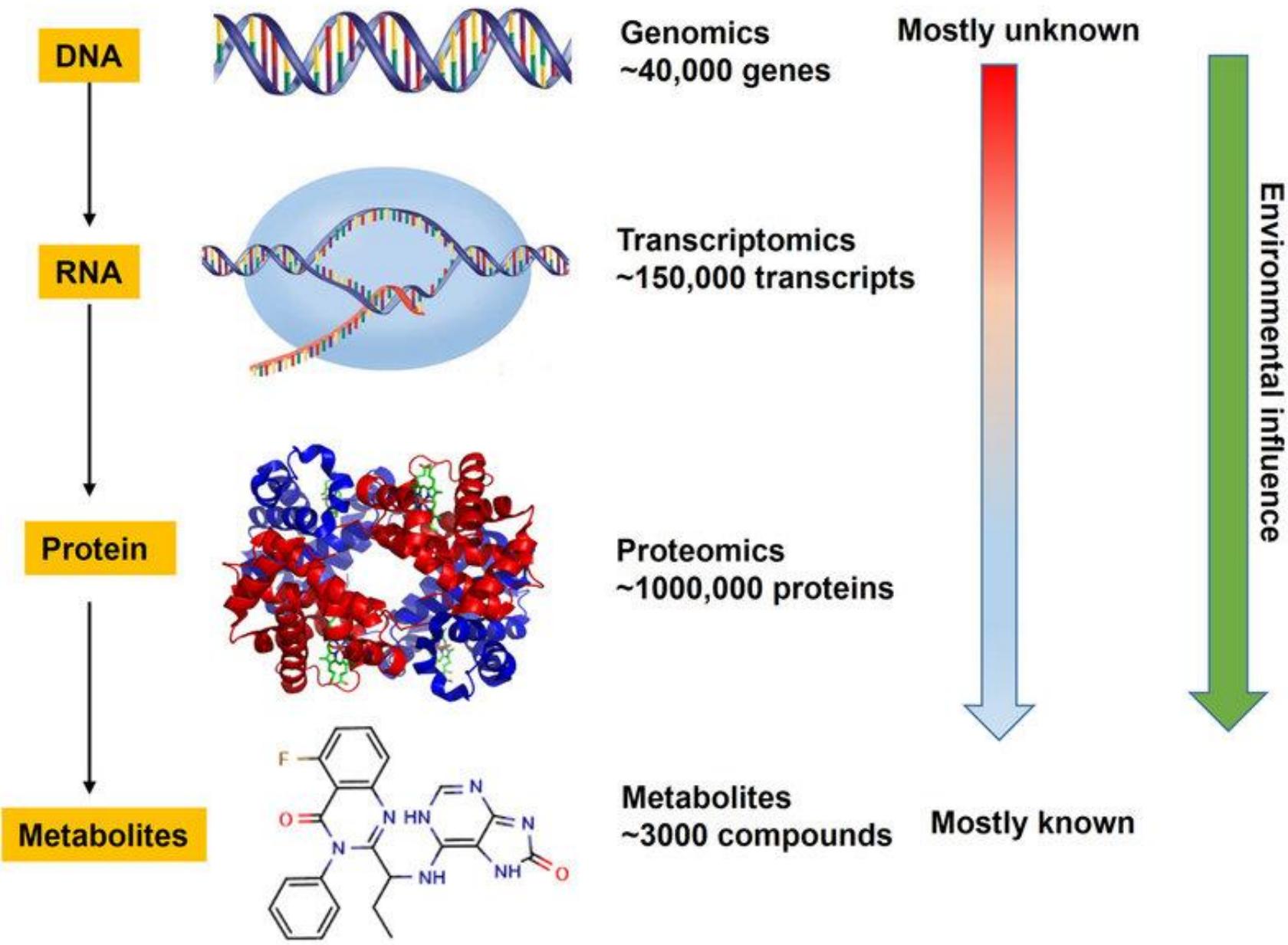


L.B.B

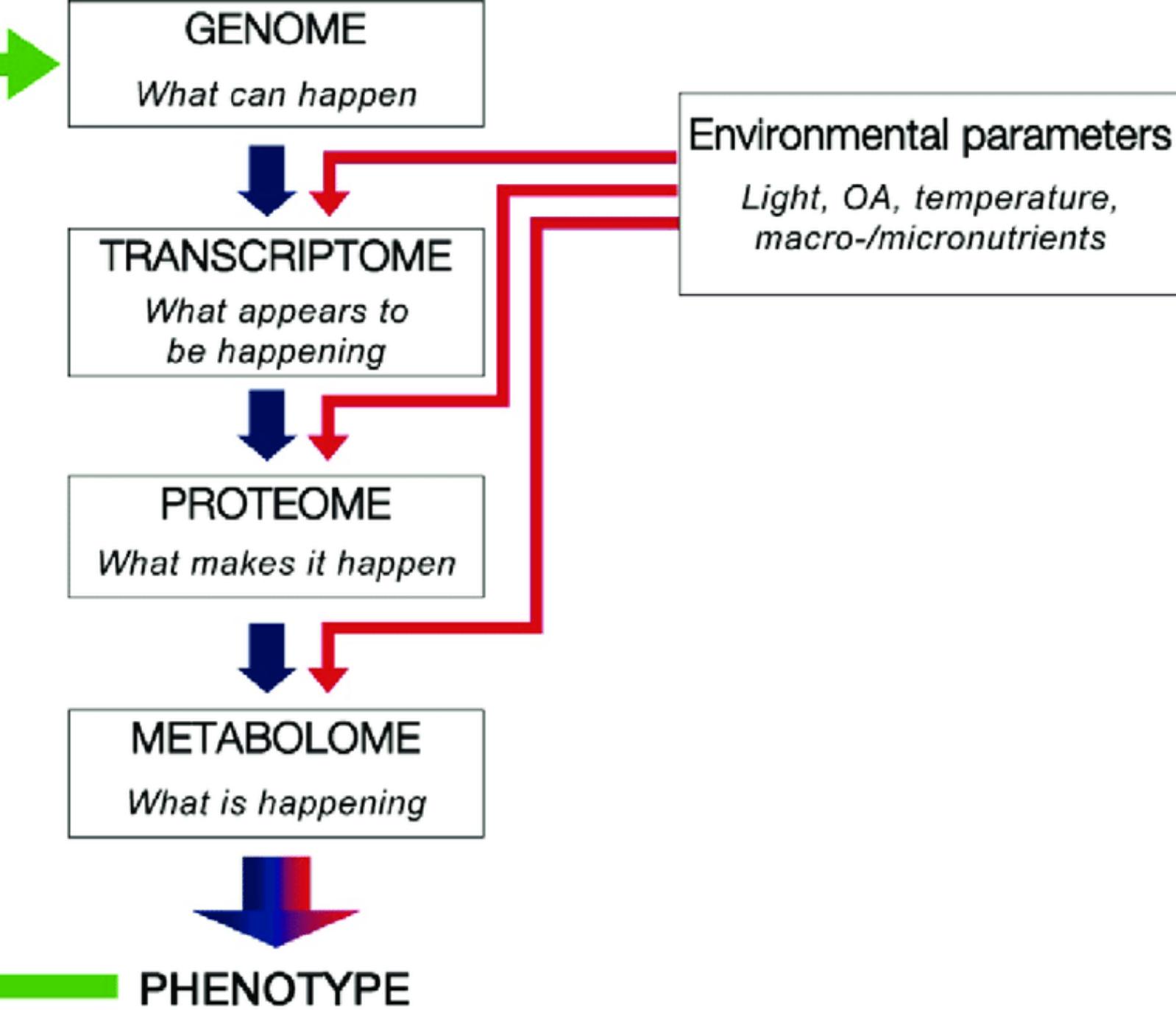
Laboratory of Systems Biology
and Bioinformatics

PYRAMID OF COMPLEXITY





Ecological performance



MULTI-OMICS

HARNESSING MULTI-OMICS TECHNIQUES FOR CANCER RESEARCH

Omics type	Technique	Application
Genomics	NGS	High throughput whole-genome and whole-exome sequencing
Epigenomics	ChIP-seq	Identification of genome-wide DNA binding sites for transcription factors and associated proteins
	DNase1-seq	Identification of the active gene regulatory elements across genome
	FAIRE-seq	Identification of the DNA regions having regulatory activity
	ChiRP-seq	Detection of genomic locations of the ncRNAs, such as lncRNAs, and their bound proteins
	WG bisulfite/array-based sequencing	Determination of the methylation pattern throughout the genome
Transcriptomics	RNA-seq	Identifications and quantification of novel transcripts- including mRNA, miRNA, and other regulatory RNAs
Proteomics	LC-MS/MS based mass-spectrometry	Identification and quantification of proteins abundances on various biological conditions
	RPPA	Quantification of proteins abundances on various biological conditions
Metabolomics	LC-MS based mass-spectrometry	Identifications and quantification of selected molecules involved in metabolic pathways

NGS BASED GENOMICS

- NGS based genomics studies are typically designed to **analyze the DNA sequence** of both **coding** and **non-coding** regions in a genome-wide manner.

NGS BASED EPIGENOMICS

- ChIP-seq (chromatin immunoprecipitation),
- DNase1-seq (DNase I hypersensitive sites – sequencing)
- FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements – sequencing) assay for mapping the **DNA-protein** interactions or chromatin accessibility.
- ChiRP-seq (Chromatin Isolation by RNA Purification) for mapping DNA-RNA interaction .
- Whole-genome bisulfite/array-based sequencing for mapping DNA methylation

NGS-DRIVEN TRANSCRIPTOMICS

NGS-driven transcriptomics (e.g., RNA-seq) techniques are used for identification and quantification of RNA molecules including mRNA, miRNA, and other regulatory RNAs in a genome-wide manner .



L.B.B

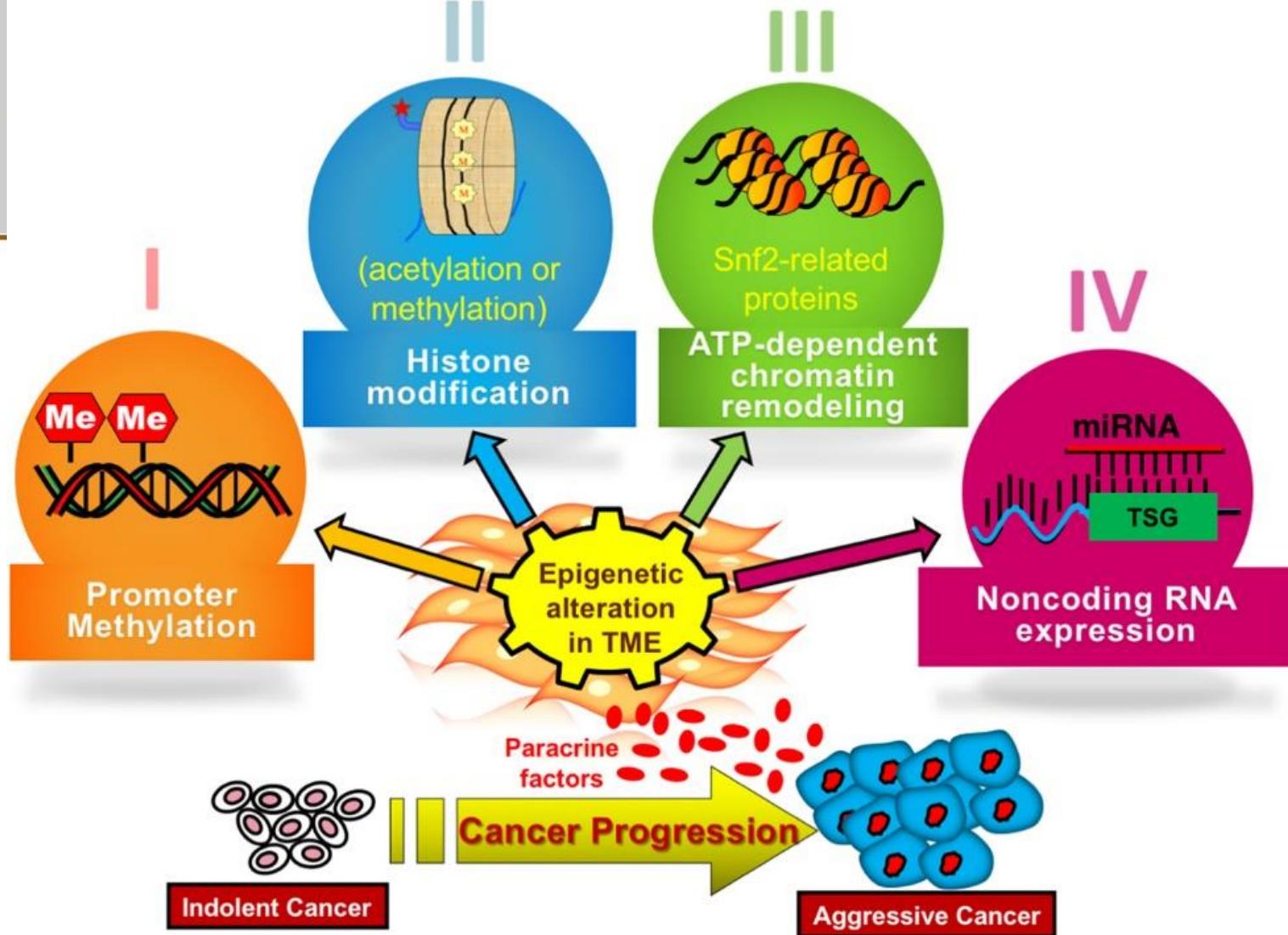
Laboratory of Systems Biology
and Bioinformatics

MASS-SPECTROMETRY BASED (LC-MS/MS) TECHNIQUES

(LC-MS/MS) based techniques

Are used to **identify** or **quantify** the proteins (proteomics) and **metabolites** (metabolomics) in a high-throughput manner .

Other techniques such as reverse-phase protein arrays (**RPPAs**) can also be used to **quantify** different protein molecules using **antibodies** .



Alterations at different regulatory layers often lead to the emergence of a cancerous phenotype .

Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry

Maria Berdasco¹ and Manel Esteller^{1,2,*}

¹Cancer Epigenetics Group, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Institute for Biomedicine, L'Hospitalet de Llobregat, 08907 Barcelona, Catalonia, Spain

²Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain

*Correspondence: mesteller@iconcologia.net

DOI 10.1016/j.devcel.2010.10.005

Genetic mutations or epigenetic alterations may drive
Tumorigenesis .

Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies

Yi-Zhou Jiang,^{1,14} Ding Ma,^{1,14} Chen Suo,^{2,3,14} Jinxiu Shi,^{4,14} Mengzhu Xue,^{5,14} Xin Hu,^{1,14} Yi Xiao,¹ Ke-Da Yu,¹ Yi-Rong Liu,¹ Ying Yu,² Yuanting Zheng,² Xiangnan Li,² Chenhui Zhang,⁴ Pengchen Hu,⁴ Jing Zhang,⁴ Qi Hua,⁴ Jiyang Zhang,² Wanwan Hou,² Luyao Ren,² Ding Bao,² Bingying Li,² Jingcheng Yang,² Ling Yao,¹ Wen-Jia Zuo,¹ Shen Zhao,¹ Yue Gong,¹ Yi-Xing Ren,¹ Ya-Xin Zhao,¹ Yun-Song Yang,¹ Zhenmin Niu,⁴ Zhi-Gang Cao,^{1,13} Daniel G. Stover,⁶ Claire Verschraegen,⁶ Virginia Kaklamani,⁷ Anneleen Daemen,⁸ John R. Benson,⁹ Kazuaki Takabe,¹⁰ Fan Bai,¹¹ Da-Qiang Li,¹ Peng Wang,^{12,*} Leming Shi,^{2,*} Wei Huang,^{4,*} and Zhi-Ming Shao^{1,15,*}

Dysregulation genes and proteins in many cancer types .

