

Nucleotide sequence database: DNA and RNA

23/05/2024

Phuc-Loi Luu, PhD

Luu.p.loi@googlemail.com

Content

- Nucleotide [<https://www.ncbi.nlm.nih.gov/nucleotide/>]
- Gene [<https://www.ncbi.nlm.nih.gov/gene>]

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

NCBI Resources ▾ How To ▾ Sign in to NCBI

Nucleotide Nucleotide ▾ Search Advanced Help



Nucleotide

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

Using Nucleotide

[Quick Start Guide](#)

[FAQ](#)

[Help](#)

[GenBank FTP](#)

[RefSeq FTP](#)

Nucleotide Tools

[Submit to GenBank](#)

[LinkOut](#)

[E-Utilities](#)

[BLAST](#)

[Batch Entrez](#)

Other Resources

[GenBank Home](#)

[RefSeq Home](#)

[Gene Home](#)

[SRA Home](#)

[INSDC](#)

Where possible, the sequences are annotated so that you can find the strings of sequences that may be functional.

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

2281 ccttcgcttg aggggggtg attgctaatc acttcacata aaaattgaag tgtaaacaca
2341 cttaaagatt ttttttattta taataatattt aaaaacaaact ctatgcacc attaatttat
2401 ccatcaccat ggcaattagg aaaaccacgt ggaagcatca aagatactat taacaaaaaa
2461 acaggaaccg ttacatcgta tagtgtcaca ttatccccac caaatgctca accaattact
2521 aaattctct ctatttctga atatggatcc caaacaatg caaaaaaaagc atcagaaaaat
2581 tggagacaaa caatgtcaga tcaatacaat ctaactcgca accaaatccg atattttagac
2641 aaggatataca ttgaagtaca attaacgca gggaaaacga tgaaaaactga cgcgaaatac
2701 atagatgaag tacaaaaata tccaataaat ataaaaaaca agaaatcgaa atccgatcca
2761 tctaaagata gacattatgc gatgtgtcaa aacattaaag aagtgttcc atttgcgat
2821 ttaatatgca aatatgatata aatagaatata attaatggag attctttgga tctaaggaaa
2881 gaaaaatataa aagaatttgg ttcaatggaa acaaataata tcaaagttagg taaaaatata
2941 gttcatttga cagataatga cattaatata cagttatggaa gttttacaaa agatattgac
3001 caacttcctg taaatatttgc gttacgcgtt aaaccagctg gaagtgttatt caaacgaaaca
3061 gaaaaatgaga atatataat tgcattgcgtg actgatcgtg acgggtttca acatgaaaaa
3121 acctttaata taacaaaata taattcagac gaagaagcaa aaagagtggc tgataaatgg
3181 caatatgaaa catcttataa attggaaatg aaaaaaaatt taattcgtgt tggtgatgat
3241 aatactatcg aagtggaaatt gacaaaggat gaaattatga aaacggataa aatatttt
3301 ccattaaatac aagaaattcc gttattcata acagaaggta gtaataaaaat aaaatacgta
3361 gccacaagta ttaataatcg aaattgtcaa ttcatgggc tgattactga ttttgatatg
3421 gtggatcata tgaatggta tactctcgat aatcggtttaa gtaatcttag agtagctgt
3481 ttctcactca acaatagtaa taggcataat gatgtaaaag gtgttaagaa agtagatgct
3541 attttcggga aatcatcacaa agcatcaatt aaaattgacg gcaaaagaata tagtaaatat
3601 attccagtag ataattacga tgaaggagaa gcactagaaa tagcaaaaaa tatgagaaaa
3661 gctgtacaa atgttgattt gcgtacatgaa aaaacccaaa tcattatgga cccaaaaatc
3721 atgaacattc aatcaaaaaa attggataag atattaaaac ttattcgaaa taatacgat
3781 tatgataata cattatattt aacccatatt aatctaccgc aagaaataaa ggatgaaata
3841 ttttattact atgtcgaaaa gcaaatggat tattatagag aatgtaaaga gaagcgaaaa

Example 1: Giant Viruses

> [Science. 2017 Apr 7;356\(6333\):82-85. doi: 10.1126/science.aal4657.](#)

- You might have read in the news or heard on the radio or TV about the new giant viruses that have been discovered.

