

NCBI

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

The screenshot shows the NCBI website homepage. At the top, there is a navigation bar with the NCBI logo, a search bar, and links for "Resources" and "How To". A prominent red banner in the center contains a COVID-19 alert: "COVID-19 is an emerging, rapidly evolving situation. Get the latest public health information from CDC: <https://www.cdc.gov/coronavirus>. Get the latest research from NIH: <https://www.nih.gov/coronavirus>." Below the banner, the main content area is divided into several sections. On the left is a "Resource List (A-Z)" sidebar with categories like "All Resources", "Chemicals & Bioassays", "Data & Software", "DNA & RNA", "Domains & Structures", "Genes & Expression", "Genetics & Medicine", "Genomes & Maps", "Homology", "Literature", "Proteins", "Sequence Analysis", "Taxonomy", and "Training & Tutorials". The central "Welcome to NCBI" section features a mission statement and a grid of six service tiles: "Submit" (Deposit data or manuscripts into NCBI databases), "Download" (Transfer NCBI data to your computer), "Learn" (Find help documents, attend a class or watch a tutorial), "Develop" (Use NCBI APIs and code libraries to build applications), "Analyze" (Identify an NCBI tool for your data analysis task), and "Research" (Explore NCBI research and collaborative projects). On the right, there are sections for "Popular Resources" (PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem) and "NCBI News & Blog" (Recalculation of prokaryotic reference and representative genome assemblies, dated 10 Apr 2020).

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.

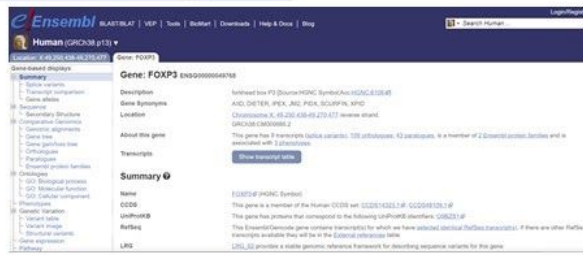
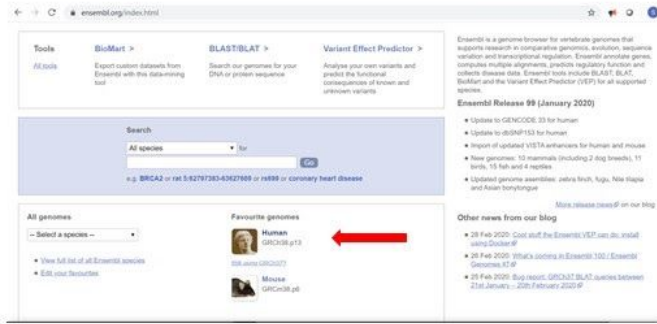
Using <https://www.ncbi.nlm.nih.gov/nucleotide/>, you can find what type of genome has the virus COVID-19

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.

Gene: FOXP3

Gene: FOXP3 ENSG00000049768

Description: forkhead box P3 [Source:HGNC Symbol;Acc:HGNC:6106]

Gene Synonyms: AIID, DIETER, IPEX, JM2, PIDX, SCURFIN, XPID

Location: Chromosome X: 49,250,438-49,270,710 reverse strand

About this gene: This gene has 5 transcripts (splice variants), 109 orthologues, 43 paralogues, is a member of 2 Ensembl protein families and is associated with 23 phenotypes.

Transcripts: [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq Match	Flags
FOXP3-203	ENST00000376207.10	2264	431aa	Protein coding	CCDS14323.9	Q9BZS1.9	NM_014009.4	TSL:1 GENCODE basic APPRIS P1 MANE Select v0.7
FOXP3-202	ENST00000376199.7	1597	396aa	Protein coding	CCDS48109.9	Q9BZS1.9	-	TSL:2 GENCODE basic
FOXP3-206	ENST00000587224.6	1443	456aa	Protein coding	-	Q9BZS1.9	-	TSL:2 GENCODE basic
FOXP3-204	ENST00000455775.7	1434	454aa	Protein coding	-	B7ZLG1.9	-	TSL:5 GENCODE basic
FOXP3-201	ENST00000376197.1	1326	441aa	Protein coding	-	A0A0C4DFW6.9	-	TSL:2 GENCODE basic
FOXP3-205	ENST00000518685.6	1215	404aa	Protein coding	-	Q9BZS1.9	-	TSL:1 GENCODE basic

Chromosome coordinate

Click on Transcript ID
Corresponding to FoxP3-203

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.

Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Human (GRCh38.p13)

Location: X:49,250,438-49,270,477 Gene: FOXP3 Transcript: FOXP3-203

Transcript: FOXP3-203 ENST00000376207.10

Description: forkhead box P3 [Source:HGNC Symbol;Acc:HGNC:6106]

Gene Synonyms: AIID, DIETER, IPEX, JM2, PIDX, SCURFIN, XPID

Location: Chromosome X: 49,250,438-49,264,710 reverse strand

About this transcript: This transcript has 5 exons, is annotated with 27 domains and features, is associated with 2756 variant alleles and maps to 446 oligo probes.

Gene: This transcript is a product of gene ENSG00000049768.16 [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq Match	Flags
FOXP3-203	ENST00000376207.10	2264	431aa	Protein coding	CCDS14323.9	Q9BZS1.9	NM_014009.4	TSL:1 GENCODE basic APPRIS P1
FOXP3-202	ENST00000376199.7	1597	396aa	Protein coding	CCDS48109.9	Q9BZS1.9	-	TSL:2 GENCODE
FOXP3-206	ENST00000587224.6	1443	456aa	Protein coding	-	Q9BZS1.9	-	TSL:2 GENCODE
FOXP3-204	ENST00000455775.7	1434	454aa	Protein coding	-	B7ZLG1.9	-	TSL:5 GENCODE
FOXP3-201	ENST00000376197.1	1326	441aa	Protein coding	-	A0A0C4DFW6.9	-	TSL:2 GENCODE

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.

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence					gcgtggtttttctctcoggtataaaagcaaagttgtttttgatacgtgac
	ENSE00003849914	49,264	49,264	-	-		AGTTTCCCAGCCAGGCTGATCCTTTCTGTGTCAGTCCATGCCA
	Intron 1-2	49,264	49,258				gtgagtgteectgtctceccctacc.....ctectttcccctattgtctacgcag
	ENSE00001316456	49,258	49,258	-	0		CCTGCTTGCACAAAGACCAGATGCCRAACCCAGGCTGCAAGCCCTCAGCCCTTC
	Intron 2-3	49,258	49,257				gtgagccctgggcccaggatgggg.....ctcaccacctgcctttctgcccag
	ENSE00000978820	49,257	49,257	0	0		TGCCCCACTGCCCCGTATGTCATGTGGCACCTCCGGCGCACGGCTGAGCCATTGCC
	Intron 3-4	49,257	49,257				gtatggacggtgaatgggcaggag.....accacagcccacatgtgccccccag
	ENSE00000670078	49,257	49,257	0	1		CTCTAACSTGGATGCACGCGGACCCCTGTGTGCAGGTGCACCCCTGGAAGCC
	Intron 4-5	49,257	49,257				gtaacacctagcccgtaccccatg.....ataggtgccctgtccccaccacag
	ENSE00000867139	49,257	49,256	1	2		GGATCAACTGGCCAGCCATGGAATGGCTGTCBGGAGCCGCACTGCTCGCACCTTCC
	Intron 5-6	49,256	49,256				gtcagtgacaggggtgggaaggat.....ccctgacaccctctgtccccccag

↑ ↑
Number of
intron and exon

Coordinate

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>.