SINGLE-OMICS VS MULTI-OMICS

These **single-omics datasets** *fail* to fully untangle the complexity of a disease like **cancer**.

Instead, cancer can be better understood by integrating the **multi-omics** datasets rather than analyzing **single-omics** datasets in isolation .

OVERVIEW

- CANCER-SPECIFIC MULTI-OMICS DATA RESOURCES
- CANCER-SPECIFIC SINGLE-OMICS RESOURCES
- STRATEGIES FOR MULTI-OMICS DATA INTEGRATION
- ALGORITHMS OF DATA-INTEGRATION
- CURRENT CHALLENGES OF MULTI-OMICS DATA INTEGRATION
- FUTURE OF CANCER-RESEARCH

ONLINE RESOURCES OF OMICS-DATA

Online single and multi-omics resources.

1 - CANCER-SPECIFIC DATA RESOURCES .

Multi-Omics

Single-Omics

2 - Generalized DATA RESOURCES .

CANCER-SPECIFIC DATA RESOURCES .

* **MULTI-OMICS**



1-800-4-CANCER Live Chat Publications Dictionary

ABOUT CANCER CANCER TYPES RESEARCH GRANTS & TRAINING NEWS & EVENTS ABOUT NCI

search



[Home](#) > [About NCI](#) > [NCI Organization](#) > [CCG](#) > [Research](#) > [Structural Genomics](#)



TCGA

Program History +

TCGA Cancers Selected for Study

Publications by TCGA

Using TCGA +

Contact

The Cancer Genome Atlas Program

The Cancer Genome Atlas (TCGA), a landmark [cancer genomics](#) program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types. This joint effort between NCI and the National Human Genome Research Institute began in 2006, bringing together researchers from diverse disciplines and multiple institutions.

Over the next dozen years, TCGA generated over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data. The data, which has already led to improvements in our ability to diagnose, treat, and prevent cancer, will remain [publicly available](#) for anyone in the research community to use.



ONLINE RESOURCES OF OMICS-DATA

- Over 20,000 primary cancers and matched normal samples 33 different cancer types.
- Over 2.5 petabytes of omics data
- Genomic, epigenomic, transcriptomic, and proteomic
- NGS based (SNVs), DNA methylation, copy number alterations (CNAs), and mRNA/miRNA expression
- The highest number of samples were collected for the Glioma cohort (GBMLGG, N = 1129) followed by the Breast invasive carcinoma cohort (BRCA, N = 1098) and pan-kidney cohort (KIPAN, N = 941)
- The lowest number of samples were obtained for cholangiocarcinoma (CHOL, N = 45), lymphoid neoplasm diffuse large B-cell lymphoma (DLBC, N = 48), and uterine carcinosarcoma (UCS, N = 57).

NON-TUMOR ADJACENT TISSUE (NT)

Understanding the mechanisms of carcinogenesis requires a rigorous comparison of the tumor multi-omics data with normal controls .

non-tumor adjacent tissue (NT)

The number of NT samples is not as high as tumor samples (TP) in TCGA.

Glioma (GBMLGG), only 51 CNV , 5 RNA , and 2 methylation

The highest number : pan-kidney cohort (KIPAN) (N = 562)

The lowest number : mesothelioma (N = 1)

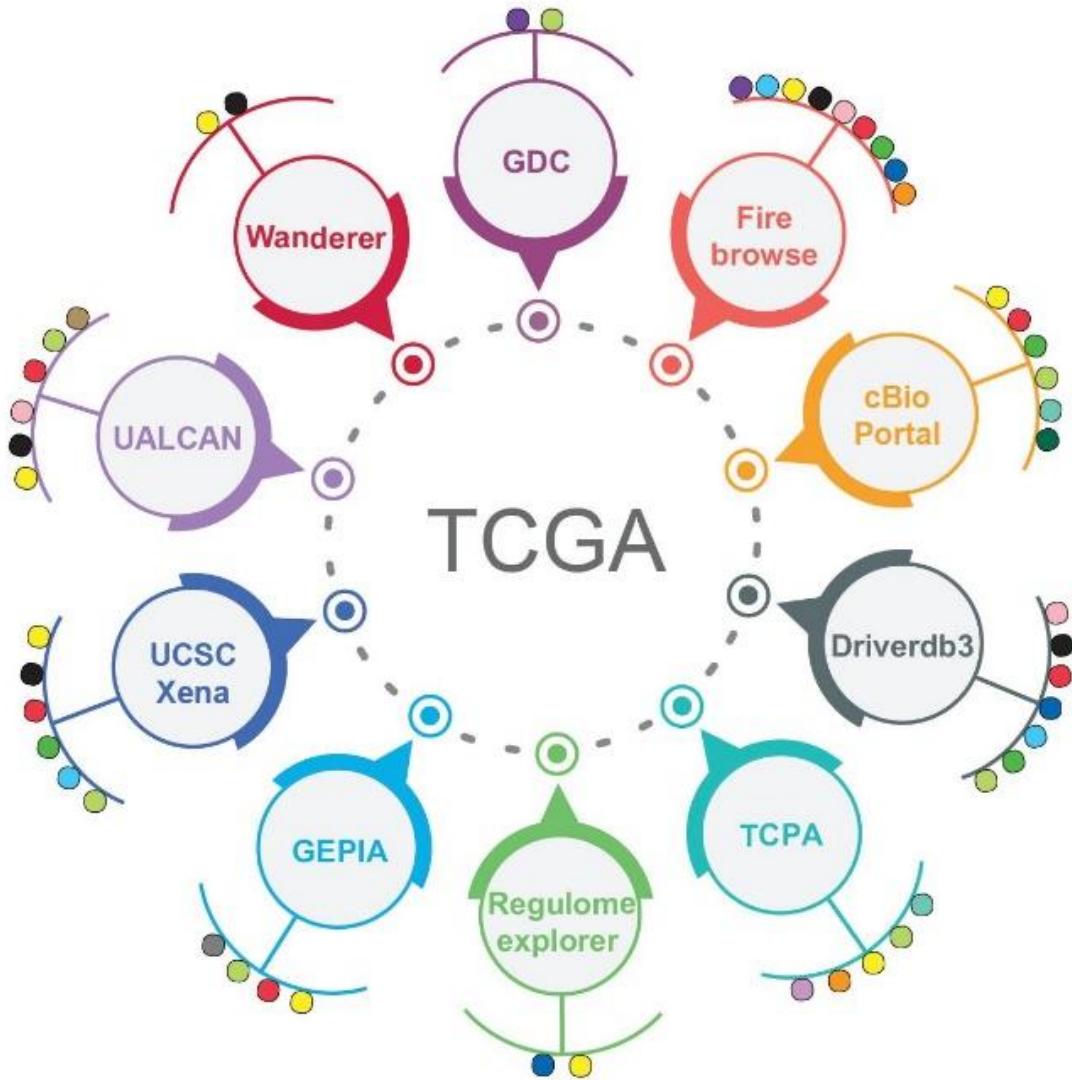


WEAKNESS

The reduced number of
patient-matched NT samples .

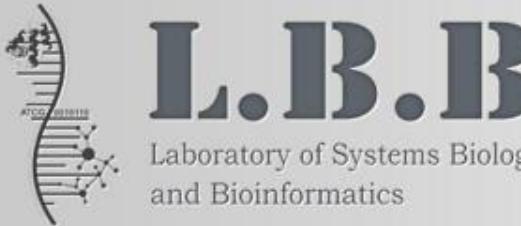


L.B.B ANALYSES AND VISUALIZATION PORTALS OF TCGA DATA



- Clinical analysis
- Copy number analysis
- Correlation analysis
- Methylation analysis
- Survival analysis
- Network analysis
- Co-expression analysis
- miRseq analysis
- Gene expression analysis
- Mutation analysis
- Pathway analysis
- RPPA analysis
- Protein expression
- Differentially expressed protein
- Principle component analysis

General information			Relative strengths and limitations					
Name	URL	Database coverage	Processed and normalized data	Clinical data	Tumor adjacent non-tumor data	Pan-cancer analysis	Multiple database integration	Multomics analysis
GDC	https://portal.gdc.cancer.gov/	TCGA, Target, GENIE		✓	✓		✓	
Firebrowse	http://firebrowse.org/	TCGA	✓	✓	✓	✓		✓
cBioPortal	https://www.cbioportal.org/	TCGA, CCLE, Target	✓	✓		✓	✓	✓
Driverdb3	http://driverdb.tms.cmu.edu.tw/cancer	TCGA			✓	✓		✓
TCPA	https://tcpaportal.org/tcpa/	TCGA		✓		✓		✓
Regulome explorer	http://explorer.cancerregulome.org/all_pairs/	TCGA		✓		✓		✓
GEPIA	http://gepia.cancer-pku.cn/	TCGA, GTEX		✓	✓	✓	✓	✓
UCSC Xena	https://xena.ucsc.edu/	TCGA, GTEX, Target	✓	✓	✓	✓	✓	✓
UALCAN	http://ualcan.path.uab.edu/	TCGA			✓	✓		✓
Wanderer	http://maplab.imppc.org/wanderer/	TCGA			✓			



TCGA PORTALS

Harmonized Cancer Datasets

Genomic Data Commons Data Portal

Get Started by Exploring:

- [Projects](#)
- [Exploration](#)
- [Analysis](#)
- [Repository](#)

e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary Data Release 30.0 - September 23, 2021

PROJECTS
 70

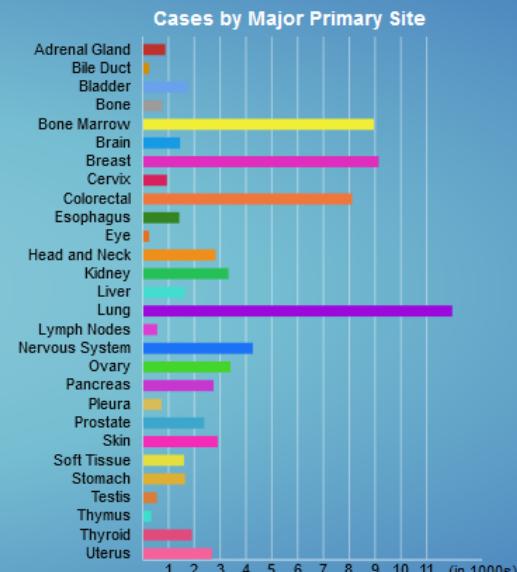
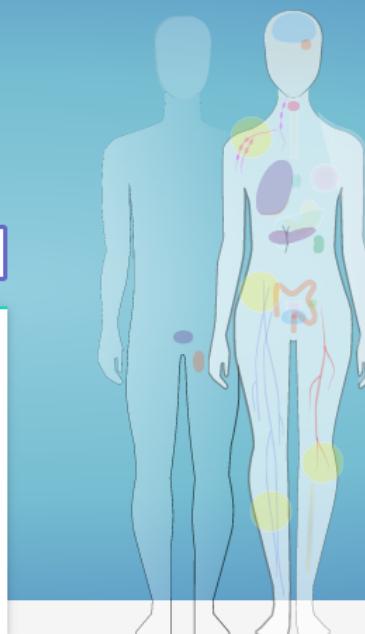
PRIMARY SITES
 67

CASES
 85,414

FILES
 619,488

GENES
 23,621

MUTATIONS
 3,599,319



GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:


[Data Portal](#)


[Website](#)


[API](#)


[Data Transfer Tool](#)


[Documentation](#)


[Data Submission Portal](#)


[Legacy Archive](#)


[Publications](#)

[Site Home](#) | [Policies](#) | [Accessibility](#) | [FOIA](#) | [Support](#)

GDC EXPLORATION

Cases Clinical Genes Mutations

 Primary Site IS bronchus and lung AND Tissue Or Organ Of Origin IN (lower lobe, lung lung, nos ...)

Cases (11,961)
Genes (21,668)
Mutations (469,493)
OncoGrid
[View Files in Repository](#)

Primary Site
Project
Disease Type
Gender
Vital Status

Showing 1 - 20 of 11,961 cases

<input type="checkbox"/> Case ID	Project	Primary Site	Gender	Files	Available Files per Data Category								# Mutations	# Genes	Slides
					Seq	Exp	SNV	CNV	Meth	Clinical	Bio				
<input type="checkbox"/> TCGA-18-3409	TCGA-LUSC	Bronchus and lung	Male	51	3	3	16	7	2	8	12	4,500	3,329	--	
<input type="checkbox"/> TCGA-90-A4ED	TCGA-LUSC	Bronchus and lung	Male	53	4	5	16	7	1	7	12	2,866	2,293	--	
<input type="checkbox"/> HCM-BROD-0027-C34	HCM-CMDC	Bronchus and lung	Female	80	12	10	38	4	6	1	1	2,812	2,256	--	
<input type="checkbox"/> TCGA-21-1079	TCGA-LUSC	Bronchus and lung	Male	53	4	5	16	7	1	7	12	2,605	2,168	--	
<input type="checkbox"/> TCGA-55-8506	TCGA-LUAD	Bronchus and lung	Female	59	4	5	16	7	1	7	18	2,339	2,020	--	
<input type="checkbox"/> TCGA-05-4382	TCGA-LUAD	Bronchus and lung	Male	61	6	5	16	7	1	8	18	2,312	1,905	--	
<input type="checkbox"/> TCGA-L9-A7SV	TCGA-LUAD	Bronchus and lung	Male	59	4	5	16	7	1	7	18	2,107	1,779	--	
<input type="checkbox"/> TCGA-MN-A4N4	TCGA-LUAD	Bronchus and lung	Male	59	4	5	16	7	1	7	18	1,969	1,686	--	
<input type="checkbox"/> TCGA-86-A4JF	TCGA-LUAD	Bronchus and lung	Male	59	4	5	16	7	1	7	18	1,852	1,675	--	
<input type="checkbox"/> TCGA-55-7994	TCGA-LUAD	Bronchus and lung	Male	60	4	5	16	9	1	7	18	1,990	1,669	--	
<input type="checkbox"/> C3L-00144	CPTAC-3	Bronchus and lung	Male	63	15	14	20	2	6	0	0	1,983	1,657	--	
<input type="checkbox"/> TCGA-49-AARE	TCGA-LUAD	Bronchus and lung	Female	59	4	5	16	7	1	7	18	1,937	1,645	--	
<input type="checkbox"/> TCGA-78-7155	TCGA-LUAD	Bronchus and lung	Male	61	6	5	16	7	1	7	18	1,963	1,627	--	

