NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

The **Nucleotide database** is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

Using https://www.ncbi.nlm.nih.gov/nucleotide/, you can find what type of genome has the virus COVID-19

Write COVID-19 in the search bar.

ENSEMBL BROWSER

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotates genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Select your favorite organism and **write in the search bar**, the name of your favourite gene, for example, FOXP3. Select human FOXP3 from the menu and you will open the FOXP3 gene page. Near the name, FOXP3, there is the Ensembl gene identifier.
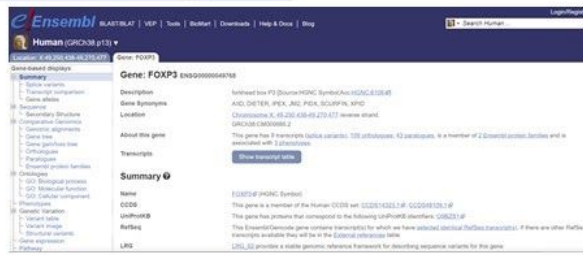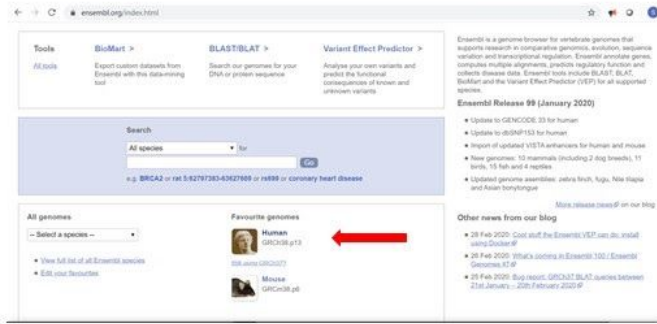
How many transcripts does FOXP3 encode?

In the FOXP3 gene page, you can find the number of transcripts, and the table of all transcripts. If you don't see the table, click on the "Show transcript table".

This table summarizes the differences between transcripts.

Indicate the chromosome coordinate.

You can see several transcripts and you have to indicate the length of the transcript and the number of amino acids about FoxP3-203.

Click on Transcript ID.

In the menu on the left, you can select Exons.

Scroll down the page and you can find the sequence of this transcript that is signed with a translated **sequence in blue**, **flanking regions in green** are genomic regions upstream or downstream of the transcript, and **untranslated regions (UTR) are indicated in red**. Can you recognize the first exon? Please, indicate the coordinate: the number of starting and ending nucleotides.

The other nucleotides marked with different colours indicate the variants, nucleotides that change in the genome and some of them may be associated with mutation involved in the pathology.

The Human Protein Atlas

The Human Protein Atlas is a European project with the aim to map all the human proteins in cell lines, tissues, and organs by integrating data from antibody-based imaging, mass spectrometry-based proteomics, and transcriptomics. The resources from this project can be accessed through the website http://www.proteinatlas.org.
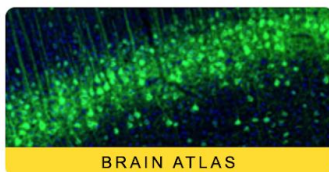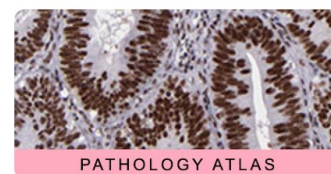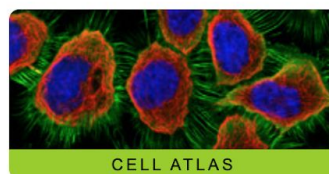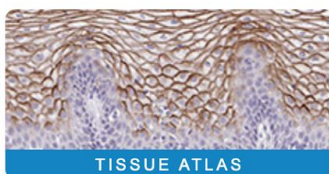
From the homepage of the website, you can access the different atlases of the project. The atlases describe results from expression analysis from physiological tissues (Tissue Atlas), cell lines (Cell Atlas), pathological tissues (Pathology Atlas), brain regions (Brain Atlas), blood cells (Blood Atlas).

From the homepage, you can search for data of a specific gene by **writing the gene symbol in the search field**. A summary page with the gene information will be provided by indicating the main characteristics of the gene expression considering the different atlases.

