

Как найти последовательность, кодирующую Ваш белок?

- Как найти последовательность ДНК,
кодирующую Ваш белок:
 - Ссылки из белковых баз данных
 - Прямой поиск в GB
 - Related Sequences
- Как читать запись GenBank:
 - Прокариотическая последовательность
 - Эукариотическая
- Как положить последовательность в GenBank
- UniGene

Поиск последовательности нужной ДНК: по ссылке в Swiss-Prot

Sequence databases	
<input type="text" value="EMBL"/> <input type="text" value="EMBL"/> <input type="text" value="GenBank"/> <input type="text" value="DDBJ"/>	<p>X00588 mRNA. Translation: CAA25240.1. U95089 mRNA. Translation: AAB53063.1. U48722 mRNA. Translation: AAC50802.1. U48723 Genomic DNA. Translation: AAC50804.1. U48724 Genomic RNA. Translation: AAC50796.1. U48725 Genomic RNA. Translation: AAC50797.1. U48726 Genomic RNA. Translation: AAC50798.1. U48727 Genomic RNA. Translation: AAC50799.1. U48728 Genomic RNA. Translation: AAC50800.1. U48729 Genomic RNA. Translation: AAC50801.1. AF288738 Genomic DNA. Translation: AAG35786.1. AF288738 Genomic DNA. Translation: AAG35787.1. AF288738 Genomic DNA. Translation: AAG35788.1. AF288738 Genomic DNA. Translation: AAG35789.1. AF288738 Genomic DNA. Translation: AAG35790.1. AY698024 mRNA. Translation: AAT97979.1. AY588246 Genomic DNA. Translation: AAS83109.1. AF277897 mRNA. Translation: AAK01080.1. AF125253 mRNA. Translation: AAG43240.1. AF125539, AF125538 Genomic DNA. Translation: AAG43243.1. X06370 Genomic DNA. Translation: CAA29668.1. X00663 mRNA. Translation: CAA25282.1. M38425 Genomic DNA. Translation: AAA63171.1. M11234 Genomic DNA. Translation: AAA52370.1.</p>
IPI	<p>IPI00018274. IPI00221346. IPI00221347. IPI00221348.</p>
PIR	<p>GQHUE. A00641.</p>
RefSeq	<p>NP_005219.2. NP_958439.1. NP_958440.1. NP_958441.1.</p>
UniGene	<p>Hs.488293</p>

GenBank

- ✓ Сначала в Los Alamos National Lab, затем (начало 1990х) - в NCBI
- ✓ Включает исходные данные сиквенса последовательностей ДНК
- ✓ Последовательности поставляются напрямую экспериментаторами, которые отсеквенировали и проаннотировали данную последовательность
- ✓ Только эти авторы отвечают за точность данных – записи не курируются
- ✓ Изменить запись впоследствии (исправить или уточнить) может только автор или тот, кому он это позволит – случается очень редко!
- ✓ (=>) Одному локусу ДНК может соответствовать несколько различных записей GenBank

Типы записей нужной ДНК в GB

- Genomic DNA: ген целиком = фланки + UTRs + экзоны + интроны
- mRNA: UTRs + кодирующая последовательность = все экзоны
- CDS: только кодирующая последовательность
- Всевозможные сегменты genomic DNA
- EST: маленькие кусочки транскриптов
- Очень длинные последовательности геномной ДНК, “контиги”, включающие много генов одновременно, вплоть до целого генома или хромосомы

International Nucleotide Sequence Database Collaboration (INSDC)

= GenBank (NCBI) + EBI (EMBL) + DNA Data Bank of Japan (DDBJ)

Обмениваются информацией ежедневно
=> общее содержание

Используют разные форматы для представления данных

RefSeq

- ✓ Курированная база данных NCBI, включающая ДНК, РНК и белковые последовательности
- ✓ Содержит одну запись для одной реальной молекулы
- ✓ Запись включает самую точную и полную аннотации из всех записей GB, соответствующих этой молекуле
- ✓ Записи довольно часто обновляются
- ✓ Широкий, но ограниченный спектр модельных организмов (4 000 из 250 000 организмов в GB)
- ✓ База пополняется за счет других организм-специфичных баз данных
- ✓ Использует тот же формат записи, что и GB

GenBank *versus* RefSeq

GenBank

RefSeq

Not curated

Curated

Author submits

NCBI creates from existing data

Only author can revise

NCBI revises as new data emerge

Multiple records for same loci common

Single records for each molecule of major organisms

Records can contradict each other

No limit to species included

Limited to model organisms

Data exchanged among INSDC members

Exclusive NCBI database

Akin to primary literature

Akin to review articles

Proteins identified and linked

Proteins and transcripts identified and linked

Access via NCBI Nucleotide databases

Access via Nucleotide & Protein databases

Другие типы поиска нужной ДНК в GB

- ✓ По accession number из других источников (статей)
- ✓ По названию белка
- ✓ По названию белка и организму
- ✓ По названию белка в названии GB записи
- ✓ По ключевым словам
- ✓ Related Sequences