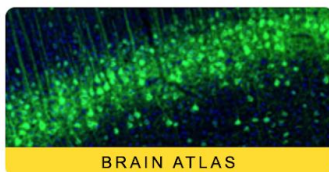
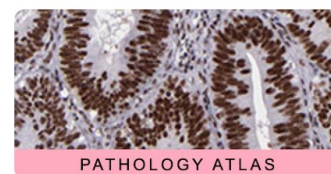
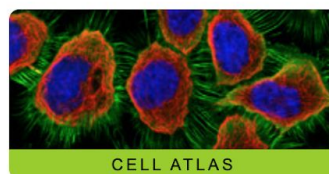
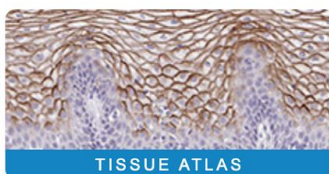
THE HUMAN PROTEIN ATLAS

≡ MENU HELP NEWS

SEARCH¹

e.g. RBM3, insulin, CD36

Search Fields »



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 the gene name

Gene	Gene description	Evidence	Tissue	Cell	Pathology	Brain	Blood
ESR1	Estrogen receptor 1	■				RNA	RNA
HEY1	Hes related family bHLH transcription factor with YRPW motif 1	■				RNA	RNA
DPH3	Diphthamide biosynthesis 3	■				RNA	RNA

Click on gene name, you can see general information

Click on Tissue

ESR1

SUMMARY TISSUE CELL PATHOLOGY BRAIN BLOOD

PROTEIN SUMMARY

RNA DATA

GENE/PROTEIN

ANTIBODIES AND VALIDATION

HUMAN PROTEIN ATLAS SUMMARY

Protein: Estrogen receptor 1

Gene name: ESR1 (ER-alpha, Era, ESR, NR3A1)

Tissue specificity: Tissue enhanced (cervix, uterine, endometrium)

Subcellular location: Vesicles and additionally in Nucleoplasm

Cancer prognostic summary: Prognostic marker in endometrial cancer (favourable)

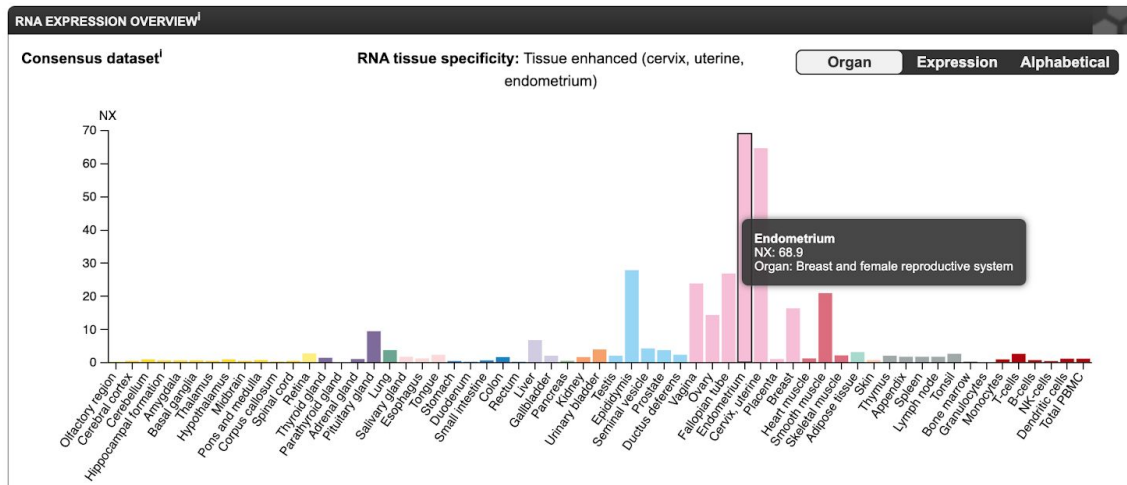
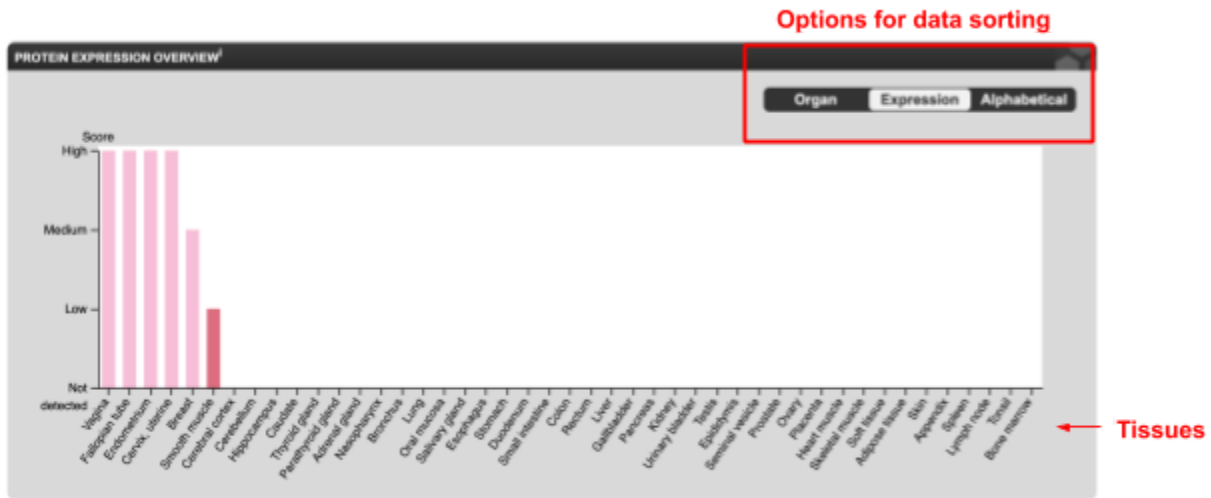
Brain specificity: Not detected in human brain

Blood specificity: Immune cell enhanced (naive CD4 T-cell)

Predicted location: Intracellular

Protein function (UniProt): Nuclear hormone receptor. The steroid hormones and their receptors are involved in the regulation of eukaryotic gene expression and affect cellular proliferation and differentiation in target tissues. Ligand-dependent nuclear transactivation involves either direct homodimer binding to a palindromic estrogen response element (ERE) sequence or association with other DNA-binding transcription factors, such as AP-1/c-Jun, c-Fos, ATF-2, Sp1 and Sp3, to mediate ERE-independent signaling. Ligand binding induces a conformational change allowing subsequent or combinatorial association with multiprotein coactivator complexes through LXXLL motifs of their respective components. Mutual transrepression occurs between the estrogen receptor (ER) and NF-kappaB in a cell-type specific manner. Decreases NF-kappaB DNA...