Click on gene name, you can see general information

In the Tissue Atlas, you can retrieve information of proteins and RNAs expression in physiological tissues. The expression levels are reported as colored histograms in which each color represents a specific tissue class. The histograms can be sorted based on specific characteristics including the tissue of origin and the expression level. The protein level of expression is qualitative while the RNA expression level is quantitative. Please read the Help section to understand how these levels are computed.
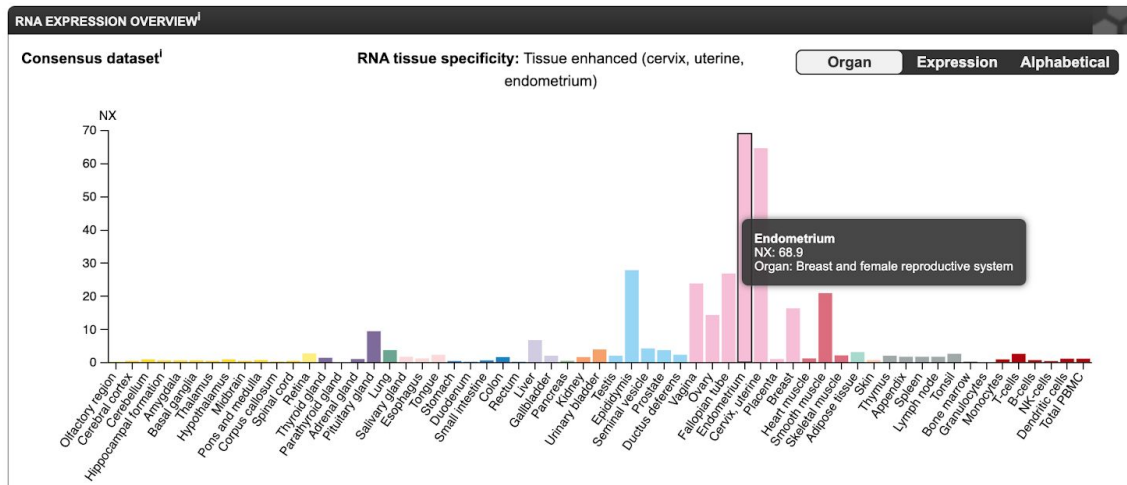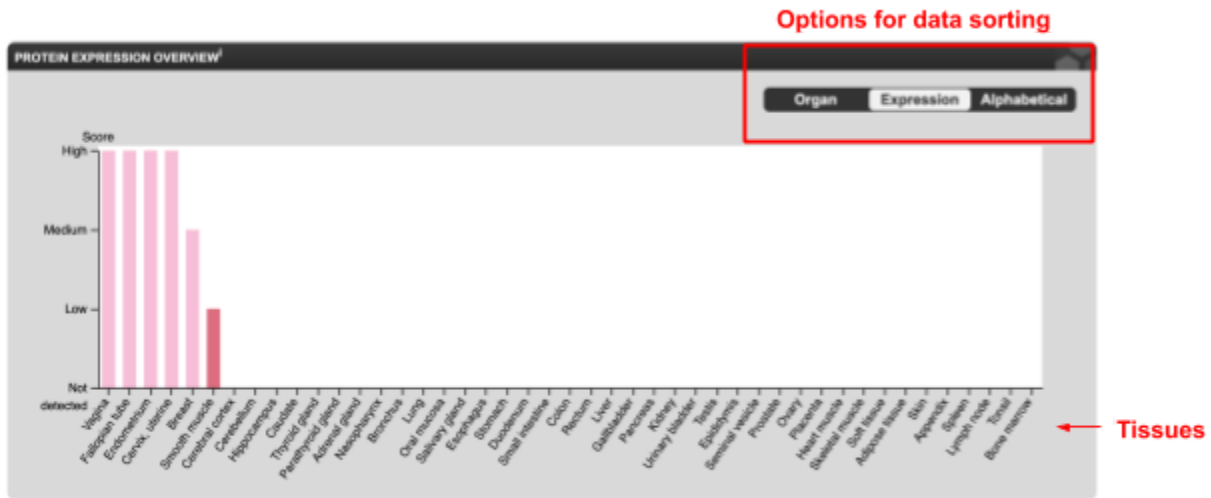
Conversely. in the **Cell Atlas** you can retrieve information of proteins and RNAs expressions in cell line models with the indication of the protein cellular localization. In the picture in green are reported the cellular localization in which the protein was verified to be localized.

In the section **Human cells** of this atlas is it possible to explore results from immunocytochemistry analyses performed using an antibody against the protein. By Clicking on the **Toggle channels** buttons it is also possible to observe the co-staining with antibodies targeting the nuclear compartment, ER, or the microtubules. The intensity of fluorescence of the target protein can be also observed by clicking on the Intensity button.

In the **Pathology Atlas** is it possible to obtain the information on the gene expression in tumor samples from the TCGA project and the information of the relationship between the gene expression and the patient's survival.

The relation between the protein expression and the patient's survival is reported in the **PROGNOSTIC SUMMARY** panel reporting only the significant association. In pink is indicated the survival of patients with a high protein expression while in blue the survival of patients with low gene expression. In the **RNA EXPRESSION OVERVIEW** pane, it is reported as a box plot

the RNA expression of the gene in different tumor types. Immunohistochemistry data are also available in the section **PROTEIN EXPRESSION.**

EBI Expression Atlas

Expression Atlas is a resource to query gene and protein expression data across species and biological conditions and to visualise down-stream analysis results to explore co-expression. It contains thousands of selected microarray and RNA-sequencing data that are manually curated and annotated with ontology terms, checked for high quality and re-analysed using standardised methods. The atlas can be accessed at https://www.ebi.ac.uk/gxa/home.

In this database, the expression levels are expressed as FPKM (fragments per kilobase of exon model per million reads mapped) and TPM (transcripts per million) which are the most common units reported to estimate gene expression based on RNA-seq data. These units normalized gene expression by considering:

1) The number of reads from a gene depends on its length. One expects more reads to be produced from longer genes.

2) The number of reads from a gene depends on the sequencing depth that is the total number of reads you sequenced. One expects more reads to be produced from the sample that has been sequenced to a greater depth.

The expression of a specific gene can be searched using from the home page. Specific filters based on the analysed species and the biological condition can be selected.

The first section of the atlas is called Baseline Atlas which reports data from good quality experiments from different conditions (e.g. tissues, cell types, developmental stages).
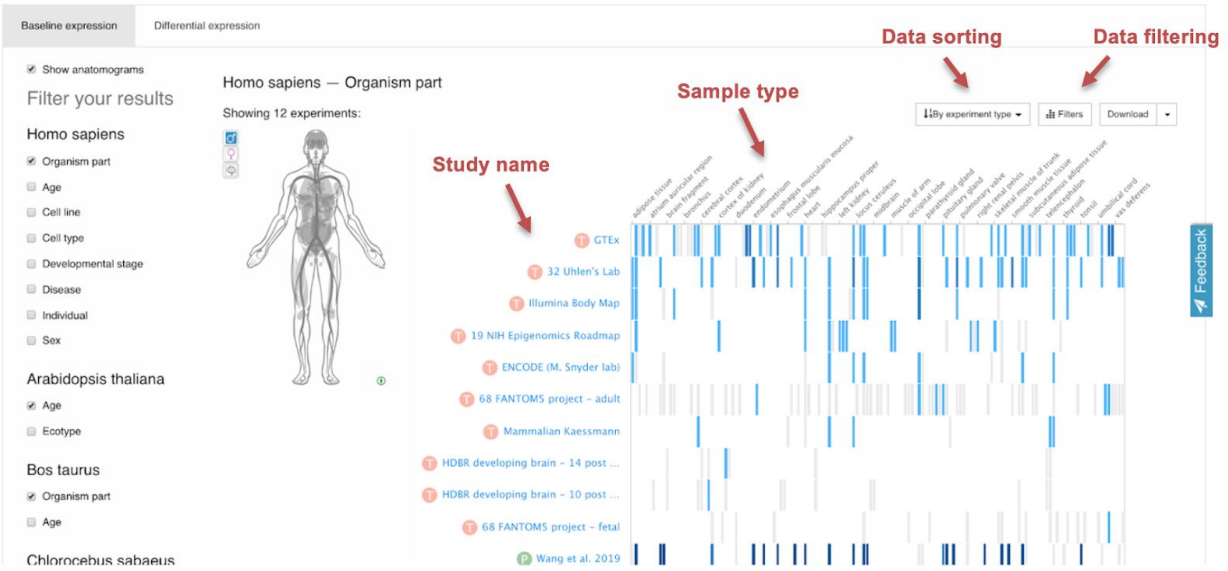
Expression levels are reported using a blue scale color code in which darker colors represent higher expression levels. Using the filter at the top of the heat map it is also possible to sort the data by expression rank or filter samples characterized by a low expression level.