Giant viruses with an expanded complement of translation system components

Frederik Schulz ¹, Natalya Yutin ², Natalia N Ivanova ³, Davi R Ortega ⁴, Tae Kwon Lee ⁵, Julia Vierheilig ⁵, Holger Daims ⁵, Matthias Horn ⁵, Michael Wagner ⁵, Grant J Jensen ^{4 6}, Nikos C Kyripides ³, Eugene V Koonin ⁷, Tanja Woyke ¹

Affiliations + expand

PMID: 28386012 DOI: [10.1126/science.aal4657](https://doi.org/10.1126/science.aal4657)

[Free article](#)

Abstract

The discovery of giant viruses blurred the sharp division between viruses and cellular life. Giant virus genomes encode proteins considered as signatures of cellular organisms, particularly translation system components, prompting hypotheses that these viruses derived from a fourth domain of cellular life. Here we report the discovery of a group of giant viruses (Klosneuviruses) in metagenomic data. Compared with other giant viruses, the Klosneuviruses encode an expanded translation machinery, including aminoacyl transfer RNA synthetases with specificities for all 20 amino acids. Notwithstanding the prevalence of translation system components, comprehensive phylogenomic analysis of these genes indicates that Klosneuviruses did not evolve from a cellular ancestor but rather are derived from a much smaller virus through extensive gain of host genes.

Example 1: Giant Viruses

- 1) Now try searching the NCBI Nucleotide database for klosneuvirus.
- 2) Results appear in the middle of the page. Search filters and other discovery tools appear on either side of results.
- 3) Each result in the database represents a record for one sequence or string of nucleic acids. The strings may overlap.
- 4) Click on the record for accession number KY684123.1 and view the full record.

[Klosneuvirus KNV1 Klosneuvirus_16 genomic sequence](#)
1. 7,441 bp linear DNA
Accession: KY684123.1 GI: 1174573961
[GenBank](#) [FASTA](#) [Graphics](#)

- 5) Record for accession number KY684123.1

[Note: Sidebar images like the above can be expanded by clicking on them.]

Example 1: Giant Viruses

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide klosneuvirus Create alert Advanced Help

Species Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

Viruses (144) Customize ...

Molecule types Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

genomic DNA/RNA (144) Customize ...

Source databases Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

INSDC (GenBank) (143) Customize ...

Sequence Type Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

Nucleotide (144)

Sequence length Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

Custom range... Database: Select

Release date Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

Custom range... Find items

Revision date Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

Custom range... Search

[Clear all](#) See more...

[Show additional filters](#)

Items: 1 to 20 of 144

<< First < Prev Page 1 of 8 Next > Last >>

[Klosneuvirus KNV1 Klosneuvirus 16 genomic sequence](#)
1. 7,441 bp linear DNA
Accession: KY684123.1 GI: 1174573961
[BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[Klosneuvirus KNV1 Klosneuvirus 15 genomic sequence](#)
2. 9,506 bp linear DNA
Accession: KY684122.1 GI: 1174573947
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[Klosneuvirus KNV1 Klosneuvirus 14 genomic sequence](#)
3. 9,663 bp linear DNA
Accession: KY684121.1 GI: 1174573937
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[Klosneuvirus KNV1 Klosneuvirus 13 genomic sequence](#)
4. 9,915 bp linear DNA
Accession: KY684120.1 GI: 1174573928

Results by taxon

Top Organisms [Tree]
Edafosvirus sp. (25)
Hyperionvirus sp. (17)
Klosneuvirus KNV1 (16)
Harfovirus sp. (16)
Homavirus sp. (16)
All other taxa (54)
More...

Find related data

Database: Select

Search details

"Klosneuvirus"[Organism] OR
klosneuvirus[All Fields]

<https://www.ncbi.nlm.nih.gov/genbank/samplerecord/>

Sample GenBank Record

This page presents an annotated sample GenBank record (accession number [U49845](#)) in its *GenBank Flat File* format. You can see the corresponding [live record for U49845](#), and see [examples of other records](#) that show a range of biological features.