GenBank: название белка

The image shows a screenshot of the National Center for Biotechnology Information (NCBI) website. At the top, the NCBI logo is on the left, and the text "National Center for Biotechnology Information" is in the center, with "National Library of Medicine" and "National Institutes of Health" below it. A navigation bar contains links for PubMed, All Databases, BLAST, OMIM, Books, TaxBrowser, and Structure. Below this is a search bar with a dropdown menu set to "All Databases" and a "Go" button. The dropdown menu is open, listing various databases: All Databases, NCBI Web Site, PubMed, Protein, Nucleotide, EST, GSS, Structure, Genome, BioSystems, Books, CancerChromosomes, Conserved Domains, 3D Domains, Gene, Genome Project, dbGaP, GENSAT, and GEO Profiles. The "All Databases" option is highlighted. On the right side of the page, there is a "Hot Spots" section with several links: Clusters of orthologous groups, Coffee Break, Genes & Disease, NCBI Handbook, Electronic PCR, Entrez Home, Entrez Tools, Gene expression omnibus (GEO), Human genome resources, Influenza Virus Resource, Map Viewer, dbMHC, Mouse genome resources, and My NCBI. In the center, there is a section titled "What's new home page!" with a "NEW" tag, followed by a paragraph about the new home page and site guide. Below that is a section titled "What does NCBI do?" with a paragraph describing the center's mission and a link to "More about NCBI...". At the bottom, there is a section titled "NLM/NCBI H1N1 Flu Resources" with a list of links: Newest H1N1 influenza sequences, Submit flu sequences to GenBank, Latest H1N1 citations in PubMed, MedlinePlus (consumer health information), and Enviro-Health links. On the left side, there are several vertical sections: "SITE MAP" with links for Alphabetical Resource and About NCBI; "GenBank" with links for Sequence submission and software; "Literature databases" with links for PubMed, Books, and Central; "Molecular databases" with links for Sequences, structures, and taxonomy; "Genomic biology" with a link for The human genome, whole genomes, and related resources; and "Tools" with a link for Data mining.

Результаты поиска

NCBI Nucleotide

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Nucleotide for Hsp27 Go Clear Save Search

Limits Preview/Index History Clipboard Details

Found 497 nucleotide sequences. Nucleotide [272] EST [225]

Display Summary Show 20 Sort By Send to

All: 272 Bacteria: 0 RefSeq: 106 mRNA: 118

This search in Gene shows 131 results, including:

- [Hsp27](#) (*Drosophila melanogaster*): Heat shock protein 27
- [Hsp27](#) (*Sus scrofa*): heat shock 27kDa protein 1
- [Hspb2](#) (*Mus musculus*): heat shock protein 2

Items 1 - 20 of 272 Page 1 of 14 Next

1: [NW_002198222](#) Reports Links
Taeniopygia guttata chromosome 28 genomic contig, reference assembly (based on Taeniopygia_guttata-3.2.4), whole genome shotgun sequence
gi|212554137|ref|NW_002198222.1|Tgu28_WGA1805_1|[212554137]

2: [NM_053612](#) Reports Order cDNA clone, Links
Rattus norvegicus heat shock protein B8 (Hspb8), mRNA
gi|54262175|ref|NM_053612.2|[54262175]

3: [NM_000488](#) Reports Links
Homo sapiens serpin peptidase inhibitor, clade C (antithrombin), member 1 (SERPINC1), mRNA
gi|254588059|ref|NM_000488.3|[254588059]

4: [NG_008995](#) Reports Links
Homo sapiens heat shock 27kDa protein 1 (HSPB1) on chromosome 7
gi|212549530|ref|NG_008995.1|[212549530]

5: [NM_001540](#) Reports Links
Homo sapiens heat shock 27kDa protein 1 (HSPB1), mRNA

My NCBI [Sign In] [Reg]

Top Organisms [Tree]
Homo sapiens (76)
Oncorhynchus mykiss (36)
Oncorhynchus tshawytscha (34)
Drosophila melanogaster (32)
Rattus norvegicus (17)
All other taxa (77)
More...

Recent Activity Turn Off C
Hsp27 (272) Nucle
Structure and function of 5S rRNA in the ribosome.
The location of mRNA in the ribosomal 30S initiation complex; site-directed cross-linking .
Three widely separated positions in the 16S R lie in or close to the ribosomal decoding ...
> See m

Поиск с указанием полей (Nucleotides)

- ✓ Hsp27 [protein name] – 6 hits
- ✓ Hsp27 [protein name] AND human [organism] – 2 hits
- ✓ Hsp27 [Title] – 58 hits
- ✓ Hsp27 [Title] human [organism] – 5 hits

Название белка может варьировать!

(например, “HSP27” или “Hsp27 ERE-TATA-binding protein”)

Результаты поиска

NCBI Nucleotide

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Nucleotide for Hsp27 [Title] human [organism] Go Clear Save Search

Limits Preview/Index History Clipboard Details

Found 144 nucleotide sequences. Nucleotide [4] EST [140]

Display Summary Show 20 Sort By Send to

All: 4 Bacteria: 0 RefSeq: 0 mRNA: 4

This search in Gene shows 1 result.

SAFB (*Homo sapiens*): scaffold attachment factor B
Chromosome 19, 19p13.3-p13.2, NC_000019.9 (5623163 .. 5668488)
Gene ID: 6294; Other Aliases: DKFZp779C1727, HAP, HET, SAFB1; Other Designations: HSP27 ERE-TATA-binding protein[Hsp27]

Items 1 - 4 of 4 One page.

