

Βιοπληροφορική

Γρηγόριος Αμούτζιας
Καθηγητής Βιοπληροφορικής με έμφαση στη Μικροβιολογία
Τμήμα Βιοχημείας και Βιοτεχνολογίας
Πανεπιστήμιο Θεσσαλίας

Προτεινόμενα συγγράμματα

- Ελληνικά συγγράμματα:
 - Jonathan Pevsner. Βιοπληροφορική και λειτουργική γονιδιωματική
 - Andreas D. Baxevanis & B.F. Francis Quelletto. Βιοπληροφορική: Ένας πρακτικός οδηγός για την ανάλυση γονιδίων και πρωτεϊνών.
 - Σοφία Κοσσίδα. Βιοπληροφορική - Δυνατότητες & Προοπτικές.
- Αγγλικά συγγράμματα:
 - Jin Xiong. Essential Bioinformatics. (Σύντομο, περιεκτικό και απλά γραμμένο σύγγραμμα).
 - David W. Mount. Bioinformatics. Sequence and genome analysis. (Εκτενές και πολύ αναλυτικό σύγγραμμα)
 - Translational bioinformatics. Open access book.
 - <http://www.ploscollections.org/article/browseIssue.action?issue=info:doi/10.1371/issue.pcol.v03.i11>

Βιοπληροφορική

Εισαγωγή

Βιοπληροφορική: τι είναι

- Η ανάπτυξη και χρήση τεχνικών και εργαλείων πληροφορικής/μαθηματικών/στατιστικής για την ανάλυση βιολογικών δεδομένων (κυρίως μοριακής βιολογίας)
- Σήμερα γίνεται διάκριση μεταξύ της βιοπληροφορικής και της υπολογιστικής βιολογίας
 - Βιοπληροφορική: Η ανάπτυξη μεθόδων και προγραμμάτων.
 - Υπολογιστική Βιολογία: Η χρήση των παραπάνω μεθόδων και προγραμμάτων για την ανάλυση βιολογικών δεδομένων.
- Συχνά συμβαίνουν και τα δύο ταυτόχρονα και τα σύνορα δεν είναι πάντα ευδιάκριτα
- Πολλές και συμπληρωματικές μεταξύ τους ειδικότητες (από Βιολογία, Βιοχημεία, Χημεία, Χημική Μηχανική, Μηχανική, Υπολογιστές, Μαθηματικά, Στατιστική κ.α.) συνεργάζονται σήμερα στο χώρο της Βιοπληροφορικής

Βιοπληροφορική: βασικοί τομείς

- Βάσεις δεδομένων (Databases)
 - Οργάνωση, αποθήκευση, αναζήτηση των δεδομένων.
- Ανάλυση ακολουθιών DNA, RNA, πρωτεϊνών. (Sequence analysis)
 - Στοιχίση ακολουθιών: Σύγκριση των αντίστοιχων/ομόλογων περιοχών, μεταξύ δύο ή περισσότερων ακολουθιών.
 - Φυλογενετική ανάλυση: Οι εξελικτικές σχέσεις μεταξύ ομοειδών αντικειμένων (γονίδια, πρωτεΐνες, οργανισμοί).
- Γονιδιακή ρύθμιση/έκφραση (Gene expression)
Ανάλυση δεδομένων από μικροσυστοιχίες, RNA-seq.
- Δομή RNA/πρωτεϊνών (structural biology):
Πρόβλεψη δευτεροταγούς και τριτοταγούς δομής. Ανάλυση πρωτεϊνικών επιφανειών που αλληλεπιδρούν μεταξύ τους.
- Εξόρυξη δεδομένων από βιβλιογραφία (text mining).
- Βιολογικά δίκτυα/μονοπάτια, Βιολογία Συστημάτων (FBA, MCA).
- Οντολογίες (Ontologies)
Η χρήση ενός ελεγχόμενου λεξιλογίου (με ιεραρχική δόμηση), για την περιγραφή των ιδιοτήτων και των λειτουργιών ομοειδών αντικειμένων (π.χ πρωτεϊνών).

Ιστορική αναδρομή

- 1965: Η πρώτη έκδοση του Atlas of protein sequence and structure (Margaret Dayhoff), πρόδρομος της βάσης δεδομένων πρωτεϊνικών ακολουθιών PIR (protein information resource).
 - Ακολουθούν και άλλες βάσεις δεδομένων. 1986: Swissprot, Geneva
- 1970: Αλγόριθμος Needleman-Wunsch για την σύγκριση ακολουθιών
- 1990: Blast
- 1990s: Αρχή του Human genome project, που 'ολοκληρώθηκε' το 2001. Κινητήριος δύναμη για την αλματώδη ανάπτυξη της Βιοπληροφορικής.

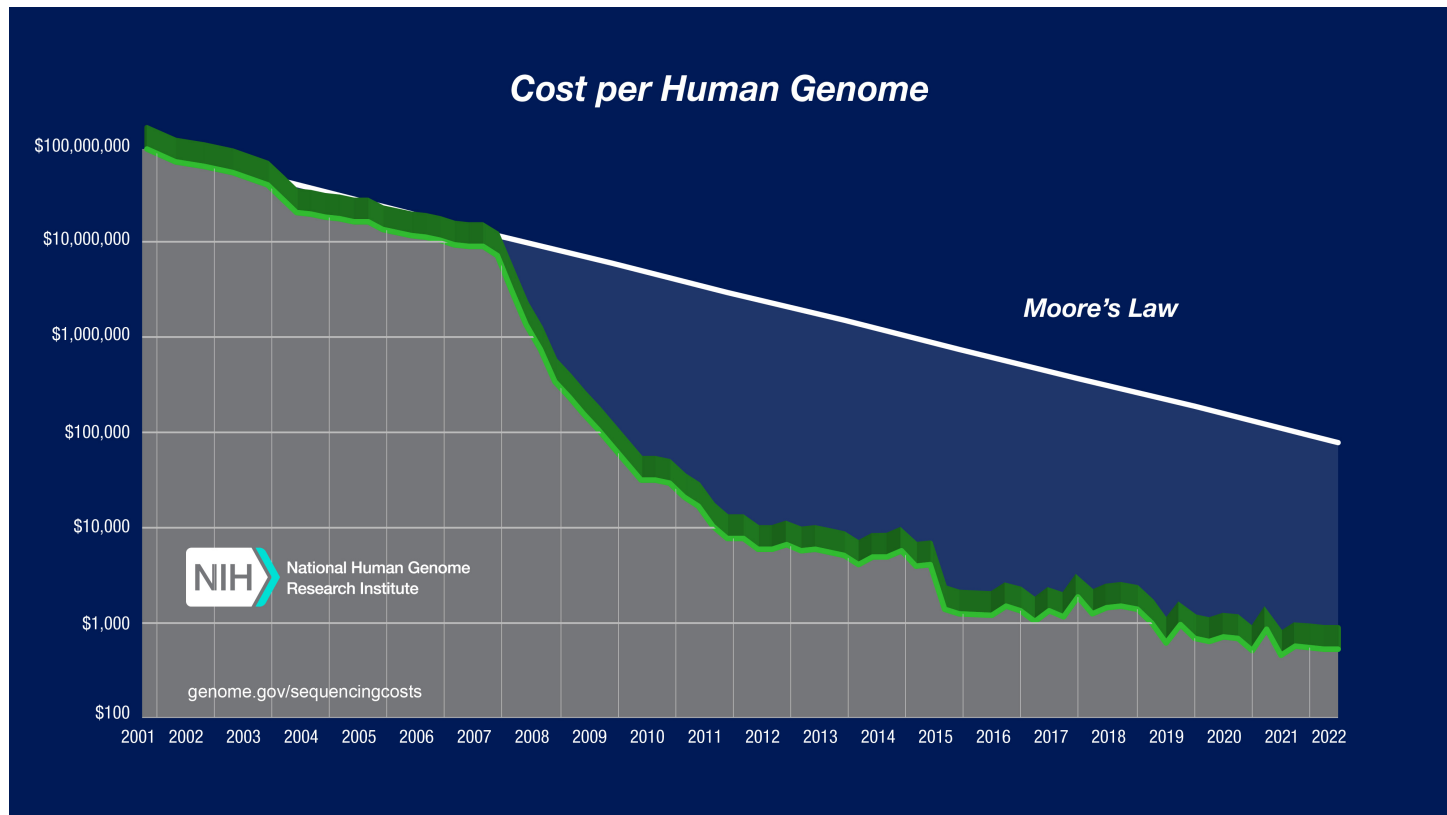


Παρόν/μέλλον

- Μέχρι το 2000, Βιοπληροφορική σήμαινε κυρίως ανάλυση ακολουθιών.
- Η γενωμική αποτέλεσε το ερέθισμα για την ανάπτυξη τεχνολογιών που κάνουν μετρήσεις ευρείας κλίμακας.
- Από το 2000 και μετά, η Βιοπληροφορική καλείται επίσης να διαχειριστεί και να αναλύσει μεγάλα και πολύπλοκα δεδομένα από το χώρο της γενωμικής, της γονιδιακής έκφρασης, της πρωτεομικής κ.α.
- Πλέον ο όρος 'Βιοπληροφορική' είναι τόσο εξειδικευμένος/γενικός, όσο και ο όρος 'Μοριακή Βιολογία'!
- Βρισκόμαστε σε μια μεταβατική περίοδο για τις Βιολογικές επιστήμες, όπως η Φυσική πριν πολλά χρόνια. Βέβαιη η εισδοχή περισσότερων μαθηματικών, στατιστικής και πληροφορικής (προγραμματισμός) μεσοπρόθεσμα στο πρόγραμμα σπουδών.

Χαμηλό κόστος γενωμικών τεχνολογιών θα οδηγήσει σε καθημερινές εφαρμογές

- Κόστος αλληλούχισης
 - <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>
 - Ο νόμος του Moore προβλέπει διπλασιασμό της υπολογιστικής ισχύς κάθε δύο χρόνια.



Βάσεις Δεδομένων

Βάσεις Δεδομένων: Εισαγωγή

Χρησιμοποιούνται για:

- Οργάνωση
- Αποθήκευση
- Επεξεργασία
- Αναζήτηση/επαναπόκτηση της βιολογικής πληροφορίας

Κύρια είδη:

Επίπεδης οργάνωσης (Flat-files:) Το ποιο απλό είδος. Ουσιαστικά είναι κατάλογοι

Σχεσιακές βάσεις. Πιο περίπλοκες και πλέον πολύ διαδεδομένες . Π.χ., SQL. Η πληροφορία οργανώνεται σε πίνακες που σχετίζονται μεταξύ τους. Έτσι αποφεύγεται η επανάληψη και συσσώρευση δεδομένων

Αντικειμενοστρεφείς βάσεις κ.α.

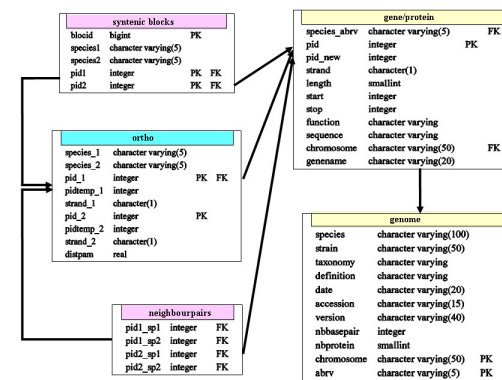
Διακρίνονται κυρίως σε αρχαιακές/πρωτεύοντες και δευτερεύοντες

Στις αρχαιακές γίνεται κατάθεση δεδομένων ενώ στις δευτερεύοντες τα δεδομένα είναι περαιτέρω επεξεργασμένα/σχολιασμένα/αλληλοσυνδεδεμένα

```

LOCUS       name of locus, length and type of sequence,
            classification of organism, data of entry
DEFINITION  description of entry
ACCESSION   accession numbers of original source
KEYWORDS    key words for cross referencing this entry
SOURCE      source organism of DNA
ORGANISM    description of organism
REFERENCE   biological function or database information
COMMENT     information about sequence by base position or range of positions
FEATURES    information about sequence by base position or range of positions
            source      range of sequence, source organism
            misc_signal range of sequence, type of function or signal
            mRNA        range of sequence, mRNA
            CDS         range of sequence, protein coding region
            intron      range of sequence, position of intron
            mutation    sequence position, change in sequence for mutation
BASE COUNT  count of A, C, G, T and other symbols
ORIGIN      text indicating start of sequence
            1 gaattcgata aatctctggt ttattgtgca gttttggtt ccaaaatcgc
            51 atatactcac agcataactg tatatacacc cagggggcgg aatgaaagcg
            //
            database symbol for end of sequence
    
```

Figure 2.5. GenBank DNA sequence entry.



Ετήσιος κατάλογος ΒΔ

- Κάθε Ιανουάριο στο Nucleic Acids Research (Special database issue)
- 2024: 1959 databases
- <http://www.oxfordjournals.org/nar/database/c/>



You are here: [NAR Journal Home](#) » Database Summary Paper Categories

NAR Database Summary Paper Category List

[Nucleotide Sequence Databases](#)
[RNA sequence databases](#)
[Protein sequence databases](#)
[Structure Databases](#)
[Genomics Databases \(non-vertebrate\)](#)
[Metabolic and Signaling Pathways](#)
[Human and other Vertebrate Genomes](#)
[Human Genes and Diseases](#)
[Microarray Data and other Gene Expression Databases](#)
[Proteomics Resources](#)
[Other Molecular Biology Databases](#)
[Organelle databases](#)
[Plant databases](#)
[Immunological databases](#)
[Cell biology](#)

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

Bionumbers

<https://bionumbers.hms.harvard.edu/>



[Home](#) [Index](#) [Book](#) [About Us](#) [Misc.](#) ▾

e.g. "ribosome coli"; "p53 human"; "transcription"



Popular BioNumbers

Recent BioNumbers

Key BioNumbers

Results

Click a row for more details

Property	Organism	Value	Units	ID	Details
Number of hairs on human head	Human Homo sapiens	90,000-150,000	unitless	101509	Amazing numbers in biology...
Average duration of a single eye blink	Human Homo sapiens	0.1-0.4	sec	100706	http://faculty.washi...
One unit of OD600 corresponds to a cell wet weight and a cell dry weight of	Bacteria Escherichia coli	Cell wet weight 1.7g/L: cell dry weight 0.39g/L		109836	Glazyrina J, Materne EM...
Length of neuron from base of spine to big toe-longest cell in human body	Human Homo sapiens	≤1	meter	104901	Fletcher DA, Theriot...
Size of glucose molecule (open chain form)	Generic	1.5	nm	106979	Daniel Minoli, Nanotechnology...