Cohort	Clinical	SNV	CNV	Methylation	mRNA-seq	miRSeq	RPPA	Proteome	Phospho-proteome	Glyco-proteome
ACC	92	92	92	80	80	80	46			
BLCA	412	412	412	412	412	409	344			
BRCA	1097	1044	1098	1095	1097	1078	887	233		233
CESC	307	305	302	307	307	307	173			
CHOL	45	51	36	36	36	36	30			
COAD	458	433	460	458	459	406	360	164		101
Cohort	Clinical	SNV	CNV	Methylation	mRNA-seq	miRSeq	RPPA	Proteome	Phospho-proteome	Glyco-proteome
ACC	92	92	92	80	80	80	46			
BLCA	412	412	412	412	412	409	344			
BRCA	1097	1044	1098	1095	1097	1078	887	233		233
CESC	307	305	302	307	307	307	173			
CHOL	45	51	36	36	36	36	30			
COAD	458	433	460	458	459	406	360	164		101
COADREAD	629					549	491			
DLBC	48	37	50	48	48	47	33			
ESCA	185	184	185	185	184	184	126			
GBM	595	396	599	423	166	0	238	100		100
OV	337	448	337	662	432	430	420	268	182	122
PAAD	185	183	185	184	178	178	123			
PCPG	179	179	179	179	179	179	80			
PRAD	499	498	498	498	498	494	352			
READ	171	158	167	165	167	143	131	30		
SARC	261	255	261	261	261	259	223			
SKCM	470	470	470	470	469	448	353			
STAD	443	441	443	443	439	436	357			
STES	628					620	483			
TGCT	134	150	150	150	150	150	118			
THCA	503	496	505	507	507	502	222			
THYM	124	123	124	124	124	124	90			
UCEC	548	542	558	559	559	538	440	104		104
UCS	57	57	57	57	57	56	48			
UVM	80	80	80	80	80	80	12			
Total	11196	10418	11124	10943	10558	10156	7429	1135	921	122

ACC, adrenocortical carcinoma;
LAML, acute myeloid leukemia;
BLCA, bladder urothelial carcinoma;
BRCA, breast invasive carcinoma;
CESC, cervical squamous cell carcinoma
and endocervical adenocarcinoma;
CHOL, cholangiocarcinoma;
COADREAD, colorectal adenocarcinoma;
DLBC, lymphoid neoplasm diffuse
large B-cell lymphoma;
ESCA, esophageal carcinoma;
GBM, glioblastoma multiforme;
HNSC, head and neck squamous cell carcinoma;
LGG, brain lower grade glioma;
LIHC, liver hepatocellular carcinoma;
LUAD, lung adenocarcinoma;
LUSC, lung squamous cell carcinoma;
SKCM, skin cutaneous melanoma;

MESO, mesothelioma;
OV, ovarian serous cystadenocarcinoma;
PCPG, pheochromocytoma and paraganglioma;
PAAD, pancreatic adenocarcinoma;
SARC, sarcoma;
STAD, stomach adenocarcinoma;
TGCT, testicular germ cell tumors;
THYM, thymoma;
THCA, thyroid carcinoma;
UCEC, uterine corpus endometrial carcinoma;
UCS, uterine carcinosarcoma;
UVM, uveal melanoma;
KIRC, kidney renal clear cell carcinoma;
KICH, kidney chromophobe;
KIRP, kidney renal papillary cell carcinoma)

FIREBROWSE



[HOME](#)
[BROAD GDAC](#)
[WEB API](#)
[FAQ](#)
[SAMPLES REPORT](#)
[AWG RESULTS](#)
[OLD RUNS](#)
[TUTORIAL](#)
[RELEASE NOTES](#)
[CONTACT](#)

[View Expression Profile](#)

[Enter cohort abbrev](#)

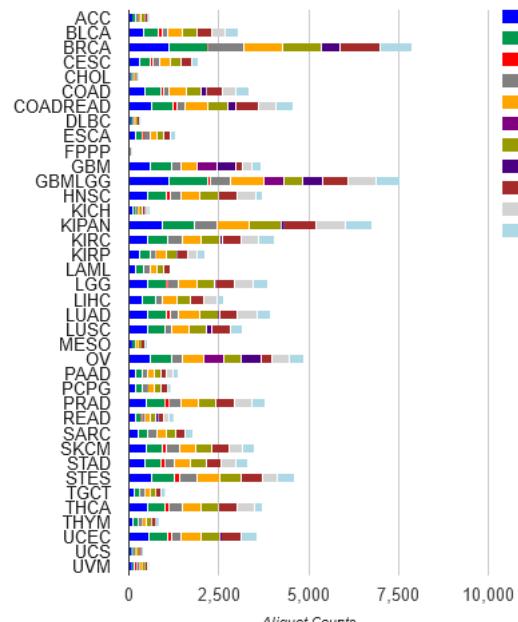


[View Analysis Profile](#)

SELECT COHORT

- █ Clinical Analyses
- █ CopyNumber Analyses
- Correlations Analyses
- █ miR Analyses
- █ miRseq Analyses
- █ mRNA Analyses
- █ mRNASeq Analyses
- Mutation Analyses
- Pathway Analyses
- █ RPPA Analyses

TCGA data version 2016_01_28



Cohort	Clinical	SNP6 CopyNum	LowPass DNaseq CopyNum	Mutation Annotation File	methylation	miR	miRSeq	mRNA	mRNASeq	raw Mutation Annotation File	Reverse Phase Protein Array
ACC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
BLCA	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
BRCA	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
CESC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
CHOL	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
COAD	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
COADREAD	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
DLBC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
ESCA	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
FPPP	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
GBMLGG	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
GBMLGG	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
HNSC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
KICH	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
KIRP	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
KIRC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
KIRP	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
LAML	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
LGG	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
LIHC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
LIHC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
LUAD	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
LUSC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
MESO	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
OV	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
PAAD	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
PCPG	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
PRAD	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
READ	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
SARC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
SKCM	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
STAD	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
STES	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
TGCT	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
THCA	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
THYM	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
UCEC	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
UCS	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100
UVM	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100	~100

Clinical Analyses

Aggregate AnalysisFeatures

Correlate Clinical vs CopyNumber Arm

Correlate Clinical vs CopyNumber Focal

Correlate Clinical vs Methylation

Correlate Clinical vs miRseq

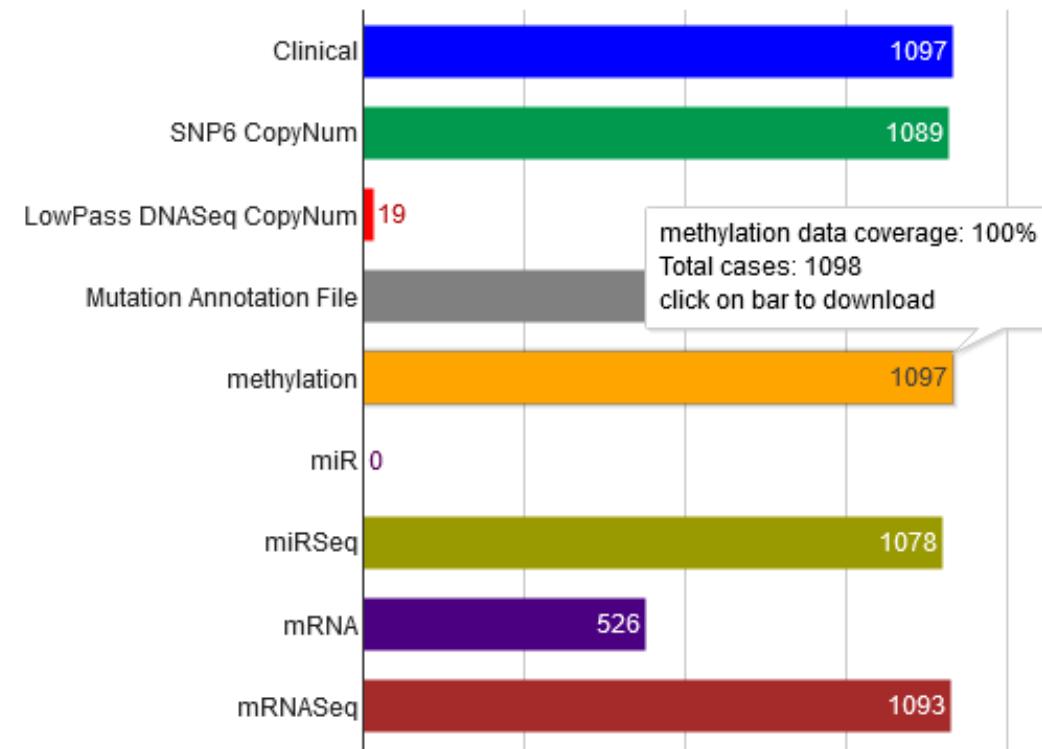
Correlate Clinical vs Molecular Subtypes

Correlate Clinical vs mRNA

Correlate Clinical vs mRNAseq

Correlate Clinical vs Mutation

Correlate Clinical vs Mutation APOBEC Categorical

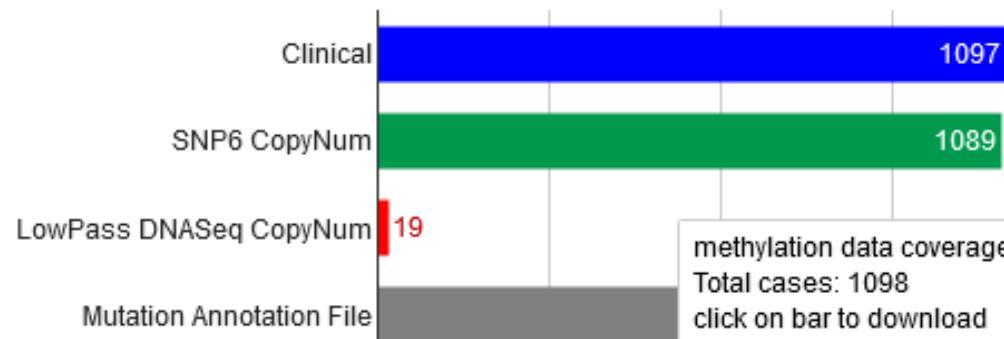


Clinical Analyses

Aggregate AnalysisFeatures

Correlate Clinical vs CopyNumber Arm

Correlate Clinical vs CopyNumber Focal



FIREBROWSE EXPANSION

- **ViewGene** : a graphical tool for polymorphism visualization and characterization .
- **iCoMut** : a web tool for visual summary of mutations in cancer cohorts .

Volume 12 Supplement 5

Selected articles from the 8th Translational Bioinformatics Conference: Medical Genomics

Research | Open Access | Published: 11 July 2019

CoMutPlotter: a web tool for visual summary of mutations in cancer cohorts

Po-Jung Huang, Hou-Hsien Lin, Chi-Ching Lee, Ling-Ya Chiu, Shao-Min Wu, Yuan-Ming Yeh, Petrus Tang, Cheng-Hsun Chiu, Ping-Chiang Lyu & Pei-Chien Tsai

BMC Medical Genomics 12, Article number: 99 (2019) | Cite this article

8353 Accesses | 3 Citations | 7 Altmetric | Metrics

viewGene: A Graphical Tool for Polymorphism Visualization and Characterization

Carl Kashuk,¹ Sanghamitra SenGupta,² Evan Eichler,² and Aravinda Chakravarti^{1,3}

¹ McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA; ² Department of Genetics, Case Western Reserve University, Cleveland 44106, Ohio, USA


[Query](#)
[Quick Search Beta!](#)
[Download](#)

Please cite: Cerami et al., 2012 & Gao et al., 2013

Select Studies for Visualization & Analysis:

0 studies selected (0 samples)

Search...

PanCancer Studies

8

Quick select:

[TCGA PanCancer Atlas Studies](#)
[Curated set of non-redundant studies](#)

Pediatric Cancer Studies

13

Immunogenomic Studies

8

Cell lines

3

Adrenal Gland

3

Ampulla of Vater

1

Biliary Tract

13

Bladder/Urinary Tract

17

Bone

2

Bowel

12

Breast

22

CNS/Brain

20

Cervix

2

Esophagus/Stomach

17

PanCancer Studies

<input type="checkbox"/> MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017)	10945 samples			
<input type="checkbox"/> Metastatic Solid Cancers (UMich, Nature 2017)	500 samples			
<input type="checkbox"/> MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018)	249 samples			
<input type="checkbox"/> SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018)	141 samples			
<input type="checkbox"/> TMB and Immunotherapy (MSKCC, Nat Genet 2019)	1661 samples			
<input type="checkbox"/> Tumors with TRK fusions (MSK, Clin Cancer Res 2020)	106 samples			
<input type="checkbox"/> Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020)	24146 samples			
<input type="checkbox"/> China Pan-cancer (OrigiMed2020)	10194 samples			

Pediatric Cancer Studies

<input type="checkbox"/> Pediatric Preclinical Testing Consortium (CHOP, Cell Rep 2019)	261 samples			
<input type="checkbox"/> Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018)	1978 samples			
<input type="checkbox"/> Pediatric Rhabdoid Tumor (TARGET, 2018)	72 samples			
<input type="checkbox"/> Pediatric Wilms' Tumor (TARGET, 2018)	657 samples			
<input type="checkbox"/> Pediatric Acute Myeloid Leukemia (TARGET, 2018)	1025 samples			
<input type="checkbox"/> Pediatric Neuroblastoma (TARGET, 2018)	1089 samples			
<input type="checkbox"/> Pediatric Pan-Cancer (DKFZ, Nature 2017)	961 samples			
<input type="checkbox"/> Pediatric Pan-cancer (Columbia U, Genome Med 2016)	103 samples			
<input type="checkbox"/> Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016)	73 samples			
<input type="checkbox"/> Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)	93 samples			

0 studies selected (0 samples)

[Query By Gene](#)

OR

[Explore Selected Studies](#)

What's New

...

@cbioportal



Example Queries

- Primary vs. metastatic prostate cancer
- RAS/RAF alterations in colorectal cancer
- BRCA1 and BRCA2 mutations in ovarian cancer
- POLE hotspot mutations in endometrial cancer
- TP53 and MDM2/4 alterations in GBM
- PTEN mutations in GBM in text format
- Patient view of an endometrial cancer case
- All TCGA Pan-Cancer
- MSK-IMPACT clinical cohort, Zehir et al. 2017
- Histone mutations across cancer types

Local Installations

[Host your own](#)


Are you running a local instance of cBioPortal, public or private? [Complete the survey here](#) to add your installation to the map.

Testimonials

"I would like to congratulate you and the team of the cBio portal. It's just an amazing tool to work with, and we at Mass General really appreciate it."

--Research Fellow at Massachusetts General Hospital



[README](#)[1. GENERAL](#)[News](#)[Frequently Asked Questions](#)[About Us](#)[List of RFCs](#)[2.1 DEPLOYMENT](#)[Architecture overview](#)[Hardware Requirements](#)[2.1.1 DEPLOY WITH DOCKER
\(RECOMMENDED\)](#)[Deploy with Docker](#)[Import data with Docker](#)

API and API Clients

cBioPortal provides a REST API for programmatic access to the data. The visualizations one can see on the website leverage the same API. By connecting to the API directly, anyone can build their own visualizations/reports.

Please see the full reference documentation for the API [here](#).

API Clients

The cBioPortal REST API is described using Swagger/OpenAPI, which allows one to generate a client in most programming languages. One can use the command line tool `curl` for downloading data on the command line or use another language such as `Python` or `R` to make visualizations. We list some common examples below, but if your language is not listed, there is likely a client generator available elsewhere (see e.g. <https://swagger.io/tools/swagger-codegen/>). Do reach out if you'd like us to add a language.

R client

There are multiple ways to access the API using R. The recommended way is to use `cBioPortalData`.

cBioPortalData (recommended)

`cBioPortalData` aims to import all cBioPortal datasets as `MultiAssayExperiment` objects in Bioconductor. Some of its key features:



- Provides visualization features like **-correlation plots for copy number alterations**, mRNA expression, gene methylation, survival analysis (Kaplan–Meier plots), coexpression analysis, and network analysis .
- cBioPortal has two unique web tools- **MutationMapper** and **OncoPrinter**.

