



Introduction to Bio-Informatics

Lecture 6

Lecture by: Ahmad R. Naghsh-Nilchi, PhD

Department of Artificial Intelligence

Faculty of Computer Engineering

University of Isfahan



In the last session, we discussed:

- Prokaryotes and Eukaryotes are introduced.
- GenBank information on Prokaryotes is explained.
- One Prokaryote gene type: dUPase with Accession code (X01714) is searched on GenBank and information is discussed one by one.



GenBank: X01714.1

[FASTA](#) [Graphics](#)



[Go to:](#) ☐

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

KEYWORDS dUTPase; unidentified reading frame.

SOURCE Escherichia coli

ORGANISM [Escherichia coli](#)

Bacteria; Pseudomonadati; Pseudomonadota; Gammaproteobacteria; Enterobacterales; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1609)

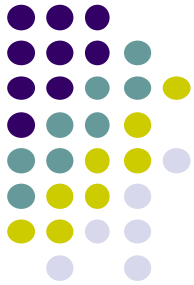
AUTHORS Lundberg,L.G., Thoresson,H.O., Karlstrom,O.H. and Nyman,P.O.

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..1689 /organism="Escherichia coli" /mol_type="genomic DNA" /db_xref="taxon:562"
regulatory	286..294 /regulatory_class="promoter" /note="-35 region"
regulatory	318..318 /regulatory_class="promoter" /note="-10 region"
misc feature	322..324 /note="put. transcription start region"
regulatory	338..338 /regulatory_class="ribosome_binding_site" /note="put. rRNA binding site"
CDS	343..798 /note="unannoted protein product; dbFP-ass (aa 1-451)" /codon_start=1 /transl_table=11 /protein_id="CAA25859.1" /db_xref="GDA:PB6068" /db_xref="InterPro:IPR088188" /db_xref="InterPro:IPR088181" /db_xref="PDB:1J3D" /db_xref="PDB:1OJ8" /db_xref="PDB:1EJ5" /db_xref="PDB:1EJ4" /db_xref="PDB:1J38" /db_xref="PDB:1J37" /db_xref="PDB:1J3H" /db_xref="PDB:1J3L" /db_xref="PDB:2J65" /db_xref="PDB:2J6N" /db_xref="UniProtKB/Swiss-Prot:PB6068" /translation="MKIDAKILDPYGRKFPPLPIYATSGAQLDLACLNDAVELAP GGTTLVPTGLADHLADPSLAAMNLIHSGLIHKGIVLQNLVGLIDSYQQQLMCSYVM HGQSFTIQGERIAQKIFPVVQAEPNLVEDPDATDRGGGFGHGRQ"