Βάσεις νουκλεοτιδικών δεδομένων (I)

Αρχειακές ΒΔ για νουκλεοτιδικές αλληλουχίες:

- EMBL-BANK. European Nucleotide Archive (ENA), EBI. Hinxton, UK.
- GENBANK. NCBI, NIH. Bethesda, USA
- DNA databank of Japan (DDBJ). National institute of Genetics, .Mishima, JP

Η ακολουθία κατατίθεται σε μία από τις ΒΔ, η οποία έχει και τη δυνατότητα να την αναθεωρήσει (μόνο αυτή, για αποτροπή 'συγκρούσεων')

Και οι 3 ΒΔ ανήκουν στο International nucleotide sequence database collection (INSDC). Κάθε μέρα ανταλλάσσουν δεδομένα. Η ίδια ακολουθία Χ3.



Βάσεις νουκλεοτιδικών δεδομένων. EMBL format

```

FT mRNA 1..1859
FT /experiment="experimental evidence, no additional details
FT recorded"
XX
SQ Sequence 1859 BP; 609 A; 314 C; 355 G; 581 T; 0 other;
aaacaaacca aatatggatt ttattgtagc catatttgct ctgtttgtta ttagctcatt 60
cacaattact tccacaaatg cagttgaagc ttctactctt cttgacatag gtaacctgag 120
tcggagcagt tttcctcgtg gtttcactct tgggtgctgga tcttcagcat accaatttga 180
aggtgcagta aacgaaggcg gtagaggacc aagtatttgg gatacettca cccataaata 240
tccagaaaaa ataagggatg gaagcaatgc agacatcacg gttgaccaat atcaccgcta 300
caaggaagat gttgggatta tgaaggatca aaatatggat tcgtatagat tctcaatctc 360
ttggccaaga atactcccaa agggaaagtt gagcggaggc ataaatcacg aaggaatcaa 420
atattacaac aaccttatca acgaactatt ggctaacggt atacaaccat ttgtaactct 480
ttttcattgg gatcttcccc aagtcttaga agatgagtat ggtggtttct taaactccgg 540
tgtaataaat gattttcgag actatacggg tctttgcttc aaggaatttg gagatagagt 600
gaggtattgg agtactctaa atgagccatg ggtgtttagc aattctggat atgcactagg 660
aacaaatgca ccaggtcgat gttcggcctc caacgtggcc aagcctggtg attctggaac 720
aggaccttat atagttacac acaatcaaat tcttgetcat gcagaagctg tacatgtgta 780
taagactaaa taccaggcat atcaaaaggg aaagataggc ataacgttgg tatctaactg 840
gttaatgcca cttgatgata atagcatacc agatataaag gctgccgaga gatcacttga 900
cttccaattt ggattgttta tgaacaatt aacaacagga gattattcta agagcatgcg 960
gcgtatagtt aaaaaccgat tacctaagtt ctcaaaattc gaatcaagcc tagtgaatgg 1020
ttcatttgat tttattggta taaactatta ctcttctagt tatattagca atgcccttc 1080
acatggcaat gccaaaacca gttactcaac aaactctatg accaatattt catttgaaaa 1140
acatgggata cccttaggtc caagggctgc ttcaatttgg atatatgttt atccatatat 1200
gcttatccaa gaggacttcg agatcttttg ttacatatta aaaataaata taacaatcct 1260
gcaattttca atcactgaaa atggtatgaa tgaattcaac gatgcaacac ttccagtaga 1320
agaagctctt ttgaatactt acagaattga ttactattac cgtcacttat actacattcg 1380
ttctgcaatc agggctggct caaatgtgaa gggtttttac gcatggtcac ttttgactg 1440
taatgaatgg tttgcaggct ttactgttcg ttttgatta aactttgtag attagaaga 1500
tggattaaaa aggtacccta agctttctgc ccaatggtag aagaactttc tcaaaagaaa 1560
ctagctagta ttattaaaag aactttgtag tagattacag tacatcgttt gaagttgagt 1620
tgggtgcacct aattaaataa aagaggttac tcttaacata tttttaggcc attcgtttgtg 1680
aagttgttag getgttattt ctattatact atgttgtagt aataagtgca ttgttgtacc 1740
agaagctatg atcataacta taggttgatc cttcattgat cagtttgatg ttgagaatac 1800
tttgaattaa aagtcttttt ttattttttt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1859

```

Βάσεις νουκλεοτιδικών δεδομένων. FASTA format

```
>gi|44890066|ref|NM_002228.3| Homo sapiens jun proto-oncogene (JUN), mRNA
GACATCATGGGCTATTTTTAGGGGTTGACTGGTAGCAGATAAGTGTTGAGCTCGGGCTGGATAAGGGCTC
AGAGTTGCACTGAGTGTGGCTGAAGCAGCGAGGCGGGAGTGGAGGTGCGCGGAGTCAGGCAGACAGACAG
ACACAGCCAGCCAGCCAGGTCGGCAGTATAGTCCGAACTGCAAATCTTATTTTTCTTTTCACCTTCTCTCT
AACTGCCCAGAGCTAGCGCCTGTGGCTCCCGGGCTGGTGTTCGGGAGTGTCAGAGAGCCTGGTCTCCA
GCCGCCCCCGGGAGGAGAGCCCTGCTGCCCAGGCGCTGTTGACAGCGGCGGAAAGCAGCGGTACCCACGC
GCCCGCCGGGGGAAGTCGGCGAGCGGCTGCAGCAGCAAAGAACTTCCCGGCTGGGAGGACCGGAGACAA
GTGGCAGAGTCCCGGAGCGAACTTTTGCAAGCCTTTCCTGCGTCTTAGGCTTCTCCACGGCGGTAAAGAC
```

Sequence Read Archive SRA - Bioproject

ΒΔ όπου κατατίθενται τα δεδομένα αλληλούχισης γονιδιωμάτων (raw sequence reads), είτε Whole genome sequencing (WGS) είτε RNA-Seq.
 Τα δεδομένα είναι οργανωμένα ανά Bioproject & Biosample

Display Settings: ▾
Send to: ▾

Bacillus cereus Accession: PRJNA574468 ID: 574468

GenomeTrakr Project: US Food and Drug Administration, Center for Food Safety and Applied Nutrition

Whole genome sequencing of cultured *Bacillus cereus* as part of the US Food and Drug Administration's WGS surveillance effort for the rapid detection of foodborne illness outbreaks.

Accession	PRJNA574468
Data Type	Genome sequencing and assembly
Scope	Multispecies
Organism	Bacillus cereus [Taxonomy ID: 1396] Bacteria; Bacillota; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group; Bacillus cereus
Submission	Registration date: 13-Jan-2021 FDA
Related Resources	<ul style="list-style-type: none"> CFSAN
Relevance	Agricultural

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (WGS master)	1
SRA Experiments	132
OTHER DATASETS	
BioSample	132
Assembly	1

See [Genome Information for Bacillus cereus](#)

NAVIGATE UP

This project is a component of the GenomeTrakr umbrella project for *Bacillus cereus*

NAVIGATE ACROSS

488 additional projects are related by organism.

7 additional projects are components of the GenomeTrakr umbrella project for *Bacillus cereus*.

Sequence Read Archive SRA - Biosample

Pathogen.env.1.0

Identifiers	BioSample: SAMN39831649; SRA: SRS20383910; CFSAN: CFSAN133898	
Organism	Bacillus cereus cellular organisms; Bacteria; Terrabacteria group; Bacillota; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group	
Package	Pathogen: environmental/food/other; version 1.0	
Attributes	collected by	CFSAN/DFPST
	collection date	2023
	geographic location	USA:IL
	isolation source	powdered infant formula
	latitude and longitude	missing
	strain	DFPST-SP1
	isolate name alias	CFSAN133898
	source type	food
	project name	GenomeTrakr
	attribute_package	environmental/food/other

Links

BioProject [PRJNA574468](#) Bacillus cereus
Retrieve [all samples](#) from this project

Submission FDA; 2024-02-05

Accession: SAMN39831649 ID: 39831649

[SRA](#)

Sequence Read Archive SRA – SRA experiment

Περιέχει πληροφορίες για την αλληλούχιση του συγκεκριμένου δείγματος

Links from BioSample

[SRX23535113](#): Whole genome Illumina MiSeq sequence of *Bacillus cereus*

1 ILLUMINA (Illumina MiSeq) run: 963,136 spots, 397.7M bases, 212.5Mb downloads

External Id: EXT00466158

Design: Whole genome library prepared from a cultured bacterial isolate.

Submitted by: FDA

Study: GenomeTrakr Project: US Food and Drug Administration, Center for Food Safety and Applied Nutrition

[PRJNA574468](#) • [SRP235182](#) • [All experiments](#) • [All runs](#)

[show Abstract](#)

Sample: Pathogen.env.1.0

[SAMN39831649](#) • [SRS20383910](#) • [All experiments](#) • [All runs](#)

Organism: [Bacillus cereus](#)

Library:

Name: Illumina DNA Prep library SEQ000139830

Instrument: Illumina MiSeq

Strategy: WGS

Source: GENOMIC

Selection: RANDOM

Layout: PAIRED

Spot descriptor:



Runs: 1 run, 963,136 spots, 397.7M bases, [212.5Mb](#)

Run	# of Spots	# of Bases	Size	Published
SRR27873194	963,136	397.7M	212.5Mb	2024-02-05

Sequence Read Archive SRA – SRA run SRR

Από εδώ μπορούμε να αποκτήσουμε τα fasta/fastq δεδομένα

Run Browser > SRR27873194

Whole genome Illumina MiSeq sequence of Bacillus cereus (SRR27873194)

Metadata

Analysis

Reads

Data access

FASTA/FASTQ download

Download for Experiment SRX23535113

<input type="checkbox"/> Accession	Total Bases	Spots	
		Total	Filtered
<input checked="" type="checkbox"/> SRR27873194	397.7Mbases	963.1k	

Filter Runs

Search by sub-sequence, spo



Filter

[What can the filter be applied to?](#)

Download

Filtered Clipped

FASTA or FASTQ

Sequence Read Archive SRA – SRR

Από το Analysis tab, μέσω του Krona view μπορούμε να δούμε σε ποιές ταξινομικές ομάδες ανήκουν οι ακολουθίες (αν π.χ. Έχουμε επιμόλυνση από άλλο είδος).

Run Browser > SRR27873194

Whole genome Illumina MiSeq sequence of *Bacillus cereus* (SRR27873194)

[Metadata](#)
[Analysis](#)
[Reads](#)
[Data access](#)
[FASTA/FASTQ download](#)

Taxonomy Analysis

97.38% IDENTIFIED READS
 2.62% UNIDENTIFIED READS

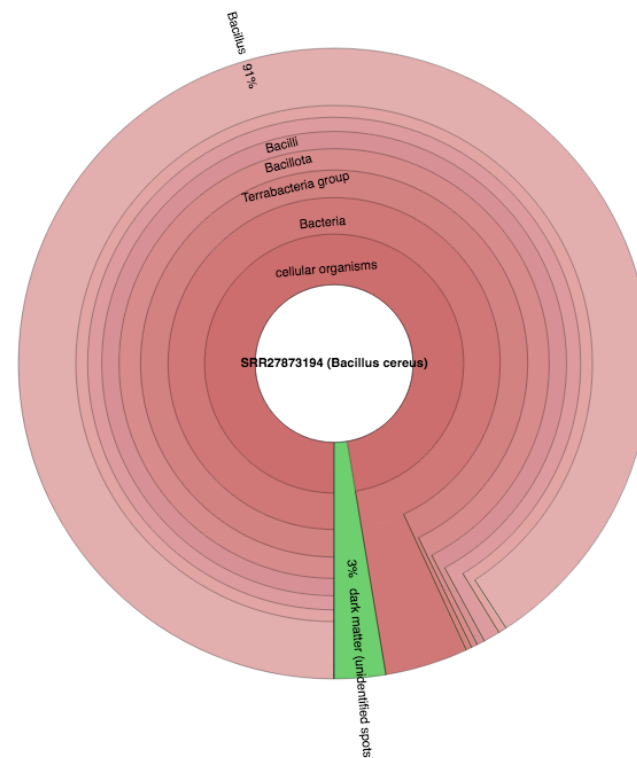
- cellular organisms: **97.38%**
 - Bacteria: **97.37%**
 - Terrabacteria group: **93.13%**
 - Bacillota: **92.83%**
 - Actinomycetota: **<0.01%**
 - Cyanobacteriota/Melainabacteria group: **<0.01%**
 - Pseudomonadota: **<0.01%** (1 Kbp)
- Viruses: **<0.01%**

View in Krona

Hide Krona View

Krona

Max depth
 Font size
 Chart size
 Color by Kbp
 Show magnitudes
 Collapse



Βάσεις πρωτεϊνικών δεδομένων

- Swissprot. 1987, Uni Geneva + SIB. Σχολιασμός των δεδομένων από επιστήμονες
- TrEMBL. 1996. SIB + EBI. Αυτόματη μετάφραση των ακολουθιών που βρίσκονται στην EMBL. Δεδομένα στην ίδια μορφή με την Swissprot. Μπορεί να είναι υποθετικές ή ο σχολιασμός να μην είναι εκτενής, όπως στην Swissprot.
- PIR. 1984, USA
- UniProt. 2002. Ενώθηκαν οι παραπάνω βάσεις.

Swissprot (I)

<https://www.uniprot.org/uniprotkb/P03372/entry>

Function

P03372 · ESR1_HUMAN

Names & Taxonomy

Proteinⁱ | Estrogen receptor

Amino acids | 595 ([go to sequence](#))


Subcellular Location

Geneⁱ | ESR1

Protein existenceⁱ | Evidence at protein level

Disease & Variants

Statusⁱ |  UniProtKB reviewed (Swiss-Prot)

Annotation scoreⁱ | 

PTM/Processing

Organismⁱ | Homo sapiens (Human)

Expression

[Entry](#) [Variant viewer](#)  [Feature viewer](#) [Genomic coordinates](#) [Publications](#) [External links](#) [History](#)

Interaction

Structure

[BLAST](#) [Align](#) [Download](#)  [Add](#) [Add a publication](#) [Entry feedback](#)


Family & Domains

Sequence & Isoforms

Similar Proteins

Functionⁱ

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-kappa-B in a cell-type specific manner. Decreases NF-kappa-B DNA-binding activity and inhibits NF-kappa-B-mediated transcription from the IL6 promoter and displace RELA/p65 and associated coregulators from the promoter. Recruited to the NF-kappa-B response element of the CCL2 and IL8 promoters and can displace CREBBP. Present with NF-kappa-B components RELA/p65 and NFkB1/p50 on ERE sequences. Can also act synergistically with NF-kappa-B to activate transcription involving respective recruitment adjacent response elements; the function involves CREBBP. Can activate the transcriptional activity of TFF1. Also mediates membrane-initiated estrogen signaling involving various kinase cascades. Essential for MTA1-mediated transcriptional regulation of BRCA1 and BCAS3 (PubMed:17922032).