Conversely, in the Differential Expression section, it is possible to identify experimental comparisons in which a specific gene is detected as significantly down-regulated or up-regulated. The expression differences with respect to a control condition are reported as log2 fold change.

EBI Single Cell Expression Atlas

A more recent atlas provided by the EBI is the Single Cell Expression Atlas which allows the analysis of public data from single-cell RNA-Sequencing experiments. The atlas can be accessed at https://www.ebi.ac.uk/gxa/sc/home.

By indicating the symbol of a gene of interest in the main form of the website it is possible to obtain a list of single cells RNA-Seq experiments in which the gene was detected.

By selecting a specific study, a t-SNE plot will be displayed reporting on the left a set of single cells clusters derived from the analysis and on the right, the same clustering result colored based on the expression level of the gene of interest.

A different color-code can be used to distinguish the cells based on specific features (eg. tissue of origin, expression of specific markers, the gender of the subjects, etc.)
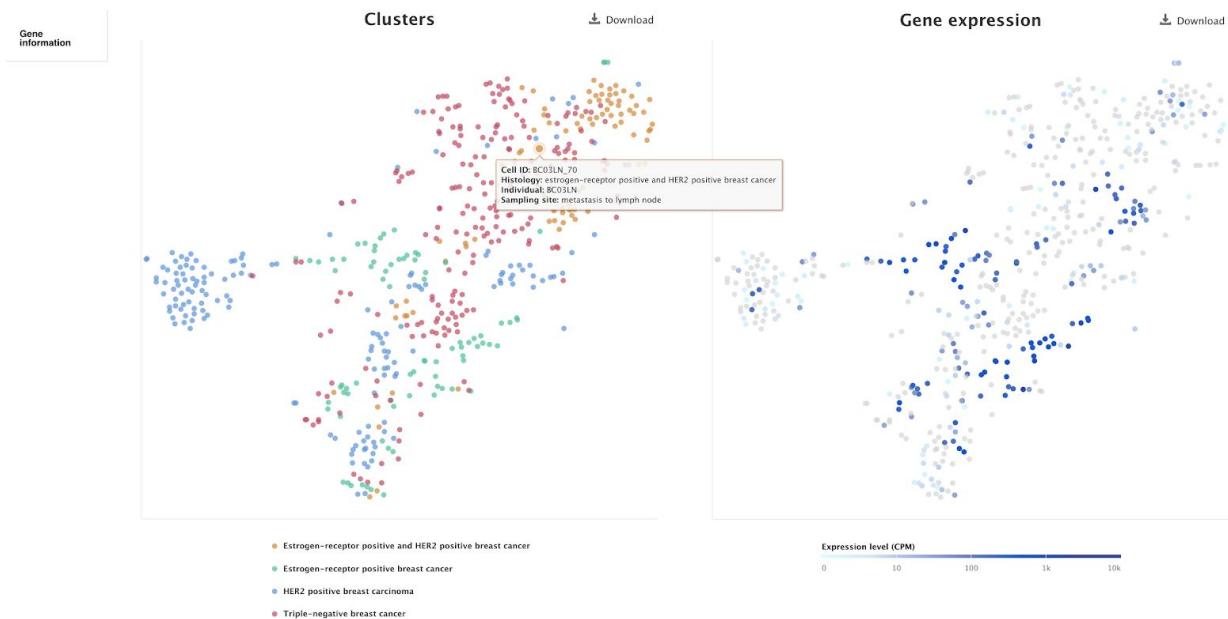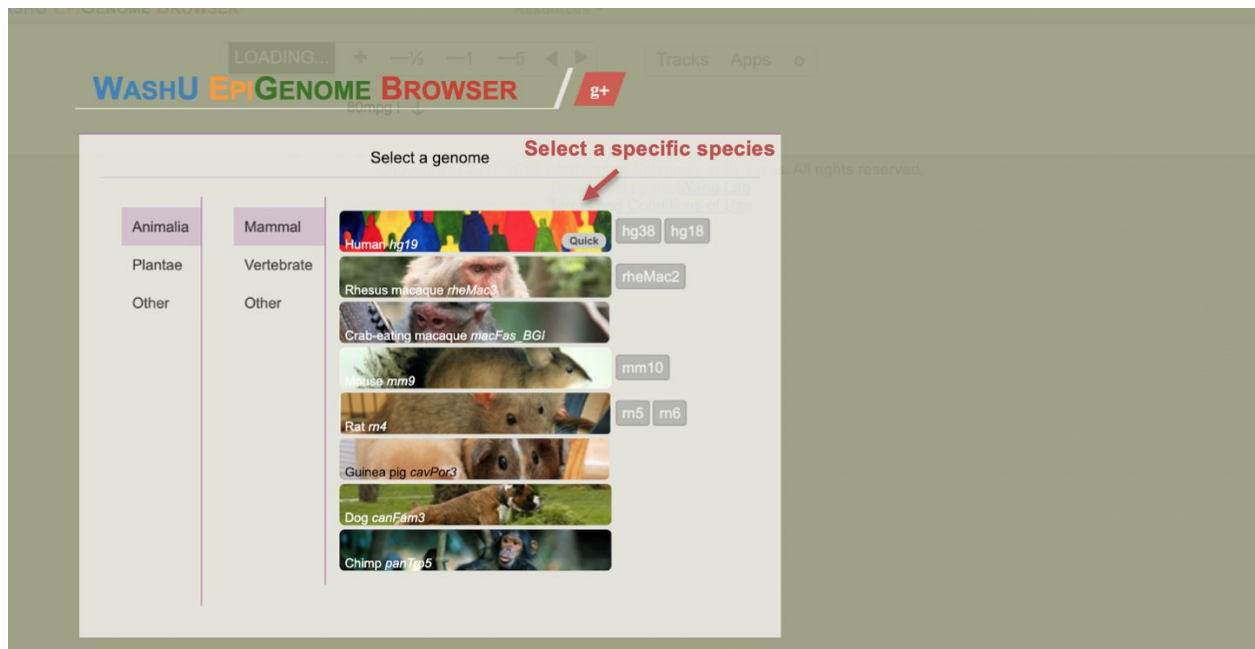
The WashU Epigenome Browser