<u>LOCUS</u>	SCU49845	5028 bp	DNA	PLN	21-JUN-1999
<u>DEFINITION</u>	Saccharomyces cerevisiae TCP1-beta gene, partial cds, and Axl2p (AXL2) and Rev7p (REV7) genes, complete cds.				
<u>ACCESSION</u>	U49845				
<u>VERSION</u>	U49845.1 GI:1293613				
<u>KEYWORDS</u>	.				
<u>SOURCE</u>	Saccharomyces cerevisiae (baker's yeast)				
<u>ORGANISM</u>	Saccharomyces cerevisiae Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
<u>REFERENCE</u>	1 (bases 1 to 5028)				
<u>AUTHORS</u>	Torpey,L.E., Gibbs,P.E., Nelson,J. and Lawrence,C.W.				
<u>TITLE</u>	Cloning and sequence of REV7, a gene whose function is required for DNA damage-induced mutagenesis in <i>Saccharomyces cerevisiae</i>				
<u>JOURNAL</u>	Yeast 10 (11), 1503-1509 (1994)				
<u>PUBMED</u>	7871890				
<u>REFERENCE</u>	2 (bases 1 to 5028)				
<u>AUTHORS</u>	Roemer,T., Madden,K., Chang,J. and Snyder,M.				
<u>TITLE</u>	Selection of axial growth sites in yeast requires Axl2p, a novel plasma membrane glycoprotein				
<u>JOURNAL</u>	Genes Dev. 10 (7), 777-793 (1996)				
<u>PUBMED</u>	8846915				
<u>REFERENCE</u>	3 (bases 1 to 5028)				
<u>AUTHORS</u>	Roemer,T.				
<u>TITLE</u>	Direct Submission				
<u>JOURNAL</u>	Submitted (22-FEB-1996) Terry Roemer, Biology, Yale University, New Haven, CT, USA				
<u>FEATURES</u>	Location/Qualifiers				
<u>source</u>	1..5028				
	/organism="Saccharomyces cerevisiae"				

/organism="Saccharomyces cerevisiae"
[/db_xref="taxon:4932"](#)
/chromosome="IX"
/map=""
≤1..206
/codon_start=3
/product="TCP1-beta"
[/protein_id="AAA98665.1"](#)
[/db_xref="GI:1293614"](#)
[/translation="SSIYNGISTSGLDLNNTIADMRLGIVESYKLKRAVSSASEA
AEVLLRVDNIIRARPRTANRQHM"](#)
687..3158
/gene="AXL2"
687..3158
/gene="AXL2"
/note="plasma membrane glycoprotein"
/codon_start=1
/function="required for axial budding pattern of *S. cerevisiae*"
/product="Axl2p"
[/protein_id="AAA98666.1"](#)
[/db_xref="GI:1293615"](#)
[/translation="MTQLQISLLTATISLLHLVVATPYEAYPIGKQYPVARVNESF
TFQISNDTYKSSVDKTAQITYNCFDLPSWLSFDSSRTSGEPPSDLLSDANTTLFVN
VILEGTDASDSTSNNNTYQFVVTNRPSISSLSSDFNLNLLKNYGYTNKGKNALKLDPN
VFNVTFDRSMFTNEESIVSYGRSQLYNAPLPNWLFDSGELKFTGTAPVINSAIAPE
TSYSFVIIATDIEGFSAVEFEVFLVIGAHQLTTSIQNSLIINVDTGNVSYDLPLNYV
YLDPPISSDKLGSIINLLDAPDWVALDNATISGSVPDELLGKNSNPANFSVSIYDTYG
DVIYFNFEVVSTTDFAISSLPNINATRGWEFSYYFLPSQFTDYVNTNVSLFTNSSQ
DHDWVKFQSSNLTLAGEVPKNFDKLSLGLKANQGSQSELYFNIIGMDSKITHSNHSA
NATSTRSSHSTSTSSYTTSYTTAKISSTSAAATSSAPAAPAANKTSSHNNKAVAIA
CGVAIPLGVILVALICFLIFWRRRENPDENLPHIASGPDLNNPANKPNQENATPLN
NPFDDDASSYDDTSIARRLAALNTLKLDNHSATESDISSVDEKRDSLGMNTYNDQFQ
SQSKHEELAKPPVQPPESPFFDPQNRSVYMDSEPAVNKSWRYTGNLSPVSDIVRDS
YGSQKTVDEKRTSRDVTMSSLDPWNSNISPSPVRKSVPSPYNVT
HRNRHLQNIQDSQSGKNGITPTTMSTSSDDFVPVKDGENFCWVHSMEPDRRPSKKRL](#)