 20 Publications

Isoform 3

Involved in activation of NOS3 and endothelial nitric oxide production (PubMed:21937726).

Isoforms lacking one or several functional domains are thought to modulate transcriptional activity by competitive ligand or DNA binding and/or heterodimerization with the full-length receptor (PubMed:10970861).

Binds to ERE and inhibits isoform 1 (PubMed:10970861).  2 Publications

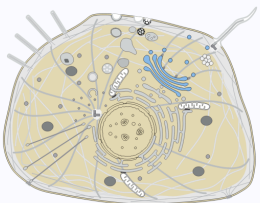
Miscellaneous

Selective estrogen receptor modulators (SERMs), such as tamoxifen, raloxifene, toremifene, lasofoxifene, clomifene, femarelle and ormeloxifene, have tissue selective agonistic and antagonistic effects on the estrogen receptor (ER). They interfere with the ER association with coactivators or corepressors, mainly involving the AF-2 domain.

Swissprot (II)

Subcellular Location

UniProt Annotation GO Annotation



- Nucleus**
- Golgi apparatus**
- Cell membrane**

Note: Coincides with ZDHHC7 and ZDHHC21 in the Golgi apparatus where most probably palmitoylation occurs. Associated with the plasma membrane when palmitoylated.

Isoform 1

- Nucleus** (Pfam:PF00144, Pfam:PF00145)
- Golgi apparatus** (Pfam:PF00144, Pfam:PF00145)
- Cell membrane** (Pfam:PF00144, Pfam:PF00145)

Note: A minor fraction is associated with the inner membrane.

Isoform 3

- Nucleus**
- Cytoplasm**
- Cell membrane** (Peripheral membrane protein)
- Cell membrane** (Single-pass type I membrane protein)

Note: Associated with the inner membrane via palmitoylation (Probable). At least a subset exists as a transmembrane protein with a N-terminal extracellular domain.

Disease & Variants¹

Involvement in disease¹

Estrogen resistance (ESTRR)

2 Publications

Note The disease is caused by variants affecting the gene represented in this entry

Description A disorder characterized by partial or complete resistance to estrogens, in the presence of elevated estrogen serum levels. Clinical features include absence of the pubertal growth spurt, delayed bone maturation, unfused epiphyses, reduced bone mineral density, osteoporosis, continued growth into adulthood and very tall adult stature. Glucose intolerance, hyperinsulinemia and lipid abnormalities may also be present.

See also MIM:615363

Natural variants in ESTRR

VARIANT ID	POSITION(S)	CHANGE	DESCRIPTION
VAR_070072	375	Q>H	in ESTRR; results in highly reduced activity; dbSNP:rs397509428 1 Publication
VAR_078516	394	R>H	in ESTRR; highly decreased estrogen receptor activity; dbSNP:rs1131692059 1 Publication

Organism-specific databases

DisGeNET | 2099

MIM | 133430 gene
615363 phenotype

MalaCards | ESR1

OpenTargets | ENSG0000091831

Orphanet | 785 Estrogen resistance syndrome

PharmGKB | PA156

Miscellaneous

Pharos | P03372 Tclin

Chemistry

ChEMBL | CHEMBL206

DrugBank | DB07567 (2R,3R,4S)-3-(4-HYDROXYPHENYL)-4-METHYL-2-[4-(2-PYRROLIDIN-1-YLETHOXY)PHENYL]CHROMAN-6-OL
DB07638 (3AS,4R,9BR)-2,2-DIFLUORO-4-(4-HYDROXYPHENYL)-1,2,3,3A,4,9B-HEXAHYDROCYCLOPENTA[C]CHROMEN-8-OL
DB08737 (3AS,4R,9BR)-4-(4-HYDROXYPHENYL)-1,2,3,3A,4,9B-HEXAHYDROCYCLOPENTA[C]CHROMEN-9-OL
DB08020 (3AS,4R,9BR)-4-(4-HYDROXYPHENYL)-6-(METHOXYMETHYL)-1,2,3,3A,4,9B-HEXAHYDROCYCLOPENTA[C]CHROMEN-8-OL
DB07678 (9ALPHA,13BETA,17BETA)-2-[(1Z)-BUT-1-EN-1-YL]ESTRA-1,3,5,10-TRIENE-3,17-DIOL
[More DrugBank links](#)

DrugCentral | P03372

GuidetoPHARMACOLOGY | 620

Genetic variation databases

BioMuta | ESR1

DMDM | 544257

ClinGen | HGNC:3467

GenCC | HGNC:3467

Swissprot (III)

Expressionⁱ

Tissue specificityⁱ

Widely expressed (PubMed:[10970861](#)).

Not expressed in the pituitary gland (PubMed:[10970861](#)). 1 Publication

Isoform 3

Widely expressed, however not expressed in the pituitary gland. 1 Publication

Gene expression databases

Bgee | [ENSG00000091831](#) [↗](#) Expressed in oviduct epithelium and 160 other cell types or tissues

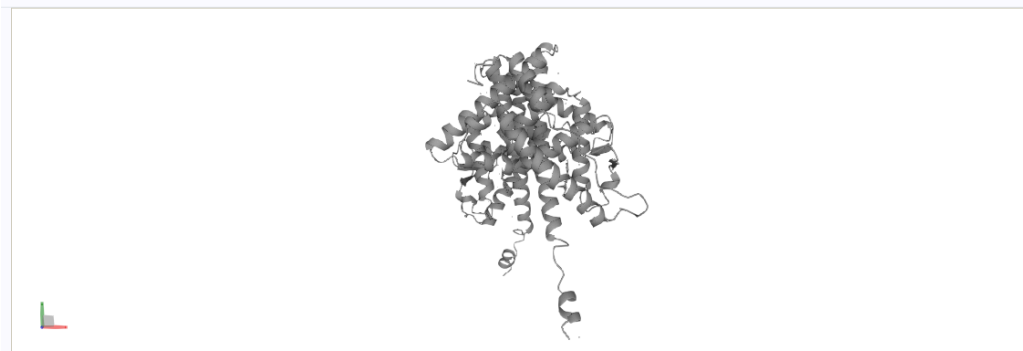
ExpressionAtlas | [P03372](#) [↗](#) baseline and differential

Genevisible | [P03372](#) [↗](#) HS

Organism-specific databases

HPA | [ENSG00000091831](#) [↗](#) Tissue enhanced (cervix, endometrium, fallopian tube)

Structureⁱ



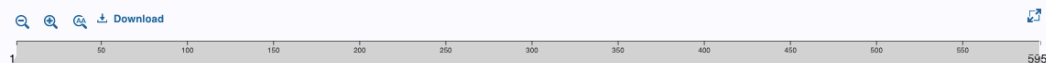
SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
-- Select --		-- Select --				
PDB	1A52	X-ray	2.80 Å	A/B	297-554	PDBe · RCSB-PDB · PDBJ · PDBsum ↗ · Foldseek ↗
PDB	1ERE	X-ray	3.10 Å	A/B/C/D/E/F	301-553	PDBe · RCSB-PDB · PDBJ · PDBsum ↗ · Foldseek ↗
PDB	1ERR	X-ray	2.60 Å	A/B	301-553	PDBe · RCSB-PDB · PDBJ · PDBsum ↗ · Foldseek ↗
PDB	1G50	X-ray	2.90 Å	A/B/C	304-550	PDBe · RCSB-PDB · PDBJ · PDBsum ↗ · Foldseek ↗
PDB	1GWQ	X-ray	2.45 Å	A/B	301-548	PDBe · RCSB-PDB · PDBJ · PDBsum ↗ · Foldseek ↗
PDB	1GWR	X-ray	2.40 Å	A/B	305-549	PDBe · RCSB-PDB · PDBJ · PDBsum ↗ · Foldseek ↗
PDB	1HCP	NMR		A	180-254	PDBe · RCSB-PDB · PDBJ · PDBsum ↗ · Foldseek ↗
PDB	1HCQ	X-ray	2.40 Å	A/B/E/F	180-262	PDBe · RCSB-PDB · PDBJ · PDBsum ↗ · Foldseek ↗

Swissprot

Family & Domains¹

Features

Showing features for region¹, zinc finger¹, domain¹, compositional bias¹.



TYPE	ID	POSITION(S)	DESCRIPTION	
-- Select --				
▶ Zinc finger		221-245	NR C4-type PROSITE-ProRule Annotation	BLAST Add
▶ Region		251-310	Hinge	BLAST Add
▶ Region		259-282	Disordered Automatic Annotation	BLAST Add
▶ Region		262-595	Interaction with AKAP13 1 Publication	BLAST Add
▶ Region		264-595	Self-association	BLAST Add
▶ Domain		311-547	NR LBD PROSITE-ProRule Annotation	BLAST Add
▶ Region		311-595	Transactivation AF-2	BLAST Add

[Expand table](#)

Domain¹

Composed of three domains: a modulating N-terminal domain, a DNA-binding domain and a C-terminal ligand-binding domain. The modulating domain, also known as A/B or AF-1 domain has a ligand-independent transactivation function. The C-terminus contains a ligand-dependent transactivation domain, also known as E/F or AF-2 domain which overlaps with the ligand binding domain. AF-1 and AF-2 activate transcription independently and synergistically and act in a promoter- and cell-specific manner. AF-1 seems to provide the major transactivation function in differentiated cells.

Family and domain databases

CDD	cd07171 NR_DBD_ER 1 hit cd06949 NR_LBD_ER 1 hit	PRINTS	PR00543 OESTROGENR PR00398 STRDHORMONER PR00047 STROIDFINGER
DisProt	DP00074	PROSITE	View protein in PROSITE PS51843 NR_LBD 1 hit PS00031 NUCLEAR_REC_DBD_1 1 hit PS51030 NUCLEAR_REC_DBD_2 1 hit
Gene3D	3.30.50.10 Erythroid Transcription Factor GATA-1, subunit A 1 hit 1.10.565.10 Retinoid X Receptor 1 hit	Pfam	View protein in Pfam PF12743 ESR1_C 1 hit PF00104 Hormone_recep 1 hit PF02159 Oest_recep 1 hit PF00105 zf-C4 1 hit
IDEAL	IID00013	SMART	View protein in SMART SM00430 HOLI 1 hit SM00399 ZnF_C4 1 hit
InterPro	View protein in InterPro IPR024178 Est_rcpt/est-rel_rcp IPR001292 Estr_rcpt IPR046944 Estr_rcpt_N IPR035500 NHR-like_dom_sf More InterPro links	SUPFAM	SSF57716 Glucocorticoid receptor-like (DNA-binding domain) 1 hit SSF48508 Nuclear receptor ligand-binding domain 1 hit
PANTHER	PTHR48092:SF16 ESTROGEN RECEPTOR 1 hit PTHR48092 KNIRPS-RELATED PROTEIN-RELATED 1 hit	MobiDB	Search...
PIRSF	PIRSF500101 ER-a 1 hit PIRSF002527 ER-like_NR 1 hit		

Swissprot

Sequence & Isoformsⁱ

[Align 4 isoforms](#)

Sequence statusⁱ | Complete

This entry describes 4 isoformsⁱ produced by **Alternative promoter usage & Alternative splicing**.

P03372-1

This isoform has been chosen as the **canonical** sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

Name 1

Synonyms Long, hER-alpha66, ER66

See also sequence in [UniParc](#) or sequence clusters in [UniRef](#)

Tools [Download](#) [Add](#) [Highlight](#) [Copy sequence](#)

Length 595

Mass (Da) 66,216

Last updated 1994-06-01 v2

Checksumⁱ 5455C57AB0CCCAA7

10	20	30	40	50	60	70	80	90	100	110	120	130
MTMTLHTKAS	GMALLHQIQG	NELEPLNRQ	LKIPLERPLG	EVYLDSSKPA	VYNYPEGAAY	EFNAAAAANA	QVYGGTGLPY	GPGEAAAFG	SNGLGGFPPL	NSVSPSPLML	LHPPQLSPF	LQPHGQQVRY
140	150	160	170	180	190	200	210	220	230	240	250	260
YLENEPSGYT	VREAGPPAFY	RPNSDNRRQG	GRERLASTND	KGSMAMESAK	ETRYCAVCND	YASGYHYGVW	SCGCKAFFK	RSIQGHNDYM	CPATNQCTID	KNRRKSCQAC	RLRKCYEVGM	MKGGIRKDRR
270	280	290	300	310	320	330	340	350	360	370	380	390
GGRMLKHKRQ	RDDGEGRGEV	GSAGDMRAAN	LWPSPLMIKR	SKKNSLALS	TADQMSALL	DAEPPILYSE	YDPTRPFSEA	SMMGLLTNLA	DRELVHMINW	AKRVPGFVDL	TLHDQVHLL	CAWLEILMIG
400	410	420	430	440	450	460	470	480	490	500	510	520
LVWRSMHPG	KLLFAPNLLL	DRNQKCEV	MVEIFDMLLA	TSSFRMMNL	QGEEFVCLKS	IILLNSGVYT	FLSSTLKSLE	EKDIHRVLD	KITDTLIHLM	AKAGTLQQQ	HQRLAQLLLI	LSHIRHMSNK
530	540	550	560	570	580	590						
GMEHLYSMKC	KNVVPLYDLL	LEMLDAHRLH	APTSRGGASV	EETDQSHLAT	AGSTSSHSLQ	KYYITGEAEG	FPATV					

```
>sp|P03372|ESR1_HUMAN Estrogen receptor OS=Homo sapiens OX=9606 GN=ESR1 PE=1 SV=2
MTMTLHTKASGMALLHQIQGNELEPLNRQLKIPLERPLGGEVYLDSSKPAVYNYPEGAAY
EFNAAAAANAQVYGGTGLPYGPGEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPQLSPF
LQPHGQQVRYYLENEPSGYTVREAGPPAFYRPNSDNRRQGGRRERLASTNDKGSMAMESAK
ETRYCAVCNDYASGYHYGVWSCGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQAC
RLRKCYEVGMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMIKR
SKKNSLALS LTADQMSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLA DRELVHMINW
AKRVPGFVDLTLHDQVHLLCAWLEILMIGLVWRSMHPGKLLFAPNLLLDRNQKCEV
MVEIFDMLLATSSFRMMNLQGEEFVCLKSIIILLNSGVYTFLSSTLKSLEEKDIHRVLD
KITDTLIHLMKAGTLQQQHQRLAQLLLLSHIRHMSNKGMEHLYSMCKNVVPLYDLL
LEMLDAHRLHAPTSRGGASVEETDQSHLATAGSTSSHSLQKYYITGEAEGFPATV
```

Uniprot

Sequence databases

CCDS | [CCDS5234.1](#) [\[P03372-1\]](#)
[CCDS87457.1](#) [\[P03372-4\]](#)

PIR | [S64737](#) [S64737](#)

RefSeq | [NP_000116.2](#) [NM_000125.3](#) [\[P03372-1\]](#)
[NP_001116212.1](#) [NM_001122740.1](#) [\[P03372-1\]](#)
[NP_001116213.1](#) [NM_001122741.1](#) [\[P03372-1\]](#)
[NP_001116214.1](#) [NM_001122742.1](#) [\[P03372-1\]](#)
[NP_001278159.1](#) [NM_001291230.1](#) [\[P03372-1\]](#)
[More RefSeq links](#)

NUCLEOTIDE SEQUENCE	PROTEIN SEQUENCE	MOLECULE TYPE	STATUS
X03635 EMBL GenBank DDBJ	CAA27284.1 EMBL GenBank DDBJ	mRNA	
M12674 EMBL GenBank DDBJ	AAA52399.1 EMBL GenBank DDBJ	mRNA	
U47678 EMBL GenBank DDBJ	AAB00115.1 EMBL GenBank DDBJ	mRNA	Sequence problems.
AY425004 EMBL GenBank DDBJ	AAQ91815.1 EMBL GenBank DDBJ	Genomic DNA	
BX640939 EMBL GenBank DDBJ	CAE45969.1 EMBL GenBank DDBJ	mRNA	
AL049821 EMBL GenBank DDBJ	-	Genomic DNA	No translation available.
AL078582 EMBL GenBank DDBJ	-	Genomic DNA	No translation available.