MutationMapper is linked to 3D protein structure databases so that any user can understand the potential effects of the mutations with respect to proteins.

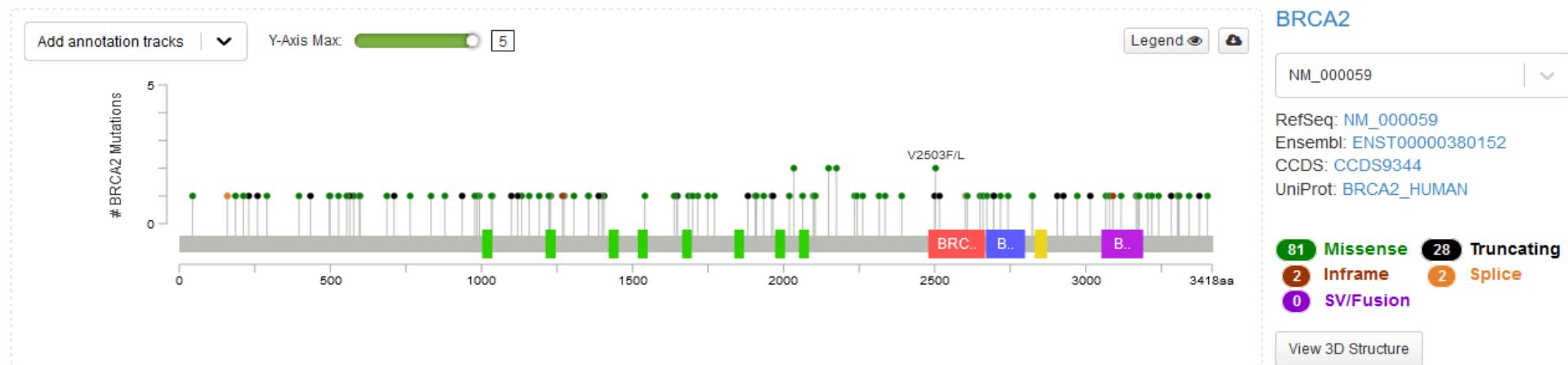
Mutation_mapper

MutationMapper

interprets mutations with protein annotations

[Modify Input](#)

EGFR BRCA1 BRCA2 PTEN



113 Mutations (page 1 of 5)

Sample ID	Cancer Type Detailed	Protein Change	Annotation	Functional Impact	Mutation Type
TCGA-04-1331-01	Serous Ovarian Cancer	C711*	1		Nonsense
TCGA-09-2050-01	Serous Ovarian Cancer	S1882*	1		Nonsense
TCGA-13-0885-01	Serous Ovarian Cancer	K1406Nfs*3	1		FS del
TCGA-13-0890-01	Serous Ovarian Cancer	S1230Lfs*9	1		FS del
TCGA-13-1481-01	Serous Ovarian Cancer	S2697Kfs*31	1		FS del
TCGA-13-0889-01	Serous Ovarian Cancer	T2517Hfs*22	1		FS ins

https://www.cbioportal.org/mutation_mapper

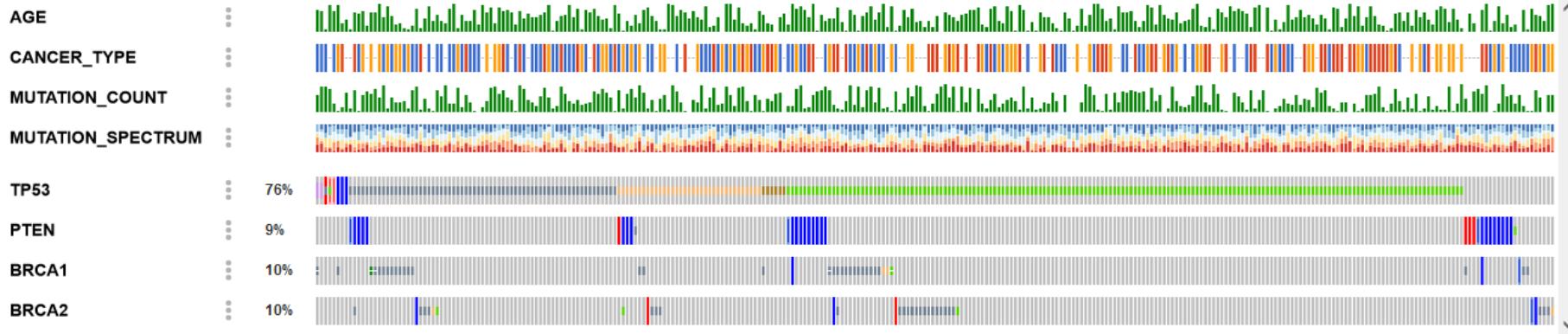
Oncoprinter

 Generate Oncoprints and perform mutual exclusivity analysis from your own data.[Modify Input](#)[Oncoprint](#)[Mutual Exclusivity](#)

 Driver annotations reflect only user-provided data. Use the Mutations menu to modify annotation settings. [Hide](#)

Sort [▼](#) Mutations [▼](#) View [▼](#) Download [▼](#)   38 % 

Altered in 300 (82%) of 368 samples.



Oncoprinter provides intuitive diagrams of genomic alterations such as somatic mutations and copy number alterations across a set of samples.

<https://www.cbiportal.org/oncoprinter>

DRIVERDB3

- Identification of driver genes or mutations .
- Mutation profiling, expression levels, copy number variations (CNV), methylation status, and miRNA-gene network across **different cancers types** .
- For the cancer-specific driver genes, **3 levels** of **functional analysis** (Gene Ontology, Pathways, and Protein/Genetics interactions) can be performed on this portal.
- For the pathway analysis Uses : KEGG, Reactome, PID, Biocarta, MsigDB, miRTar, and miRWalk

DriverDBv3: A database for human cancer driver gene research

[Home](#)[Cancer](#)[Gene](#)[Customized-analysis](#)[Download](#)[Help](#)[FAQs](#)

What is DriverDBv3?

DriverDBv3 is a cancer omics database which incorporates somatic mutation, RNA expression, miRNA expression, methylation, copy number variation and clinical data in addition to annotation bases. This database also uses published bioinformatics algorithms to identify driver genes and present them with different molecular features; there are three functions, '**Cancer**', '**Gene**', and '**Customized-Analysis**', to help researchers visualize the relationships between cancers and driver genes.

Search by Functions

Cancer

The '**Cancer**' function summarizes the calculated results of driver genes in different molecular features by using published bioinformatics algorithms or tools for a specific cancer type or dataset.

Gene

The '**Gene**' function visualizes different features of a user-selected gene within cancer types or datasets.

Example: EGFR

Customized- Analysis

The '**Customized-Analysis**' function offers insights into survival analysis by providing two approaches to stratify patients - by expression or by mutation, and driver gene identification, which is the 'Meta-analysis' function in the previous database. These functions allow the researchers to select a sub-group of well-defined cancer samples

DriverDBv3: A database for human cancer driver gene research

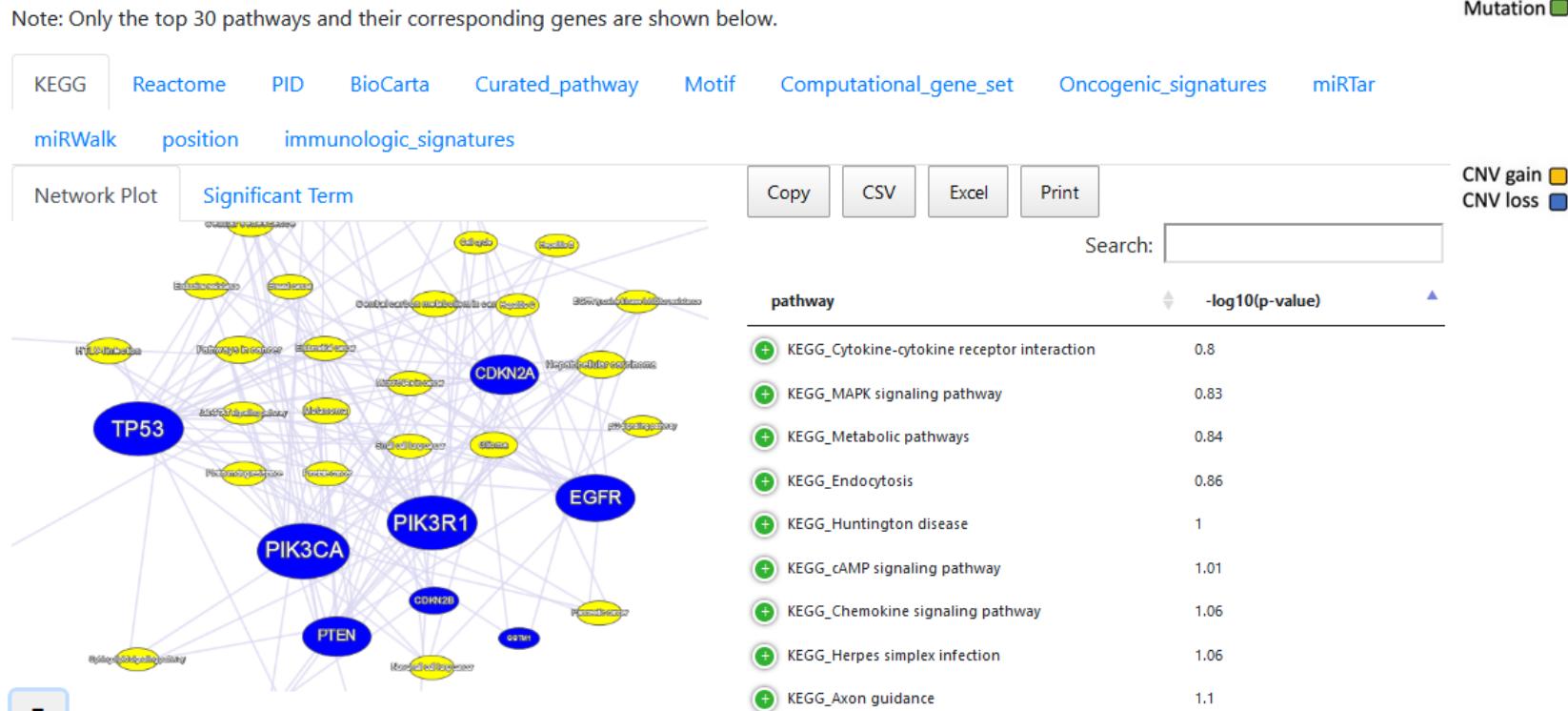
[Home](#)[Cancer](#)[Gene](#)[Customized-analysis](#)[Download](#)[Help](#)[FAQs](#)[Summary](#)[Mutation](#)[CNV](#)[Methylation](#)[Survival](#)[miRNA](#)

Cancer Type: **Glioblastoma multiforme(TCGA,US)**

The **Cancer Summary** section provides a **Summary network** which integrates cancer dysfunction and dysregulation events in a multi-omics level, and a **Functional Annotation** analysis of these cancer driver genes.

Pathway

The **Pathway** section includes 12 collections of gene sets sourced from 7 public databases: KEGG, Reactome, PID, BioCarta, Curated_pathway, Motif, Computational_gene_set, Oncogenic_signatures, miRTar, miRWALK, position, immunologic_signatures.





L.B.B THE CANCER PROTEOME ATLAS (TCPA)

- Focuses on the Reverse Phase Protein Array (**RPPA**)
- Offers a comparison between two different types of tumors and enables the **identification of proteins** that are the most differentially expressed between tumor Types
- RPPA-linked survival analysis, protein-drug analysis, and network-visualization modules are also provided by this portal .



L.B.B THE CANCER PROTEOME ATLAS (TCPA)

Laboratory of Systems Biology
and Bioinformatics

THE UNIVERSITY OF TEXAS
MD Anderson
~~Cancer Center~~

Making Cancer History®

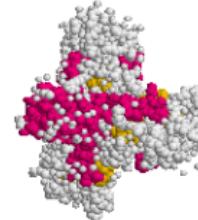
- Home
- Overview
- About
- News
- FAQ
- Resources
- Credits
- Forum
- Contact
- Download

Summary



»

My Protein



»

Visualization



»

Analysis



»

The Cancer Proteome Atlas (TCPA) is a joint project of the Departments of [Systems Biology](#) and [Bioinformatics & Computational Biology](#) at [The University of Texas MD Anderson Cancer Center](#).

This website is for educational and research purposes only.



REGULOME EXPLORER

TCGA Genome Data Analysis Center (GDAC) for Systems Analysis of the Cancer Regulome

Institute for Systems Biology and MD Anderson Cancer Center

The Cancer Genome Atlas (TCGA) provides an unprecedented opportunity to take an integrated approach toward a systems level understanding of regulatory disruptions in cancer. Such disruptions and their consequences are intertwined within complex dynamical networks through a multitude of interactions among different types of molecules. Understanding such relationships requires multivariate analysis methods that can be effective in the context of highly heterogeneous data, measurement uncertainty, and missing data.

News & Events

[← Newer](#)

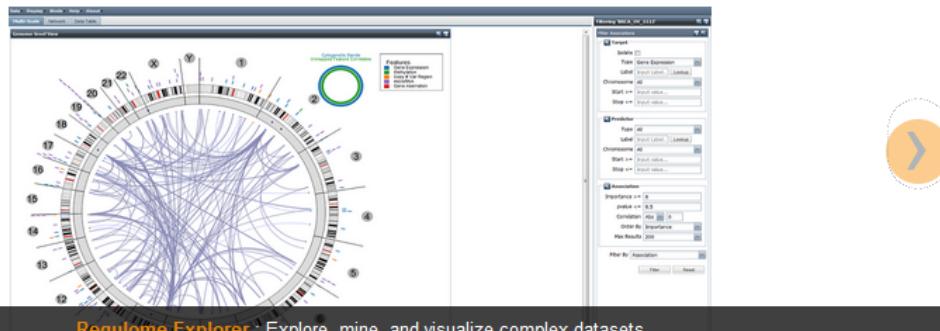
[Older →](#)

May 2016

The TCGA Research Network published a study on adrenocortical carcinoma (ACC) in [Cancer Cell](#). The study describes the comprehensive analysis of 91 ACC specimens from four continents using state of the art genomic technologies and computational methods. In addition to identification of novel ACC driver genes, pathways and refined subtypes, the AWG analysis revealed whole genome doubling (WGD) as a milestone in disease progression. We performed integrated analysis across different data types for this study and captured statistical associations in [Regulome Explorer](#).