Protein Sequence





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 acgcattaca ctttttctgg caggacggag ataaagtctt accgcttgag cgcaaagagc
301 tccttggcca attattactc gacgagatcg tgacccgtta tgatgaaaaa aatcgacgtt
361 aagattctgg acccgcgcgt tgggaaggaa tttccgctcc cgacttatgc cacctctggc
421 tctgccggac ttgacctgcg tgcctgtctc aacgacgccg tagaactggc tccgggtgac
481 actacgctgg ttccgaccgg gctggcgatt catattgccg atccttcaact ggcggcaatg
541 atgctgccgc gctccggatt gggacataag cacggtatcg tgcttggtaa cctggtagga
601 ttgatcgatt ctgactatca gggccagttg atgatttccg tgtggaaccg tggtcaggac
661 agcttcacca ttcaacctgg cgaacgcacg gccagatga tttttgttcc ggtagtacag
721 gctgaattta atctggtgga agatttcgac gccaccgacc gcggtgaagg cggctttggt
781 cactctggtc gtcagtaaca catacgcacg cgaataacgt cataacatag ccgcaaacat
841 ttcgtttgcg gtcatagcgt ggggtgccgc tggcaagtgc ttattttcag gggatatttg
901 taacatggca gaaaaacaaa ctgcgaaaag gaaccgtcgc gaggaaatac ttcagtctct
961 ggcgctgatg ctggaatcca gcgatggaag ccaacgtatc acgacggcaa aactggccgc
1021 ctctgtcggc gtttccgaag cggcactgta tcgccacttc ccagtaaga cccgcatggt
1081 cgatagcctg attgagttta tcgaagatag cctgattact cgcatcaacc tgattctgaa
1141 agatgagaaa gacaccacag cgcgcctgcg tctgattgtg ttgctgcttc tcggttttgg
1201 tgagcgtaat cctggcctga cccgcatcct cactggtcac gcgctaagt ttgaacagga
1261 tcgcctgcaa gggcgcatca accagctggt cgagcgatatt gaagcgagc tgcccgagg
1321 attgcgtgaa aagagaatgc gtgaggggtg aggttacacc accgatgaaa ccctgctggc
1381 aagccagatc ctggccttct gtgaagggtat gctgtcacgt tttgtccgca gcgaatttaa
1441 ataccgcccg acggatgatt ttgacgcccg ctggccgcta attgcggcca gttgcagtaa
1501 tatgacgccg gatgactttt catccggcga gtttctttaa acgccaaact cttcgcgata
1561 ggccttaacc gccgccagat gttccgcat ttccggcttc tcttccagg
```





GenBank Entry of an eukaryotic mRNA



- Continue with the dUPase gene, but an eukaryotic!
- dUPase presents both prokaryotes and eukaryotes.
- Human have it too!
- A simple eukaryote dUPase gene version is U90223.
- Selecting dUPase for both types, allows us to better feel and understands the complexity differences



National Library of Medicine
National Center for Biotechnology Information

[Log in](#)

Nucleotide

Nucleotide

U90223

[Search](#)

[Advanced](#)

[Help](#)

GenBank

Send to

Change region shown

Customize view

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

GenBank: U90223.1

[FASTA](#) [Graphics](#)

[Go to:](#)

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

KEYWORDS

SOURCE Homo sapiens (human)

[Analyze this sequence](#)

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

[Find in this Sequence](#)

[Related information](#)

[Protein](#)



Reading the GenBank header of a Eukaryotic entry



- Although it is related to a human gene, GenBank entry U90223 doesn't look very different from entry X01714, which describes its bacterial homologue.
- The top part of the entry follows the general information keywords order: Locus, Accession, Definition, and Version!
- The Keyword line, which is supposed to list relevant and searchable terms (such as dUPase), is empty for U90223!!!)



What? How that is possible?

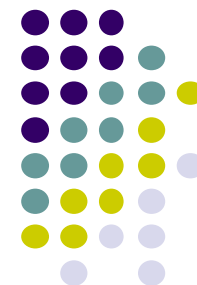
- Unfortunately, this is not an accident!
- It shows a common problem in sequence databases: Annotations may be incomplete!
- Information on databases may be missing or incomplete!
- A word to the wise: You should never expect GenBank (or any other sequence databases) annotations to be up-to-date!



Feature Section of U90223 Entry



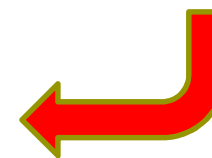
- The CDS keyword indicates a coding region (63-821) sequence that corresponds to the mitochondrial form of human dUTPase.
- Mitochondria are referred to as the "powerhouses" part of the cells of most eukaryotes because this part of the cell generates most of the chemical energy required for cellular functions.
- Following the conceptual amino-acid translation of the open reading frame, ORF, the *sig peptide* keyword indicates the location of a mitochondrial targeting sequence
- The *mat peptide* keyword provides the exact boundaries of the *mature peptide* (A molecule that contains two or more amino acids)



```
FEATURES             Location/Qualifiers
     source            1..968
                        /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"
                        /translation="MTPLCPRPALCYHFLTSLLRSAMQNARGTAEGRSGTLRARPA
                        RPPAAQHGIPLSLSSAGRLSQGCRGASTVGAAGWKGELPKAGGSPAPGPETPAISPSK
                        RARPAEVGGMQLRFARLSEHATAPTRGSARAAGYDLYSAYDYTIIPMEKAVVKTDIQI
                        ALPSGICYGRVAPRSGLAAXHFIDVGAGVIDEDYRGNVGVVLFNFGKEKFEVKKGDRIA
                        QLICERIFYPEIEEVQALDDTERGSGGFGSTGKN"
     sig_peptide       63..269
                        /note="mitochondrial targeting presequence"
     mat_peptide       270..818
                        /product="deoxyuridine triphosphate nucleotidohydrolase"