Expand table

Genome annotation databases

Ensembl | [ENST00000206249.8](#) [ENSP00000206249.3](#)
[ENSG00000091831.25](#) [\[P03372-1\]](#)
[ENST00000338799.9](#) [ENSP00000342630.5](#)
[ENSG00000091831.25](#) [\[P03372-1\]](#)
[ENST00000427531.6](#) [ENSP00000394721.2](#)
[ENSG00000091831.25](#) [\[P03372-4\]](#)
[ENST00000440973.5](#) [ENSP00000405330.1](#)

GeneID | [2099](#)

KEGG | [hsa:2099](#)

MANE-Select | [ENST00000206249.8](#) [ENSP00000206249.3](#) [NM_000125.4](#) [NP_000116.2](#)

UCSC | [uc003qom.5](#) [human](#) [\[P03372-1\]](#)

European Nucleotide Archive (ENA)

<https://www.ebi.ac.uk/ena/browser/view/X03635>



Enter text search terms

Search

Examples: histone, BN000065

X03635

View

Examples: Taxon:9606, BN000065, PRJEB402

Home [Submit](#) [Search](#) [Rulespace](#) [About](#) [Support](#)

Sequence: X03635.1

Homo sapiens mRNA for oestrogen receptor

Organism: [Homo sapiens \(human\)](#)
Mol Type: mRNA
Topology: linear
Base Count: 6450
Dataclass: STD
Tax Division: HUM
Accession: X03635

View: [EMBL](#) [FASTA](#)
Download: [EMBL](#) [FASTA](#)
Navigation: [Show](#)
Publications: [Show](#)
Sequence Versions: [View](#)

ΒΔ Γονιδιακής έκφρασης

NAR Database Summary Paper Category List

- Nucleotide Sequence Databases
- RNA sequence databases
- Protein sequence databases
- Structure Databases
- Genomics Databases (non-vertebrate)
- Metabolic and Signaling Pathways
- Human and other Vertebrate Genomes
- Human Genes and Diseases
- Microarray Data and other Gene Expression Databases

- 4DXpress
- ABA - Ascidian Body Atlas
- Allen Brain Atlas
- BarleyBase
- BioStudies
- BloodExpress
- BloodSpot
- CAGE
- CATMA - Complete Arabidopsis Transcriptome MicroArray
- CEBS
- CircaDB
- CleanEx
- CODEX
- Coexpedia
- COLOMBOS v2.0: An ever expanding collection of bacterial expression compendia
- CycleBase
- dbERGEII
- EMAGE
- Expression Atlas
- Expression Atlas
- GEISHA
- Gene Expression Barcode
- GeneFriends
- GeneNote
- GenePaint
- GeneSigDB
- GeneTrap
- GeneWeaver
- GenomeRNAi
- GENSAT
- GEO - Gene Expression Omnibus

- › [Compilation Paper](#)
- › [Category List](#)
- › [Alphabetical List](#)
- › [Category/Paper List](#)
- › [Search Summary Papers](#)

<https://www.ncbi.nlm.nih.gov/geo/>

The screenshot shows the Gene Expression Omnibus (GEO) website. At the top, there is a navigation bar with links for 'GEO Home', 'Documentation', 'Query & Browse', and 'Email GEO'. The main heading is 'Gene Expression Omnibus'. Below this, a brief description states: '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.' To the right of this text is a search box with the placeholder 'Keyword or GEO Accession' and a 'Search' button. The page is organized into three main columns: 'Getting Started', 'Tools', and 'Browse Content'. The 'Getting Started' column includes links for Overview, FAQ, About GEO DataSets, About GEO Profiles, About GEO2R Analysis, How to Construct a Query, and How to Download Data. The 'Tools' column includes links for 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, and ENCODE Data Listings and Tracks. The 'Browse Content' column includes a Repository Browser and a table of statistics: DataSets: 4348, Series: 222877, Platforms: 25886, and Samples: 7052566. At the bottom, there is an 'Information for Submitters' section with links for Login to Submit, Submission Guidelines, Update Guidelines, MIAME Standards, Citing and Linking to GEO, Guidelines for Reviewers, and GEO Publications.

ΒΔ τρισδιάστατων δομών



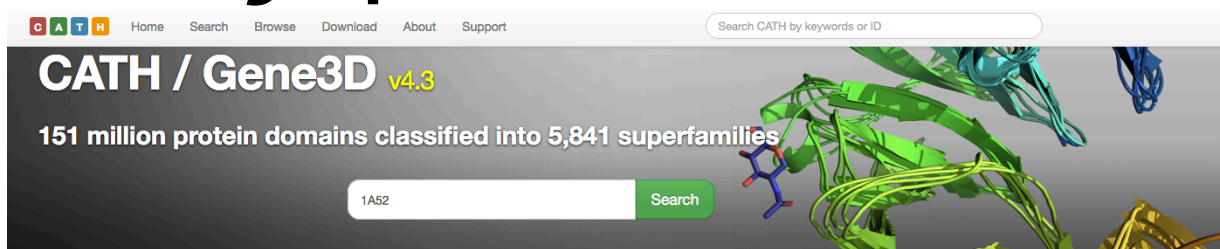
<https://www.rcsb.org/>

- Protein Data Bank (PDB)
 - Πρωτεΐνες
 - Νουκλεϊκά οξέα
 - Σύμπλοκα των παραπάνω

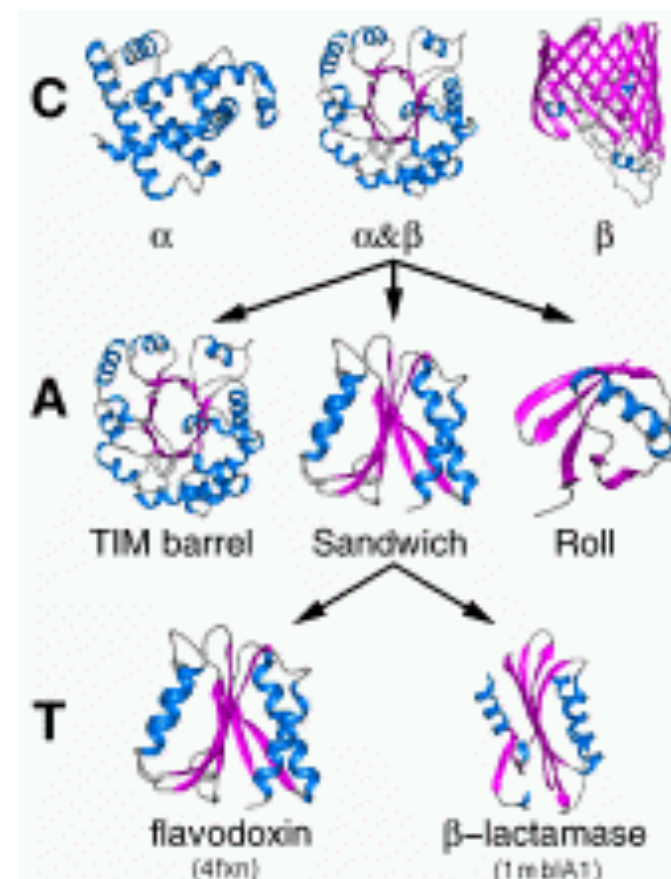
- Μέθοδοι
 - X-ray
 - NMR
 - Κρύο-ηλεκτρονική μικροσκοπία

Οι παραπάνω μέθοδοι βρίσκουν τις συντεταγμένες (3D) των ατόμων του βιολογικού μορίου. Τα αρχεία με τις συντεταγμένες διαβάζονται από ειδικά προγράμματα (π.χ Rasmol) που απεικονίζουν την δομή στο χώρο

Βάσεις τρισδιάστατων δομών



- CATH: κατηγοριοποιεί τις τρισδιάστατες δομές των πρωτεϊνικών επικρατειών ιεραρχικά, σε 4 βασικά επίπεδα.
- Η κατηγοριοποίηση γίνεται με ένα συνδυασμό αυτόματων μεθόδων και ανθρώπινης κρίσης.



Βάσεις τρισδιάστατων δομών

C A T H Home Search Browse Download About Support Search CATH by keywords or ID

Browse CATH-Gene3D Hierarchy

BROWSE LINKS

- Browse Hierarchy**
- Highly Diverse Superfamilies
- Superfamily Comparison

Select a CATH node...

Tree Sunburst

Top of CATH Hierarchy (4 Classes)

- 1 Mainly Alpha *5 Architectures, 404 Folds, 2033 Superfamilies, 103788 Domains*
- 2 Mainly Beta *21 Architectures, 244 Folds, 1290 Superfamilies, 124032 Domains*
- 3 Alpha Beta *14 Architectures, 634 Folds, 2337 Superfamilies, 262275 Domains*
- 4 Few Secondary Structures *1 Architectures, 108 Folds, 181 Superfamilies, 5716 Domains*
- 6 Special *2 Architectures, 82 Folds, 790 Superfamilies, 4427 Domains*

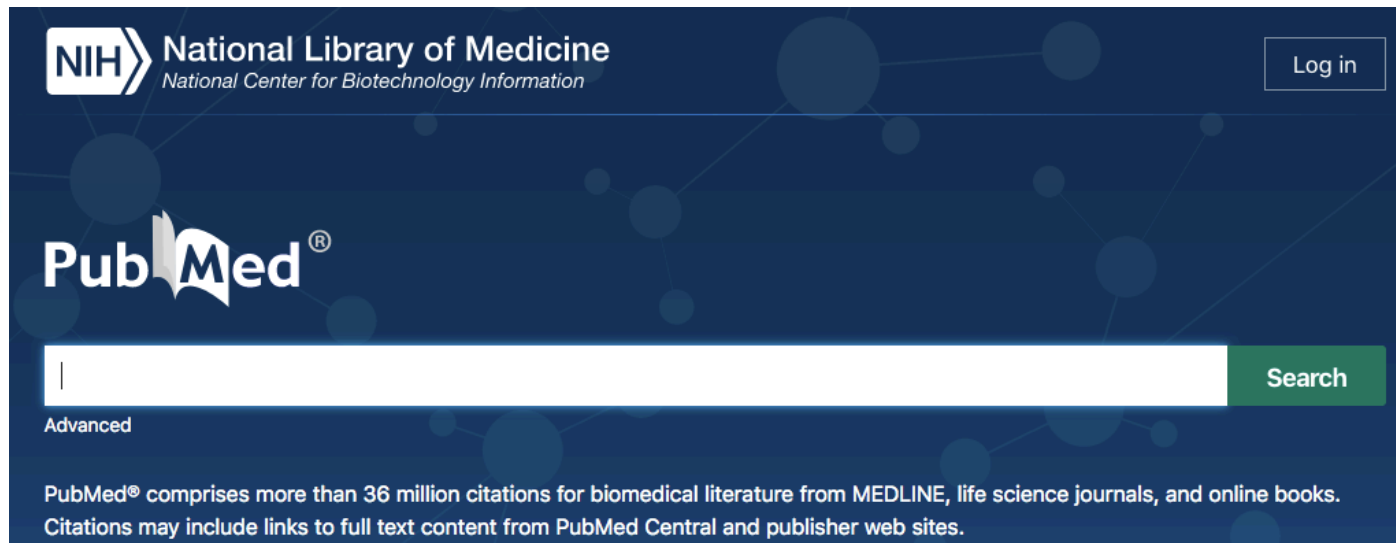
CATH Domain: 1cukA01 [XML](#)

PDB 1cuk, Chain A, Domain 1

CATH Code	Level Description	Links
2	Mainly Beta	
A 2.40	Beta Barrel	
T 2.40.50	OB fold (Dihydropoamide Acetyltransferase, E2P)	
H 2.40.50.140	Nucleic acid-binding proteins	Gene3D
S 2.40.50.140.47		
C 2.40.50.140.47.1		
L 2.40.50.140.47.1.1		
I 2.40.50.140.47.1.1.1		
D 2.40.50.140.47.1.1.1.1		



<https://pubmed.ncbi.nlm.nih.gov/>




The screenshot shows the PubMed website homepage. At the top left is the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". At the top right is a "Log in" button. Below the NIH logo is a yellow and green banner that says "FREE full text article in PubMed Central". The main logo "PubMed" is prominently displayed. Below the logo is a search bar with a green "Search" button. Under the search bar, the word "Advanced" is visible. At the bottom, there is a paragraph of text: "PubMed® comprises more than 36 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites."

- ΒΔ του NCBI. Ξεκίνησε τον Ιανουάριο του 1996.
- Καταχωρεί όλες τις δημοσιευμένες εργασίες που προέρχονται από τον ευρύτερο χώρο της βιοϊατρικής
- ~36 εκατομύρια εργασίες καταχωρημένες (2024)
- Όταν μια εργασία γίνεται δεκτή από το περιοδικό, κατατίθεται και στην Pubmed
- Η Pubmed δίνει ένα μοναδικό κωδικό εγγραφής (PMID) και λέξεις κλειδιά που χαρακτηρίζουν το περιεχόμενο της εργασίας (MeSH terms).
- Από το 2007, το NIH απαιτεί όποιες ερευνητικές εργασίες έχουν χρηματοδοτηθεί από αυτό, τα αποτελέσματά τους να γίνονται προσβάσιμα σε όλους, μέσω του PubMed Central (εντός 12 μηνών από την ημερομηνία δημοσίευσης). (~ 1 εκατομύριο εργασίες)

<https://pubmed.ncbi.nlm.nih.gov/>

PubMed Advanced Search Builder


[User Guide](#)

Add terms to the query box

All Fields

Enter a search term

ADD

Show Index

Query box

Enter / edit your search query here

Search

ADD

Add with AND

Add with OR

Add with NOT

PubMed Advanced Search Builder

- issue
- Journal
- Language
- Location ID
- MeSH Major Topic
- MeSH Subheading
- MeSH Terms
- Other Term
- Pagination
- Pharmacological Action
- Publication Type
- Publisher
- Secondary Source ID
- Subject - Personal Name
- Supplementary Concept
- Text Word
- Title
- Title/Abstract
- Transliterated Title
- Volume

Pubmed

[Review](#) > [Science](#). 1996 Oct 25;274(5287):546, 563-7. doi: 10.1126/science.274.5287.546.