<input type="checkbox"/> 1: X54079	Reports	Links
Human mRNA for heat shock protein HSP27 gi 32477 emb X54079.1 [32477]		
<input type="checkbox"/> 2: AK095736	Reports	Links
Homo sapiens cDNA FLJ38417 fis, clone FEBRA2009731, highly similar to Human Hsp27 ERE-TATA-binding protein (HET) mRNA gi 21755061 dbj AK095736.1 [21755061]		
<input type="checkbox"/> 3: AB020027	Reports	Links
Homo sapiens HSP27 mRNA, complete cds gi 11036356 dbj AB020027.1 [11036356]		
<input type="checkbox"/> 4: U72355	Reports	Links
Human Hsp27 ERE-TATA-binding protein (HET) mRNA, complete cds gi 2828536 gb U72355.1 [HSU72355][2828536]		

Recent Activity

Turn Off

- Hsp27[Title] AND human[or... (4)
- Hsp27 (272)
- Structure and function of 5S rRNA in the ribosome.
- The location of mRNA in the ribosomal 30S initiation complex; site-directed cross-linking
- Three widely separated positions in the 16S ribosomal RNA gene that lie in or close to the ribosomal decoding ...

» See

Результаты поиска (другой белок)

The screenshot shows a web browser window with the title "dUTPase [protein name] AND human [organism] - Nucleotide Results - Mozilla Firefox". The address bar shows the URL "http://www.ncbi.nlm.nih.gov/sites/entrez". The browser's menu bar includes "Файл", "Правка", "Вид", "Журнал", "Закладки", "Инструменты", and "Справка". The address bar also shows navigation buttons and a search bar with the text "http://www.ncbi.nlm.nih.gov/sites/entrez". Below the browser window, there is a navigation bar with "All Databases", "PubMed", "Nucleotide", "Protein", and "Genome". The search bar contains the text "dUTPase [protein name] AND human [organism]" and buttons for "Go", "Clear", and "Save Search". Below the search bar, there are buttons for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The search results section shows "Found 5 nucleotide sequences. Nucleotide [5]". Below this, there are options for "Display" (Summary), "Show" (20), "Sort by", and "Send to". There are also buttons for "All: 5", "Bacteria: 0", "RefSeq: 0", and "mRNA: 0". The search results list five items, each with a checkbox, a link to the accession number, and a "Reports" link. The items are: 1. AH005568: Homo sapiens dUTPase (DUT) and dUTPase (DUT)s, complete cds; 2. AF018432: Homo sapiens dUTPase (DUT) gene, exons 5, 6, 7, and complete alternatively spliced cds, nuclear gene encoding mitochondrial and nuclear proteins; 3. AF018431: Homo sapiens dUTPase (DUT) gene, exon 4; 4. AF018430: Homo sapiens dUTPase (DUT) gene, exon 3; 5. AF018429: Homo sapiens dUTPase (DUT) gene, exons 1 and 2.

CGCTCAGGATATGACTTCGGCGCTAGAGATCGGATCCCGGATATTATATAGCTCGATCGATCT
TTCTCTATATCCCGGATATGGGATATACACACACAGGCGCGGATAGCATGACTGATCT
CCCCATCT
CACAGACTCTACCGCT

All Databases PubMed Nucleotide Protein Genome

Search Nucleotide for dUTPase [protein name] AND human [organism] Go Clear Save Search

Limits Preview/Index History Clipboard Details

Found 5 nucleotide sequences. Nucleotide [5]

Display Summary Show 20 Sort by Send to

All: 5 Bacteria: 0 RefSeq: 0 mRNA: 0

Items 1 - 5 of 5

1: [AH005568](#) Reports
Homo sapiens dUTPase (DUT) and dUTPase (DUT)s, complete cds
gi|2443579|gb|AH005568.1|SEG_HSDUT[2443579]

2: [AF018432](#) Reports
Homo sapiens dUTPase (DUT) gene, exons 5, 6, 7, and complete alternatively spliced cds, nuclear gene encoding mitochondrial and nuclear proteins
gi|2443578|gb|AF018432.1|HSDUT4[2443578]

3: [AF018431](#) Reports
Homo sapiens dUTPase (DUT) gene, exon 4
gi|2443577|gb|AF018431.1|HSDUT3[2443577]

4: [AF018430](#) Reports
Homo sapiens dUTPase (DUT) gene, exon 3
gi|2443576|gb|AF018430.1|HSDUT2[2443576]

5: [AF018429](#) Reports
Homo sapiens dUTPase (DUT) gene, exons 1 and 2
gi|2443575|gb|AF018429.1|HSDUT1[2443575]

Как добиться исчерпывающего списка?

hsp27 [Title] AND human [organi... NCBI Sequence Viewer v2.0

My NCBI [Sign In] [Register]

es PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

leotide for hsp27 [Title] AND human [organism] Go Clear Save Search

Limits Preview/Index History Clipboard Details

Found 145 nucleotide sequences. CoreNucleotide [5] EST [140]

Display Summary Show 20 Sort by Send to

All: 5 Bacteria: 0 RefSeq: 0 mRNA: 5

Items 1 - 5 of 5 One page.