ORIGIN
      1 ggtggaagcc tggcgcacgt ccggaggtgc cgaggacca accagcccaa actctggggg
     61 aaatgactcc cctctgccct cgccccgcgc tctgtacca ttctctacg tctctgcttc
    121 gctcagcgat gcaaaacgcg cgaggcacgg cagagggccg aagccgcggt actctccggg
    181 ccaggccccg cctcggccg ccggcggcgc agcacgggat tccccggccg ctgtccagcg
    241 ctggccgcct gagccaaggc tgccgaggag ccagtacagt cggggccgct ggctggaagg
    301 gcgagcttcc taaggcgggg ggaagcccg cgccggggcc ggagacaccc gccatttcac
    361 ccagtaagcg ggccggcct gcggaggtgg gcggcatgca gctccgctt gcccggtctt
    421 ccgagcacgc cacggccccc acccggggct ccgcgcgcgc cgcgggctac gacctgtaca
```

Protein Sequence





Not much difference between: X01714 – U90223



- Human GenBank entry is not more complex than its bacterial homologous – Because U90223 is mRNA sequence, not a **genomic** one!
- The relationship between a eukaryotic protein and its nucleotide sequences is as collinear as its prokaryotic counterpart.
- Next, we are showing **genomic data**, which is a bit trickier!



Eukaryotic Genomic Entry on GenBank



- **AH005568**, a GenBank entry related to the gene sequence (as it is on the chromosome) is examined.
- This human genomic entry is mRNA-originated by dUTPase (DUT gene).
- This human genomic entry contains some of the specifications only encountered in eukaryotic entries.



National Library of Medicine
National Center for Biotechnology Information

Log in

Nucleotide

Nucleotide ▼

AH005568

Search

Advanced

Help

GenBank ▼

Send to: ▼

Change region shown

Homo sapiens dUTPase (DUT) gene, complete cds, alternatively spliced

GenBank: AH005568.2

[FASTA](#) [Graphics](#)

Go to: ☑

LOCUS AH005568 5507 bp DNA linear PRI 10-JUN-2016
DEFINITION Homo sapiens dUTPase (DUT) gene, complete cds, alternatively spliced.
ACCESSION AH005568 AF018429 AF018430 AF018431 AF018432
VERSION AH005568.2
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the DUT gene

Multisite phosphorylation dictates selective [Mol Cell. 202

Discovery of two new isoforms



Reading AH005568



- LOCUS: the locus name is AH005568. The rest of this line says it is 5507 bp, and it is Linear DNA. Its last update was in 2016.
- Note that this is a complete version of previous genes developed since 1997.
- DEFINITION: the gene is DUT.
- ACCESSION line is standard.
- VERSION is 2.
- KEYWORDS lines are standard. No Keywords means one can not retrieve this entry using keyword searches.



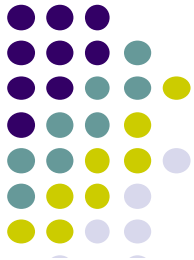
Reading AH005568 Continued ...



- **SOURCE: *Homo sapiens***
 - **ORGANISM: *It*** includes varieties of human organisms, such as Mammalia; Eutheria; Primate, Homo.
- **REFERENCE:**
 - Authors: given
 - Title: given
 - Journal: unpublished
- **COMMENT:** This sequence version replaced many Genomes listed in 2016.