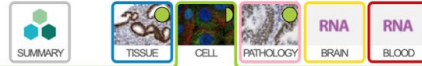
In the Tissue Atlas, you can retrieve information of proteins and RNAs expression in physiological tissues. The expression levels are reported as colored histograms in which each color represents a specific tissue class. The histograms can be sorted based on specific characteristics including the tissue 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.

ESR1

Click here to access to the atlas



CELL ATLAS

- RNA EXPRESSION
- HUMAN CELLS**
- GENE/PROTEIN
- ANTIBODIES AND VALIDATION

Dictionary

- Nucleoplasm
- Vesicles

Human cell

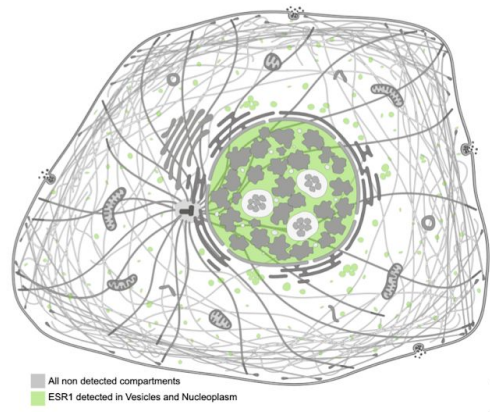
- Nucleoplasm
- Vesicles

GENERAL INFORMATION¹

Gene name¹ ESR1
Gene description¹ Estrogen receptor 1
Protein class¹ Cancer-related genes
Disease related genes
FDA approved drug targets
Nuclear receptors
Transcription factors
Predicted location¹ Intracellular
Number of transcripts¹ 11

HUMAN PROTEIN ATLAS INFORMATION¹

Summary¹ Mainly localized to vesicles. In addition localized to the nucleoplasm.
RNA cell specificity¹ Group enriched (MCF7, T-47d)
RNA cell distribution¹ Detected in some
Protein evidence¹ Evidence at protein level
Localized to the Vesicles (approved)
Main location¹ [View proteome in REACTOME](#)
Additional location¹ In addition localized to the Nucleoplasm (approved)
[View proteome in REACTOME](#)



Protein localization

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.

CELL ATLAS

- RNA EXPRESSION
- HUMAN CELLS**
- GENE/PROTEIN
- ANTIBODIES AND VALIDATION

Dictionary

- Nucleoplasm
- Vesicles

Human cell


- Nucleoplasm
- Vesicles

HPA000449: MCF7 HPA000449: RT4 HPA000449: U-2 OS

Compare selected

Toggle channels¹

- Antibody
- Nucleus
- Intensity
- Microtubules
- ER
- Objects

Thumbnail ¹	Antibody ¹	Cell line ¹	Cell line RNA Expression (NX) ¹	Location ¹	Single-cell variation ¹	Cell cycle dependent variation ¹
	HPA000449	MCF7	24.3	Nucleoplasm Vesicles		

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.

[Click here to access to the atlas](#)

ESR1

SUMMARY

TISSUE

CELL

PATHOLOGY

RNA
BRAN

RNA
BLOOD

PATHOLOGY ATLAS

GENE/PROTEIN

ANTIBODIES AND VALIDATION

Dictionary

GENERAL INFORMATION¹

Gene name¹ ESR1

Gene description Estrogen receptor 1

Protein class¹ Cancer-related genes
Disease related genes
FDA approved drug targets
Nuclear receptors
Transcription factors

Predicted location¹ Intracellular

Number of transcripts¹ 11

HUMAN PROTEIN ATLAS INFORMATION¹

RNA category¹ Consensus (human tissue): Tissue enhanced (cervix, uterine, endometrium)
Detected in many
Group enriched (MCF7, T-47d)

HPA (cell line): Detected in some
Group enriched (breast cancer, endometrial cancer, ovarian cancer)

TCGA (cancer tissue): Detected in many

Protein evidence¹ Evidence at protein level

Protein expression normal tissue¹ Selective nuclear expression in female genitalia.

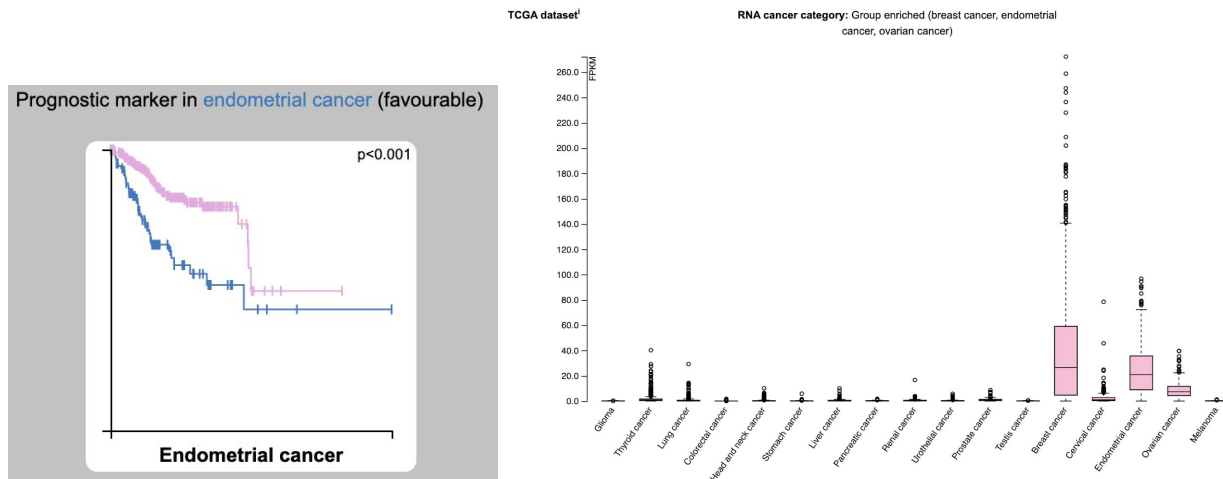
IMMUNOHISTOCHEMISTRY DATA RELIABILITY

Data reliability description¹ Antibody staining consistent with RNA expression data.

Reliability score - normal tissues¹ ● Enhanced

Antibodies¹ HPA000449, HPA000450, CAB000037, CAB055099, CAB072858

[SHOW MORE](#)



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.

Expression Atlas
Gene expression across species and biological conditions

Query single cell expression
[To Single Cell Expression Atlas](#)

Home Browse experiments Download Release notes FAQ Help Licence About Support

Search across 63 species, 3,744 studies, 122,669 assays Ensembl 99, Ensembl Genomes 46, WormBase ParaSite 14, EFO 3.10.0

Search Gene set enrichment **Insert here the gene name**

Gene / Gene properties Species Biological conditions

Enter gene query... Any Enter condition query...