1: [AK131242](#) Reports
Homo sapiens cDNA FLJ16159 fis, clone BRCAN2002854, weakly similar to Human Hsp27 ERE-TATA-binding protein
[gi|47076974|dbj|AK131242.1|](#)[47076974]

2: [AK095736](#) Reports
Homo sapiens cDNA FLJ38417 fis, clone FEBRA2009731, highly similar to Human Hsp27 ERE-TATA-binding protein
[gi|21755061|dbj|AK095736.1|](#)[21755061]

3: [AB020027](#) Reports
Homo sapiens HSP27 mRNA, complete cds
[gi|11036356|dbj|AB020027.1|](#)[11036356]

4: [U72355](#) Reports
Human Hsp27 ERE-TATA-binding protein (HET) mRNA, complete cds
[gi|2828536|gb|U72355.1|](#)[HSU72355][2828536]

5: [X54079](#) Reports
Human mRNA for heat shock protein HSP27
[gi|32477|emb|X54079.1|](#)[32477]

Links

- Gene
- Gene Genotype
- GeneView in dbSNP
- Probe
- Protein
- PubMed
- Taxonomy
- Related Sequences
- Map Viewer
- OMIM
- GEO Profiles
- UniGene
- UniSTS
- LinkOut

Стандартная запись GB: типы презентации данных

PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for **hsp27 [Title] AND human [organism]** [Save Search](#)

Limits Preview/Index History Clipboard Details

Found 145 nucleotide sequences. CoreNucleotide [5] EST [140]

Display Summary Show 20 Sort by Send to

All: 5 Bacteria: 0 RefSeq: 0 mRNA: 5

Items 1 - 5 of 5 One page.

- 1:** [AK131242](#) Reports Links
Homo sapiens cDNA FLJ16159 fis, clone BRCAN2002854, weakly similar to Human Hsp27 ERE-TATA-binding protein (HET) mRNA
[gi|47076974|dbj|AK131242.1|\[47076974\]](#)
- 2:** [AK095736](#) Reports Links
Homo sapiens cDNA FLJ38417 fis, clone FEBRA2009731, highly similar to Human Hsp27 ERE-TATA-binding protein (HET) mRNA
[gi|21755061|dbj|AK095736.1|\[21755061\]](#)
- 3:** [AB020027](#) Reports Links
Homo sapiens HSP27 mRNA
[gi|11036356|dbj|AB020027.1|\[11036356\]](#)
- 4:** [U72355](#) Reports Links
Human Hsp27 ERE-TATA-binding protein (HET) mRNA, complete cds
[gi|2828536|gb|U72355.1|HS012828536](#)
- 5:** [X54079](#) Reports Links
Human mRNA for heat shock protein HSP27
[gi|32477|emb|X54079.1|\[32477\]](#)

Reports

- ASN.1
- XML
- Summary
- Brief
- FASTA
- TinySeq XML
- GenBank
- INSDSeq XML
- GenBank(Full)
- GI List
- Graphic
- Revision History

Стандартная запись для прокариотического гена

NCBI Nucleotide

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Nucleotide for [] Go Clear

Limits Preview/Index History Clipboard Details

Format: GenBank FASTA Graphics More Formats

Download Save Links

GenBank: X01714.1

E. coli dut gene for dUTPase (EC 3.6.1.23) (deoxyuridine 5'-triphosphate nucleotidohydrolase)

[Comment](#) [Features](#) [Sequence](#)

LOCUS X01714 1609 bp DNA linear BCT 23-OCT-2008

DEFINITION E. coli dut gene for dUTPase (EC 3.6.1.23) (deoxyuridine 5'-triphosphate nucleotidohydrolase).

ACCESSION X01714

VERSION X01714.1 GI:41296

KEYWORDS dUTPase; unidentified reading frame.

SOURCE Escherichia coli

ORGANISM [Escherichia coli](#) ←
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1609)

AUTHORS Lundberg,L.G., Thoresson,H.O., Karlstrom,O.H. and Nyman,P.O.

TITLE Nucleotide sequence of the structural gene for dUTPase of Escherichia coli K-12

JOURNAL EMBO J. 2 (6), 967-971 (1983)

PUBMED [6139280](#) ←

COMMENT Data kindly reviewed (25-NOV-1985) by L. Lundberg.

FEATURES

Location/Qualifiers

source 1..1609
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"

[promoter](#) 286..291
/note="-35 region"

[promoter](#) 310..316
/note="-10 region"

[misc feature](#) 322..324
/note="put. transcription start region"

[RBS](#) 330..333

Change Region Shown

Whole sequence
 Selected Region
from: begin to: end

Update View

Customize View

Basic Features

Default features
 Gene, RNA, and CDS features only

Sequence display options

Show sequence
 Show minus strand

Update View

Sequence Analysis Tools

- ▶ BLAST Sequence
- ▶ Pick Primers

Recent Activity

Turn On Clear

Activity recording is turned off.

[Turn recording back on](#)

All links from this record

Стандартная запись GB: Sequence

```
/protein_id="CAA25860.1"  
/db_xref="GI:41298"  
/db_xref="GOA:P0C093"  
/db_xref="UniProtKB/Swiss-Prot:P0C093"  
/translation="MAEKQTAKRNRREEILQSLALMLESSDGSQRITTAKLAASVGVSE  
EAALYRHFPSKTRMFDLIEFIEDSLITRINLILKDEKDTTARLRLIVLLLLGFGERN  
PGLTRILTGHALMFEQDRLQGRINQLFERIEAQLRQVLREKRMREGEYTTDETLLAS  
QILAFCEGMLSRFVRSEFKYRPTDDFDARWPLIAASCSNMPDDFSSGEFL"
```