Life with 6000 genes

A Goffeau ¹, B G Barrell, H Bussey, R W Davis, B Dujon, H Feldmann, F Galibert, J D Hoheisel, C Jacq, M Johnston, E J Louis, H W Mewes, Y Murakami, P Philippsen, H Tettelin, S G Oliver

Affiliations + expand

PMID: 8849441 DOI: [10.1126/science.274.5287.546](https://doi.org/10.1126/science.274.5287.546)

Abstract

The genome of the yeast *Saccharomyces cerevisiae* has been completely sequenced through a worldwide collaboration. The sequence of 12,068 kilobases defines 5885 potential protein-encoding genes, approximately 140 genes specifying ribosomal RNA, 40 genes for small nuclear RNA molecules, and 275 transfer RNA genes. In addition, the complete sequence provides information about the higher order organization of yeast's 16 chromosomes and allows some insight into their evolutionary history. The genome shows a considerable amount of apparent genetic redundancy, and one of the major problems to be tackled during the next stage of the yeast genome project is to elucidate the biological functions of all of these genes.

[PubMed Disclaimer](#)

Comment in

[Funding sequencing efforts.](#)

Bond E, Austin MJ.

Science. 1997 Feb 21;275(5303):1051-2. doi: 10.1126/science.275.5303.1049f.

PMID: 9054002 No abstract available.

Similar articles

[Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII.](#)

Johnston M, Andrews S, Brinkman R, Cooper J, Ding H, Dover J, Du Z, Favello A, Fulton L, Gattung S, et al.

Science. 1994 Sep 30;265(5181):2077-82. doi: 10.1126/science.8091229.

PMID: 8091229

FULL TEXT LINKS



ACTIONS

“ Cite

📖 Collections

SHARE



PAGE NAVIGATION

< Title & authors

Abstract

Comment in

Similar articles

Cited by

Publication types

MeSH terms

Substances

Related information

ΒΔ πρωτεϊνικών επικρατειών

- Πρωτεϊνική επικράτεια: Μια περιοχή της πρωτεΐνης με συγκεκριμένη λειτουργία/δομή και καλά συντηρημένη.
- Διάφορες βάσεις δεδομένων, όπως:
 - PROSITE
 - Pfam
 - PRINTS
 - ProDom
 - SMART
 - TIGRFAMs
 - PIR superfamily
 - Superfamily
- Έχουν ενσωματωθεί στο INTERPRO
- Το INTERPRO περιέχει πρωτεϊνικές επικράτειες. Το πρόγραμμα INTERPROscan ανιχνεύει αυτές τις επικράτειες στις πρωτεΐνες.

INTERPRO

https://www.ebi.ac.uk/interpro/protein/UniProt/P03372/

InterPro Classification of protein families
🔍 ☰

Home Search Browse Results Release notes Download Help About Contact us

🏠 / Browse / By Protein / UniProt / P03372 / Overview

P03372 Estrogen receptor

UniProtKB/Swiss-Prot protein

Overview

- Entries 9
- Structures 433
- Sequence
- Similar Proteins 545
- AlphaFold 1

Short name	ESR1_HUMAN
Length	595 amino acids
Species	Homo sapiens (Human)
Proteome	UP000005640
Gene	ESR1
Function	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... Show More
Family membership	<ul style="list-style-type: none"> 📄 Nuclear hormone receptor (IPR001723) 📄 Estrogen receptor/oestrogen-related receptor (IPR024178) 📄 Estrogen receptor (IPR001292)
Isoforms	Select an isoform <input type="button" value="⌵"/>

External Links

[UniProt](#)

[Search sequence with InterProScan](#)

📄 [Download matches \(TSV\)](#)

📄 [Download sequence \(FASTA\)](#)

Entry matches to this protein

▼ AlphaFold Confidence

▼ Representative Domains

▼ Family

▼ Domain

Options

📄 [pLDDT](#)

📄 [Estr_rcpt - IPR001292](#)
ESTROGENR - PRO0543
ER-α - PIRSF00101

📄 [Nuclear_hrmn_rcpt - IPR001723](#)
STRDHORMONER - PRO0398

📄 [Est_rcpt/est-rel_rcp - IPR024178](#)
ER-like_NR - PIRSF02527

📄 [Nucl_hrmn_rcpt_lig-bd - IPR0005](#)
HOLL - SM0043D
NR_LBD - PSS1843
Hormone_recep - PF00104

NCBI/Entrez

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

- ✓ All Databases
- Assembly
- Biocollections
- BioProject
- BioSample
- Books**
- ClinVar
- Conserved Domains
- dbGaP
- dbVar
- Gene
- Genome
- GEO DataSets
- GEO Profiles
- GTR
- Identical Protein Groups
- MedGen
- MeSH
- NLM Catalog
- Nucleotide

Search

NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical information.
[Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Manuscripts
es

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



Research
Explore NCBI research and collaborative projects



Popular Resources


- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

- Submit Your Somatic Variation Data to ClinVar 29 Feb 2024
- NCBI's ClinVar is now accepting submissions for somatic variants
- Join NCBI at TAGC 2024 22 Feb 2024
- March 6-10 in Washington, D.C. We look forward to seeing you in person at The Allied Genetics Conference (TAGC)

Μεταβολικά μονοπάτια KEGG

KEGG
Databases
Tools
Auto annotation
Kanehisa Lab



KEGG ▾

[» Japanese](#)

KEGG Home

- [Release notes](#)
- [Current statistics](#)

KEGG Database

- [KEGG overview](#)
- [KEGG mapping](#)
- [Color codes](#)

KEGG Objects

- [KEGG Weblinks](#)
- [Entry format](#)

KEGG Web Apps

- [Coloring GUI](#)
- [Coloring URL](#)

KEGG Software

- [KEGG API](#)
- [KGML](#)

KEGG FTP

- [Subscription](#)
- [Background info](#)

[GenomeNet](#)

[DBGET/LinkDB](#)

[Feedback](#)

[Copyright request](#)

[Kanehisa Labs](#)

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. See [Release notes](#) (March 1, 2024) for new and updated features.

[New article](#) [KEGG tools for classification and analysis of viral proteins](#)

- **Main entry point to the KEGG web service**
 - KEGG2** [KEGG Table of Contents](#) [[Update notes](#) | [Release history](#)]
- **Data-oriented entry points**

KEGG PATHWAY	KEGG pathway maps	Pathway Brite Brite table Module Network KO (Function) Organism Virus Compound Disease (ICD) Drug (ATC) Drug (Target) Antimicrobials
KEGG BRITE	BRITE hierarchies and tables	
KEGG MODULE	KEGG modules	
KEGG ORTHOLOGY	KO functional orthologs [Annotation]	
KEGG GENES	Genes and proteins [SeqData]	
KEGG GENOME	Genomes [Taxonomy KEGG Virus]	
KEGG COMPOUND	Small molecules	
KEGG GLYCAN	Glycans	
KEGG REACTION	Biochemical reactions [RModule]	
KEGG ENZYME	Enzyme nomenclature	
KEGG NETWORK	Disease-related network variations	
KEGG DISEASE	Human diseases	
KEGG DRUG	Drugs [New drug approvals]	
KEGG MEDICUS	Health information resource [Drug labels search]	

KEGG pathways

- Kyoto encyclopedia of genes and genomes
- 2010: 374 μεταβολικά μονοπάτια



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

[KEGG2](#)
[PATHWAY](#)
[BRITE](#)
[DISEASE](#)
[DRUG](#)
[KO](#)
[GENES](#)
[GENOME](#)
[LIGAND](#)
[DBGET](#)

Select prefix:
 Enter keywords:
 [Help](#)

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see [new maps](#), [change history](#), and [last updates](#)) representing our knowledge on the molecular interaction and reaction networks for:

- 0. Global Map**
- 1. Metabolism**
 - [Carbohydrate](#) [Energy](#) [Lipid](#) [Nucleotide](#) [Amino acid](#) [Other amino acid](#) [Glycan](#)
 - [Cofactor/vitamin](#) [Terpenoid/PK](#) [Other secondary metabolite](#) [Xenobiotics](#) [Overview](#)
- 2. Genetic Information Processing**
- 3. Environmental Information Processing**
- 4. Cellular Processes**
- 5. Organismal Systems**
- 6. Human Diseases**

and also on the structure relationships (KEGG drug structure maps) in:

- 7. Drug Development**

KEGG Atlas may now be used to examine any of the KEGG pathway maps.



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions and relations

KEGG2 PATHWAY BRITE MODULE KO GENES COMPOUND NETWORK DISEASE DRUG

Select prefix

map

Organism

Enter keywords

Go

Help

[[New pathway maps](#) | [Update history](#)]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn [pathway maps](#) representing our knowledge of the molecular interaction, reaction and relation networks for:

1. Metabolism

[Global/overview](#) [Carbohydrate](#) [Energy](#) [Lipid](#) [Nucleotide](#) [Amino acid](#) [Other amino](#) [Glycan](#)
[Cofactor/vitamin](#) [Terpenoid/PK](#) [Other secondary metabolite](#) [Xenobiotics](#) [Chemical structure](#)

2. Genetic Information Processing

3. Environmental Information Processing

4. Cellular Processes

5. Organismal Systems

6. Human Diseases

7. Drug Development

The pathway map viewer linked from this page contains features of [KEGG mapping](#), especially for coloring map objects as described [here](#).

1. Metabolism

1.0 Global and overview maps

- 01100 M [Metabolic pathways](#)
- 01110 M [Biosynthesis of secondary metabolites](#)
- 01120 M [Microbial metabolism in diverse environments](#)
- 01200 M R [Carbon metabolism](#)
- 01210 M R [2-Oxocarboxylic acid metabolism](#)
- 01212 M R [Fatty acid metabolism](#)
- 01230 M R [Biosynthesis of amino acids](#)
- 01232 M R [Nucleotide metabolism](#)
- 01250 M R [Biosynthesis of nucleotide sugars](#)
- 01240 M R [Biosynthesis of cofactors](#)
- 01220 M R [Degradation of aromatic compounds](#)

1.1 Carbohydrate metabolism

- 00010 M N [Glycolysis / Gluconeogenesis](#)
- 00020 M N [Citrate cycle \(TCA cycle\)](#)
- 00030 M [Pentose phosphate pathway](#)
- 00040 M [Pentose and glucuronate interconversions](#)
- 00051 M [Fructose and mannose metabolism](#)
- 00052 M N [Galactose metabolism](#)
- 00053 M [Ascorbate and aldarate metabolism](#)
- 00500 M N [Starch and sucrose metabolism](#)
- 00520 M N [Amino sugar and nucleotide sugar metabolism](#)
- 00620 M N [Pyruvate metabolism](#)
- 00630 M [Glyoxylate and dicarboxylate metabolism](#)
- 00640 M [Propanoate metabolism](#)
- 00650 M [Butanoate metabolism](#)
- 00660 M [C5-Branched dibasic acid metabolism](#)
- 00562 M [Inositol phosphate metabolism](#)

1.2 Energy metabolism

- 00190 M [Oxidative phosphorylation](#)
- 00195 M [Photosynthesis](#)
- 00196 [Photosynthesis - antenna proteins](#)
- 00710 M [Carbon fixation in photosynthetic organisms](#)
- 00720 M [Carbon fixation pathways in prokaryotes](#)
- 00680 M [Methane metabolism](#)
- 00910 M [Nitrogen metabolism](#)
- 00920 M [Sulfur metabolism](#)

KEGG Metabolic pathways - Reference pathway

[Pathway menu | Pathway entry | Download | Help]