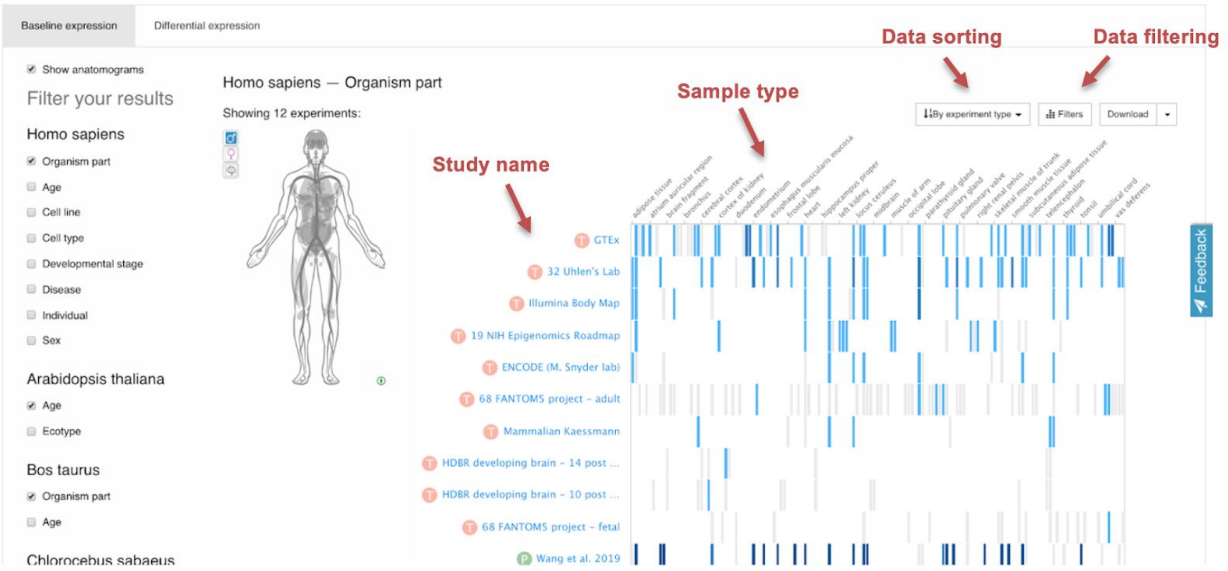
Examples: REG1B, zinc finger, O14777, (UniProt), GO:0010468 (regulation of gene expression) Examples: lung, leaf, valproic acid, cancer

Search Clear

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).

Results for ESR1



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.

Filter your results

- Kingdom
- Animals
 - Plants
- Species
- Mus musculus
 - Homo sapiens
 - Sus scrofa
 - Arabidopsis thaliana
 - Rattus norvegicus
 - Danio rerio
 - Gallus gallus
- Experiment type
- Microarray 1-colour mRNA differential
 - RNA-seq mRNA differential

Expression change as log₂FC

Hide log₂-fold change
Download results

Log ₂ -fold change	Species	Gene name	Comparison	Experimental variables	Experiment name
-10		ESR1	'estrogen receptor alpha shRNA' vs 'scrambled shRNA'	RNA interference	RNA-seq of the human breast cancer ER1-suppressed MCF-7(MCF-7/SP10+) cells and of their internal control MCF-7 (MCF-7/C) cells
-9.6		ESR1	'Snail overexpression' vs 'control'	treatment	Expression data from breast cancer cell line MCF-7 with ectopic expression of the transcription factor Snail
-8		ESR1	'estrogen receptor alpha knockdown' vs 'control'	phenotype	Expression data from MCF7 cell line after silencing of Estrogen receptor
6.8		Esr1	'beta cell specific Pax6 knockout' vs 'wild type'	phenotype	RNA-Seq of pancreatic islets from beta cell-specific Pax6 knockout mice
-6.1		ESR1	'erythroleukemia; ZRSR2 shRNA' vs 'normal'	disease, genotype	Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome

Compared conditions

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.

Marker genes¹

Experiments with marker genes

Species

- Arabidopsis thaliana
- Callithrix jacchus
- Danio rerio
- Gallus gallus
- Homo sapiens
- Mus musculus
- Rattus norvegicus

Inferred cell type¹

Select...

Organism part¹

Select...

ESR1 is expressed in:

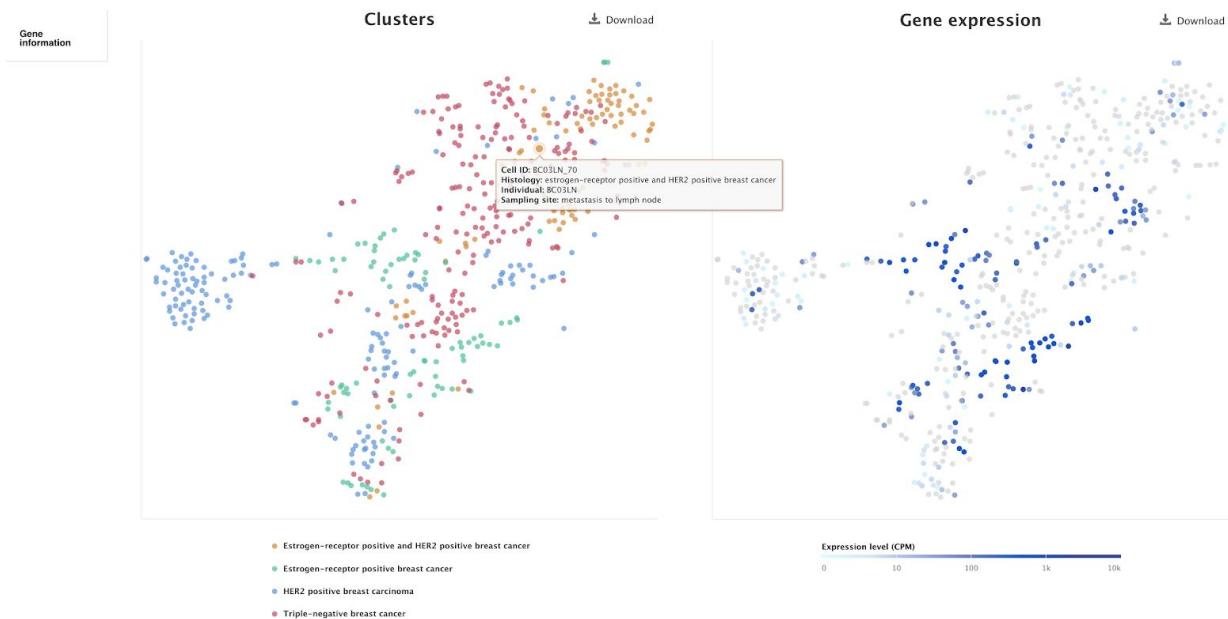
Species	Marker genes	Title	Experimental variables	Number of assays
Mus musculus	• See cluster 1 for k = 3	Deciphering the relationship between polycomb repression and stochastic gene expression from single-cell RNA-seq data	phenotype cell line single cell identifier	288
Homo sapiens	• See cluster 21 for k = 29	Single cell RNA-seq of primary breast cancer cells and lymph node metastases from 11 patients representing the four subtypes of breast cancer: luminal A, luminal B, HER2 and triple negative breast cancer	single cell identifier histology sampling site	540
Rattus norvegicus	✘	Single cell RNA-seq of female rat ventral mesenchymal pad and adjacent urethra	single cell identifier organism part	115
Mus musculus	✘	Origin and differentiation trajectories of fibroblastic reticular cells in the splenic white pulp - single cell RNAseq dataset 1	genotype	2,993

Feedback

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.)