The WashU Epigenome Browser is a web tool which allows the visualization of the results from genomic and transcriptomic experiments from international projects and single studies. The tool can be accessed at http://epigenomegateway.wustl.edu/legacy/.

Video tutorial on the use of this browser can be found at: http://epigenomegateway.wustl.edu/support/video.html

From the homepage is it possible to select a specific species of interest and the following results will be the representation of a genomic region whose coordinates will be reported on the top. In the genome browser, each row represents specific information, including genomic positions, gene annotations, annotations of repetitive elements, and coverage signals from sequencing experiments, particularly ChIP-Seq and RNA-Seq.
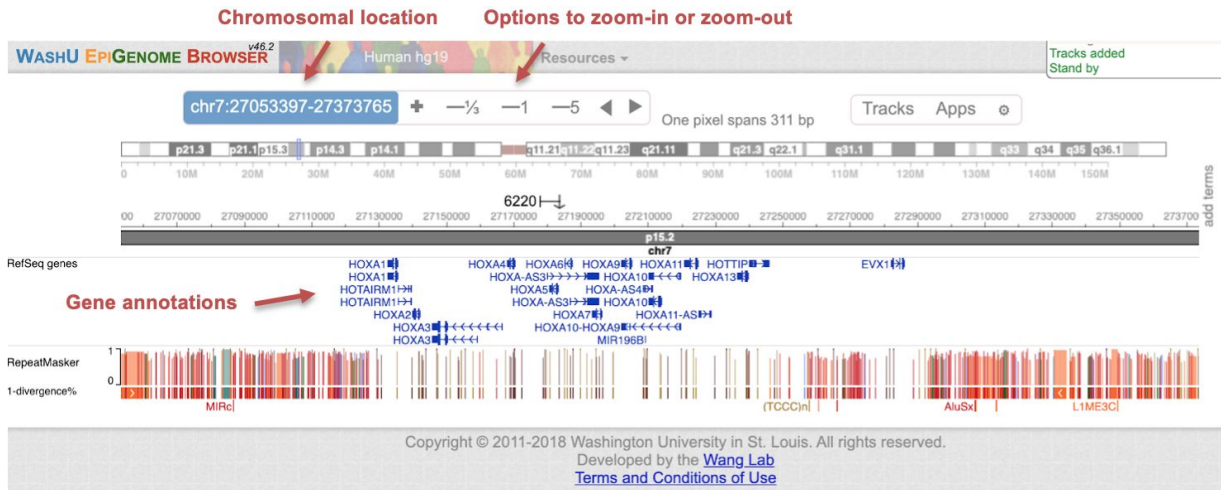