Features Table



- Features table is what makes an eukaryotic genomic entry different!
- It is much longer than the one for prokaryotic organisms.
- Its GenBank page is shown here and on the next slide.
- It contains many elements described as follows.

FEATURES	Location/Qualifiers
source	1..5507 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /map="15q15-q21.1"
gene	<1..>5236 /gene="DUT"
mRNA	join(<282..561,1034..1172,2395..2486,3113..3157, 4447..4521,4673..4743,5180..>5236) /gene="DUT" /product="dUTPase" /note="alternatively spliced; encodes mitochondrial form of the protein"
CDS	join(282..561,1034..1172,2395..2486,3113..3157,4447..4521, 4673..4743,5180..5236) /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" /translation="MTPLCPRPALCYHFLTSLRSAMQNARGTAEGRSRGTLRARPAP RPPAAQHGIPRPLSSAGRLSQGCRGASTVGAAGNKGELPKAGGSPAPGPETPAISPSK RARPAEVGGMQLRFARLSEHATAPTRGSARAAGYDLYSAYDYTIIPMEKAVVKTDIQI ALPSGCYGRVAPRSLAAKHFIDVGAGVIDEDYRGNVGVVLFNFGKEKFEVKKGDRIA QLICERIFYPEIEEVQALDDTERGSGGFGSTGKN"



Features Table Continued ...



exon <282..561
/gene="DUT"
/note="exon used only in the mitochondrial form of the protein"
/number=1

protein_bind 898..903
/gene="DUT"
/bound_moiety="SP-1"

protein_bind 935..943
/gene="DUT"
/bound_moiety="E2F"

mRNA join(<1018..1172,2395..2486,3113..3157,4447..4521,4673..4743,5180..5236)
/gene="DUT"
/product="dUTPase"
/note="alternatively spliced; encodes nuclear form of the protein"

CDS join(1018..1172,2395..2486,3113..3157,4447..4521,4673..4743,5180..5236)
/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"
/translation="MPCSEETPAISPSKRARPAEVGGMQLRFARLSEHATAPTRGSAR AAGYDLYSAYDYTIIPMEKAVVKTDIQIALPSGCYGRVAPRSGLAAKHFIDVGAGVID EDYRGNVGVVLFNFGKEKFEVKKGDRIAQLICERIFYPEIEEVQALDDTERGSGGFGS TGKN"

exon <1018..1172
/gene="DUT"
/note="alternative exon used in nuclear form of the protein"
/number=2

exon 1034..1172
/gene="DUT"
/note="alternative exon used in mitochondrial form of the protein"
/number=2

gap 1736..1835
/estimated_length=unknown

exon 2395..2486
/gene="DUT"
/number=3

gap 3013..3112
/estimated_length=unknown

exon 3113..3157
/gene="DUT"
/number=4

gap 3690..3789
/estimated_length=unknown

exon 4447..4521
/gene="DUT"
/number=5

exon 4673..4743
/gene="DUT"
/number=6

exon 5180..5236
/gene="DUT"
/number=7



AH005568 Features Table



- Features table element includes source section.
- The source section contains a /map section.
- For this genome, it indicates that the sequence belongs to chromosome 15, and is mapped on the long arm (q) of this chromosome, within the q21.1 cytogenetic band.
- The mRNA: The AH005568 entry has two mRNA fields, which are interesting.
- This is a common property of higher eukaryotic gene expression.