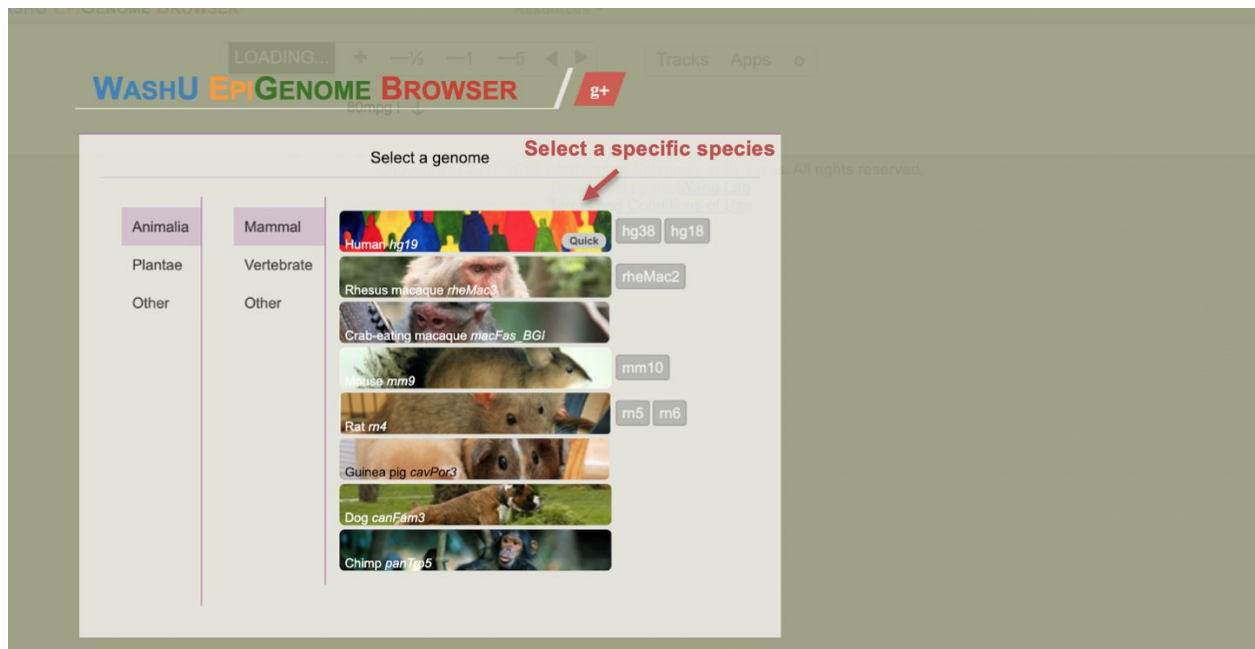


The WashU Epigenome Browser

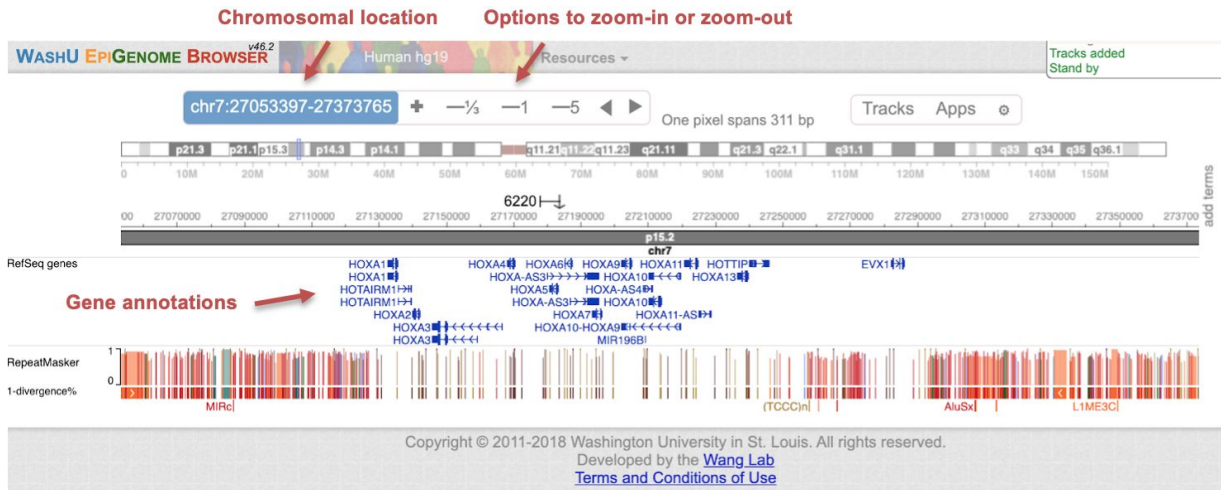
The WashU Epigenome Browser is a web tool which allows the visualization of the results from genomic and transcriptomic experiments from international projects and single studies. The tool can be accessed at <http://epigenomegateway.wustl.edu/legacy/>.