It is possible to access the information from public experiments using the "**Public Track Hubs**" function in the "**Tracks**" section. Then, data from different projects can be selected by clicking on the "**Load**" button. Finally, it is possible to display the data of specific experiments from the selected project by clicking on the "**Tracks > Click for track table**" section. In this section data from different assays (columns) generated from different samples (rows) can be selected.
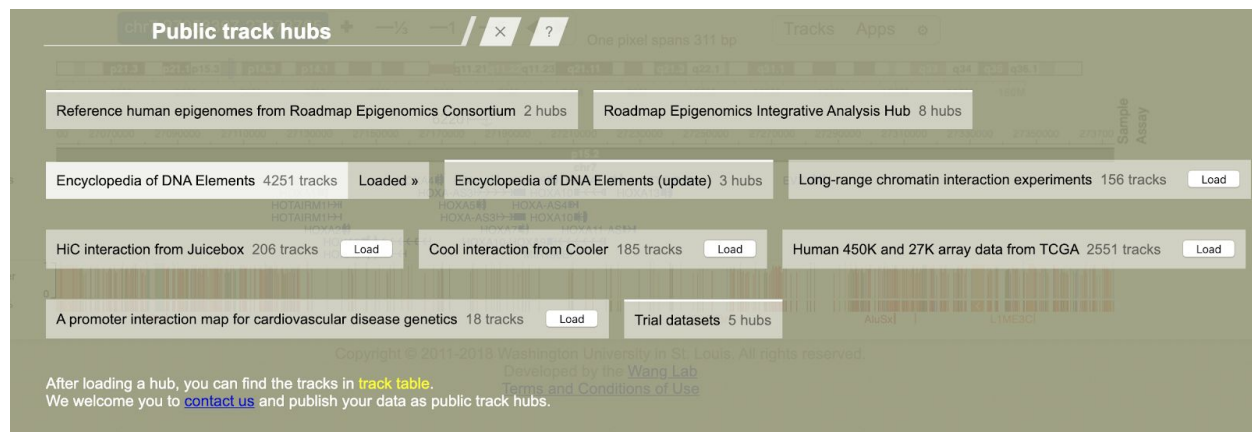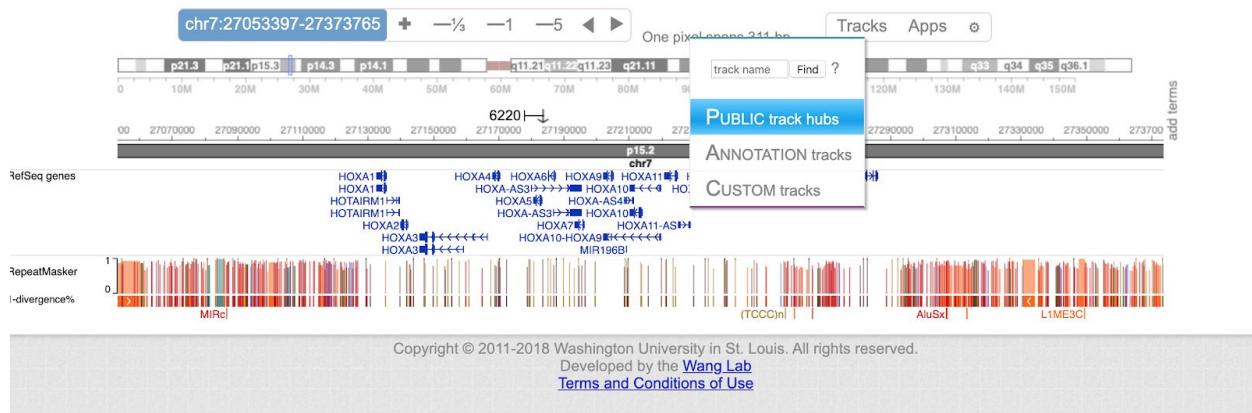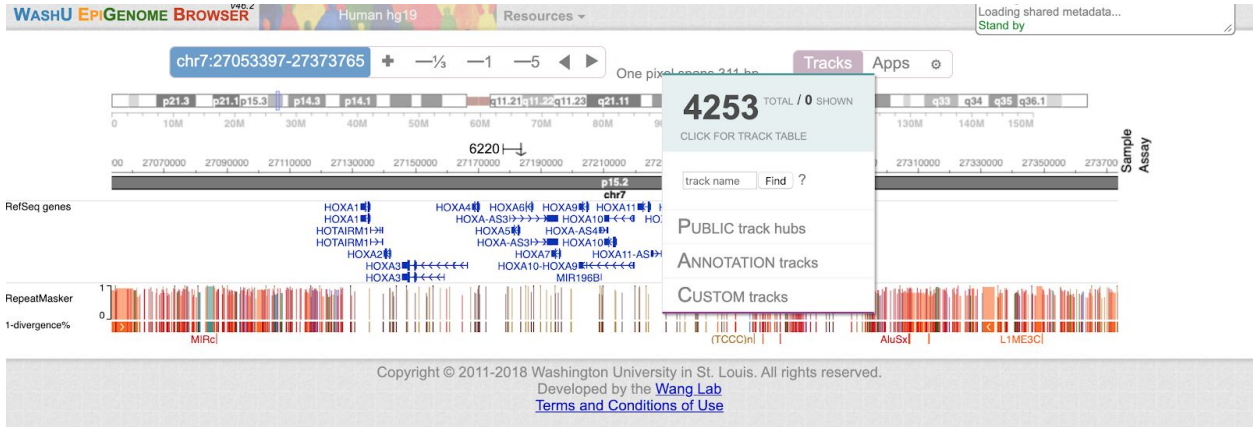
To add the experiment of a specific sample click on its name and press the green button "**Add 1 track**". Multiple tracks can be selected and added together.
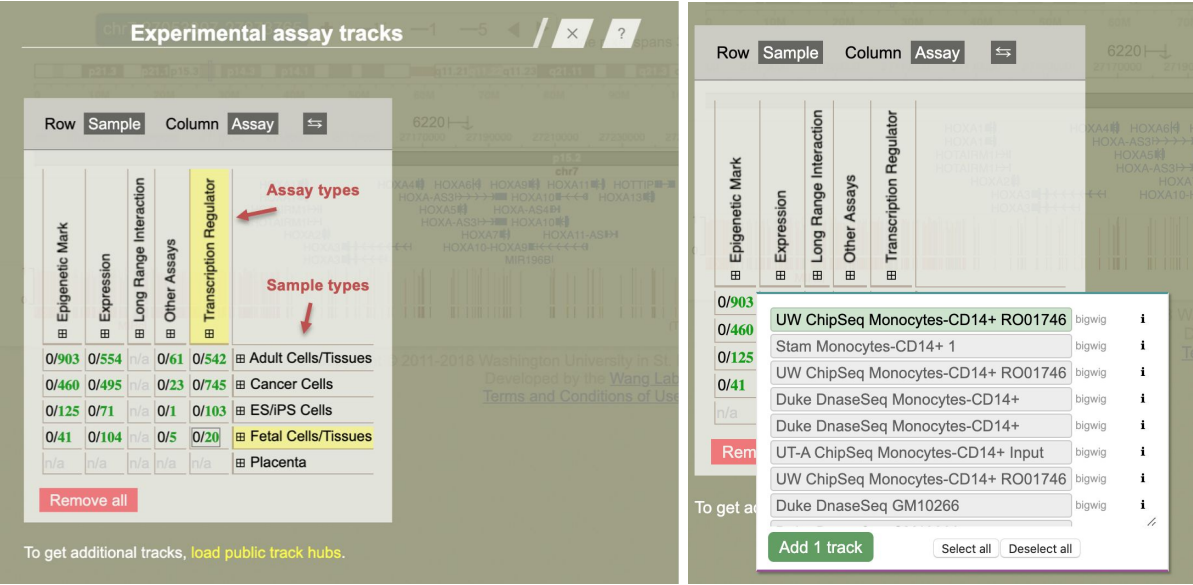
Each selected track will be initially shown in a compact mode but by **right-clicking** on the track and clicking on **Configure**, is it possible to visualize the peak coverage signals by clicking two times on the button "**+**" in the "**Height**" section.