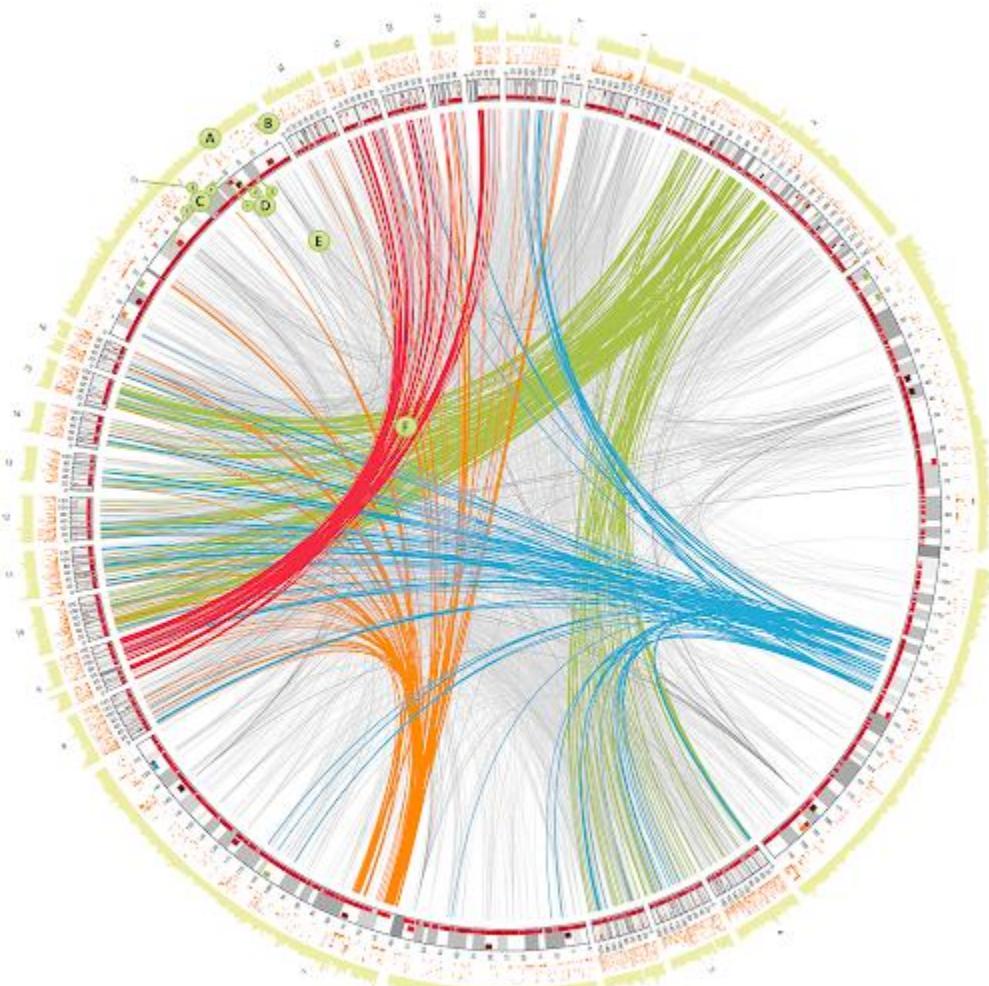
November 2015

The TCGA Research Network published a study on primary Prostate Cancer in [Cell](#). The study described a molecular taxonomy in which 74% of the tumors belonged to one of seven subtypes. This study revealed additional molecular heterogeneity among primary prostate cancers and molecular defects which may be therapeutically actionable. We performed integrated analysis across different data types for this study. Associations for the 333 tumor dataset can be viewed in [Regulome Explorer](#).



REGULOME EXPLORER

- Interrelation between **clinical** and **molecular features** of TCGA



analysis between different
tic copy number/gene
ylation/miRNA expression
Explorer is the mapping of
with genomic coordinates .



L.B.B GENE EXPRESSION PROFILING INTERACTIVE ANALYSIS (GEPIA)

- This **online tool** can be used for the rapid retrieval of customizable functionalities that depends on the data from The Genotype-Tissue Expression (**GTEX**) and **TCGA**.
- Allows the **comparison** of the expression profile of a gene of **normal tissues (GTEX)** to the corresponding **tumors samples (TCGA)** .
- Correlation analysis, patient survival analysis, co-expression analysis, and dimensionality reduction analysis.



GEPIA
Gene Expression Profiling Interactive Analysis

Single Gene Analysis

Cancer Type Analysis

Multiple Gene Analysis

Enter gene name:

The indicators in search box are "symbol" or "alias (newest symbol)".

e.g. ERBB2/ENSG00000141736/2064

GoPIA!

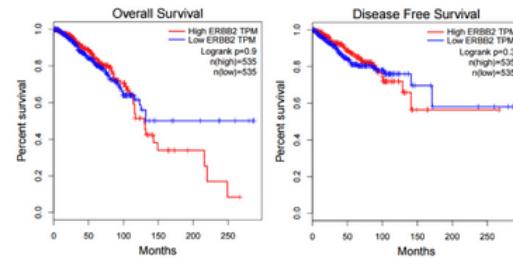
Profile

Boxplots

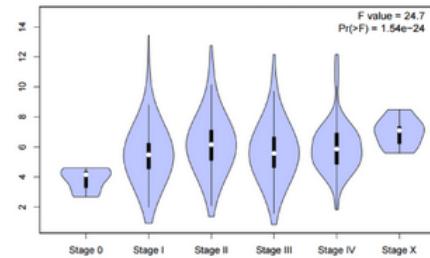
Stage Plots

Survival Analysis

Similar



Survival analysis



Analyze expression by stage





FUNCTIONS



EXTRAS



GEPIA2

Gene Expression Profiling Interactive Analysis

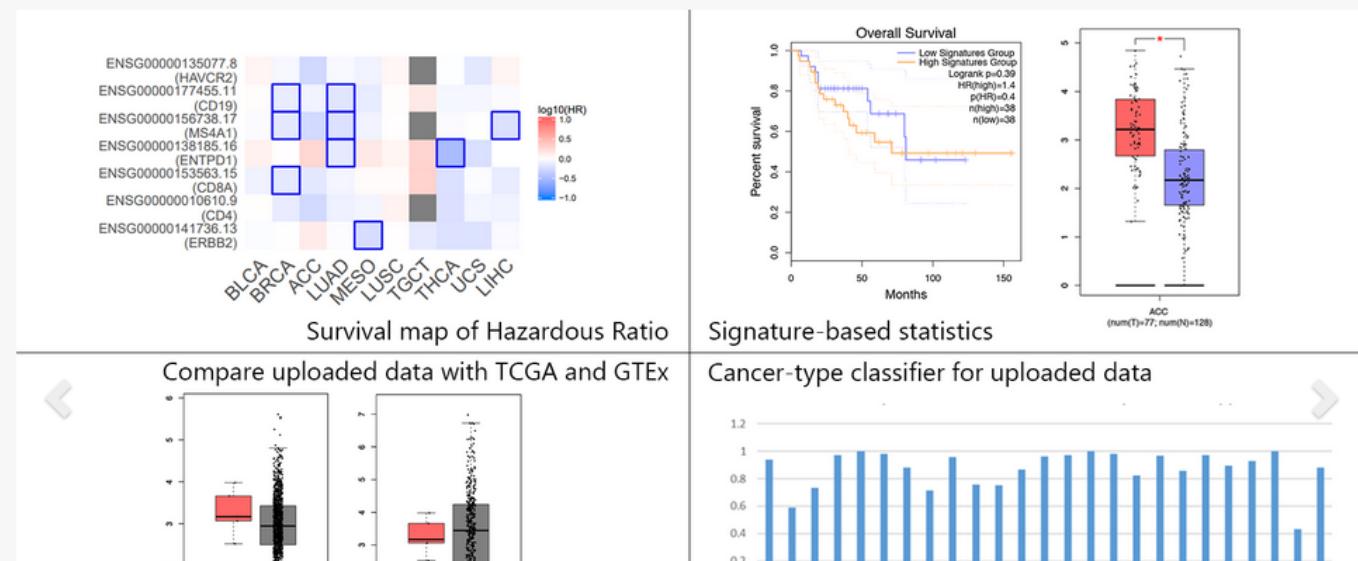
Single Gene Analysis Cancer Type Analysis Custom Data Analysis Multiple Gene Analysis

Enter gene/isoform name:

The indicators in search box are "symbol" or "alias (newest symbol)".

e.g. ERBB2/ENSG00000141736 or ERBB2-001/ENST00000584601

Profile Boxplots Stage Plots Survival Analysis Similar



UCSC Xena

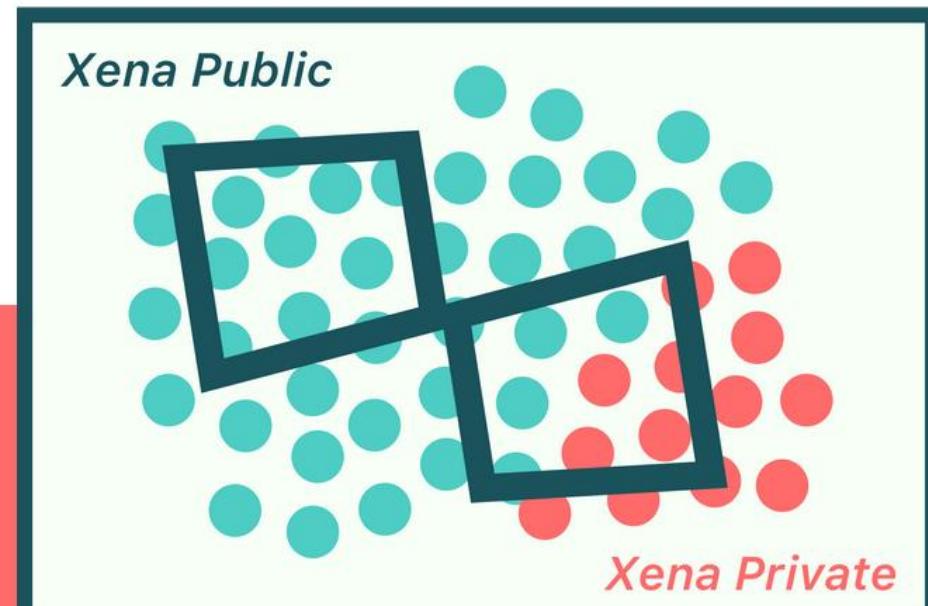
See the bigger picture

An online exploration tool for public and private, multi-omic and clinical/phenotype data

[Launch Xena](#)

Tutorials and walkthroughs

Don't know where to start? Jump in with one of our tutorials or "How do I ..." walkthroughs



- ✿ Explore public data resources such as - TCGA, GDC, ICGC, GTEx, TARGET, and PARADIGM pathway inference .
- ✿ Somatic mutation, copy number, methylation, gene/protein expression, and phenotypic data .
- ✿ Visualization of the gene expression, DNA copy number methylation, and somatic mutational data for a user-defined gene and allows comparison across omics datasets.

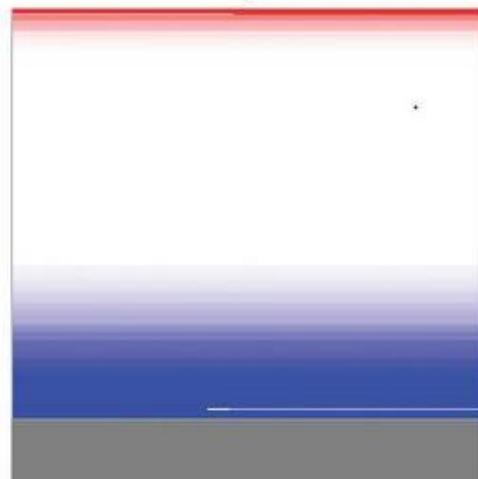
COMPARISON ACROSS OMICS DATASETS

UCSC Xena

APC gene expression and copy number in READ

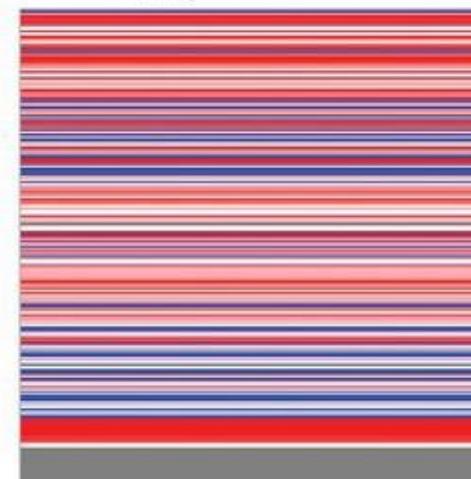
Gene expression

samples



14 16
 $\log_2(\text{fpkm}-\text{uq}+1)$

Copy number



null

-0.49 0.49
 $\log_2(\text{copy-number}/2)$

Xena can help you ...



[Learn More](#)

Perform a Kaplan-Meier survival analysis

Wondering if a gene (or probe, or clinical value, etc) affects survival? We have survival analyses complete with p-values, adjustable time frames, and multiple survival endpoints.



Compare tumor vs normal within or across tissue types

Use Xena to compare TCGA tumor samples to GTEx normal samples to see if your gene or transcript is up- or down-regulated in one or more cancer types.



[Learn More](#)

Discover relationships among genomic and clinical data

Wondering if an increase in gene expression is correlated with a promoter hypomethylation, an increase in chromatin accessibility, or hotspot missense mutations? Xena lines up multiple types of data side-by-side to help you find out.



Create subgroups based any genomic data

With Xena you can compare expression of a gene between wild type vs mutant samples, or between samples with normal copy number vs amplifications, or between any other subgroups.

- ✿ Web portal for analysis and **visualization** of the association between altered gene expression pattern, Kaplan-Meier based , survival curves, of a particular **TCGA cancer type** .
- ✿ Compare the relative expression pattern of any given gene between **tumor** and **NT adjacent tissues** in a paired analysis.
- ✿ Filtered based on different categories such as tumor grade, cancer stage, race, and other clinical features and the results can be exported in different output formats .

- Provide access to cancer OMICS data (TCGA, MET500 and CPTAC)
- Perform **PAN-CANCER** gene expression analysis.
- Provide additional information about the selected genes/targets by linking to HPRD, GeneCards, Pubmed, TargetScan, The human protein atlas, DRUGBANK, Open Targets and the GTEx.



Analyze, Integrate, Discover

Home

Tutorial

TCGA

CPTAC

CBTTC

Welcome to UALCAN

UALCAN is a comprehensive, user-friendly, and interactive web resource for analyzing cancer OMICS data. It is built on PERL-CGI with high quality graphics using javascript and CSS. UALCAN is designed to, a) provide easy access to publicly available cancer OMICS data (TCGA, MET500 and CPTAC), b) allow users to identify biomarkers or to perform in silico validation of potential genes of interest, c) provide graphs and plots depicting expression profile and patient survival information for protein-coding, miRNA-coding and lincRNA-coding genes, d) evaluate epigenetic regulation of gene expression by promoter methylation, e) perform pan-cancer gene expression analysis, and f) Provide additional information about the selected genes/targets by linking to HPRD, GeneCards, Pubmed, TargetScan, The human protein atlas, DRUGBANK, Open Targets and the GTEx. These resources allow researchers to gather valuable information and data about the genes/targets of interest.