Change pathway type

Option

Scale: 30%

Coloring: Yes No

Link: Normal None

Search

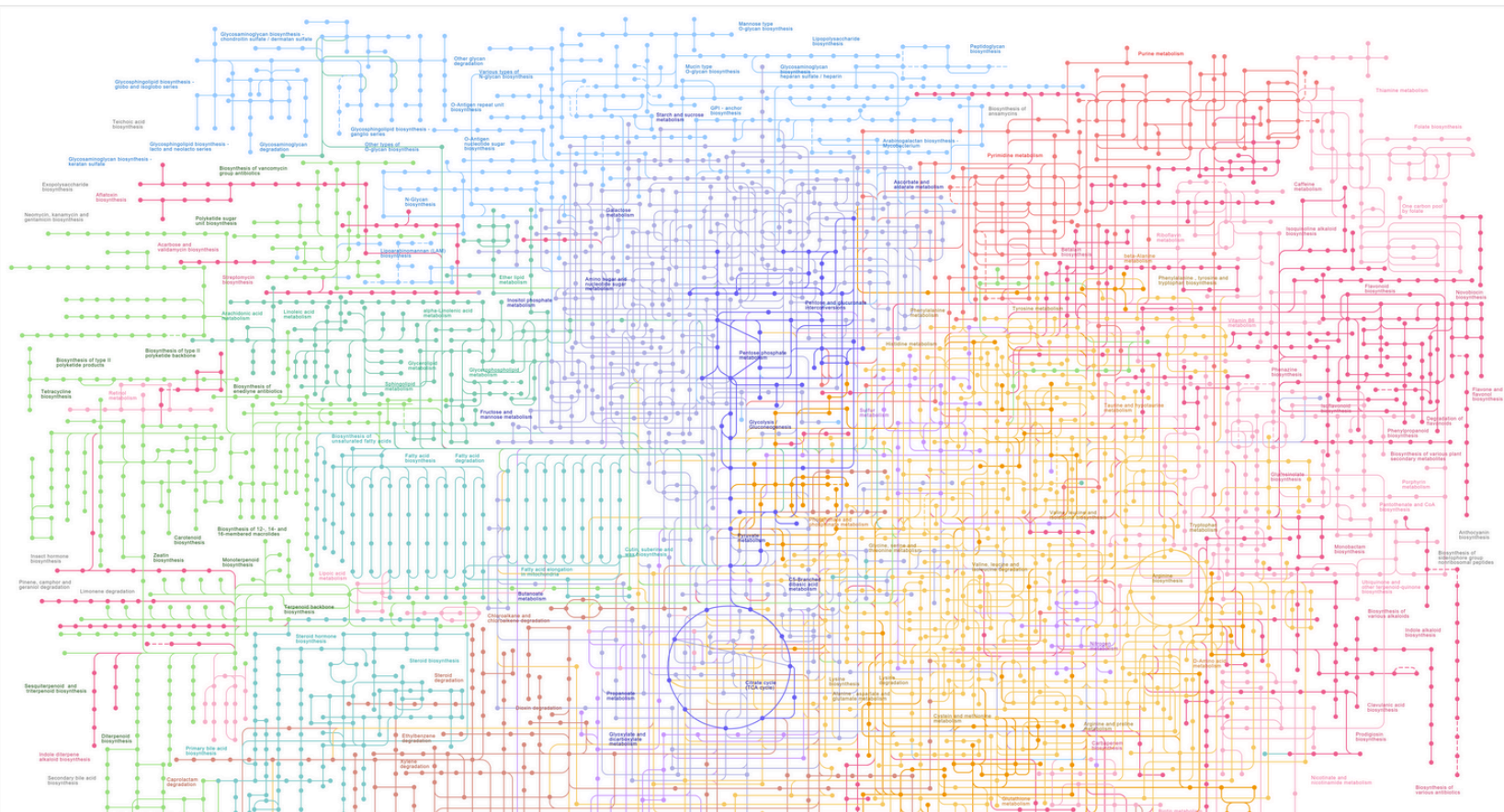
ID search

Color

Module

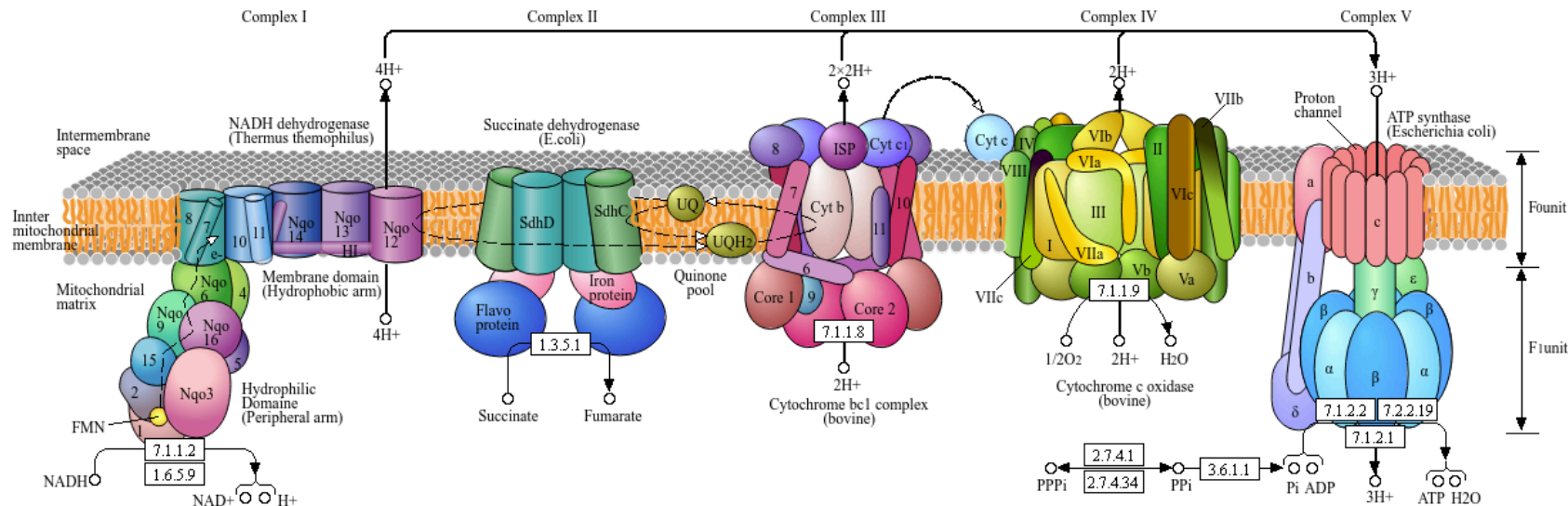
Pathway modules

- Carbohydrate metabolism
- Central carbohydrate met
 - M00001 Glycolysis (Em
 - M00002 Glycolysis, con
 - M00003 Gluconeogenes
 - M00307 Pyruvate oxida
 - M00009 Citrate cycle (T
 - M00010 Citrate cycle, f
 - M00011 Citrate cycle, s
 - M00004 Pentose phosph
 - M00006 Pentose phosph
 - M00007 Pentose phosph
 - M00580 Pentose phosph
 - M00005 PRPP biosynthr
 - M00008 Entner-Doudor
 - M00308 Semi-phospho
 - M00633 Semi-phospho
 - M00309 Non-phosphor
- Other carbohydrate meta



<https://www.genome.jp/pathway/map00190>

OXIDATIVE PHOSPHORYLATION



NADH dehydrogenase

E	ND1	ND2	ND3	ND4	ND4L	ND5	ND6										
E	Ndufs1	Ndufs2	Ndufs3	Ndufs4	Ndufs5	Ndufs6	Ndufs7	Ndufs8	Ndufv1	Ndufv2	Ndufv3						
B/A	NuoA	NuoB	NuoC	NuoD	NuoE	NuoF	NuoG	NuoH	NuoI	NuoJ	NuoK	NuoL	NuoM	NuoN			
E/B	NdhC	NdhK	NdhJ	NdhH	NdhA	NdhI	NdhG	NdhE	NdhF	NdhD	NdhB	NdhL	NdhM	NdhN	HoxE	HoxF	HoxU
E	Ndufa1	Ndufa2	Ndufa3	Ndufa4	Ndufa5	Ndufa6	Ndufa7	Ndufa8	Ndufa9	Ndufa10	Ndufab1	Ndufa11	Ndufa12	Ndufa13			
E	Ndubf1	Ndubf2	Ndubf3	Ndubf4	Ndubf5	Ndubf6	Ndubf7	Ndubf8	Ndubf9	Ndubf10	Ndubf11	Ndufc1	Ndufc2				

Succinate dehydrogenase / Fumarate reductase

E	SDHC	SDHD	SDHA	SDHB				
B/A	SdhC	SdhD	SdhA	SdhB	FrdA	FrdB	FrdC	FrdD

Cytochrome c reductase

E/B/A	ISP	Cyt b	Cyt 1				
E	COR1	QCR2	QCR6	QCR7	QCR8	QCR9	QCR10

Cytochrome c oxidase

E	COX10	COX3	COX1	COX2	COX4	COX5A	COX5B	COX6A	COX6B	COX6C	COX7A	COX7B	COX7C	COX8	E/B/A	COX11	COX15	COX17
B/A	CyoE	CyoD	CyoC	CyoB	CyoA	CoxD	CoxC	CoxA	CoxB	QoxD	QoxC	QoxB	QoxA	SoxD	SoxC	SoxB	SoxA	

Cytochrome c oxidase, cbb3-type

B	I	II	IV	III
---	---	----	----	-----

Cytochrome bd complex

B/A	CydA	CydB	CydX
-----	------	------	------

Cytochrome c

CYC

F-type ATPase (Bacteria)

alpha	beta	gamma	delta	epsilon
a	b	c		

F-type ATPase (Eukaryotes)

alpha	beta	gamma	delta	epsilon	
OSCP	a	b	c	d	e
f	g	f6/h	j	k	g

V/A-type ATPase (Bacteria, Archaea)

A	B	C	D	E	F	G/H
I	K					

V-type ATPase (Eukaryotes)

A	B	C	D	E	F	G	H
a	c	d	e	S1			

https://www.genome.jp/pathway/map00190



Search for

Database: KEGG - Search term: Estrogen receptor

KEGG PATHWAY

map07226
Progesterone, androgen and estrogen receptor agonists/antagonists

KEGG ORTHOLOGY

K04246
GPER1, GPR30; G-protein coupled estrogen receptor 1

K08550
ESR1, NR3A1; estrogen receptor alpha

K08551
ESR2, NR3A2; estrogen receptor beta

K08552
NR3B1, ESRR1; estrogen-related receptor alpha

K08553
NR3B2, ESRR2; estrogen-related receptor beta

... » display all

KEGG GENES

hsa:9687
no KO assigned | (RefSeq) GREB1; growth regulating estrogen receptor binding 1

hsa:2852
K04246 G-protein coupled estrogen receptor 1 | (RefSeq) GPER1, CEPR, CMKRL2, DRY12, FEG-1, GPCR-Br, GPER, GPR30, LERGU, LERGU2, LyGPR, mER; G protein-coupled estrogen receptor 1

hsa:2104
K08554 estrogen-related receptor gamma | (RefSeq) ESRRG, ERR-gamma, ERR3, ERRg, ERRgamma, NR3B3; estrogen related receptor gamma

hsa:2101
K08552 estrogen-related receptor alpha | (RefSeq) ESRR1, ERR1, ERRa, ERRalpha, ESRL1, NR3B1; estrogen related receptor alpha

hsa:2103
K08553 estrogen-related receptor beta | (RefSeq) ESRR2, DFNB35, ERR_beta-2, ERR2, ERRb, ERRbeta2, ESRL2, NR3B2; estrogen related receptor beta

... » display all

<https://www.genome.jp/entry/hsa:2099>



Homo sapiens (human): 2099

Help

Entry	2099	CDS	T01001
Symbol	ESR1, ER, ESR, ESRA, ESTRR, Era, NR3A1		
Name	(RefSeq) estrogen receptor 1		
KO	K08550 estrogen receptor alpha		
Organism	hsa Homo sapiens (human)		
Pathway	hsa01522 Endocrine resistance hsa04915 Estrogen signaling pathway hsa04917 Prolactin signaling pathway hsa04919 Thyroid hormone signaling pathway hsa04961 Endocrine and other factor-regulated calcium reabsorption hsa05200 Pathways in cancer hsa05205 Proteoglycans in cancer hsa05207 Chemical carcinogenesis - receptor activation hsa05224 Breast cancer		
Network	nt06210 ERK signaling (cancer) nt06214 PI3K signaling (cancer) nt06227 Nuclear receptor signaling (cancer) nt06270 Breast cancer nt06323 KISS1-GnRH-LH/FSH-E2 signaling		
Element	N00286 Nuclear-initiated estrogen signaling pathway N00287 ESR1-positive to nuclear-initiated estrogen signaling pathway N01350 NNK/NNN to PI3K signaling pathway N01351 E2-ER-RAS-ERK signaling pathway N01352 BPA to RAS-ERK signaling pathway N01353 E2 to RAS-ERK signaling pathway N01364 E2 to nuclear-initiated estrogen signaling pathway		
Disease	H00031 Breast cancer H01730 Myocardial infarction H02061 Estrogen resistance syndrome		
Drug target	Acolbifene hydrochloride: D02758 Afimoxifene: D06551 Alfatradiol: D07121 Amcenestrant: D12145 Arzoxifene hydrochloride: D02993 Bazedoxifene: D03062<JP> Brilanestrant: D11264 Camizestrant: D12049 Chlorotrianisene: D00269		

All links
Ontology (3)
KEGG BRITE (3)
Pathway (9)
KEGG PATHWAY (9)
Network (12)
KEGG NETWORK (12)
Disease (8)
KEGG DISEASE (3)
OMIM (5)
Drug (84)
KEGG DRUG (84)
Genome (1)
KEGG GENOME (1)
Gene (38)
KEGG ORTHOLOGY (1)
RefGene (2)
NCBI-PROTEINID (1)
NCBI-Gene (1)
HGNC (1)
Ensembl (1)
RIKEN BRC-DNA (2)
OC (1)
PRONIT (28)
Protein sequence (58)
UniProt (2)
SWISS-PROT (1)
RefSeq(pep) (55)
DNA sequence (253)
RefSeq(nuc) (55)
GenBank (99)
EMBL (99)
3D Structure (405)
PDB (405)
Protein domain (4)
Pfam (4)
All databases (875)
Download RDF

<https://www.genome.jp/entry/H00031>



DISEASE: Breast cancer

Help

Entry	H00031 Disease
Name	Breast cancer
Description	Breast cancer is the leading cause of cancer death among women worldwide. The vast majority of breast cancers are carcinomas that originate from cells lining the milk-forming ducts of the mammary gland. The molecular subtypes of breast cancer, which are based on the presence or absence of hormone receptors (estrogen and progesterone subtypes) and human epidermal growth factor receptor-2 (HER2), include: hormone receptor positive and HER2 negative (luminal A subtype), hormone receptor positive and HER2 positive (luminal B subtype), hormone receptor negative and HER2 positive (HER2 positive), and hormone receptor negative and HER2 negative (basal-like or triple-negative breast cancers (TNBCs)). Hormone receptor positive breast cancers are largely driven by the estrogen/ER pathway. In HER2 positive breast tumours, HER2 activates the PI3K/AKT and the RAS/RAF/MAPK pathways, and stimulate cell growth, survival and differentiation. In patients suffering from TNBC, the deregulation of various signalling pathways (Notch, Wnt/beta-catenin, and EGFR) have been confirmed.
Category	Cancer
Brite	<p>Human diseases in ICD-11 classification [BR:br08403]</p> <ul style="list-style-type: none"> 02 Neoplasms <ul style="list-style-type: none"> Malignant neoplasms, except primary neoplasms of lymphoid, haematopoie Malignant neoplasms, stated or presumed to be primary, of specified s Malignant neoplasms of breast <ul style="list-style-type: none"> 2C61 Invasive carcinoma of breast <ul style="list-style-type: none"> H00031 Breast cancer <p>Pathway-based classification of diseases [BR:br08402]</p> <ul style="list-style-type: none"> Replication and repair <ul style="list-style-type: none"> nt06506 Double-strand break repair <ul style="list-style-type: none"> H00031 Breast cancer nt06508 Interstrand crosslink repair <ul style="list-style-type: none"> H00031 Breast cancer Signal transduction <ul style="list-style-type: none"> nt06526 MAPK signaling <ul style="list-style-type: none"> H00031 Breast cancer nt06530 PI3K signaling <ul style="list-style-type: none"> H00031 Breast cancer Tumor markers [br08442.html] <ul style="list-style-type: none"> H00031 Cancer-associated carbohydrates [br08441.html] <ul style="list-style-type: none"> H00031 <p>BRITE hierarchy</p>
Disease pathway	hsa05224 Breast cancer