ORIGIN

```
1  cagagaaaat  caaaaagcag  gccacgcagg  gtgatgaatt  aacaataaaa  atggttaaaa  
61  accccgatat  cgtcgcaggc  gttgccgcac  taaaagacca  tcgaccctac  gtcgttggat  
121  ttgccgccga  aacaaataat  gtggaagaat  acgcccggca  aaaacgtatc  cgtaaaaacc  
181  ttgatctgat  ctgcgcgaac  gatgtttccc  agccaactca  aggatttaac  agcgacaaca  
241  acgattaca  ccttttctgg  caggacggag  ataaagtctt  accgcttgag  cgaaagagc  
301  tccttggcca  attattactc  gacgagatcg  tgaccctgta  tgatgaaaaa  aatcgacgtt  
361  aagattctgg  acccgcgcgt  tgggaaggaa  tttccgctcc  cgacttatgc  cacctctggc  
421  tctgccggac  ttgacctgcg  tgacctgtct  aacgacgccg  tagaactggc  tccgggtgac  
481  actacgctgg  ttccgaccgg  gctggcgatt  catattgccg  atccttcact  ggcggcaatg  
541  atgctgccgc  gctccggatt  gggacataag  cacggatcgc  tgcttggtaa  cctggtagga  
601  ttgatcgatt  ctgactatca  gggccagtgg  atgatttccg  tgtggaaccg  tggtcaggac  
661  agcttcacca  ttcaacctgg  cgaacgcctc  gccagatga  tttttgttcc  ggtagtacag  
721  gctgaattta  atctggtgga  agatttcgac  gccaccgacc  gcggtgaagg  cggctttggt  
781  cactctggtc  gtcagtaaca  catacgcctc  cgaataacgt  cataacatag  cgcgaaacat  
841  ttcgtttgcg  gtcatagcgt  gggtgccgcc  tggcaagtgc  ttattttcag  gggatatttg  
901  taacatggca  gaaaaacaaa  ctgcgaaaag  gaaccgtcgc  gaggaaatac  ttcagtctct  
961  ggcgctgatg  ctggaatcca  gcgatggaag  ccaacgtatc  acgacggcaa  aactggccgc  
1021  ctctgtcggc  gtttccgaag  cggcactgta  tcgccacttc  cccagtaaga  cccgcatggt  
1081  cgatagcctg  attgagttta  tcgaagatag  cctgattact  cgcatacaacc  tgattctgaa  
1141  agatgagaaa  gacaccacag  cgcgcctgcg  tctgattgtg  ttgctgcttc  tcggttttgg  
1201  tgagcgtaat  cctggcctga  cccgcctcct  cactggtcat  gcgctaattg  ttgaacagga  
1261  tcgcctgcaa  gggcgcctca  accagctggt  cgagcgtatt  gaagcgcagc  tcgcgccagg  
1321  attgctgtaa  aagagaatgc  gtgaggggtg  aggttacacc  accgatgaaa  ccctgctggc  
1381  aagccagatc  ctggccttct  gtgaaggtat  gctgtcacgt  tttgtccgca  gcgaatttaa  
1441  ataccgcccg  acggatgatt  ttgacgcccg  ctggccgcta  attgcccgca  gttgcagtaa  
1501  tatgacgccg  gatgactttt  catccggcga  gtttctttaa  acgccaact  cttcgcgata  
1561  ggccttaacc  gccgccagat  gttccgccat  ttccggcttc  tttccagg
```

//

Стандартная эукариотическая запись GB (mRNA)

GenBank - Поиск в Google NCBI Sequence Viewer v2.0

LOCUS HSU90223 960 bp mRNA linear PRI 03-JAN-1998

DEFINITION Human deoxyuridine triphosphate nucleotidohydrolase precursor mRNA, nuclear gene encoding mitochondrial protein, complete cds.

ACCESSION U90223

VERSION U90223.1 GI:2735291

KEYWORDS .

SOURCE Homo sapiens (human) ←

ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 960)

AUTHORS Ladner, R.D. and Caradonna, S.J.

TITLE The Human dUTPase Gene Encodes Both Nuclear and Mitochondrial Isoforms: Differential Expression of the Isoforms and Characterization of a cDNA Encoding the Mitochondrial Species

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 960) ←

AUTHORS Ladner, R.D. and Caradonna, S.J.

TITLE Direct Submission

JOURNAL Submitted (19-FEB-1997) Dept. of Molecular Biology, Univ. of Med. and Dent. of NJ-School of Osteopathic Medicine, 2 Medical Center Drive, Stratford, NJ 08084, USA

FEATURES

Location/Qualifiers

source 1..960
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

[CDS](#) 63..821
/note="mitochondrial dUTPase isoform; DUT-M"
/codon_start=1
/product="deoxyuridine triphosphate nucleotidohydrolase precursor"
/protein_id="AAB94642.1"
/db_xref="GI:2735292"
/translation="MTPLCPRPALCYHFLTSLLRSAMQNARGTAEGRSRGTLRARPAPRPPAAQHGI PRPLSSAGRLSQGCRGASTVGAAGWKGE LPKAGGSPAPGPETPAISPSK RARPAEVGGMQLRFARLSEHATAPTRGSARAAGYDLYSAYDYTI PPMKAVVKTDIQI ALPSGCYGRVAPRSGLA AKHFIDVGAGVIDEDYRGNVGVVLFNFGKEKEFEVKKGDRIA QLICERIFYPEIEEVQALDDTTERSGGGFGSTGKN"

[sig peptide](#) 63..269
/note="mitochondrial targeting presequence"