A selected track

Right-click to configure the track

Increase the high to see the the coverage signal
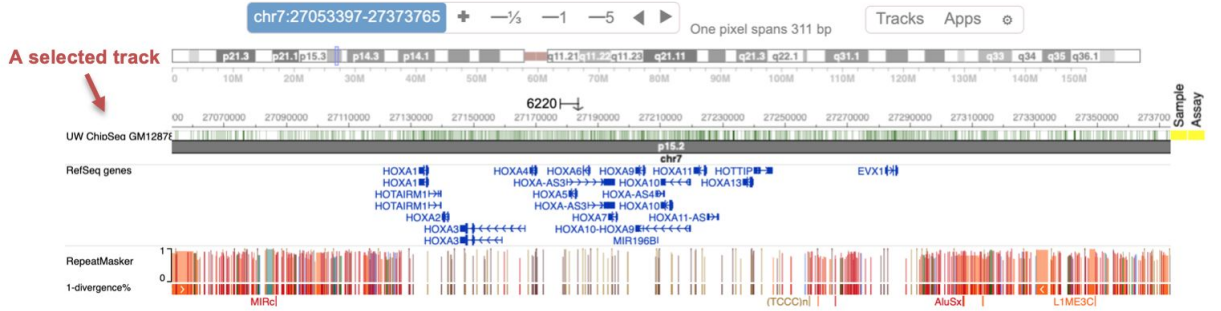
CBioPortal

The CBioPortal is a web tool which allows the exploration and analysis of cancer-related data, particularly from the TCGA consortium. The website is accessible at https://www.cbioportal.org/.

From the homepage is it possible to select a specific study and the main features of this study can be obtained using the button "**Explore Selected Studies**".

Furthermore, it is possible to analyse the genomic features of single or multiple genes as measured in specific studies by using the button "**Query by gene**". Insert the gene names in the main form and click on **Submit Query**.

The result will be a summary infographic reporting the number and types of alterations observed for the query genes in a selected tumor type. Each rectangle is a subject and a specific color-code is used to report the presence of a molecular alteration at the query genes.
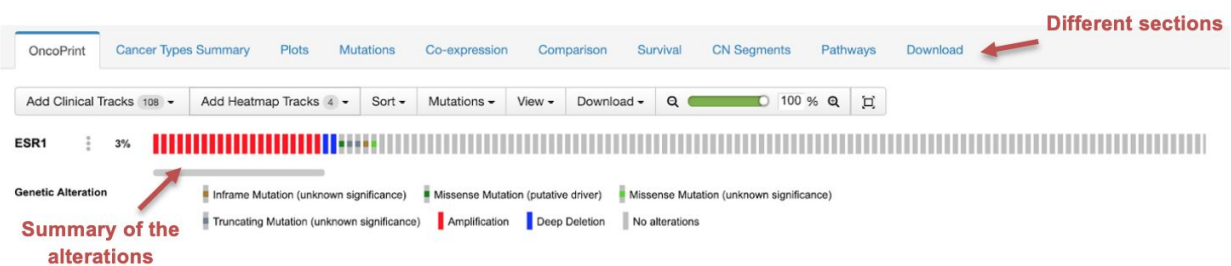
Other relevant information can be retrieved from this section, including the relation between gene alterations and the patient survival, the details on the localization of the identified mutations, the relation between different types of alterations, or the co-presence of a specific alteration affecting different genes.

Site of the mutation

ESR1

OncoPrint	Cancer Types Summary	Plots	**Mutations**	Co-expression	Comparison	Survival	CN Segments	Pathways	Download

Add annotation tracks Y-Axis Max: 5
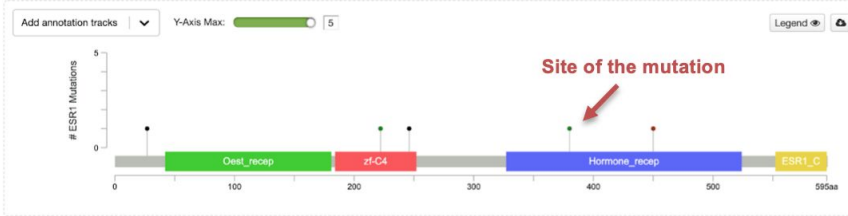
ESR1
RefSeq: NM_000125
Ensembl: ENST00000206249
CCDS: CCDS5234
UniProt: ESR1_HUMAN

Somatic Mutation Frequency ⓘ 0.6%

2 Missense **2** Truncating
1 Inframe **0** Other

View 3D Structure

5 Mutations (page 1 of 1)

Sample ID	Cancer Type	Protein Change	Annotation ▼	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
TCGA-D8-A27V-01	Breast Invasive Lobular Carcin...	E380Q		Missense	Gain	1	0.30	166
TCGA-BH-A0DS-01	Breast Invasive Ductal Carcinoma	P222S		Missense	Diploid	1	0.35	34
TCGA-E2-A10A-01	Breast Invasive Ductal Carcinoma	P29Sfs*79		FS del	Diploid	1	0.30	21
TCGA-B6-A1KI-01	Breast Invasive Ductal Carcinoma	I451_I452del		IF del	Diploid		0.11	17
TCGA-C8-A12T-01	Breast Invasive Ductal Carcinoma	Y246*		Nonsense	Diploid		0.17	129

Showing 1-5 of 5 Mutations

Effect of the mutations

OncoPrint	Cancer Types Summary	Plots	Mutations	Co-expression	**Comparison**	Survival	CN Segments	Pathways	Download

Groups: (drag to reorder) Altered group (28) Unaltered group (788) ESR1 (28) Select all | Deselect all

Overlap Survival Clinical Mutations Copy-number mRNA Protein

Survival types: Overall Disease Free

Overall
Overall patient survival status.