- These two mRNA include type 1 (for mitochondria) and type 2 (for the nucleus).
- The type 1 contains the first exon, but type 2 does not.
- Type 2 uses a part of the second exon in a different reading frame from its type 1 counterpart.
- The two protein sequences become identical on the third exon.
- The exon section indicates the position of the sole exon present in this sequence.
- The page ends with ORIGIN, which is the genomic sequence of the entry.



Exon in the Feature table



- The term "exon" holds significant importance in the context of gene structure and function.
- Exons are defined as the segments of a gene that are transcribed into messenger RNA (mRNA) and subsequently translated into proteins.
- In eukaryotic organisms, genes are often composed of both exons and introns.
- The latter are non-coding regions merged out during the post-transcriptional processing of mRNA.
- The presence of exons enables the coding sequences required for protein synthesis, serving as the blueprint for the biological activities essential for life.



Working with Related Entries



- So far, we know the accession number for the entries we wanted to study: X01714, U90223, HA005568.
- Normally, we get these numbers by reading articles mentioning them.
- A good strategy is to start with an accession number of a known genome entry out of these articles, and continue exploring by clicking the links of articles, related information, etc. on the right side of the GenBank page.



Retrieving without Accession Number



- Searching GenBank by using gene or protein keywords rather than accession numbers is possible.
- Suppose you want to find the nucleotide sequence encoding the human dUTPase, without accession number.
- Type human [organism] and dUTPase (protein name) in the search window.
- The next slide shows the search results. It gives 6752 outputs, the first three are shown.
- You may see the accession number HA005568.2 as the second result.



Release date
Custom range...

Revision date
Custom range...

[Clear all](#)
[Show additional filters](#)

Items: 1 to 20 of 6752

<< First < Prev Page 1 of 338 Next > Last >>

- ☐ 1. [Human betaherpesvirus 6A tp/cap, mdbp, U42 orf, U43 orf, U44 orf, and dUTPase genes](#)
10,601 bp linear DNA
Accession: X92436.1 GI: 1044869
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ 2. [Homo sapiens dUTPase \(DUT\) gene, complete cds, alternatively spliced](#)
5,507 bp linear DNA
Accession: AH005568.2 GI: 1036029938
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ 3. [Mus musculus endogenous retroviral sequence MuERV-L gag, pol and dUTPase genes](#)
6,471 bp linear DNA
Accession: Y12713.1 GI: 2065208
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Search details

("Homo sapiens"
[Organism] OR human[All