[mat peptide](#) 270..818
/product="deoxyuridine triphosphate nucleotidohydrolase"

ORIGIN

1 qqtqaaqcc tqqcqcacqt ccqaaqqtgc cqaqqaccca accaqcccaa actctqaaa

Стандартная эукариотическая запись GB (genomic)

1: [AF018431](#). Reports *Homo sapiens* dUTP...[gi:2443577]

[Features](#) [Sequence](#)

```
LOCUS       HSDUT3                      577 bp    DNA     linear   PRI 28-SEP-1997
DEFINITION  Homo sapiens dUTPase (DUT) gene, exon 4.
ACCESSION  AF018431
VERSION    AF018431.1  GI:2443577
KEYWORDS   .
SEGMENT    3 of 4 ←
SOURCE     Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 577)
  AUTHORS  Pearlman, R. E.
  TITLE    Human genomic nuclear and mitochondria dUTPase gene
  JOURNAL  Unpublished ←
REFERENCE  2 (bases 1 to 577)
  AUTHORS  Pearlman, R. E.
  TITLE    Direct Submission
  JOURNAL  Submitted (11-AUG-1997) Biology, York University, 4700 Keele St.,
            North York, ONT M3J 1P3, Canada
FEATURES   Location/Qualifiers
     source          1..577
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /map="15q15-q21.1"
     gene             order(AF018429.1:<1..1735,AF018430.1:1..1177,1..45,
                    AF018432.1:658..732,AF018432.1:884..954,
                    AF018432.1:1391..>1447)
                    /gene="DUT"
     mRNA            join(AF018429.1:<282..561,AF018429.1:1034..1172,
                    AF018430.1:560..651,1..45,AF018432.1:658..732,
                    AF018432.1:884..954,AF018432.1:1391..>1447)
                    /gene="DUT"
```

Features only

```
FEATURES             Location/Qualifiers
     source            1..577
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /map="15q15-q21.1"
     gene                order(AFO18429.1:<1..1735,AFO18430.1:1..1177,1..45,
                        AFO18432.1:658..732,AFO18432.1:884..954,
                        AFO18432.1:1391..>1447)
                        /gene="DUT"
     mRNA              join(AFO18429.1:<282..561,AFO18429.1:1034..1172,
                        AFO18430.1:560..651,1..45,AFO18432.1:658..732,
                        AFO18432.1:884..954,AFO18432.1:1391..>1447)
                        /gene="DUT"
                        /product="dUTPase"
                        /note="alternatively spliced; encodes mitochondrial form
                        of the protein"
     CDS                join(AFO18429.1:282..561,AFO18429.1:1034..1172,
                        AFO18430.1:560..651,1..45,AFO18432.1:658..732,
                        AFO18432.1:884..954,AFO18432.1:1391..1447)
                        /gene="DUT"
                        /note="DUT-M; alternatively spliced; mitochondrial form of
                        the protein; similar to H. sapiens dUTPase encoded by
                        GenBank Accession Number U90224"
                        /codon_start=1
                        /product="dUTPase"
                        /protein_id="AAB71393.1"
                        /db_xref="GI:2443580"
                        /translation="MTPLCPRPALCYHFLTSLLRSA MQNARGTAEGRSRGTLRARPA P
                        RPPAAQHGI PRPLSSAGRLSQGCRGASTVGAAGWKGELPKAGSPAPGPETPAISPSK
                        RARPAE VGMQLRFARLSEHATAPTRGSARAAGYDLYSAYDYTI PPMKAVVKTDIQI
                        ALPSGCGYGRVAPRSGLA AKHFIDVGAGVIDEDYRGNVGVVLFNFGKEKFEVKKGDRIA
                        QLICERIFYPEIEEVQALDDTERGSGGFSTGKN"
     mRNA              join(AFO18429.1:<1018..1172,AFO18430.1:560..651,1..45,
                        AFO18432.1:658..732,AFO18432.1:884..954,
                        AFO18432.1:1391..>1447)
                        /gene="DUT"
                        /product="dUTPase"
                        /note="alternatively spliced; encodes nuclear form of the
                        protein"
     CDS                join(AFO18429.1:1018..1172,AFO18430.1:560..651,1..45,
                        AFO18432.1:658..732,AFO18432.1:884..954,
                        AFO18432.1:1391..1447)
                        /gene="DUT"
                        /note="DUT-N; alternatively spliced; nuclear form of the
                        protein; similar to H. sapiens dUTPase encoded by GenBank
                        Accession Number U90224"
                        /codon_start=1
                        /product="dUTPase"
                        /protein_id="AAB71394.1"
                        /db_xref="GI:2443581"
                        /translation="MPCSEETPAISPSKRARPAE VGMQLRFARLSEHATAPTRGSAR
                        AAGYDLYSAYDYTI PPMKAVVKTDIQIALPSGCGYGRVAPRSGLA AKHFIDVGAGVID
                        EDYRGNVGVVLFNFGKEKFEVKKGDRIAQLICERIFYPEIEEVQALDDTERGSGGFSGS
                        TGKN"
     exon              1..45
                        /gene="DUT"
                        /number=4
```