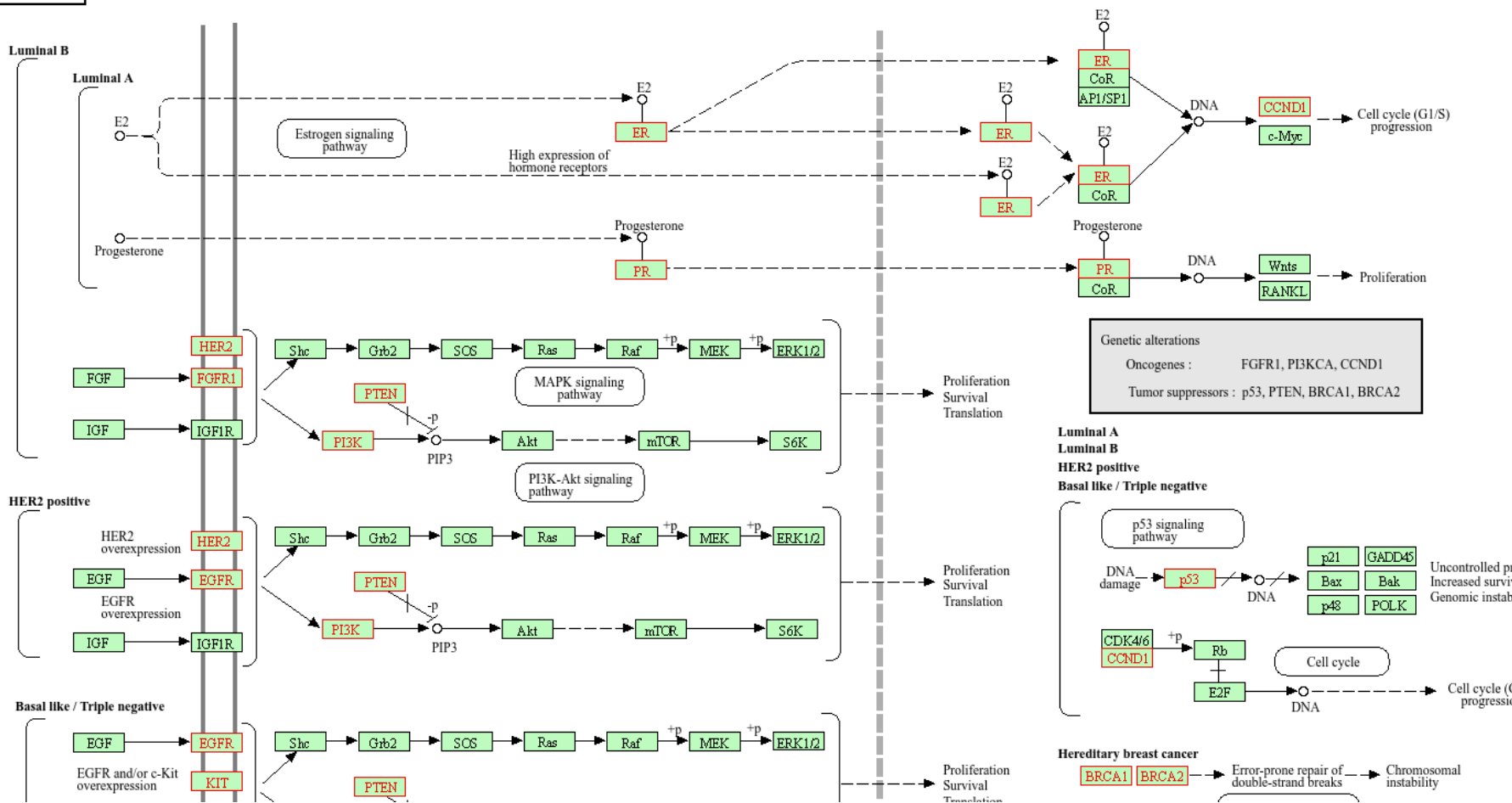
All links
<ul style="list-style-type: none"> Ontology (2) <ul style="list-style-type: none"> KEGG BRITE (2) Pathway (1) <ul style="list-style-type: none"> KEGG PATHWAY (1) Network (26) <ul style="list-style-type: none"> KEGG NETWORK (4) KEGG VARIANT (22) Disease (1) <ul style="list-style-type: none"> OMIM (1) Drug (75) <ul style="list-style-type: none"> KEGG DRUG (75) Gene (56) <ul style="list-style-type: none"> KEGG ORTHOLOGY (28) KEGG GENES (28) Literature (20) <ul style="list-style-type: none"> PubMed (20) All databases (181) <p>Download RDF</p>

<https://www.genome.jp/entry/H00031>

Gene	BRCA1 (germline mutation, hypermethylation) [HSA:672] [KO:K10605] BRCA2 [HSA:675] [KO:K08775] BARD1 [HSA:580] [KO:K10683] BRIP1 [HSA:83990] [KO:K15362] PALB2 [HSA:79728] [KO:K10897] RAD51 [HSA:5888] [KO:K04482] RAD54L [HSA:8438] [KO:K10875] XRCC3 [HSA:7517] [KO:K10880] ERBB2/HER2 (overexpression) [HSA:2064] [KO:K05083] ESR1/ER1 [HSA:2099] [KO:K08550] PGR [HSA:5241] [KO:K08556] GATA3 [HSA:2625] [KO:K17895] PIK3CA [HSA:5290] [KO:K00922] TP53 [HSA:7157] [KO:K04451] PPM1D [HSA:8493] [KO:K10147] RB1CC1 [HSA:9821] [KO:K17589] HMMR [HSA:3161] [KO:K06267] NQO2 [HSA:4835] [KO:K08071] SLC22A18 [HSA:5002] [KO:K08214] PTEN [HSA:5728] [KO:K01110] EGFR (overexpression) [HSA:1956] [KO:K04361] KIT (overexpression) [HSA:3815] [KO:K05091] NOTCH1 (overexpression) [HSA:4851] [KO:K02599] NOTCH4 (overexpression) [HSA:4855] [KO:K20996] FZD7 (overexpression) [HSA:8324] [KO:K02432] LRP6 (overexpression) [HSA:4040] [KO:K03068] FGFR1 (amplification) [HSA:2260] [KO:K04362] CCND1 (amplification) [HSA:595] [KO:K04503]
Drug	Fluoxymesterone [DR:D00327] Methyltestosterone [DR:D00408] Testosterone enanthate [DR:D00958] Cyclophosphamide [DR:D00287] Thiotepa [DR:D00583] Methotrexate sodium [DR:D02115] Gemcitabine hydrochloride [DR:D01155] Capecitabine [DR:D01223] Vinblastine sulfate [DR:D01068] Paclitaxel [DR:D00491] Docetaxel [DR:D07866] Docetaxel [DR:D02165] Doxorubicin hydrochloride [DR:D01275] Epirubicin hydrochloride [DR:D02214] Ixabepilone [DR:D04645] Palbociclib [DR:D10372] (HR positive, HER2 negative) Ribociclib succinate [DR:D10979] (HR positive, HER2 negative) Abemaciclib [DR:D10688] (HR positive, HER2 negative) Everolimus [DR:D02714] (HR positive, HER2 negative) Lapatinib ditosylate [DR:D04024] (HER2 overexpressing) Neratinib maleate [DR:D10898] (HER2 positive) Tucatinib [DR:D11141] (HER2 positive)

<https://www.genome.jp/pathway/hsa05224+H00031>

BREAST CANCER



KEGG pathways



ENZYME: 3.1.3.9

Help

Entry	EC 3.1.3.9	Enzyme
Name	glucose-6-phosphatase; glucose 6-phosphate phosphatase	
Class	Hydrolases; Acting on ester bonds; Phosphoric-monoester hydrolases BRITE hierarchy	
Synsname	D-glucose-6-phosphate phosphohydrolase	
Reaction (IUBMB)	D-glucose 6-phosphate + H ₂ O = D-glucose + phosphate [RN:R00303]	
Reaction (KEGG)	R00303 > R01788 Show all	
Substrate	D-glucose 6-phosphate [CPD:C00092]; H ₂ O [CPD:C00001]	
Product	D-glucose [CPD:C00031]; phosphate [CPD:C00009]	
Comment	Wide distribution in animal tissues. Also catalyses potent transphosphorylations from carbamoyl phosphate, hexose phosphates, diphosphate, phosphoenolpyruvate and nucleoside di- and triphosphates, to D-glucose, D-mannose, 3-methyl-D-glucose or 2-deoxy-D-glucose [cf. EC 2.7.1.62 (phosphoramidate---hexose phosphotransferase), EC 2.7.1.79 (diphosphate---glycerol phosphotransferase) and EC 3.9.1.1 (phosphoamidase)].	
Pathway	ec00010 Glycolysis / Gluconeogenesis ec00052 Galactose metabolism ec00500 Starch and sucrose metabolism ec01100 Metabolic pathways	
Orthology	K01084 glucose-6-phosphatase	
Genes	HSA: 2538(G6PC) 57818(G6PC2) PTR: 741431(G6PC2) MCC: 712053 MMU: 14377(G6pc) RNO: 25634(G6pc) CFA: 403492(G6PC) BTA: 538710(G6PC) SSC: 100134959(G6PC)	

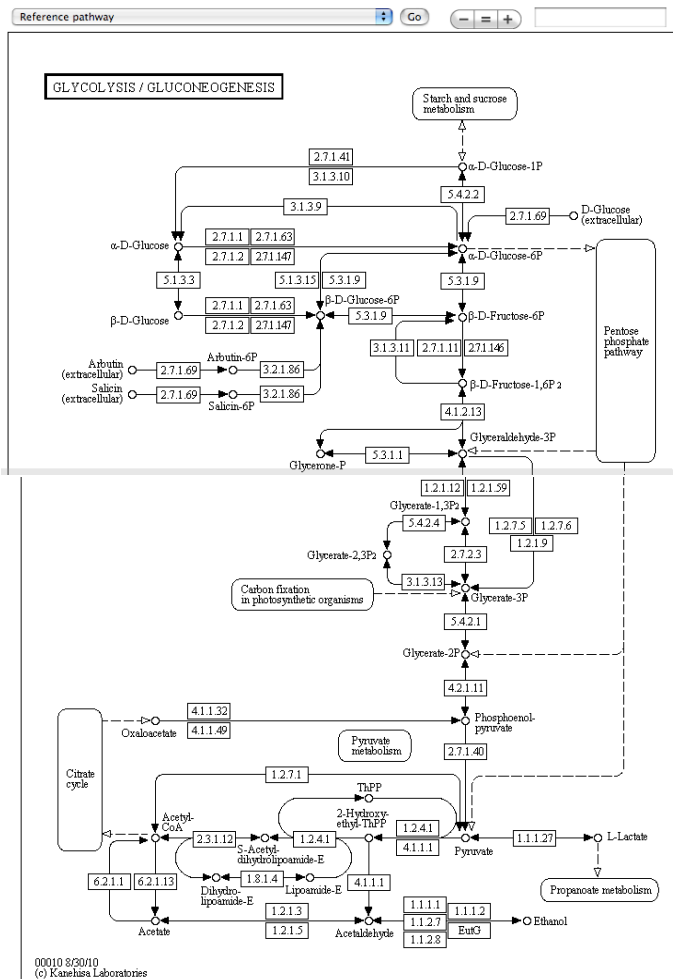
All links

- Ontology (5)
- KEGG BRITE (5)
- Pathway (89)
- KEGG PATHWAY (89)
- Disease (1)
- OMIM (1)
- Chemical substance (6)
- KEGG COMPOUND (6)
- Chemical reaction (12)
- KEGG ENZYME (4)
- KEGG REACTION (2)
- KEGG RPAIR (5)
- KEGG RCLASS (1)
- Genome (2)
- KEGG GENOME (2)
- Gene (58)
- KEGG ORTHOLOGY (1)
- KEGG GENES (20)
- KEGG DGENES (8)
- KEGG EGENES (29)
- Protein sequence (64)
- UniProt (25)
- FRP (3)
- RefSeq(pep) (25)
- PMD (11)
- DNA sequence (50)
- RefSeq(nuc) (30)
- GenBank (10)
- EMBL (10)
- Protein domain (2)
- InterPro (1)
- Pfam (1)
- Literature (3)
- PubMed (3)
- Enzyme (4)
- BRENDA (1)
- EXPASY-ENZYME (1)
- EXPLORZ (1)
- IUBMB (1)
- All databases (296)

KEGG Glycolysis / Gluconeogenesis - Reference pathway

[Pathway menu | Pathway entry | Hide description]

Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (reducing power). It is a central pathway that produces important precursor metabolites: six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glyceraldehyde-3P, glyceralate-3P, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [MD:M00079]. When the enzyme genes of this pathway are examined in completely sequenced genomes, the reaction steps of three-carbon compounds from glyceraldehyde-3P to pyruvate form a conserved core module [MD:M00002], which is found in almost all organisms and which often corresponds to operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway of glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003].