Fields]) AND dUTPase[All
Fields]

Search

[See more...](#)

Recent activity

[Turn Off](#) [Clear](#)

- human dUTPase (6752)
Nucleotide
- RefSeq RNA Links for Gene
(Select 34529) (4) Nucleotide
- human UTPase (4)
Nucleotide
- human duppies accession
(513353) Nucleotide
- human dUTPase (6752)
Nucleotide

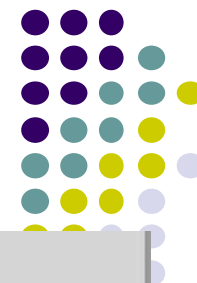
[See more...](#)



Using a Gene-Centric Database



- Besides GenBank, NCBI has developed other databases more adapted to gene-centric searches.
- Making a gene-centered query involves asking a question that relates directly to a specific gene, rather than going through all known pieces of sequences related to that gene.
- Main advantage: it returns results that are more synthetic than a long list of GenBank entries.
- For that search NCBI “**Gene database**” rather than GenBank.



Gene [Advanced](#) [Help](#)

[Full Report](#) [Send to:](#) [Hide sidebar >>](#)

DUT deoxyuridine triphosphatase [*Homo sapiens* (human)]

Gene ID: 1854, updated on 10-Dec-2024

[Download Datasets](#)

Table of contents

[Summary](#)

[Genomic context](#)

[Genomic regions, transcripts, and products](#)

[Expression](#)

[Bibliography](#)

[Phenotypes](#)

[Variation](#)

[HIV-1 interactions](#)

[Pathways from PubChem](#)

[Interactions](#)

[General gene information](#)

[Markers, Related pseudogene\(s\), Clone Names, Homology, Gene Ontology](#)

[General protein information](#)

[NCBI Reference Sequences \(RefSeq\)](#)

[Related sequences](#)

Summary

Official Symbol	DUT <small>provided by HGNC</small>
Official Full Name	deoxyuridine triphosphatase <small>provided by HGNC</small>
Primary source	HGNC:HGNC:3078
See related	Ensembl:ENSG00000128951 MIM:601266 AllianceGenome:HGNC:3078
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	BMFDMS; dUTPase
Summary	This gene encodes an essential enzyme of nucleotide metabolism. The encoded protein forms a ubiquitous, homotetrameric enzyme that hydrolyzes dUTP to dUMP and pyrophosphate. This reaction serves two cellular purposes: providing a precursor (dUMP) for the synthesis of thymine nucleotides needed for DNA replication, and limiting intracellular pools of dUTP. Elevated levels of dUTP lead to increased incorporation of uracil into DNA, which induces extensive excision repair mediated by uracil glycosylase. This repair process, resulting in the removal and reincorporation of dUTP is self-defeating and leads to DNA fragmentation and



mitochondrion or nucleus. A related pseudogene is located on chromosome 19. [provided by RefSeq, Jul 2008]

Expression Ubiquitous expression in bone marrow (RPKM 16.5), lymph node (RPKM 14.4) and 25 other tissues [See more](#)

Orthologs [mouse](#) [all](#)

NEW

Try the new [Gene table](#)