<https://www.ncbi.nlm.nih.gov/genbank/samplerecord/>

SQSKEELLAKPPVQPPESPFFDPQRSSSVYMDSEPAVNKSWRYTGNLSPVSDIVRDS
YGSQKTVDEKLFLEAPEKEKRTSRDVTMSSLDPWNSNISPSPVRSVPSPVNVTK
HRNRHLQNIQDSQSGKNGITPTTMSTSSSDDFVPVKDGENFCWVHSMEPDRRPSKKRL
VDFSNKSNVNQVGQVKDIHGRIPLEM"

gene complement (3300..4037)
/gene="REV7"
CDS complement (3300..4037)
/gene="REV7"
/codon_start=1
/product="Rev7p"
/protein_id="AAA98667.1"
/db_xref="GI:1293616"
/translation="MNRWVEKWLRYVLKYIINLILFYRNVYPPQSFDYTTYQSFNLPO
FVPINRHPALIDYELIDVLSKLTHVYRFSCIIINKKNDLCIEKYVLDSELQHVD
KDDQIITETEVFDEFRSSLNSLIMHLEKLPKVNDDTITFEAVINAIELELGHKLDRN
RVDSLEEKAEIERDSNWVKCQEDENLPDNNNGFQPPKIKLTSLVGSVGPLIIHQFSEK
LISGDDKILNGVYSQEEGESIFGSLF"

ORIGIN

1 gatcctccat atacaacggat atctccacct caggtaga tctcaacaac ggaaccattg
61 ccgacatgag acagttaggt atcgtcgaga gttacaagct aaaacgagca gtagtcagct
121 ctgcatactga agccgcgtgaa gttctactaa gggtggataa catcatccgt gcaagaccaa
181 gaaccgccaa tagacaacat atgttaacata tttaggatata acctcgaaaa taataaaacc
241 ccacactgtc attattataa ttagaaacag aacgaaaaaa ttatccacta tataattcaa
301 agacgcgaaa aaaaaaaaac aacgcgtcat agaacttttg gcaattcgcg tcacaaataa
361 attttggcaa cttatgtttc ctcttcgagc agtactcgag ccctgttca agaatgtaat
421 aatacccccgtatgtatgg ttaaagatag catctccaca acctcaaagc tccttgcgc
481 gagtcgcct cctttgtcga gtaatttca ctttccat gagaacttat tttcttattc
541 ttatctctca catcctgttag tgattgacac tgcaacagcc accatcacta gaagaacaga
601 acaattactt aataaaaaaa ttatatcttc ctcgaaacgta tttcctgttccaacatcta
661 cgtatatacaa gaagcattca cttaccatga cacagcttca gatbtcatta ttgctgacag
721 ctactatatac actactccat ctatgttaggg ccacgcccata tgaggcatat cctatcgaa
781 aacaataaccc cccagttggca agagtcaatg aatcggtttac atttcaatatttccaaatgata
841 cctataaaatc gtctgttagac aagacagctc aaataacata caattgttcc gacttaccga
901 gctggcttgc gtttgactct agttcttagaa cggttctcagg tgaaccttct tctgacttac
961 tatctgtatgc gaacaccacg ttgttattca atgttaatact cgagggtacg gactctgccg

GenBank format

Klosneuvirus KNV1 Klosneuvirus_16 genomic sequence

GenBank: KY684123.1

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

Go to:

LOCUS KY684123 7441 bp DNA linear ENV 07-APR-2017
DEFINITION Klosneuvirus KNV1 Klosneuvirus_16 genomic sequence.
ACCESSION KY684123
VERSION KY684123.1
DBLINK BioProject: [PRJNA245600](#)
BioSample: [SAMN06458143](#)
KEYWORDS ENV.
SOURCE Klosneuvirus KNV1
ORGANISM [Klosneuvirus KNV1](#)
Viruses; Varidnaviria; Bamfordvirae; Nucleocytoparvoviricota;
Megaviricetes; Imitervirales; Mimiviridae; unclassified
Mimiviridae; Klosneuvirinae; Klosneuvirus.
REFERENCE 1 (bases 1 to 7441)
AUTHORS Schulz,F., Yutin,N., Ivanova,N.N., Ortega,D.R., Lee,T.K.,
Vierheilig,J., Daims,H., Horn,M., Wagner,M., Jensen,G.J.,
Kyripides,N.C., Koonin,E.V. and Woyke,T.
TITLE Giant viruses with an expanded complement of translation system
components
JOURNAL Science 356 (6333), 82-85 (2017)
PUBMED [28386012](#)
REFERENCE 2 (bases 1 to 7441)
AUTHORS Schulz,F. and Woyke,T.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2017) Prokaryotic Superprogram, DOE JGI, 2800
Mitchell Dr, Walnut Creek, CA 94598, USA
COMMENT GenBank Accession Numbers KY684108-KY684123 represent sequences
from Klosneuvirus KNV1.