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

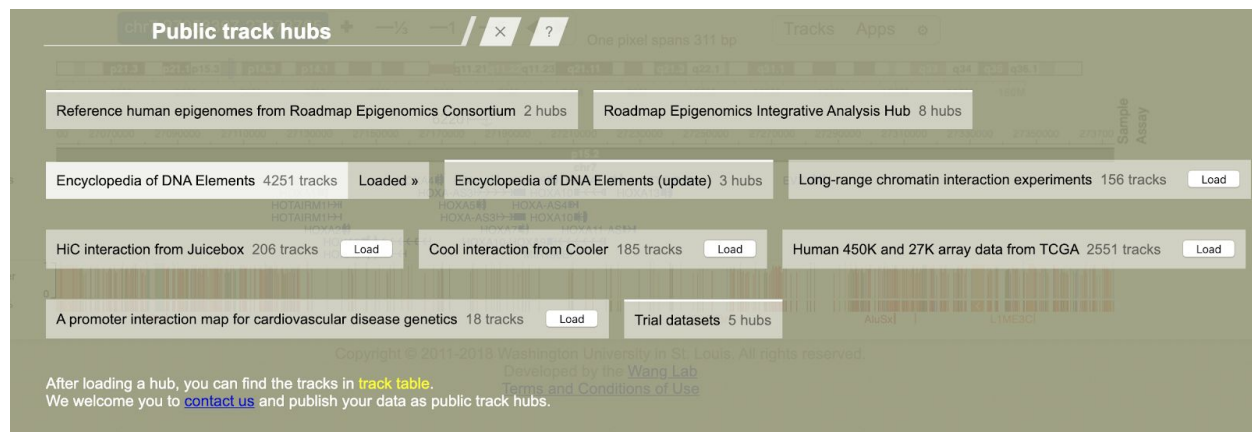
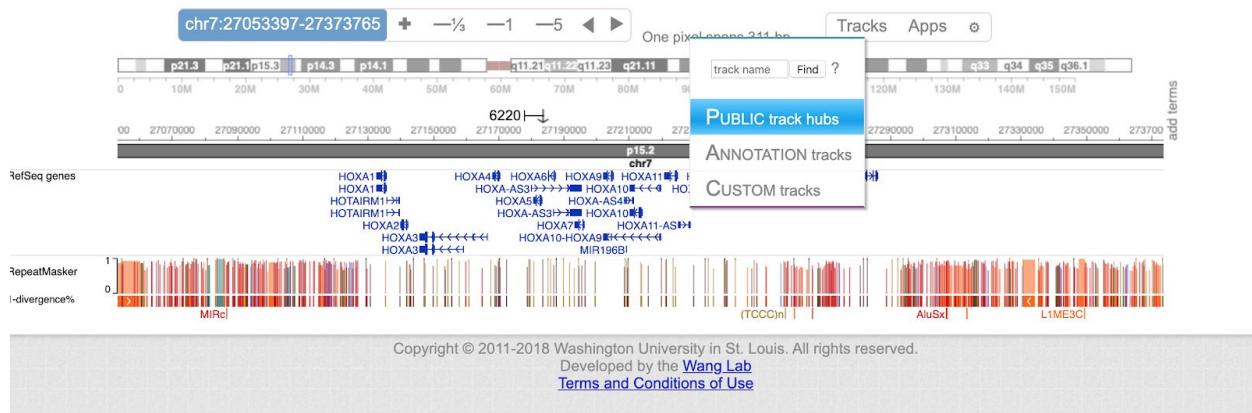
From the homepage is it possible to select a specific species of interest and the following results will be the representation of a genomic region whose coordinates will be reported on the top. In the genome browser, each row represents specific information, including genomic positions, gene annotations, annotations of repetitive elements, and coverage signals from sequencing experiments, particularly ChIP-Seq and RNA-Seq.

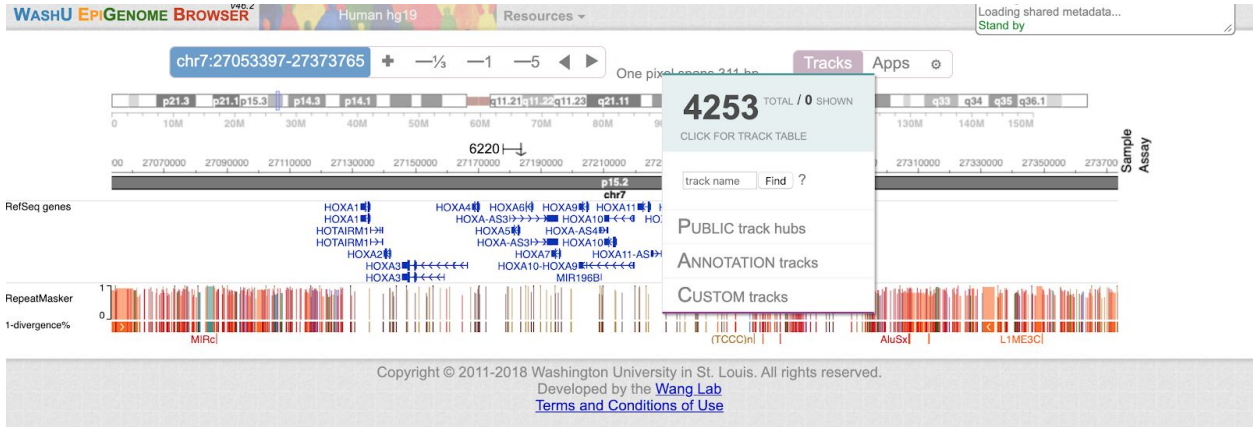


The screenshot displays the WashU Epigenome Browser interface. At the top, there is a navigation bar with a "LOADING" indicator, zoom controls, and a "Tracks Apps" menu. The main header reads "WASHU EPIGENOME BROWSER" with a "g+" logo. Below the header, there are two tabs: "Select a genome" and "Select a specific species". A red arrow points to the "Select a specific species" tab. The interface is divided into a left sidebar with taxonomic categories (Animalia, Mammal, Plantae, Vertebrate, Other) and a main content area. The main content area shows a list of species with their respective genome versions and a "Quick" button. The species listed are: Human hg19, Rhesus macaque rheMac2, Crab-eating macaque macFas_BGI, Mouse mm9, Rat m4, Guinea pig cavPor3, Dog canFam3, and Chimp panTro5. To the right of the species list, there are buttons for hg38, hg18, rheMac2, mm10, m5, and m6. The text "All rights reserved" is visible in the bottom right corner.

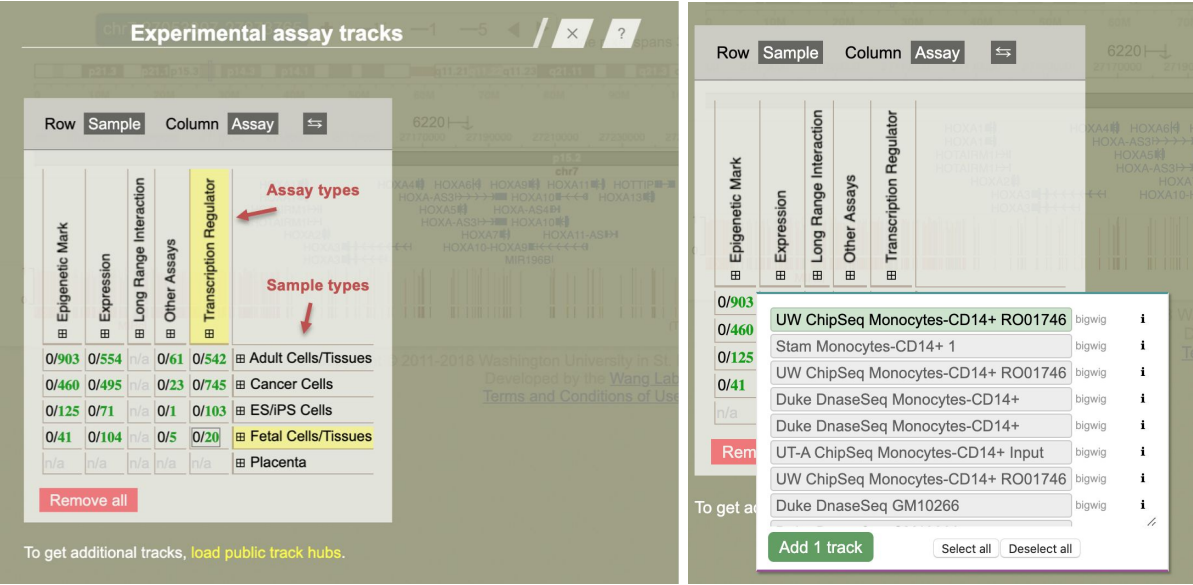


It is possible to access the information from public experiments using the “**Public Track Hubs**” function in the “**Tracks**” section. Then, data from different projects can be selected by clicking on the “**Load**” button. Finally, it is possible to display the data of specific experiments from the 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.