Try the new [Transcript table](#)

Additional links

Genome Browsers

[Genome Data Viewer](#)

[Variation Viewer \(GRCh37.p13\)](#)

[Variation Viewer \(GRCh38\)](#)

[Ensembl](#)

[UCSC](#)

Genomic context



Location: 15q21.1

See DUT in [Genome Data Viewer](#)

Exon count: 8

Annotation release	Status	Assembly	Chr	Location
RS_2024_08	current	GRCh38.p14 (GCF_000001405.40)	15	NC_000015.10 (48331095..48343373)
RS_2024_08	current	T2T-CHM13v2.0 (GCF_009914755.1)	15	NC_060939.1 (46139356..46151634)
RS_2024_09	previous assembly	GRCh37.p13 (GCF_000001405.25)	15	NC_000015.9 (48623292..48635570)

Related information

[3D structures](#)

[BioAssay by Target \(List\)](#)

[BioAssay by Target \(Summary\)](#)

[BioAssay, by Gene target](#)

[BioAssays, RNAi Target, Active](#)

[BioAssays, RNAi Target, Tested](#)

[BioProjects](#)

[BioSystems](#)

[CCDS](#)

[ClinVar](#)

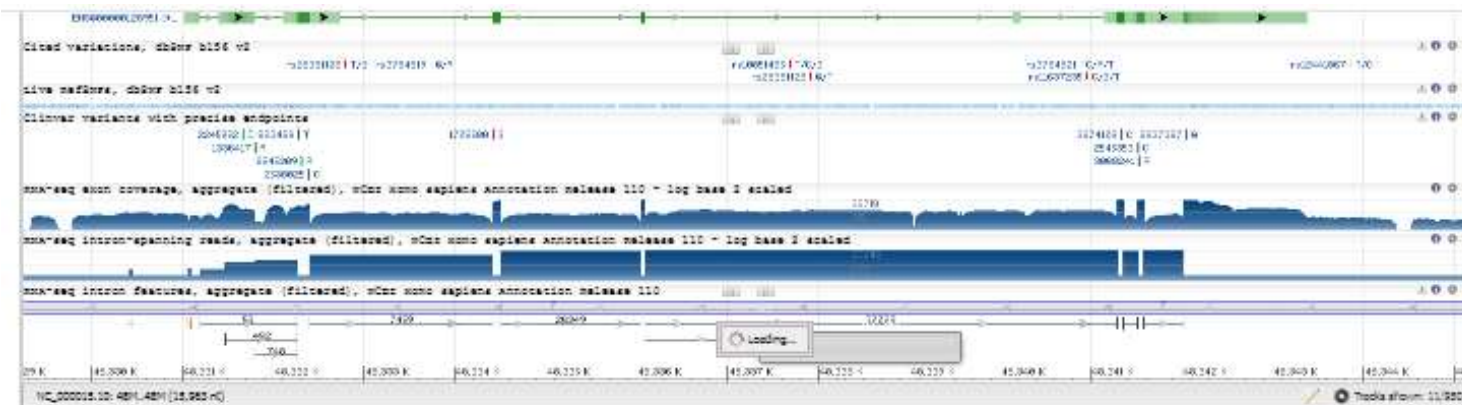
[Conserved Domains](#)

Chromosome 15 - NC_000015.10





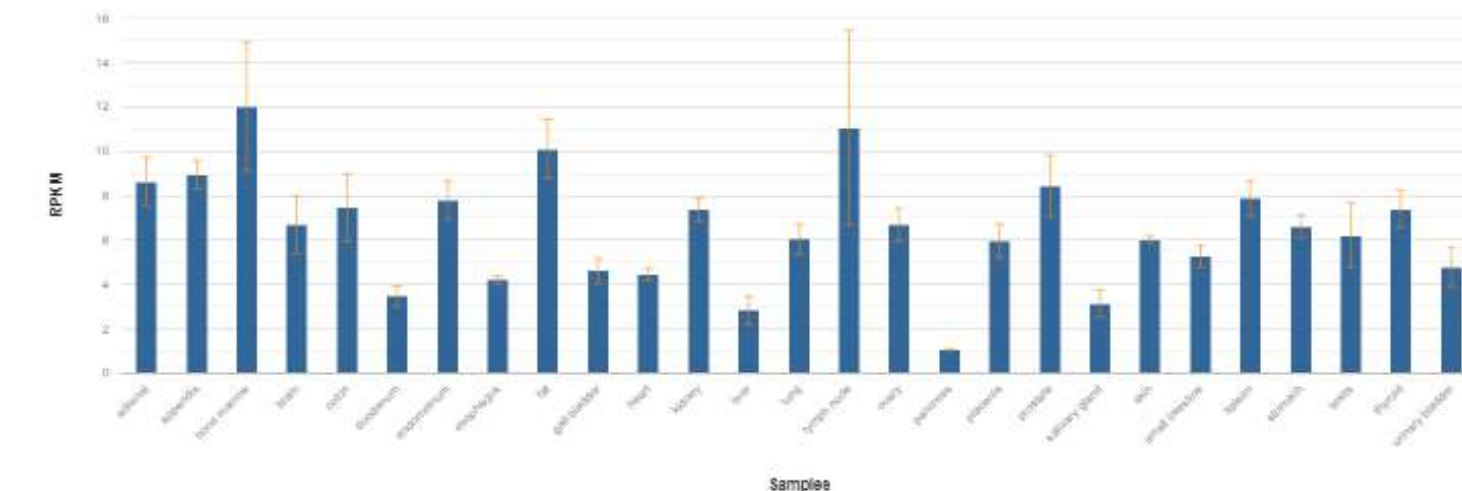
RefSeq Proteins



Expression

HPA RNA-seq normal tissues

- Project title: HPA RNA-seq normal tissues
- Description: RNA-seq was performed of tissue samples from 95 human individuals representing 27 different tissues in order to determine tissue-specificity of all protein-coding genes
- BioProject: [PRJEB4337](#)
- Publication: [PMID 24309898](#)
- Analysis date: Wed Apr 4 07:08:55 2018



MedGen
Nucleotide
OMIM
Probe
Protein
PubChem Compound
PubChem Substance
PubMed
PubMed (GeneRIF)
PubMed (OMIM)
PubMed(nucleotide/PMC)
RefSeq Proteins
RefSeq RNAs
RefSeqGene
Related gene-specific medical variations
SNP
Taxonomy
Variation Viewer

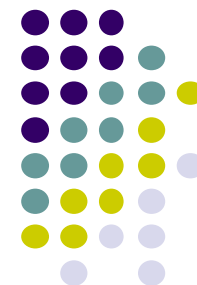
Links to other resources
HGNC
Ensembl
AllianceGenome
AceView
PharmGKB
MGC
HuGE Navigator
KEGG



Guanine (G) and Cytosine (C)

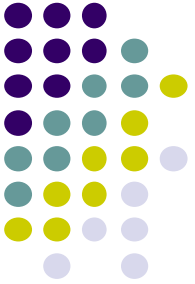


- "GC" typically refers to the GC content of a DNA sequence, which is the proportion of guanine (G) and cytosine (C) bases in the DNA.
- High GC content can influence the stability and structure of DNA, as well as affect gene expression



تکلیف دوم

- یک ارگانسیم پروکاریوتی و یک ارگانسیم یوکاریوتی را انتخاب کنید بطوریکه هر دو متعلق به یک منشاء باشند. با استفاده از GenBank ، دنباله نوکلئوتیدی یک ژن خاص از هر موجود زنده را پیدا و تجزیه و تحلیل کنید. برای موجودات انتخابی خود، موارد زیر را انجام دهید:
- نام علمی و سلسله مراتب طبقه بندی هر دو موجود زنده را مشخص کنید.
- توالی نوکلئوتیدی ژن انتخاب شده را از GenBank بازیابی کنید و Accession number ها را برای مرجع ارائه دهید.
- طول و ویژگی های قابل توجه (مانند اینترون / اگزون در ژن یوکاریوتی) توالی های به دست آمده را مقایسه کنید.
- در مورد اهمیت بیولوژیکی ژن ها در رابطه با ارگانسیم های مربوطه بحث کنید.



Thank you

for your attention