REACTION: R01788

Help

Entry	R01788	Reaction
Name	alpha-D-Glucose 6-phosphate phosphohydrolase	
Definition	alpha-D-Glucose 6-phosphate + H ₂ O <=> alpha-D-Glucose + Orthophosphate	
Equation	C00668 + C00001 <=> C00267 + C00009	
RPair	RP00216 RP05676 RP06709	C00267_C00668 main C00001_C00009 leave C00009_C00668 leave
Enzyme	3.1.3.9	
Pathway	rn00010 Glycolysis / Gluconeogenesis rn00052 Galactose metabolism rn00500 Starch and sucrose metabolism rn01100 Metabolic pathways	
Orthology	K01084 glucose-6-phosphatase [EC:3.1.3.9]	

All links

- Ontology (2)
- KEGG BRITE (2)
- Pathway (8)
- KEGG PATHWAY (8)
- Chemical substance (4)
- KEGG COMPOUND (4)
- Chemical reaction (5)
- KEGG ENZYME (1)
- KEGG RPAIR (3)
- KEGG RCLASS (1)
- Gene (1)
- KEGG ORTHOLOGY (1)
- All databases (20)

OMIM

<https://www.omim.org/entry/114480>

ICD+

External Links

▶ Protein

Clinical Resources

[Clinical Trials](#)
[EuroGentest](#)
[Gene Reviews](#)
[Genetic Alliance](#)
[GTR](#)
[OrphaNet](#)

▶ Animal Models

▶ Cell Lines

#114480

[Table of Contents](#)

114480

BREAST CANCER


Alternative titles; symbols

BREAST CANCER, FAMILIAL

Other entities represented in this entry:

BREAST CANCER, FAMILIAL MALE, INCLUDED

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
1p34.1	{Breast cancer, invasive ductal}	114480	AD , SMu	3	RAD54L	603615
2q33.1	{Breast cancer, protection against}	114480	AD , SMu	3	CASP8	601763
2q35	{Breast cancer, susceptibility to}	114480	AD , SMu	3	BARD1	601593
3q26.32	Breast cancer, somatic	114480		3	PIK3CA	171834
5q34	{Breast cancer, susceptibility to}	114480	AD , SMu	3	HMMR	600936
6p25.2	{?Breast cancer susceptibility} 	114480	AD , SMu	1	NQO2	160998
6q25.1-q25.2	Breast cancer, somatic	114480		3	ESR1	133430
8q11.23	Breast cancer, somatic	114480		3	RB1CC1	606837
11p15.4	Breast cancer, somatic	114480		3	SLC22A1L	602631
11q22.3	{Breast cancer, susceptibility to}	114480	AD , SMu	3	ATM	607585
12p12.1	Breast cancer, somatic	114480		3	KRAS	190070
13q13.1	{Breast cancer, male, susceptibility to}	114480	AD , SMu	3	BRCA2	600185
14q32.33	{Breast cancer, susceptibility to}	114480	AD , SMu	3	XRCC3	600675
14q32.33	Breast cancer, somatic	114480		3	AKT1	164730
15q15.1	{Breast cancer, susceptibility to}	114480	AD , SMu	3	RAD51	179617
16q22.1	Breast cancer, lobular, somatic	114480		3	CDH1	192090
17p13.1	Breast cancer, somatic	114480		3	TP53	191170
17q21.33	{Breast cancer, susceptibility to}	114480	AD , SMu	3	PHB1	176705
17q23.2	Breast cancer, somatic	114480		3	PPM1D	605100

[Title](#)

[Phenotype-Gene Relationships](#)

[Clinical Synopsis](#)

[Text](#)

[Description](#)

[Clinical Features](#)

[Other Features](#)

[Inheritance](#)

[Diagnosis](#)

[Clinical Management](#)

[Mapping](#)

[Cytogenetics](#)

[Molecular Genetics](#)

[Pathogenesis](#)

[Animal Model](#)

[History](#)

[See Also](#)

[References](#)

[Contributors](#)

[Creation Date](#)

[Edit History](#)

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

THE HUMAN PROTEIN ATLAS

[SECTIONS](#) [ABOUT](#) [NEWS](#) [LEARN](#) [DATA](#) [HELP](#)

The open access resource for human proteins

Search for specific genes/proteins or explore the 12 different sections

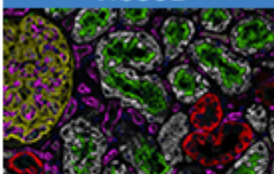
e.g. [ACE2](#), [GFAP](#), [EGFR](#)

Search

[Fields »](#)

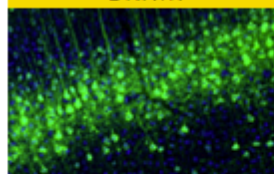
[Search help¹](#)

TISSUE




Protein and RNA profiles in tissues based on antibodies and transcriptomics

BRAIN




Protein and RNA profiles in brain based on microdissected regions

SINGLE CELL TYPE



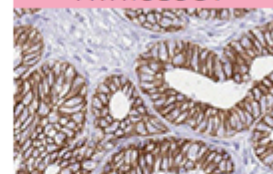
Single cell RNA profiles in tissues based on single cell RNA analysis

TISSUE CELL TYPE



Cell type profiles in tissues based on deconvolution of bulk transcriptomics

PATHOLOGY



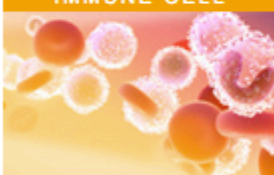
Protein and RNA profiles in human cancers based on antibodies and transcriptomics

DISEASE




Protein levels in blood in patients with diseases

IMMUNE CELL



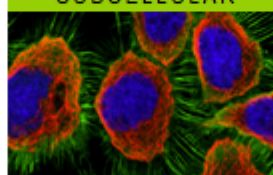
RNA profiles in human immune cells

BLOOD PROTEIN



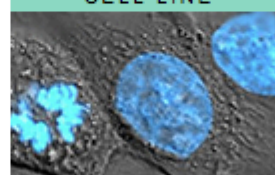
Blood protein levels in healthy individuals and the human secretome

SUBCELLULAR




Spatial, subcellular protein profiles in human cells based on antibodies

CELL LINE




RNA profiles in human cell lines with best models for human cancers

STRUCTURE



3D-structures of human proteins with antibody-binding sites and genetic variants

INTERACTION



Human protein-protein interactions and metabolic enzyme profiles

https://www.proteinatlas.org/ENSG00000091831-ESR1

ESR1

Search [Fields »](#)

Search result (3 genes): [ESR1](#) | [HEY1](#) | [DPH3](#)

ESR1



RNA

BRAIN

RNA

SINGLE CELL

RNA

TISSUE CELL



N/A

DISEASE

RNA

IMMUNE

N/A

BLOOD



RNA

CELL LINE



PROTEIN SUMMARY

SECTION OVERVIEW

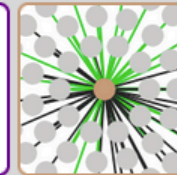
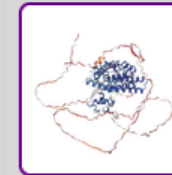
GENE INFORMATION

RNA DATA

ANTIBODY DATA 

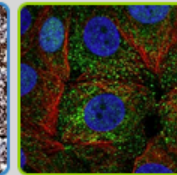
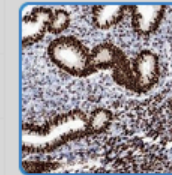
ESR1 INFORMATION

Protein ¹	Estrogen receptor 1
Gene name ¹	ESR1 (ER-alpha, Era, ESR, NR3A1)
Protein class ¹	Cancer-related genes Disease related genes FDA approved drug targets Human disease related genes Nuclear receptors Transcription factors
Protein evidence	Evidence at protein level (all genes)
Number of transcripts ¹	10
Protein interactions	Interacting with 57 proteins



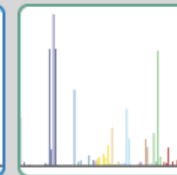
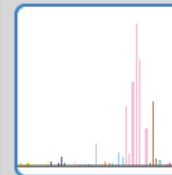
PROTEIN EXPRESSION AND LOCALIZATION

Tissue profile ¹	Selective nuclear expression in female genitalia.
Subcellular location ¹	Localized to the Vesicles In addition localized to the Nucleoplasm
Predicted location ¹	Intracellular



TISSUE RNA EXPRESSION

Tissue specificity ¹	Tissue enhanced (cervix, endometrium, fallopian tube)
Tissue expression cluster ¹	Connective tissue - ECM organization (mainly)
Brain specificity ¹	Low human brain regional specificity
Brain expression cluster ¹	Hypothalamus - Neuropeptide signaling (mainly)
Single cell type specificity ¹	Cell type enhanced (Glandular and luminal cells, Secretory cells, Breast glandular cells, Endometrial stromal cells, Hepatocytes, Early spermatids)
Single cell type expression cluster ¹	Glandular & Luminal cells - Unknown function (mainly)
Tissue cell type classification ¹	Cell type enriched (Liver - Hepatocytes, Testis - Early spermatids, Testis - Late spermatids, Thyroid - Fibroblasts)



ESR1

ESR1

Search [Fields »](#)


Search result (3 genes): [ESR1](#) | [HEY1](#) | [DPH3](#)

[SUMMARY](#) [TISSUE](#) [BRAIN](#) [SINGLE CELL](#) [TISSUE CELL](#) [PATHOLOGY](#) [DISEASE](#) [IMMUNE](#) [BLOOD](#) [SUBCELL](#) [CELL LINE](#) [STRUCTURE](#) [INTERACTION](#)


TISSUE

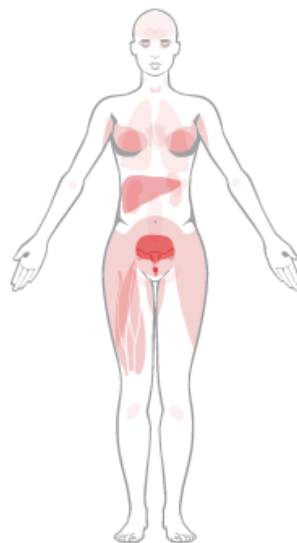
PRIMARY DATA

TISSUES 

ANTIBODIES AND VALIDATION 

Dictionary 

Tissue proteome 



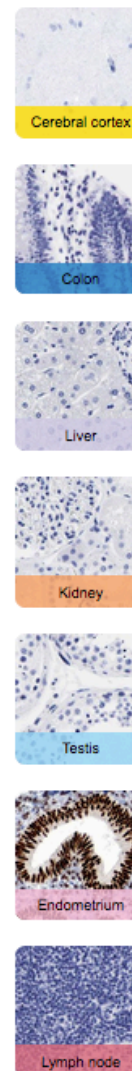
Expression **Detection** **All organs**



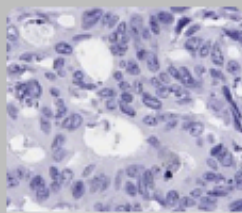
RNA expression (nTPM)ⁱ

Protein expression (score)^j

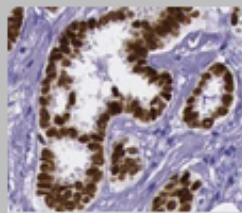
- Brain
- Eye
- Endocrine tissues
- Respiratory system
- Proximal digestive tract
- Gastrointestinal tract
- Liver & Gallbladder
- Pancreas
- Kidney & Urinary bladder
- Male tissues
- Female tissues
- Muscle tissues
- Connective & Soft tissue
- Skin
- Bone marrow & Lymphoid tissues



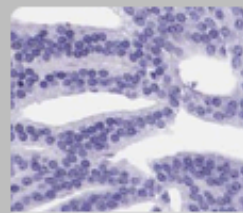
PROTEIN EXPRESSION¹



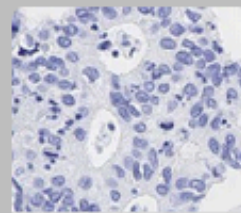
Colorectal cancer



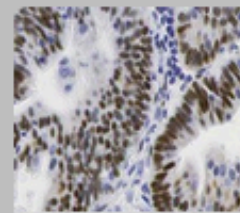
Breast cancer



Prostate cancer



Lung cancer



Endometrial cancer

PROTEIN EXPRESSION SUMMARY¹

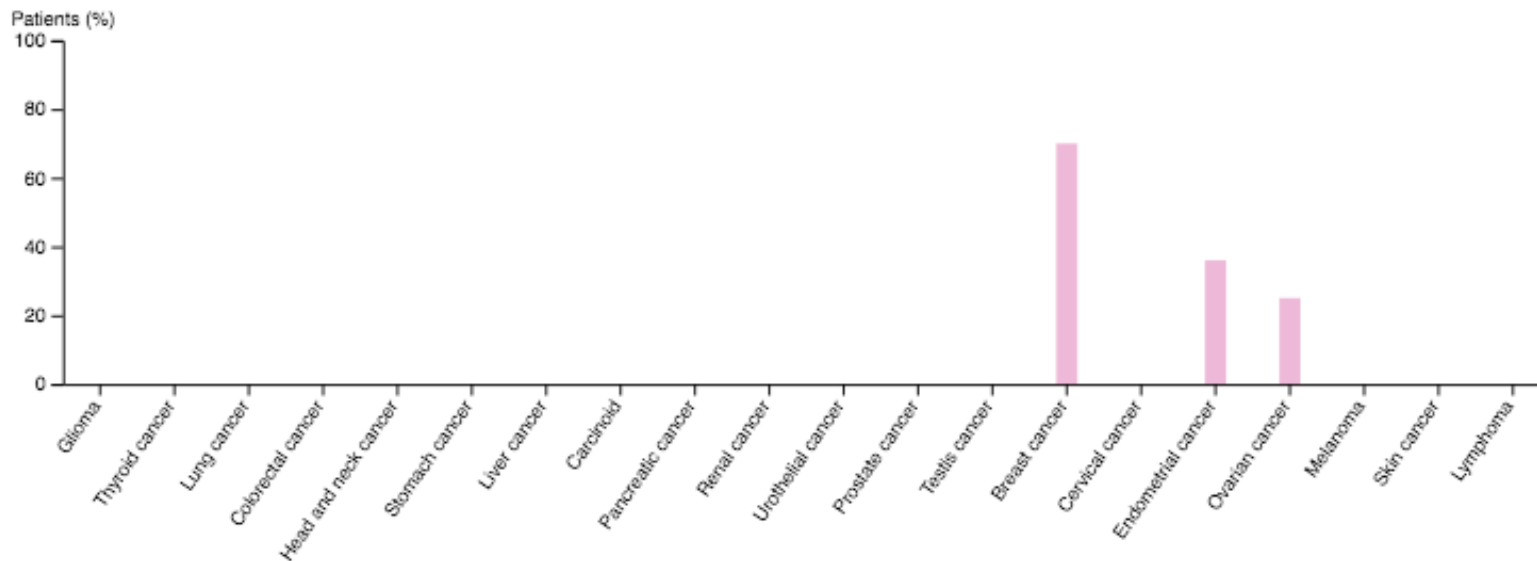
HPA000449 HPA000450 **CAB000037** CAB055099 CAB072858

Organ

Expression

Alphabetical

Breast, ovarian and endometrial cancers displayed strong nuclear positivity. Remaining cancers were negative.



Καμπύλη επιβίωσης Kaplan Meier

THE HUMAN PROTEIN ATLAS

SECTIONS ABOUT NEWS LEARN DATA HELP

ESR1

Search

Fields >

Search result (3 genes): **ESR1** | HEY1 | DPH3

ESR1



PATHOLOGY

CANCER

ANTIBODIES AND VALIDATION

Dictionary

Human pathology

GENERAL INFORMATION¹

Gene name ¹	ESR1
Gene description ¹	Estrogen receptor 1
Protein class ¹	Cancer-related genes Disease related genes FDA approved drug targets Human disease related genes Nuclear receptors Transcription factors
Predicted location ¹	Intracellular
Number of transcripts ¹	10

HUMAN PROTEIN ATLAS INFORMATION¹

RNA category ¹	Tissue enhanced (cervix, endometrium, fallopian tube)
Consensus (human tissue):	Detected in many
Cell line (cancer):	Cancer enhanced (Breast cancer) Detected in some
TCGA (cancer tissue):	Cancer enhanced (breast cancer, endometrial cancer) Detected in many
Protein evidence ¹	Evidence at protein level
Protein expression normal tissue ¹	Selective nuclear expression in female genitalia.

IMMUNOHISTOCHEMISTRY DATA RELIABILITY

Data reliability description ¹	High consistency between antibody staining and RNA expression data.
Reliability score - normal tissues ¹	● Enhanced
Antibodies ¹	HPA000449 , HPA000450 , CAB000037 , CAB055099 , CAB072858

SHOW MORE

PROGNOSTIC SUMMARY¹

Prognostic marker in [endometrial cancer](#) (favorable)