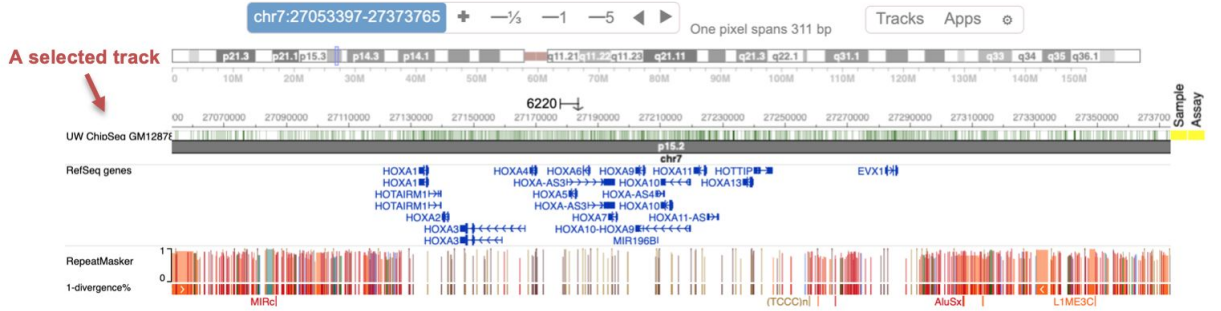




To add the experiment of a specific sample click on its name and press the green button “**Add 1 track**”. Multiple tracks can be selected and added together.



Each selected track will be initially shown in a compact mode but by **right-clicking** on the track and clicking on **Configure**, it is possible to visualize the peak coverage signals by clicking two times on the button “+” in the “**Height**” section.



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/>.

The screenshot shows the CBioPortal interface. At the top, there are navigation links like 'Data Sets', 'Web API', 'R/MATLAB', 'Tutorials', 'FAQ', 'News', 'Visualize Your Data', and 'About'. A 'Login' button is in the top right. Below the navigation is a search bar and a 'Quick Search Beta' button. The main section is titled 'Select Studies for Visualization & Analysis:'. It features a list of study categories on the left (e.g., PanCancer Studies, Cell lines, Adrenal Gland) and a central area for selecting specific studies. A red arrow points to the 'Adrenocortical Carcinoma' study, with a red text label 'Select a specific study' next to it. Below the study selection, there are buttons for 'Query By Gene' and 'Explore Selected Studies'.

Select between analysing a list of gene or explore the dataset information

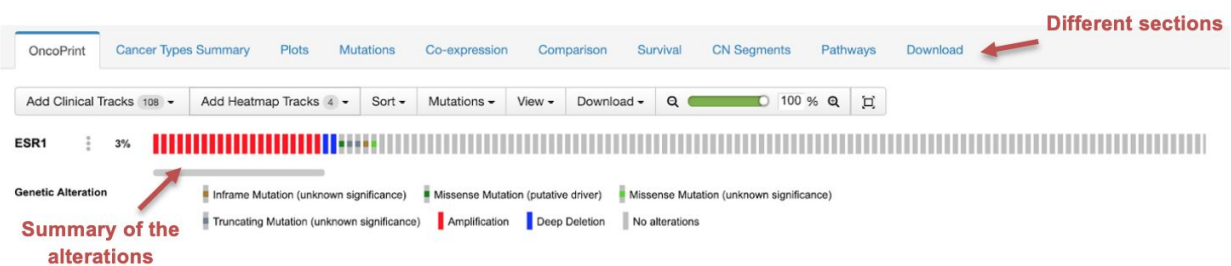
From the homepage is it possible to select a specific study and the main features of this study can be obtained using the button “Explore Selected Studies”.

The screenshot shows the CBioPortal page for 'Breast Invasive Carcinoma (TCGA, Cell 2015)'. The page is divided into several sections:

- Cancer Type Detailed:** A table showing the distribution of cancer types. A red arrow points to 'Breast Invasive Ductal Carcinoma'.
- Molecular Profile:** A table showing various molecular profiles. A red arrow points to 'mRNA expression (RNA Seq V2 R...)'.
- Overall Survival:** A Kaplan-Meier survival plot showing survival percentage over time.
- Disease Free Survival:** A Kaplan-Meier survival plot showing disease-free survival percentage over time.
- Number of Samples Per Patient:** A pie chart showing the distribution of samples per patient.
- Overall Survival Status:** A pie chart showing the distribution of overall survival status.
- Main alterations:** A table showing the most frequent gene alterations. The top entries include PIK3CA, TP53, CDH1, GATA3, MAP3K1, KMT2C, PTEN, NCOA3, and NCOR1.

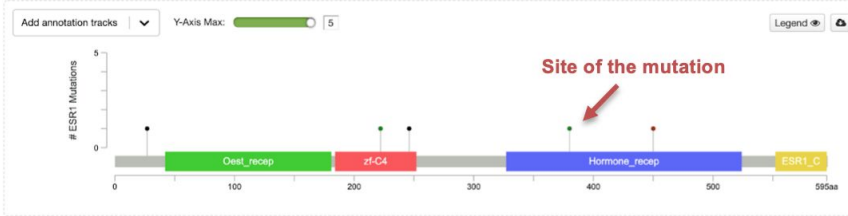
Furthermore, it is possible to analyse the genomic features of single or multiple genes as measured in specific studies by using the button “**Query by gene**”. Insert the gene names in the main form and click on **Submit Query**.

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.

ESR1



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

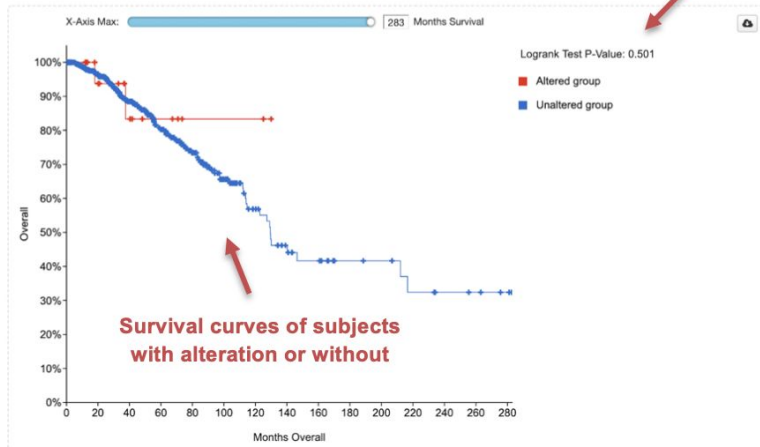
Groups: (drag to reorder) Altered group (28) Unaltered group (788) ESR1 (28) Select all | Deselect all

Overlap Survival Clinical Mutations Copy-number mRNA Protein

Survival types: Overall Disease Free

Overall

Overall patient survival status.



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/>.

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.