Significance of the difference in survival

X-Axis Max: 283 Months Survival

Logrank Test P-Value: 0.501

■ Altered group
■ Unaltered group

Survival curves of subjects with alteration or without

Months Overall
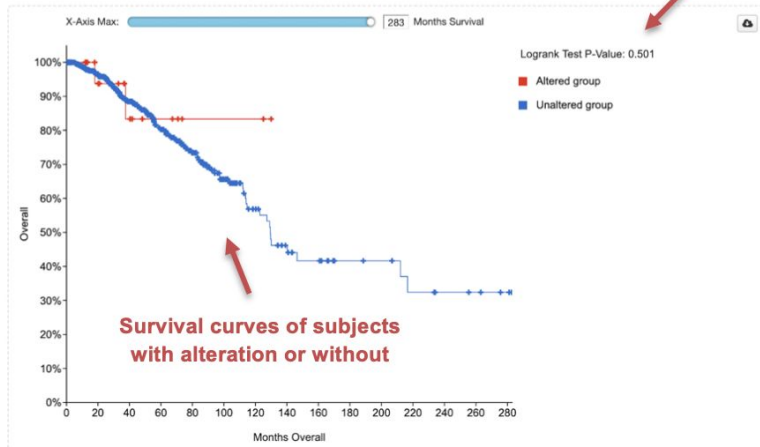
Gene Expression Omnibus (GEO)

The Gene Expression Omnibus (GEO) is an NCBI curated repository of data from public high-throughput experiments. In this website is it possible to retrieve information specific experiments of interest with the possibility to obtain the raw data as well as results from the analysis performed by the authors of the experiment. The website is accessible at https://www.ncbi.nlm.nih.gov/geo/.

From the homepage is it possible to search a specific experiment based on a keyword or its accession number. The result will be a list of GEO datasets with a description and information, including the indication of the organism considered, the type of experiments, the high-throughput platform used and the number of samples.

By selecting a specific dataset a set of information will be displayed, including the summary of the experiment, the overall design, the authors and the related publication, some contact information, the used platform and the page related to every single sample. At the bottom, additional data including results from analyses performed by the authors of the experiments will be reported.

Series GSE118432   Query DataSets for GSE118432

Status	Public on Apr 10, 2020
Title	SFRP1-regulated gene expression in premalignant breast lesions
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	• Atypical hyperplasias (AH) provide insights into early changes that may predispose breast epithelial cells to oncogenic transformation. • Of genes associated with premalignancy in prior studies, only mRNA levels of ESR1 and SFRP1 were detected in the present study. • Transcriptional profiling defined signatures distinguishing atypical hyperplasias. The patterns of expression were similar among hyperplastic lesions of lobular and ductal phenotype suggesting a common set of alterations underlying both lesions. Pathway analyses identified elevated expression of estrogen receptor alpha, androgen receptor and EGFR receptors and Rho signaling as central events nodes in the pathways altered in AH. • A set of 43 genes were identified as common targets using 2 different algorithms to detect signatures associated with AH. Knockdown of SFRP1 in a TERT immortalized breast epithelial cell line resulted in 14 genes from this signature being either up-regulated or down-regulated as observed in the expression profiles from AH. • The results demonstrate a signature of genes representing alterations that are common to the development of hyperplasias in both ductal and lobular epithelium. Loss of SFRP1 expression is a key player underlying the transcriptional changes in AH that directs a module of genes that can be used to improve reproducibility of diagnosis of AH.
Overall design	In the present study, patients with atypical hyperplasia (AH) but no history of breast cancer were selected. Laser capture microdissection was used to collect both histologically normal benign epithelium (HNB) as well as AH tissues from each patient. The complete transcriptome was evaluated using microarrays and used to define signatures that distinguish AH lesions from the HNB tissues.
Contributor(s)	Gregory K, Roberts A, Mayfield J, Conlon E, Crisi GM, Makari-Judson G, Mason H, Wang J, Schneider B, Zhu J, Simin K, Schneider S, Jerry DJ
Citation(s)	Gregory KJ, Roberts AL, Conlon EM, Mayfield JA et al. Gene expression signature of atypical breast hyperplasia and regulation by SFRP1. *Breast Cancer Res* 2019 Jun 27;21(1):76. PMID: 31248446

Submission date	Aug 10, 2018
Last update date	Apr 10, 2020
Contact name	D. Joseph Jerry
E-mail(s)	jjerry@vasci.umass.edu
Phone	413-545-5335
Organization name	University of Massachusetts
Department	Veterinary and Animal Sciences
Lab	Jerry
Street address	661 North Pleasant St
City	Amherst
State/province	MA
ZIP/Postal code	01003
Country	USA

Platforms (1)	GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]
Samples (42) ⊞ More...	GSM3330068 Breast_Atypia_S-10-3
	GSM3330069 Breast_Benign_S-10-3
	GSM3330070 Breast_Atypia_S-10-10

Information on specific samples

Relations
BioProject   PRJNA485537

[Analyze with GEO2R]