Follow us on Twitter 

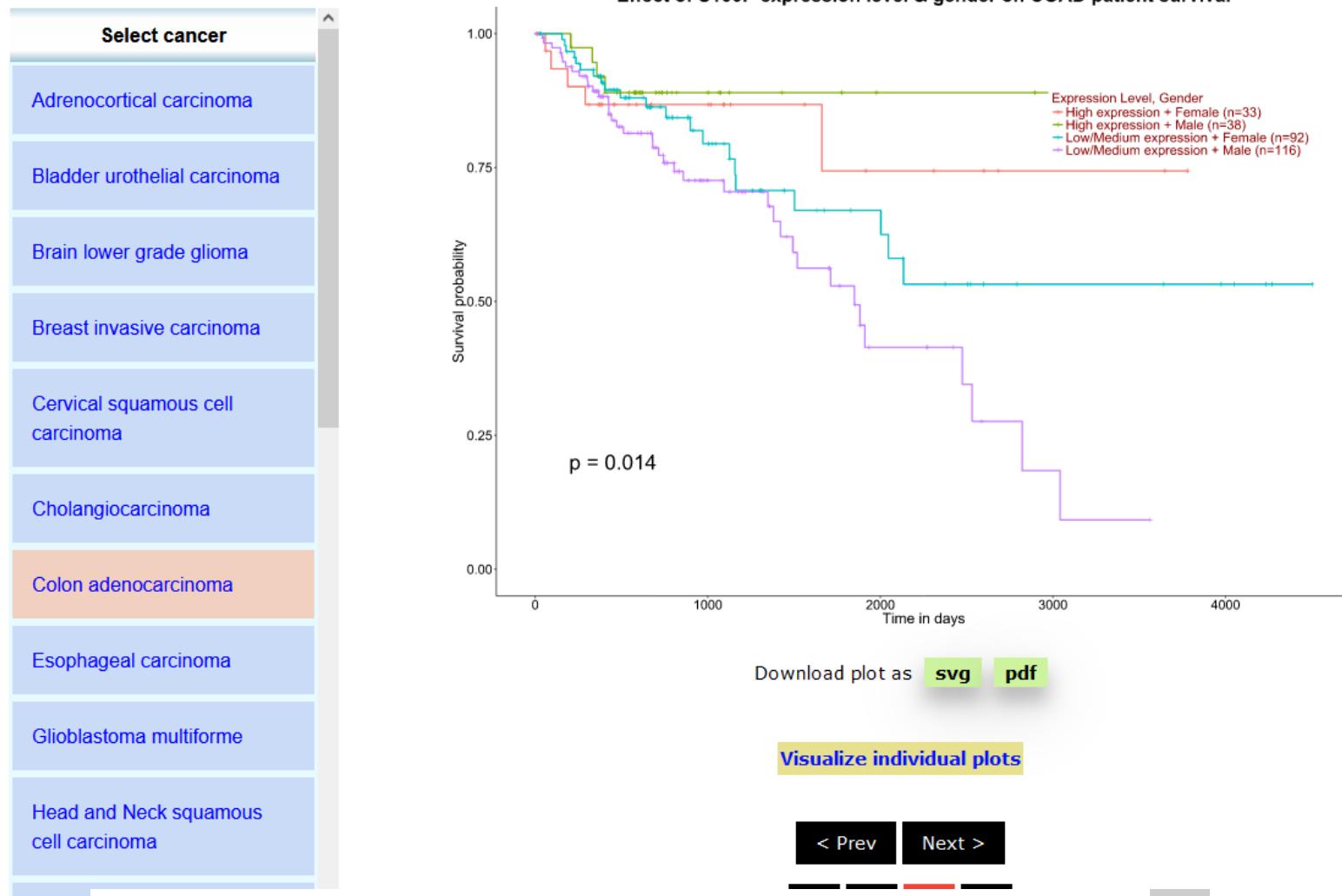
Please cite: Chandrashekhar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Rodriguez IP, Chakravarthi BVSK and Varambally S. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. 2017 Aug;19(8):649-658. doi: 10.1016/j.neo.2017.05.002 [PMID:28732212]

Updated slides from a presentation about UALCAN cancer database at Dana-Farber Cancer Institute [Download](#)

Updates

10/08/2021

UALCAN has been now visited over 700,000 times by cancer researchers from over 100 countries and cited over 1900 times



- Expression Level, Gender**
- High expression + Female (n=33)
 - High expression + Male (n=38)
 - Low/Medium expression + Female (n=92)
 - Low/Medium expression + Male (n=116)

WANDERER

- Offers exploration and interpretation of gene-specific expression profiles and DNA methylation patterns for almost all of the **cancer types available in TCGA** .
- Offers the exploration of the **DNA methylation alterations** in a gene-specific manner **normal–tumor** paired comparisons through comprehensive tables and graphs as well as **correlation analysis** can be performed .



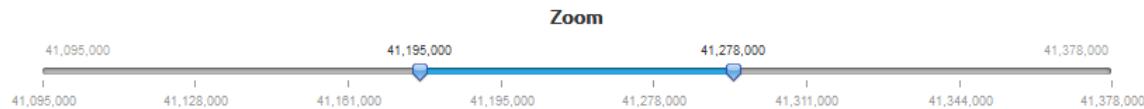
Wanderer

An interactive viewer to explore DNA methylation
and gene expression data in human cancer

Gene Symbol or Ensembl Gene Id

Examples: BRCA1 or ENSG00000141510

Important: Press refresh after entering a new gene name



Specify a region

Dataset

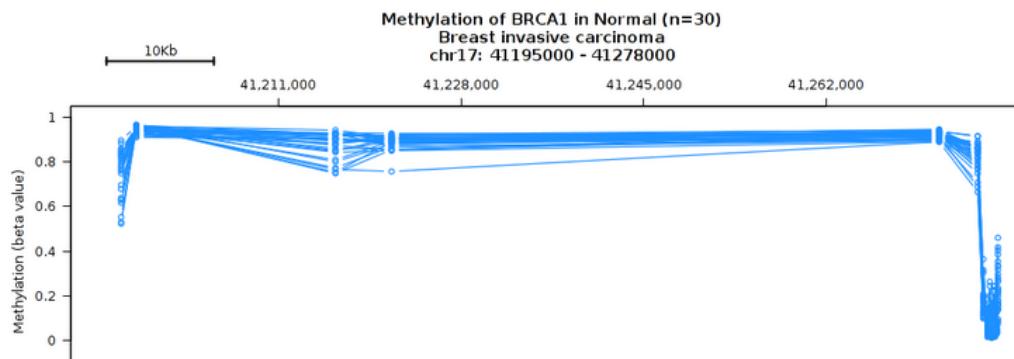
Project:

Breast invasive carcinoma
(BRCA)

Data Type:

450k Methylation Array

We offer level 3 TCGA data for methylation arrays (450k Infinium chip) and expression





CANCER-SPECIFIC MULTI-OMICS

TCGA (THE CANCER GENOME ATLAS)

ICGC (INTERNATIONAL CANCER GENOME CONSORTIUM)



L.B.B

Laboratory of Systems Biology
and Bioinformatics

INTERNATIONAL CANCER GENOME CONSORTIUM

ICGC

- ✿ Datasets encompassing **90 different cancer projects** involving **16 different countries** and **22 primary cancer sites** .
- ✿ 1.69 petabytes dataset , from a total of 24,289 donors .
- ✿ **PCAWG** : pancancer , with TCGA . (Article)
- ✿ Population-wise .
- ✿ **Highest** number of samples : **neuroblastoma** (NBL-US, N = **798**),
Acutelymphoblastic leukemia (ALL-US, N = **615**)
- ✿ **Lowest** number of samples is for **liver cancer** (LIHM-FR, N = 4),
Renal cancer (RECA-CN, N = 10)
Lung cancer (LUSC-CN, N = 10)



An open access paper

NATURE

www.nature.com

Pan-cancer analysis of whole genomes

<https://doi.org/10.1038/s41586-020-1969-6>

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium

Received: 29 July 2018

Accepted: 11 December 2019

Published online: 5 February 2020

Open access

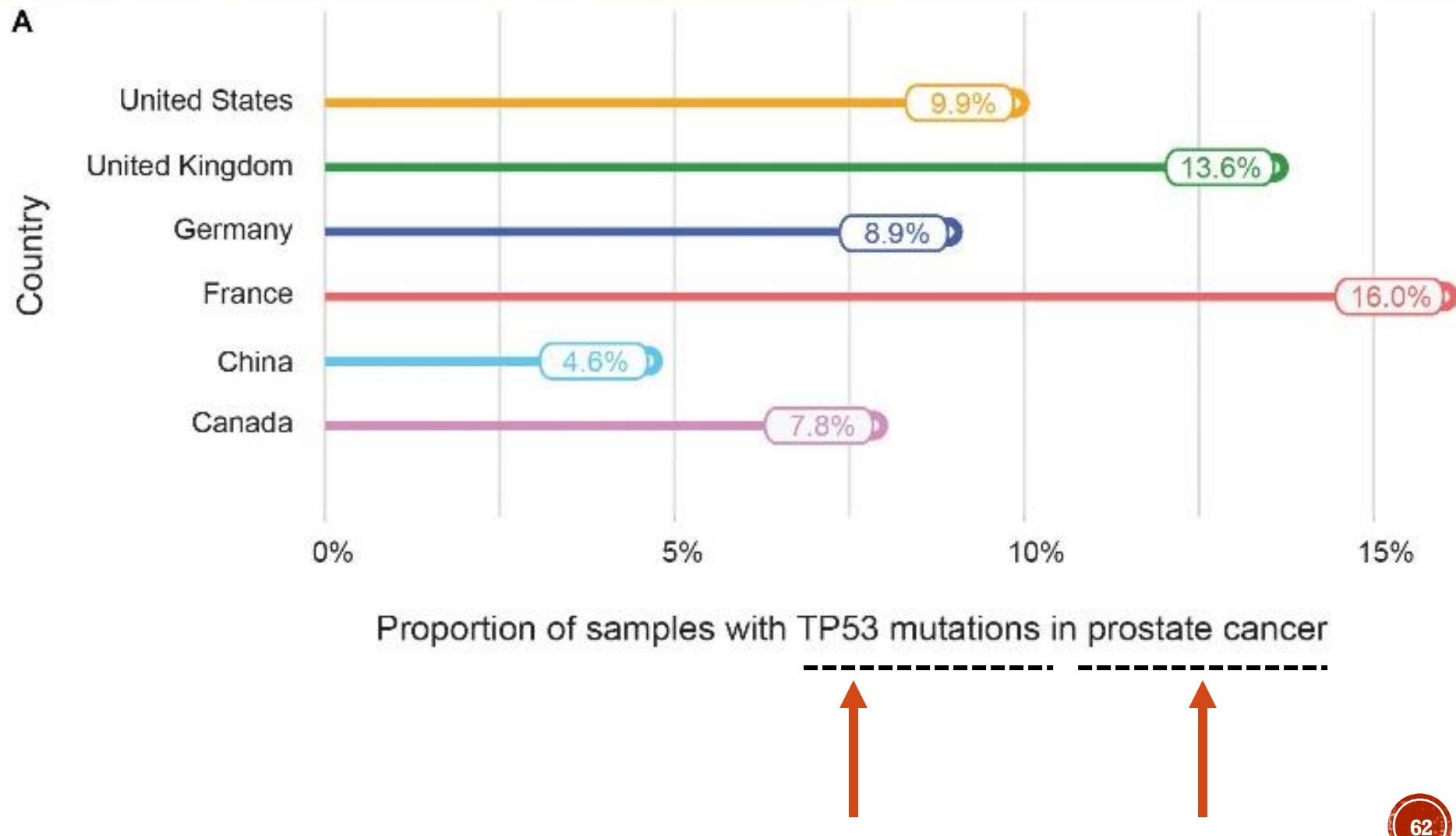
Cancer is driven by genetic change, and the advent of massively parallel sequencing has enabled systematic documentation of this variation at the whole-genome scale^{1–3}. Here we report the integrative analysis of 2,658 whole-cancer genomes and their matching normal tissues across 38 tumour types from the Pan-Cancer Analysis of Whole Genomes

Genomic variations by focusing on both coding and non-coding regions of 2,658 whole-cancer genomes

many clustered structural variants arise in a single catastrophic event, is frequently an early event in tumour evolution; in acral melanoma, for example, these events precede most somatic point mutations and affect several cancer-associated genes

POPULATION-WISE .

A





L.B.B

Laboratory of Systems Biology
and Bioinformatics

INTERNATIONAL CANCER GENOME CONSORTIUM (ICGC)



ICGC Data Portal

[Cancer Projects](#)[Advanced Search](#)[Data Analysis](#)[DCC Data Releases](#)[Data Repositories](#)

Cancer genomics data sets visualization, analysis
and download.

[Quick Search](#)[Search](#)

e.g. BRAF, KRAS G12D, DO35100, MU7870, Fl998, apoptosis, Cancer Gene Census, imatinib, GO:0016049

[Advanced Search](#)[By donors](#)[By genes](#)[By mutations](#) [Download Release](#)

Data Release 28

March 27th, 2019

Cancer projects	86
Cancer primary sites	22
Donor with molecular data in DCC	22,330
Total Donors	24,289
Simple somatic mutations	81,782,588

WEAK POINT

- Not all **omics-data types** are available for different **populations of the world**.

For example :

Malignant lymphoma- Germany and Ovarian cancer- Australia .

miRNA expression data .

ONLINE RESOURCES OF OMICS-DATA

Online single and multi-omics resources.

1 - CANCER-SPECIFIC DATA RESOURCES .

Multi-Omics

Single-Omics

2 - Generalized DATA RESOURCES .

CANCER-SPECIFIC SINGLE-OMICS RESOURCES

- **COSMIC**

(Catalogue of Somatic Mutations in Cancer)

- **The Pathology Atlas**

- COSMIC offers the exploration of genomic data focusing on **mutational types** and **frequency statistics** for a user-defined gene or cancer type .
- COSMIC provides a **graphical representation** of the **mutational frequencies** of a given gene in a particular cancer type.

Projects

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:

[COSMIC](#)

The core of COSMIC, an expert-curated database of somatic mutations

[Cell Lines Project](#)

Mutation profiles of over 1,000 cell lines used in cancer research

[COSMIC-3D](#)

An interactive view of cancer mutations in the context of 3D structures

[Cancer Gene Census](#)

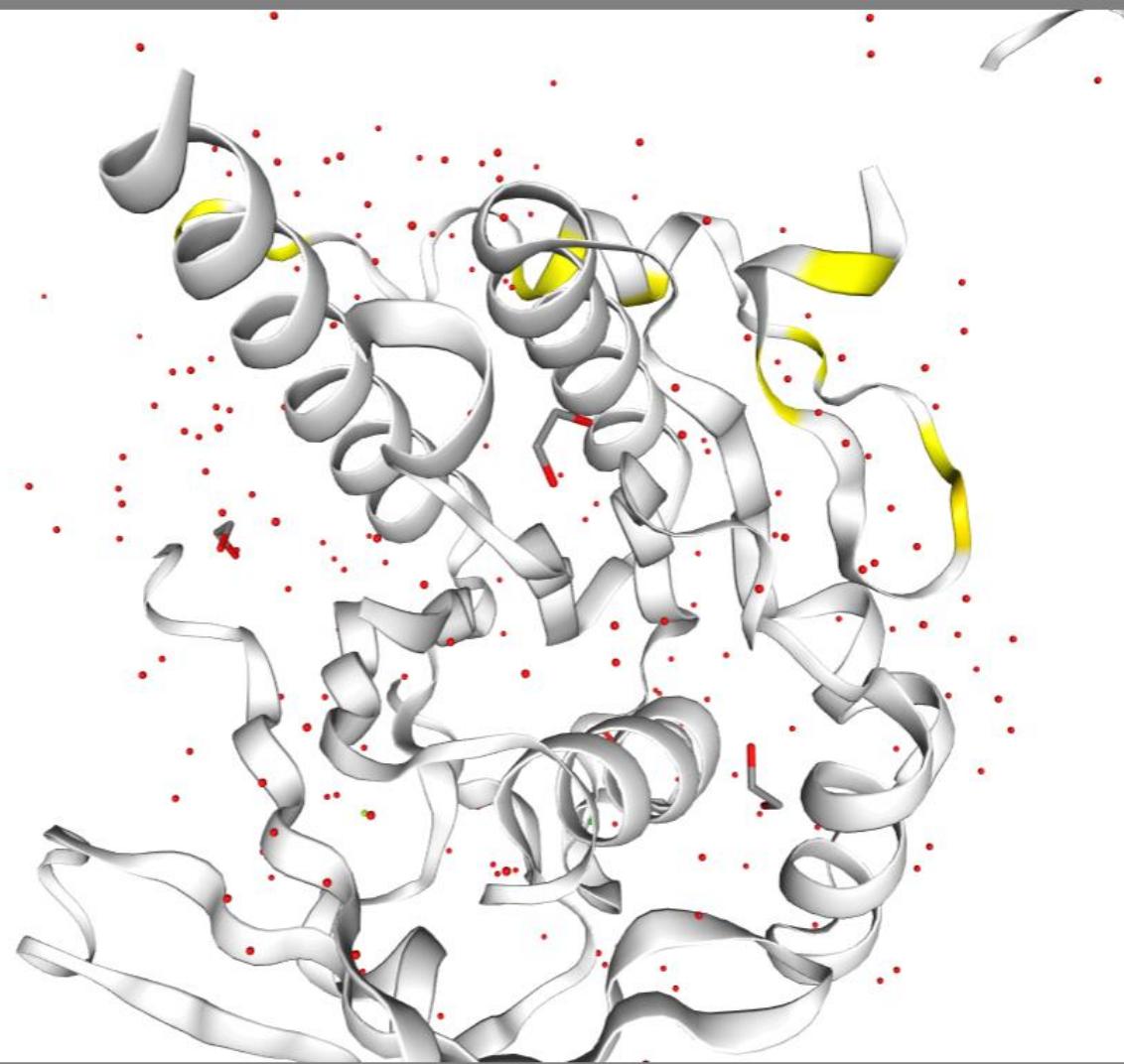
A catalogue of genes with mutations that are causally implicated in cancer

[Cancer Mutation Census](#)

Classification of genetic variants driving cancer

[Actionability](#)

Mutations actionable in precision oncology



1N0W a Rad51-brca2 Brc Repeat Complex



Terms and Conditions have been updated and include important changes. Please check the [Licensing page](#) for details.