##Assembly-Data-START##
Assembly Method :: spades v. 3.0.0

##Assembly-Data-START##
Assembly Method :: spades v. 3.0.0
Coverage :: 5x
Sequencing Technology :: Illumina
##Assembly-Data-END##

FEATURES	source	Location/Qualifiers
gene		1..7441 /organism="Klosneuvirus KNV1" /mol_type="genomic DNA" /isolate="KNV1" /isolation_source="wastewater" /db_xref="taxon: 1977640 " /clone="Klosneuvirus_16" /environmental_sample /lab_host="Nitrosomonas sp. DOME CR02B12" /country="Austria: Klosterneuburg" /collection_date="25-Mar-2014" /note="metagenomic"
CDS		189..785 /locus_tag="Klosneuvirus_16_1" 189..785 /locus_tag="Klosneuvirus_16_1" /codon_start=1 /product="hypothetical protein" /protein_id=" ARF12682.1 " /translation="MGRYCGILNRTKKQSISYYWKGDNCNCYEVMHQFHWNKDDQIE SGGGYEWYEIKYDTNTNTMESFEIDIMAEMSKAEEQDNNTPEVFTKYGFDTRLVYGN ALPKSYKLGSKSYTFVDEDDEDDGNVEEFKSLAYITDDEDDGIVNELNNFEEVNDK YNHLPEWDGDKCTKCRYVYDEKMLAKYAKWFDQTFHFN" complement(789..1454) /locus_tag="Klosneuvirus_16_2" complement(789..1454) /locus_tag="Klosneuvirus_16_2" /codon_start=1 /product="hypothetical protein" /protein_id=" ARF12683.1 " /translation="MTESDKTTKKCTSCEEELDKFSILNSKGTRNSQCKNCRNTK RKANKHDRKKEGTKECLECEQELDVKEFSANKSVSDGLQSYCKNCKFGQMOSCMSDFD

Example 1: Giant Viruses

- Look at the record for Klosneuvirus 16 and answer the following questions
- What is the sequence length or number of base pairs?
 - 7441
 - 5028

LOCUS KY684123 7441 bp DNA linear ENV 07-APR-2017
DEFINITION Klosneuvirus KNV1 Klosneuvirus_16 genomic sequence.
ACCESSION KY684123
VERSION KY684123.1

Example 1: Giant Viruses

- Scroll down the feature table and click on the first link for "gene"

FEATURES	Location/Qualifiers
source	1..7441 /organism="Klosneuvirus KNV1" /mol_type="genomic DNA" /isolate="KNV1" /isolation_source="wastewater" /db_xref="taxon: 1977640 " /clone="Klosneuvirus_16" /environmental_sample /lab_host="Nitrosomonas sp. DOME CR02B12" /country="Austria: Klosterneuburg" /collection_date="25-Mar-2014" /note="metagenomic"
gene	189..785 /locus_tag="Klosneuvirus_16_1"
CDS	189..785 /locus_tag="Klosneuvirus_16_1" /codon_start=1 /product="hypothetical protein" /protein_id=" ARF12682.1 " /translation="MGRYCGILNRTKKQSISSYWKGDNWCNCYEVMHQFWNKDDQIE SGGGYEIKYDTNTNTMESFEIDIMAEMSKAEEQDNNPTEVFTKYGFDTRLVYGKN ALPKSYKLGKSQYTFVDEDDEDDGNVEFKSLAYITDDEDGIVNELNNFEEEVNDK YNHLPEDGDKCTKCRYVYDEKMLAKYAKWFDQTFHFN"

Example 1: Giant Viruses

- This link "jumps" you to the raw sequence data at the bottom of the record, highlighting the part of the sequence identified as a gene. New navigation features appear at the bottom of your screen, allowing you to jump from feature to feature.
- How many genes are annotated ('defined') on this accession number?
 - 5
 - 47
 - 9



Example 1: Giant Viruses

- What are the first ten nucleotides in the HNH endonuclease gene?
 - ATGCAACCAT
 - MQLPIYPSPW
 - GCGTAATGAA
- Hint: Look through the feature table to find a part of the sequence with a product labeled, "HNH endonuclease." Follow the link in the feature table to the sequence at the bottom of the record to see the first ten nucleotides.