Как восстановить последовательность мРНК?

```
mRNA      /gene="DUT"  
          join(AF018429.1:<282..561,AF018429.1:1034..1172,  
          AF018430.1:560..651,1..45,AF018432.1:658..732,  
          AF018432.1:884..954,AF018432.1:1391..>1447)  
          /gene="DUT"  
          /product="dUTPase"  
          /note="alternatively spliced; encodes mitochondrial form  
          of the protein"  
CDS      join(AF018429.1:282..561,AF018429.1:1034..1172,  
          AF018430.1:560..651,1..45,AF018432.1:658..732,
```

Это можно воспринимать, как алгоритм:

1. Взять последовательность с 282 nt (или раньше) по 561 nt из AF018429.1
2. Приписать к ней справа последовательность с 1034 nt по 1172 nt из AF018429.1
3. Добавить к тому, что получилось (снова справа) последовательность с 560 nt по 651 nt из AF018430.1
4. Затем добавить нуклеотиды с 1ого по 45ый этой записи
5. Приписать справа нуклеотиды с 658 по 732 из AF018432
6. ...

Или ... просто нажать на ссылку по надписи mRNA 😊

Представление альтернативно сплайсируемых изоформ данного белка

```

mRNA      join(AF018429.1:<1018..1172,AF018430.1:560..651,1..45,
              AF018432.1:658..732,AF018432.1:884..954,
              AF018432.1:1391..>1447)
              /gene="DUT"
              /product="dUTPase"
              /note="alternatively spliced; encodes nuclear form of the
              protein"
CDS      join(AF018429.1:1018..1172,AF018430.1:560..651,1..45,

```

```

              /gene="DUT"
              /product="dUTPase"
              /note="alternatively spliced; encodes mitochondrial form
              of the protein"
CDS      join(AF018429.1:282..561,AF018429.1:1034..1172,

```

Митохонриальная и ядерная dUTPase мРНК

mRNA	AF018429	AF018430	AF018431	AF018432
Mitochondrial	<282-561 1034-1172	560-651	1-45	658-732 884-954; 1391-1447>
Nuclear	<1018-1172	560-561	1-45	658-732 884-954; 1391-1447>

Как добавить свои данные в GB?

GenBank - Поиск в Google Submit to GenBank NCBI Sequence V

NCBI **Submit to GenBank**

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

NCBI

Submitting Sequence Data to GenBank [Submit now!!](#)

The most important source of new data for GenBank® is direct submissions from scientists. GenBank depends on its contributors to help keep the database as comprehensive, current, and accurate as possible. NCBI provides timely and accurate processing and biological review of new entries and updates to existing entries, and is ready to assist authors who have new data to submit.

[Sequin](#)
Stand-alone sequence submission tool

[BankIt](#)
For quick and simple submissions

[tbl2asn](#)
Command-line sequence submission tool

[dbEST](#)
[dbGSS](#)
[dbSTS](#)
Submit to GenBank divisions

Receiving an Accession Number for your Manuscript

Most journals now expect that DNA and amino acid sequences that appear in articles will be submitted to a sequence database before publication. Soon after submission, you will receive an accession number from the database which you will be able to use in your article to refer to the sequence. Please be aware that it is only necessary to submit the sequence to one database, whichever one is most convenient, without regard for where the sequence may be published. Data exchange between GenBank, EMBL and DDBJ occurs daily. Sequence data submitted in advance of publication can be kept confidential if requested.

[GenBank](#)
Overview of the database

[Search GenBank](#)
Explore the data

[GenBank](#)
Overview of the database

[Search GenBank](#)
Explore the data

Зачем?

- информация в community;
- Журналы требуют это ДО публикации

Долго ли это?

2 рабочих дня

Данные могут быть закрыты до выхода статьи (по запросу)

Что нужно?

Последовательность, ее описание (аннотация), описание источника

<http://www.ncbi.nlm.nih.gov/Genbank/submit.html>

UniGene – база данных кластеризованных транскриптов

UniGene

[Preview/Index](#) [History](#) [Clipboard](#) [Details](#)

UGID:199560 UniGene Hs.371889 *Homo sapiens* ATP1A1 [Order cDNA clone, Links](#)

ATPase, Na⁺/K⁺ transporting, alpha 1 polypeptide (ATP1A1)

SELECTED PROTEIN SIMILARITIES

Comparison of sequences in UniGene with selected protein reference sequences. The alignments can suggest function of a gene.

	Reference Protein	Species	Id(%)	Len(aa)
NP_000692.2	Na ⁺ /K ⁺ -ATPase alpha 1 subunit isoform a proprotein	<i>H. sapiens</i>	100.0	1022
NP_659149.1	Na ⁺ /K ⁺ -ATPase alpha 1 subunit	<i>M. musculus</i>	99.0	1022
NP_036636.1	ATPase, Na ⁺ /K ⁺ transporting, alpha 1 polypeptide	<i>R. norvegicus</i>	99.0	1022
NP_001084064.1	Na ⁺ -K ⁺ -ATPase alpha 1 subunit	<i>X. laevis</i>	97.3	1024
NP_990852.1	ATPase, Na ⁺ /K ⁺ transporting, alpha 1 polypeptide	<i>G. gallus</i>	97.2	1020
NP_571761.1	ATPase, Na ⁺ /K ⁺ transporting, alpha 1a.1 polypeptide	<i>D. rerio</i>	94.9	1027
NP_996247.1	CG5670-PH, isoform H	<i>D. melanogaster</i>	88.5	999
NP_506269.1	EATing: abnormal pharyngeal pumping family member (eat-6)	<i>C. elegans</i>	85.7	994

GENE EXPRESSION

Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

[Expression Profile:](#) View expression levels using UniGene's EST ProfileViewer
[\[Show more entries with profiles like this\]](#)

Note: Highly represented in many libraries

[GEO profiles:](#) Gene expression profiles in the NCBI Gene Expression Omnibus database

MAPPING POSITION

Genomic location specified by transcript mapping, radiation hybrid mapping, genetic mapping or cytogenetic mapping.