COSMIC v94, released 28-MAY-21

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

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below.

eg Braf, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell

SEARCH

Projects

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:



[COSMIC](#)

The core of COSMIC, an expert-curated database of somatic mutations



[Cell Lines Project](#)

Mutation profiles of over 1,000 cell lines used in cancer research



[COSMIC-3D](#)

An interactive view of cancer mutations in the context of 3D structures



[Cancer Gene Census](#)

A catalogue of genes with mutations that are causally implicated in cancer

<https://cosmic-blog.sanger.ac.uk/Release-v94/>

COSMIC News

[Follow @cosmic_sanger](#)



[Curating the future of precision oncology: An interview with Steve Jupe](#)

Lean about the curation process, background to Actionability, and innovative uses of COSMIC data in our interview with Steve Jupe. [More...](#)



[COSMIC Release v94 is live!](#)

a focus on rare lung cancers and rare pancreatic cancers, and curation of somatic mutations in 12 hallmark apoptosis genes. Along with this, 9 cancer hallmark genes data are also updated. Find out more before exploring the v94 release. [More...](#)



[COSMIC Release v94 is nearly complete!](#)

After our very successful and exciting Actionability release v93, we are now gearing up for COSMIC v94 release in the coming weeks. This release is focused on our high quality manual curations including 2 new expertly curated genes, a focus on rare | [More...](#)

Tools

- ❖ [Cancer Browser](#) — browse COSMIC data by tissue type and histology
- ❖ [Genome Browser](#) — browse the human genome with COSMIC annotations
- ❖ [GA4GH Beacon](#) — access COSMIC data through the GA4GH Beacon Project

THE PATHOLOGY ATLAS

- Is a part of the Human Protein Atlas .
- Data repository including antibody-based imaging, mass spectrometry-based proteomics, and transcriptomics .
- The Pathology Atlas consists of the detailed analyses of protein-immunohistochemistry, mass-spectrometry, and TCGA-derived transcriptome data of 17 different cancer types from 8000 patients .
- The transcriptomics data TCGA , more than 5 million pathology-based images obtained through tissue microarray (TMA)-based immunohistochemistry (IHC) analysis of the corresponding proteins.

THE HUMAN PROTEIN ATLAS

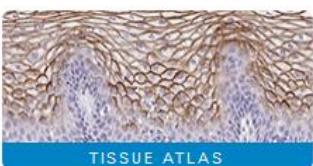
THE HUMAN PROTEIN ATLAS

≡ MENU HELP NEWS

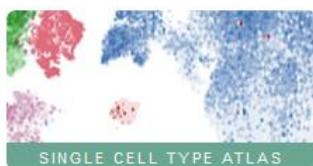
SEARCHⁱ

e.g. ACE2, GFAP, EGFR

Search Fields »



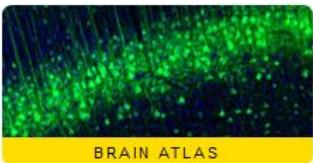
TISSUE ATLAS



SINGLE CELL TYPE ATLAS



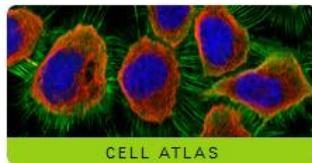
PATHOLOGY ATLAS



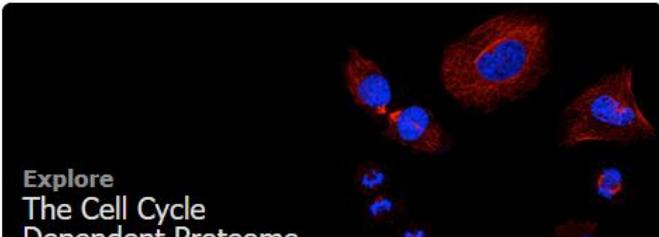
BRAIN ATLAS



BLOOD ATLAS



CELL ATLAS



Explore
The Cell Cycle
Dependent Proteome

Recent news

Wed, 28 Jul 2021
A single cell type map of human tissues

Thu, 22 Jul 2021
Movie of the month: The Fatty Liver

Tue, 15 Jun 2021
Movie of the month: The Spleen

- HPA consortium analyzed the transcriptome in 17 different cancer types of TCGA and correlated the altered-gene expression to clinical outcomes .

to enhance the usability of our website. If you continue, we'll assume that you are happy to receive all cookies. More information. Don't show this again



L.B.B

Laboratory of Systems Biology

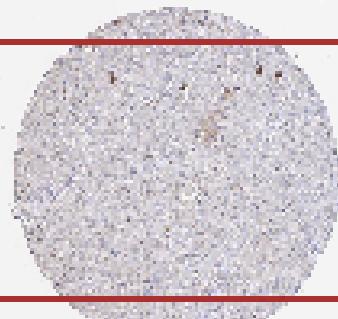
+

-

×



ANTIBODY-BASED IMAGING



Melanoma

HPA060655

Male, age 62

Lymph node (T-08000)

Malignant melanoma,
Metastatic site (M-87206)

Patient id: 3140

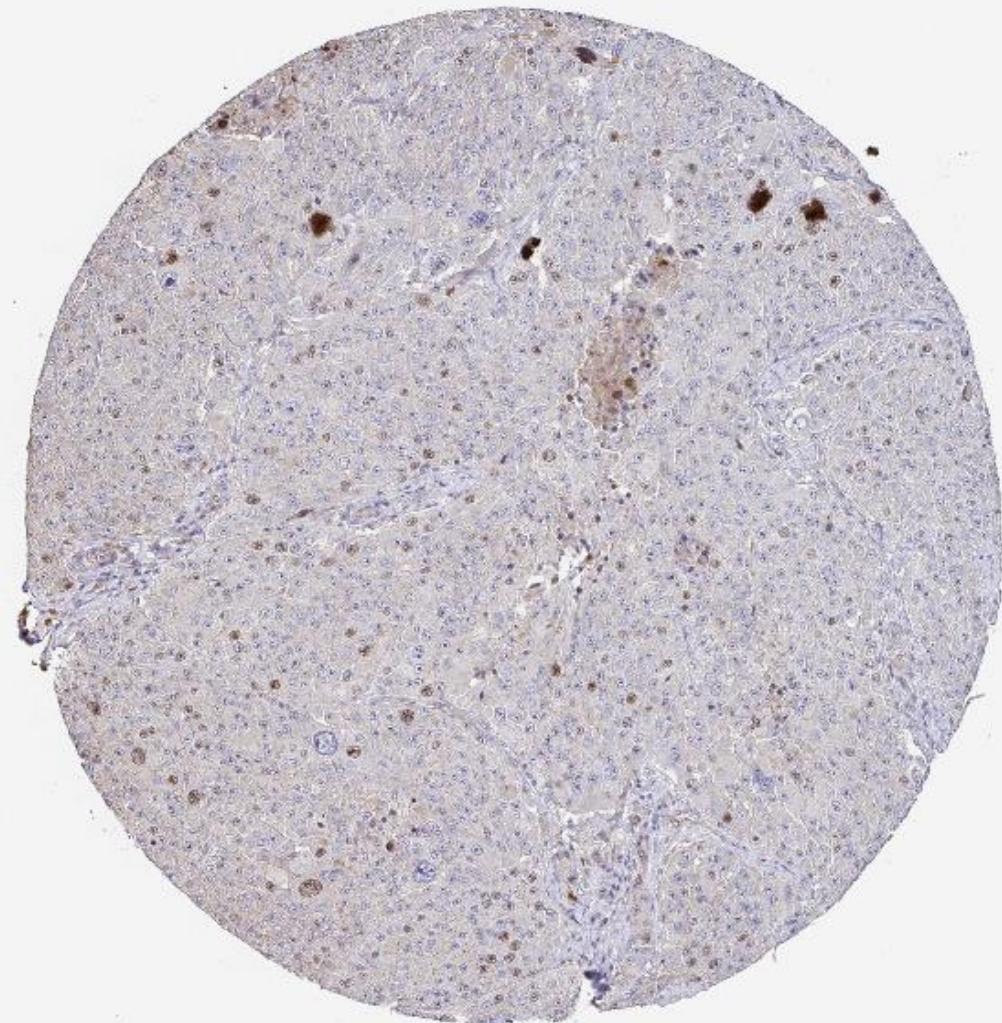
Tumor cells

Staining: Low

Intensity: Moderate

Quantity: <25%

Location: Nuclear



THE PATHOLOGY ATLAS

THE HUMAN PROTEIN ATLAS

≡ MENU HELP NEWS

THE HUMAN PROTEOME : THE PATHOLOGY ATLAS

Search Fields »

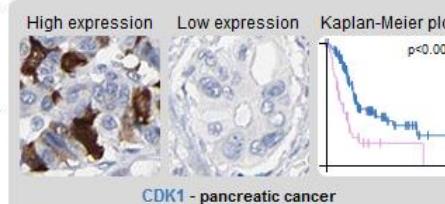
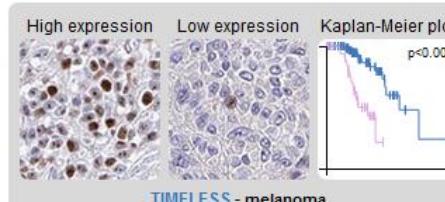
The Pathology Atlas

The Pathology Atlas contains mRNA and protein expression data from 17 different forms of human cancer. Correlation analyses based on mRNA expression levels of human genes in cancer tissue and the clinical outcome for almost 8000 cancer patients is presented in a gene-centric manner, including more than 18000 Kaplan-Meier plots with high significance ($p<0.001$).

Analysis of each protein and its corresponding cancer type in patients, using immunohistochemistry (IHC) analysis based on tissue microarrays (TMAs), is presented for a majority of the protein-coding genes. The resulting 5 million IHC cancer tissue images are presented here.

All transcriptomics data has been retrieved from the Cancer Genome Atlas and all proteomics data has been generated in-house using the same antibodies as in protein expression profiling in normal human tissues.

Cancer statistics from relevant international and Swedish databases are summarized [here](#), and hallmarks of cancer are described [here](#).



THE CANCER PROTEOMEⁱ

Interactive chapters regarding each of the 17 cancer types explore mRNA and protein expression data, including genes associated with prognosis.

THE HUMAN PROTEIN ATLAS

 Search Fields »[☰ MENU](#) [HELP](#) [NEWS](#)

THE HUMAN PROTEOME : THE PATHOLOGY ATLAS

THE CANCER PROTEOMESⁱ

Interactive chapters regarding each of the 17 cancer types explore mRNA and protein expression data, including genes associated with prognosis.

Brain (Glioma) 

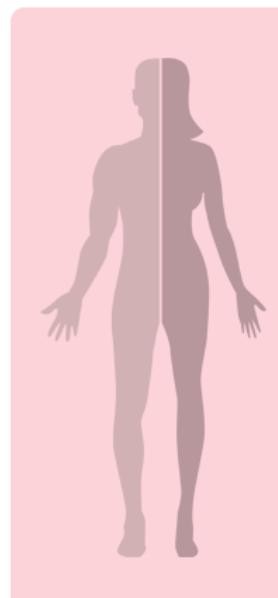
Head and neck 

Thyroid gland 

Lung 

Liver 

Testis
Prostate 



Stomach
Colon/Rectum 

Breast
Endometrium
Ovary
Cervix 

Pancreas 

Kidney 

Urinary bladder 

Skin (Melanoma) 

CORRELATION ANALYSIS AND PROGNOSTIC GENESⁱ

We use cookies to enhance the usability of our website. If you continue, we'll assume that you are happy to receive all cookies. [More information](#). [Don't show this again](#).

ONLINE RESOURCES OF OMICS-DATA

Online single and multi-omics resources.

1 - CANCER-SPECIFIC DATA RESOURCES .

Multi-Omics

Single-Omics

2 - Generalized DATA RESOURCES .

GENERALIZED SINGLE-OMICS DATA RESOURCES

- Gene Expression Omnibus (GEO)
- PRoteomics IDEntifications Database (PRIDE)



L.B.B

Laboratory of Systems Biology
and Bioinformatics

GENE EXPRESSION OMNIBUS (GEO)

- GEO harbors **4348** datasets derived from more than **100 different Organisms** .
- Data including **cell line** and **non-cancer samples** .
- **291 patient datasets** encompassing **48 different cancer types** .

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.



Getting Started

- [Overview](#)
- [FAQ](#)
- [About GEO DataSets](#)
- [About GEO Profiles](#)
- [About GEO2R Analysis](#)
- [How to Construct a Query](#)
- [How to Download Data](#)

Tools

- [Search for Studies at GEO DataSets](#)
- [Search for Gene Expression at GEO Profiles](#)
- [Search GEO Documentation](#)
- [Analyze a Study with GEO2R](#)
- [Studies with Genome Data Viewer Tracks](#)
- [Programmatic Access](#)
- [FTP Site](#)
- [ENCODE Data Listings and Tracks](#)

Browse Content

- [Repository Browser](#)
- [DataSets:](#) 4348
- [Series:](#)  161951
- [Platforms:](#) 22659
- [Samples:](#) 4676244

Information for Submitters

[Login to Submit](#)

- [Submission Guidelines](#)
- [Update Guidelines](#)

[MIAME Standards](#)

- [Citing and Linking to GEO](#)
- [Guidelines for Reviewers](#)

ggle-analytics.com

- Originated from the European Bioinformatics Institute .
- Mined for cancer-specific **transcriptomics** and **proteomics** datasets.
- The mined datasets classified into **solid tumor** and **blood cancer** categories .
- Among the **1028** proteomics studies focusing on different types of **solid tumors**,
breast cancer has the highest number (**N = 233**) of datasets .