Download family		Format
SOFT formatted family file(s)		SOFT ⍰
MINiML formatted family file(s)		MINiML ⍰
Series Matrix File(s)		TXT ⍰
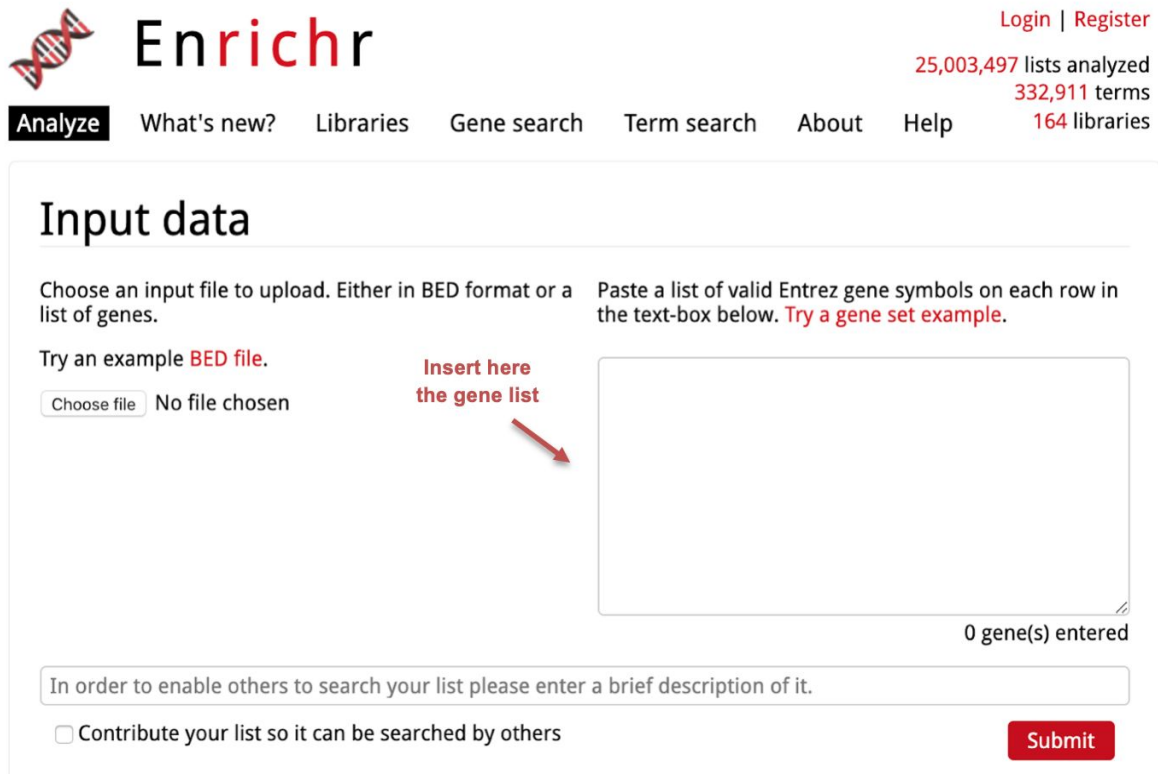
Additional data

Supplementary file	Size	Download	File type/resource
GSE118432_RAW.tar	157.2 Mb	(http)(custom)	TAR (of CEL)

Enrichr

Enrichr is a web tool for the functional enrichment analysis of a list of genes based on the gene sets annotations from multiple databases, including Gene Ontology, KEGG, and Reactome. The website is accessible at https://amp.pharm.mssm.edu/Enrichr/. From the homepage is it possible to indicate the list of genes in the main form. Then, by clicking on the button "**Submit**" the tool will display the list of enriched terms separated based on the database of origin and on the type of information stored in the database.

A colored bar plot represents the extent of enrichment but it is possible to select a specific result by clicking on the barplot. Furthermore, in the **Table** section it is possible to retrieve the exact information on the analysis significance as well as the number and name of enriched genes.

The enrichment represents the statistical significance of observing a specific overlap between the input gene list and the list of genes annotated to a specific term. Please refer to http://amp.pharm.mssm.edu/Enrichr/help#basics to further understand how the statistical significance is computed.

Enrichr

Transcription   **Pathways**   **Ontologies**   **Diseases/Drugs**   **Cell Types**   **Misc**   **Legacy**   **Crowd**

Description Sample gene list (375 genes)

GO Biological Process 2018 ❶
- alpha-amino acid metabolic process (GO:19
- glucan biosynthetic process (GO:0009250)
- glycogen biosynthetic process (GO:0005978
- mitochondrion organization (GO:0007005)
- organic cyclic compound biosynthetic proce

GO Molecular Function 2018 ❶
- retinal dehydrogenase activity (GO:0001758
- tRNA (cytosine) methyltransferase activity (C
- oxidoreductase activity, acting on the aldehy
- fatty acid synthase activity (GO:0004312)
- 4 iron, 4 sulfur cluster binding (GO:0051539

GO Cellular Component 2018 ❶
- mitochondrion (GO:0005739)
- mitochondrial matrix (GO:0005759)
- mitochondrial inner membrane (GO:000574
- nuclear exosome (RNase complex) (GO:0000
- peroxisome (GO:0005777)

MGI Mammalian Phenotype Level 4 2019 ❶
- MP:0010400 increased liver glycogen level
- MP:0010956 abnormal mitochondrial ATP sy
- MP:0008918 microgliosis
- MP:0013741 absent thymus corticomedullar
- MP:0000607 abnormal hepatocyte morpho

Human Phenotype Ontology ❶
- Abnormality of methionine metabolism (HP:
- Abnormality of aspartate family amino acid
- Gaze-evoked nystagmus (HP:0000640)
- Tubulointerstitial abnormality (HP:0001969)
- Abnormality of sulfur amino acid metabolis

Jensen TISSUES ❶
- Mesocarp
- Imaginal disc
- Seedling
- Mesophyll
- Myeloid leukemia cell

Enrichr

GO Biological Process 2018

Bar Graph   **Table**   Clustergram   ⚙ ❶

Hover each row to see the overlapping genes.

10 ⬍ entries per page

Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	alpha-amino acid metabolic process (GO:1901605)	0.002565	1.000	6.88	41.06
2	glucan biosynthetic process (GO:0009250)	0.004328	1.000	8.89	48.38
3	glycogen biosynthetic process (GO:0005978)	0.004328	1.000	8.89	48.38
4	mitochondrion organization (GO:0007005)	0.004352	1.000	2.87	15.63
5	organic cyclic compound biosynthetic process (GO:1901362)	0.004929	1.000	5.77	30.63
6	NADH metabolic process (GO:0006734)	0.005003	1.000	17.78	94.18
7	lipoprotein catabolic process (GO:0042159)	0.005003	1.000	17.78	94.18
8	NADP metabolic process (GO:0006739)	0.005069	1.000	8.42	44.50
9	aromatic amino acid family catabolic process (GO:0009074)	0.005882	1.000	8.00	41.09
10	molybdopterin cofactor biosynthetic process (GO:0032324)	0.006918	1.000	15.24	75.79

Showing 1 to 10 of 1,228 entries | Export entries to table

Terms marked with an * have an overlap of less than 5

Previous   Next