Chromosome: 1
 Map position: 1p21

Expression profiles (EST Profile)

Expression profile suggested by analysis of EST counts.

Hs.371889- ATP1A1: ATPase, Na+/K+ transporting, alpha 1 polypeptide

[See Legend](#)

Note: Please mouseover the Tissue criterion to view complete details

Breakdown by Body Sites

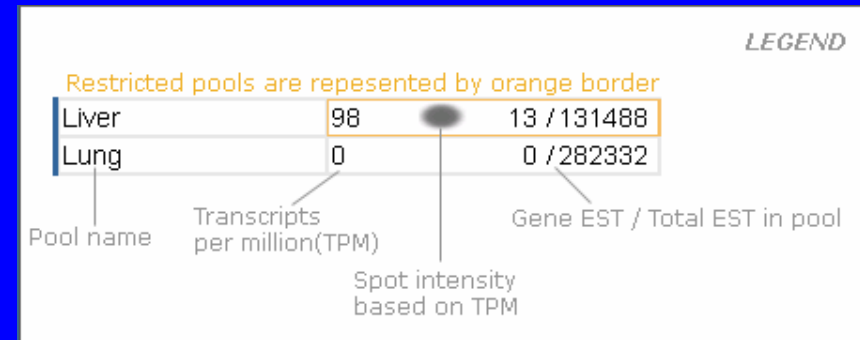


Body Sites

Health State

Developmental Stage

adrenal gland	151	●	5 / 33014
ascites	300	●	12 / 39902
bladder	306	●	9 / 29360
blood	263	●	32 / 121264
bone	420	●	30 / 71411
bone marrow	166	●	8 / 48127
brain	543	●	595 / 1094005
cervix	272	●	13 / 47770
connective tissue	208	●	31 / 148681
ear	0	●	0 / 16068
embryonic tissue	395	●	85 / 214648
esophagus	198	●	4 / 20120
eye	363	●	76 / 209221
heart	113	●	10 / 88288
intestine	1081	●	249 / 230298
kidney	594	●	125 / 210247
larynx	784	●	18 / 22941
liver	277	●	57 / 205712
lung	305	●	102 / 333432
lymph	113	●	5 / 44115
lymph node	154	●	14 / 90644
mammary gland	637	●	95 / 148094



UniGene – последовательности

SEQUENCES

Sequences representing this gene; mRNAs, ESTs, and gene predictions supported by transcribed sequences.

mRNA sequences (11)

BC050359.1	Homo sapiens ATPase, Na+/K+ transporting, alpha 1 polypeptide, mRNA (cDNA clone MGC:51750 IMAGE:6065209), complete cds		PA
NM_000701.6	Homo sapiens ATPase, Na+/K+ transporting, alpha 1 polypeptide (ATP1A1), transcript variant 1, mRNA		PA
NM_001001586.1	Homo sapiens ATPase, Na+/K+ transporting, alpha 1 polypeptide (ATP1A1), transcript variant 2, mRNA		PA
CR622871.1	full-length cDNA clone CS0CAP006YH18 of Thymus of Homo sapiens (human)		P
CR622256.1	full-length cDNA clone CS0DC023YA22 of Neuroblastoma Cot 25-normalized of Homo sapiens (human)		
CR614469.1	full-length cDNA clone CS0DD003YI18 of Neuroblastoma Cot 50-normalized of Homo sapiens (human)		P
AY946013.1	Homo sapiens Na+/K+ transporting ATPase alpha 1 polypeptide (ATP1A1) mRNA, partial cds		P
BC001330.1	Homo sapiens, clone IMAGE:3457513, mRNA, partial cds		PA
BC003077.2	Homo sapiens ATPase, Na+/K+ transporting, alpha 1 polypeptide, mRNA (cDNA clone MGC:3285 IMAGE:3507685), complete cds		PA
U16798.1	Human Na,K-ATPase alpha-1 subunit mRNA, complete cds		P
J03007.1	Human Na+,K+ ATPase alpha-subunit mRNA, partial cds		P

EST sequences (10 of 3153) [[Show all sequences](#)]

AA972529.1	Clone IMAGE:1435477	3' read	A
AA975936.1	Clone IMAGE:1567995	3' read	P
AA984376.1	Clone IMAGE:1629877	3' read	P
AA984433.1	Clone IMAGE:1630006	3' read	P
AA984644.1	Clone IMAGE:1630038	3' read	P
AA987203.1	Clone IMAGE:1603287	3' read	A
AA987775.1	Clone IMAGE:1601681	3' read	PA
AA989683.1	Clone IMAGE:1577568	3' read	A
AA991293.1	Clone IMAGE:1608864	3' read	A
AA994455.1	Clone IMAGE:1628592	3' read	P

[Download Sequences](#)

Key to Symbols

- P** Has similarity to known **P**roteins (after translation)
- A** Contains a poly-**A**denylation signal
- S** Sequence is a **S**uboptimal member of this cluster
- M** Clone is putatively CDS-complete by **MGC** criteria