Search a specific term here



Keyword or GEO Accession Search

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

Browse Content

Repository Browser
DataSets: 4348
Series: 128101
Platforms: 20780
Samples: 3540511

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.

The screenshot shows the search results page for the query "ESR1". The page is annotated with red arrows and text:

- Filters:** An arrow points to the left sidebar containing various filter categories like "Entry type", "Series", "Organism", "Study type", "Author", "Attribute name", and "Publication dates".
- Search results:** The main content area shows a list of results. An arrow points to the first result, "SFRP1-regulated gene expression in premalignant breast lesions", with the text "Link to the dataset and description".
- Organisms, Assay type, Platform, and Number of samples:** An arrow points to the details of the first result, specifically the "Expression profiling by array" section, which lists "Organism: Homo sapiens", "Platform: GPL6244", and "42 Samples".
- Right sidebar:** The right sidebar contains sections for "Top Organisms", "Find related data", "Search details", and "Recent activity".

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)

- [GSM3330068](#) Breast_Atypia_S-10-3
- [GSM3330069](#) Breast_Benign_S-10-3
- [GSM3330070](#) Breast_Atypia_S-10-10

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 ?

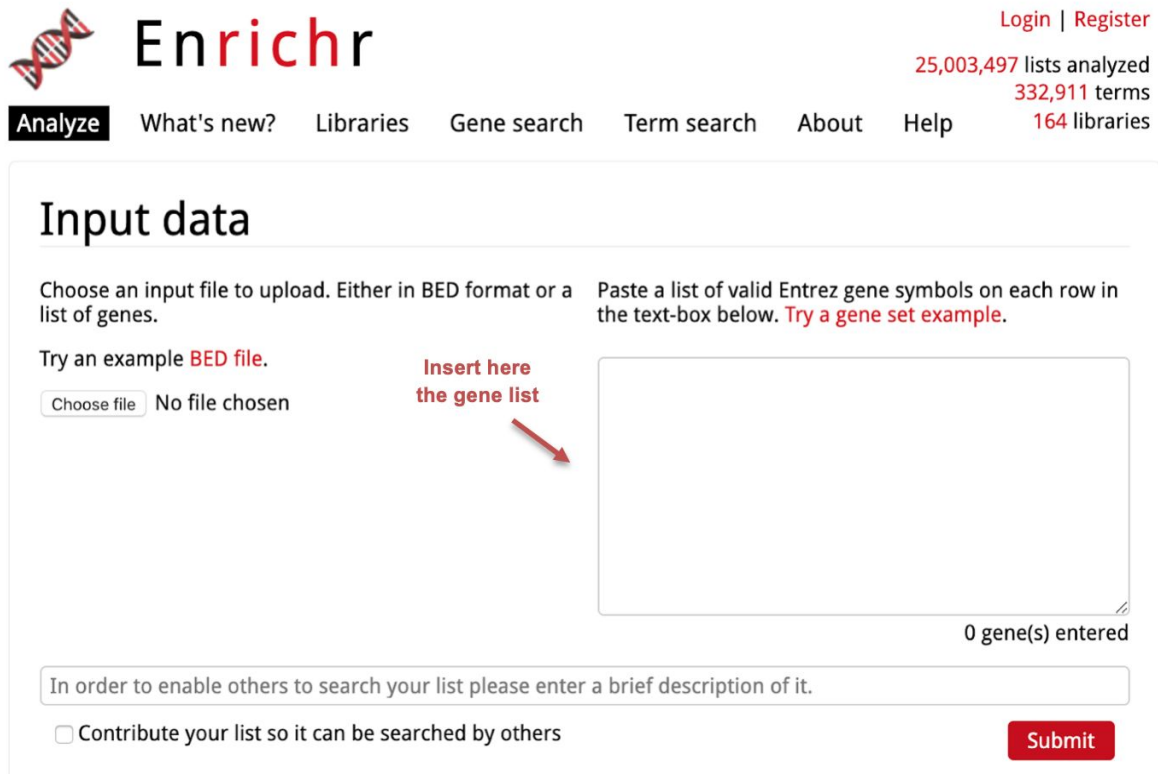
Additional data

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

Information on specific samples

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.



The screenshot shows the Enrichr website's input data form. At the top left is the Enrichr logo, a DNA double helix. To the right of the logo is the text "Enrichr". Further right are links for "Login | Register" and statistics: "25,003,497 lists analyzed", "332,911 terms", and "164 libraries". Below the logo is a navigation bar with the following items: "Analyze" (highlighted in a black box), "What's new?", "Libraries", "Gene search", "Term search", "About", and "Help". The main content area is titled "Input data" and contains two columns of instructions. The left column says "Choose an input file to upload. Either in BED format or a list of genes." and includes a "Choose file" button and the text "No file chosen". The right column says "Paste a list of valid Entrez gene symbols on each row in the text-box below. Try a gene set example." and features a large text input box. A red arrow points to the text box with the label "Insert here the gene list". Below the text box is the text "0 gene(s) entered". At the bottom of the form is a text input field with the placeholder "In order to enable others to search your list please enter a brief description of it." and a checkbox labeled "Contribute your list so it can be searched by others". A red "Submit" button is located at the bottom right of the form.

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.

Description Sample gene list (375 genes)  

GO Biological Process 2018

- alpha-amino acid metabolic process (GO:1901605)
- glucan biosynthetic process (GO:0009250)
- glycogen biosynthetic process (GO:0005978)
- mitochondrion organization (GO:0007005)
- organic cyclic compound biosynthetic process (GO:1901362)

GO Molecular Function 2018

- retinal dehydrogenase activity (GO:0001758)
- tRNA (cytosine) methyltransferase activity (GO:0003537)
- oxidoreductase activity, acting on the aldehyde or oxo group of donors (GO:0004010)
- fatty acid synthase activity (GO:0004312)
- 4 iron, 4 sulfur cluster binding (GO:0051539)

GO Cellular Component 2018

- mitochondrion (GO:0005739)
- mitochondrial matrix (GO:0005759)
- mitochondrial inner membrane (GO:000574)
- nuclear exosome (RNase complex) (GO:0005777)
- peroxisome (GO:0005777)

Enriched gene sets

MGI Mammalian Phenotype Level 4 2019

- MP:0010400 increased liver glycogen level
- MP:0010956 abnormal mitochondrial ATP synthesis
- MP:0008918 microgliosis
- MP:0013741 absent thymus corticomedullary ratio
- MP:000607 abnormal hepatocyte morphology

Human Phenotype Ontology



- Abnormality of methionine metabolism (HP:0000001)
- Abnormality of aspartate family amino acid metabolism (HP:0000002)
- Gaze-evoked nystagmus (HP:0000640)
- Tubulointerstitial abnormality (HP:0001969)
- Abnormality of sulfur amino acid metabolism (HP:0000003)

Jensen TISSUES

- Mesocarp
- Imaginal disc
- Seedling
- Mesophyll
- Myeloid leukemia cell

Description Sample gene list (375 genes)  

GO Biological Process 2018

Bar Graph **Table** Clustergram  

Hover each row to see the overlapping genes. **Enrichment significance**

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](#) [Click here to download the results](#) Previous Next

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