In contrast, with **188** proteomics studies, **leukemia** has the highest number of studies among different types of **blood cancer studies** .



L.B.B

Laboratory of Systems Biology
and Bioinformatics

PROTEOMICS IDENTIFICATIONS DATABASE (PRIDE)

EMBL-EBI Services Research Training About us EMBL-EBI Hinxton

Due to unforeseen issues in the EBI IT infrastructure, we cannot release datasets to the public at this point. For this reason, we are disabling the automatic publication feature for now. We kindly request all the submitters to wait till our next announcement. In cases when an urgent public release of datasets is required (e.g. when required for the formal acceptance of manuscripts), please send an email to pride_support@ebi.ac.uk.

PRIDE
PRoteomics IDEntifications Database

Home Resources Tools Docs About Contact Log in Register

Archive search

Search

PXD005011> cancer> human> P02768> YLEVEPVS>



PRIDE ARCHIVE

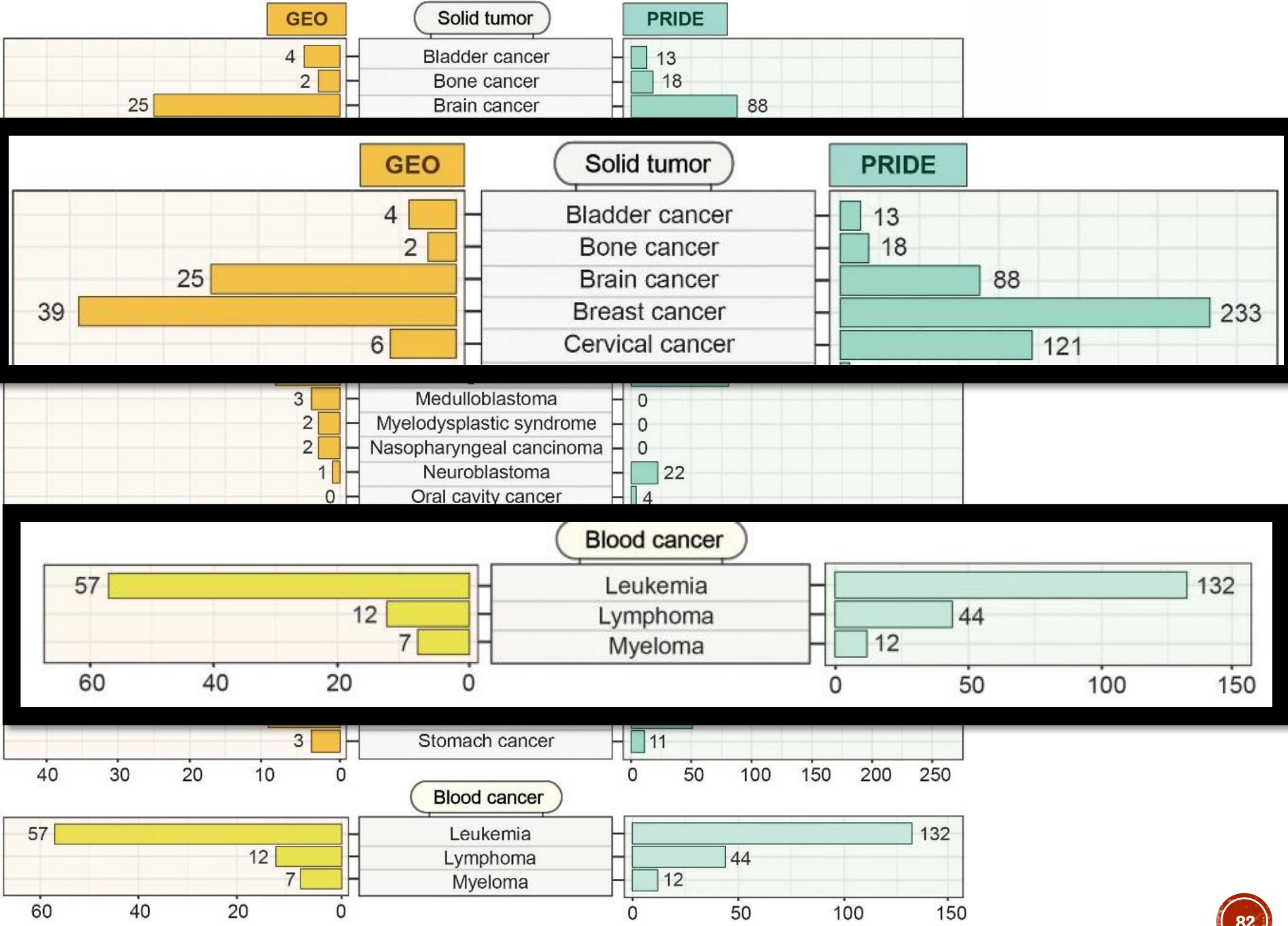


PRIDE PEPTIDOME



SPECTRA ARCHIVE

Generalized DATA RESOURCES .



ONLINE RESOURCES OF OMICS-DATA

Online single and multi-omics resources.

1 - CANCER-SPECIFIC DATA RESOURCES .

Multi-Omics

Single-Omics

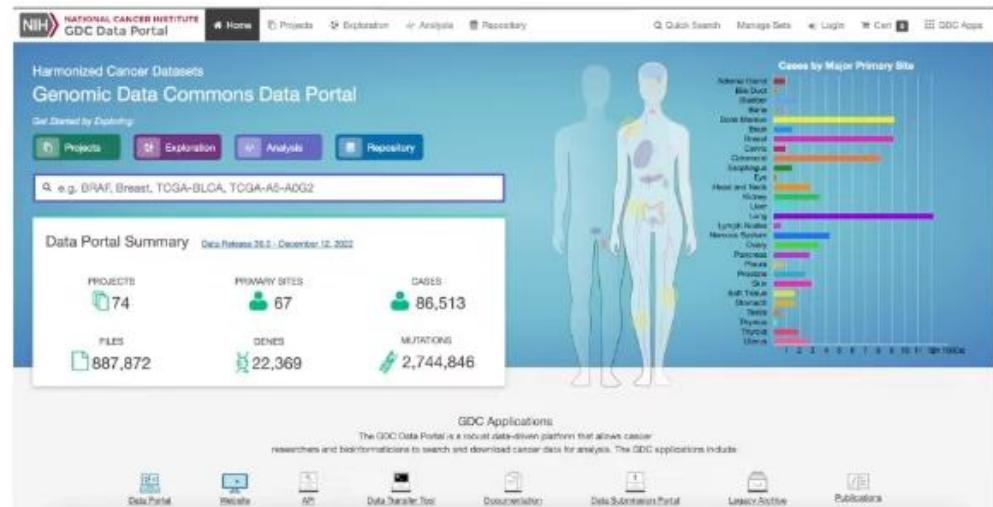
2 - Generalized DATA RESOURCES .

Download data from GDC Portal using TCGAbiolinks R package



What is NCI's GDC portal?

- Genomic Data Commons (GDC) is National Cancer Institute's (NCI's) data sharing platform
- The GDC contains NCI-generated data from some of the largest and most comprehensive cancer genomic datasets, including The Cancer Genome Atlas (TCGA) and Therapeutically Applicable Research to Generate Effective Therapies (TARGET).
- Each of these projects contains a variety of processed and unprocessed molecular data types, including genomics, epigenomics, proteomics, imaging, clinical, and others.
- Portal not only provides a large number of cancers but also data from a large number of experimental platforms.

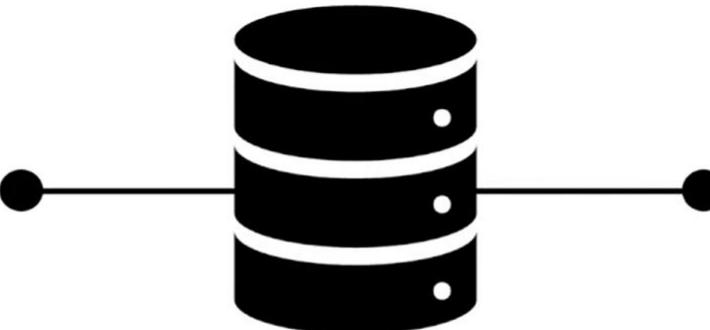


<https://portal.gdc.cancer.gov/>

Harmonized

Processed using standardized pipelines and aligned to/lifted over to the reference genome GRCh38 (hg38)

- Harmonized data represents a more recent and **standardized** phase of TCGA data.
- It involves the reprocessing and reanalysis of TCGA data to provide consistent and standardized results across different cancer types.
- The harmonization process aims to address issues and variations that might have existed in the earlier, "legacy" data.
- Harmonized data is intended to provide a more cohesive and integrated view of the genomic and clinical information from TCGA.



Legacy

Data originally aligned and processed, mostly aligned with hg19 (GRCh37)

- Legacy data refers to the **original data** and analysis from the early stages of TCGA.
- This data may have been generated using different platforms, technologies, and analysis methods, leading to some inconsistencies or variations in results between different cancer types.
- While legacy data is still valuable and has been widely used in research, it may require additional effort to harmonize and integrate it with other data sources.

Data Access

- **Open:** includes high level genomic data that is not individually identifiable, as well as most clinical and all biospecimen data elements.
- **Controlled:** includes individually identifiable data such as low-level genomic sequencing data, germline variants, SNP6 genotype data, and certain clinical data elements

TCGAbiolinks Bioconductor Package

- An R package that facilitates data retrieval and analysis
- Provides different options to query and download relevant data
- Allows to integrate different data types
- Can be used for different types of analyses dealing with all platforms such as differential expression, network inference or survival analysis, etc.
- Allows visualization of the obtained results.



3 Main Functions

GDCquery

- Search the data for a given project and data category and filters the results by samples, sample type, file type and others features if requested by the user.
- Returns an object with a summary table with the results found (samples, files and other useful information) and the arguments used in the query.

GDCdownload

- Users will be able to download the data using this function.
- The downloaded data will be saved in a directory with the project name and a sub-folder with the data.category, for example “TCGA-GBM/DNA_methylation”.

GDCprepare

- Transforms the downloaded data into a SummarizedExperiment object or a data frame.
- For SummarizedExperiment object, TCGAbiolinks will add to the object sub-type information, which was defined by The Cancer Genome Atlas (TCGA) Research Network reports and clinical information.

Requirement

- R (v4.1.1)
- RStudio (v2022.12.0+353)

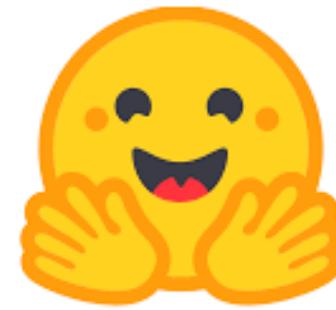
R packages:

- TCGAbiolinks (v2.25.0)
- SummarizedExperiment (v1.24.0)
- maftools (v2.8.05)
- pheatmap (v1.0.12)
- tidyverse (v1.3.1)





Let's
Code
More.



Thank you.

Article

Cancers in Agreement? Exploring the Cross-Talk of Cancer Metabolomic and Transcriptomic Landscapes Using Publicly Available Data

Derek van Tilborg  and Edoardo Saccenti *

Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng,
6708 WE Wageningen, The Netherlands; derek.vantilborg@wur.nl

* Correspondence: edoardo.saccenti@wur.nl; Tel.: +31-(0)317-486948

Review

Proteomics, Personalized Medicine and Cancer

Miao Su ^{1,†}, Zhe Zhang ^{1,†}, Li Zhou ¹, Chao Han ¹, Canhua Huang ^{1,*} and Edouard C. Nice ^{2,*} 

¹ State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, and West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, and Collaborative Innovation Center for Biotherapy, Chengdu 610041, China; 2019224065144@stu.scu.edu.cn (M.S.); scuzz@stu.scu.edu.cn (Z.Z.); 2015224060079@stu.scu.edu.cn (L.Z.); 2017141494205@stu.scu.edu.cn (C.H.)

² Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia

* Correspondence: hcanhua@scu.edu.cn (C.H.); ed.nice@monash.edu (E.C.N.)

† These authors contributed equally to this paper.

Computational resources for identification of cancer biomarkers from omics data

Harpreet Kaur[†], Rajesh Kumar[†], Anjali Lathwal[†] and Gajendra P.S. Raghava

Corresponding author: Gajendra P.S. Raghava, Department of Computational Biology, Indraprastha Institute of Information Technology, Okhla Industrial Estate, Phase III, New Delhi 110020, India. Tel.: +91 011 26907444; E-mail address: raghava@iiitd.ac.in

[†]These authors have contributed equally.



Review

A Detailed Catalogue of Multi-Omics Methodologies for Identification of Putative Biomarkers and Causal Molecular Networks in Translational Cancer Research

Efstathios Iason Vlachavas ^{1,*†}, Jonas Bohn ^{1,†}, Frank Ückert ^{1,2} and Sylvia Nürnberg ^{1,2,*}

¹ Medical Informatics for Translational Oncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; j.bohn@dkfz-heidelberg.de (J.B.); f.ueckert@dkfz-heidelberg.de (F.U.)

² Applied Medical Informatics, University Hospital Hamburg-Eppendorf, 20251 Hamburg, Germany

* Correspondence: Efstathios-Iason.Vlachavas@dkfz-heidelberg.de (E.I.V.); s.nuernberg@uke.de (S.N.)

† These authors contribution is equally to this work.

ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions

[Terrence S. Furey](#)

[Nature Reviews Genetics](#) 13, 840–852 (2012) | [Cite this article](#)

42k Accesses | 438 Citations | 26 Altmetric | [Metrics](#)

NGS based epigenomics techniques include

ChIP-seq (chromatin immunoprecipitation),

DNase1-seq (DNase I hypersensitive sites – sequencing) or

FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements –sequencing) assay for mapping the DNA-protein interactions or chromatin accessibility,

ChiRP-seq (Chromatin Isolation by RNA Purification) for mapping DNA-RNA interaction and whole-genome bisulfite/array-based sequencing for mapping DNA methylation



Integration of Online Omics-Data Resources for Cancer Research

Tonmoy Das¹, Geoffroy Andrieux^{2,3}, Musaddeque Ahmed⁴ and Sajib Chakraborty^{1}*

¹ Molecular Systems Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh, ² Medical Center – University of Freiburg, Faculty of Medicine, Institute of Medical Bioinformatics and Systems Medicine, University of Freiburg, Freiburg, Germany, ³ German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Partner Site Freiburg, Freiburg, Germany, ⁴ Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada

Received: 30 June 2020

Accepted: 05 October 2020

Published: 23 October 2020

Presented by : Siamak Salimi

More Is Better: Recent Progress in Multi-Omics Data Integration Methods

Sijia Huang^{1,2}, Kumardeep Chaudhary¹ and Lana X. Garmire^{1,2,3*}

¹ Epidemiology Program, University of Hawai'i Cancer Center, Honolulu, HI, United States, ² Molecular Biosciences and Bioengineering Graduate Program, University of Hawai'i at Manoa, Honolulu, HI, United States, ³ Department of Obstetrics, Gynecology, and Women's Health, John A. Burns School of Medicine, University of Hawai'i at Manoa, Honolulu, HI, United States

Volume 9, Issue 7, July 2023, e17653

A deep learning-based framework for predicting survival-associated groups in colon cancer by integrating multi-omics and clinical data

Siamak Salimy^a, Hossein Lanjanian^b, Karim Abbasi^c, Mahdieh Salimi^d, Ali Najafi^e, Leili Tapak^f,
Ali Masoudi-Nejad^{a1}  