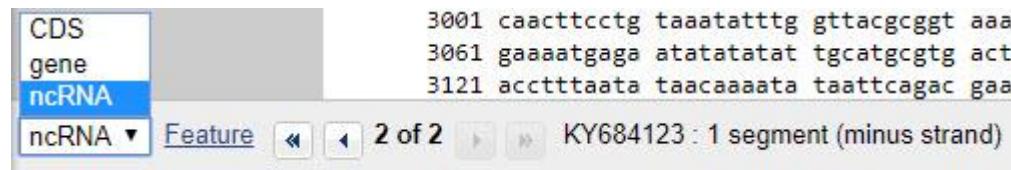
Example 1: Giant Viruses

```
2281 ctttcgcttg agggggggtg attgctaatc acttcacata aaaattgaag tgtaaacaca
2341 cttaaagatt tggtttattta taataatattt aaaaccaaact ctatgcacc attaatttat
2401 ccatcacccat ggcaattagg aaaaccacgt ggaagcatca aagatactat taacaaaaaa
2461 acaggaaccg ttacatcgta tagtgtcaca ttatccccac caaatgctca accaattact
2521 aaattcttct ctatttctga atatggatcc caaacaatg caaaaaaaagc atcagaaaaat
2581 tggagacaaa caatgtcaga tcaataacaat ctaactcgca accaaatccg atattttagac
2641 aaggatacaat ttgaagtaca attaacgcaa gggaaaaacga tgaaaaactga cgcgaaatac
2701 atagatgaag tacaaaaata tccaataaaat ataaaaacaa agaaatcgaa atccgatcca
2761 tctaaagata gacattatgc gatgtgtcaa aacattaaag aagtgttcc atttgcgcatt
2821 ttaatatgca aatatgatata aatagaatata attaatggag attctttggat tgtaaggaaaa
2881 gaaaaatataa aagaatttggtt tcagttaaa acaaataata tcaaagttagg taaaaatata
2941 gttcatttgc cagataatga cattaatata cagttatgaa gttttacaaa agatattgac
3001 caactccctg taaatatttg gttacgcgtt aaaccagctg gaagtgttattt caaacgaaca
3061 gaaaatgaga atatatataat tgcattgcgtg actgtatgcgtg acgggtttca acatgcaaaa
3121 acctttaata taacaaaata taattcagac gaagaagcaa aaagagtggc tgataatgg
3181 caatatgaaa catttataa attggaaatg acaaaaaattt taattcgtgt tggtatgtat
3241 aatactatcg aagtggaaattt gacaaaggat gaaattatgaa aaacggataaa aatattttt
3301 ccattaaatac aagaattcc gttattcata acagaaggta gtaataaaaat aaaatacgtt
3361 gcccacaagta ttaataatcg aaattgtcaa tttcatgggc tgattactga ttttgatatg
3421 gtggatcata tgaatggta tactctcgat aatcgatggaa gtaatcttag agtagctgtat
3481 ttctcactca acaatagtaa taggcataat gatgtaaaag gtgttaagaa agtagatgct
3541 attttcggga aatcatacaa agcatcaattt aaaatttgacg gcaaagaata tagtaatata
3601 attccatgtatgataatttca tgaaggagaa gcactagaaa tagcaaaaaa tatgagaaaa
3661 gctgtacaa atgttggattt gctgtacaa aaaacccaaa tcattatggc cccaaaaatc
3721 atgaacattc aaatcaaaaaa attggataag atattaaaac ttattcgaaa taatacgat
3781 tatgataata cattatattt aacccatattt aatctaccgc aagaaaataaa ggatgaaata
3841 ttttattact atgtcgaaaa gcaaatggat tattatagag aatgtaaaga gaagcgaaaa
3901 aaaaatcgatgg atattatgaa taagaaaaacg gatgttataa tttaaaaaaaa ttttttttatc
```

2383..3945
/locus_tag="Klosneuvirus_16_3"

Example 1: Giant Viruses

- Are any RNA molecules annotated ('defined') on this record (accession number KY684123)?
 - Yes
 - No
- Hint: check the drop down menu at the bottom of the screen



- What is ncRNA?
 - "ncRNA" stands for "non-coding RNA". It's a sequence that may regulate gene expression, but doesn't code for a protein.

Example 1: Giant Viruses

```
complement(2009..2237)
/ncRNA_class="autocatalytically_spliced_intron"
/locus_tag="Klosneuvirus_Intron_32"
/product="Intron_gpi"
```

Example 2: Mitochondrial Transporter SLC25A3

- Start a new search at the top of the screen with gene SLC25A3.
- What does this gene do?
- The protein encoded by this gene catalyzes the transport of phosphate into the mitochondrial matrix, either by proton cotransport or in exchange for hydroxyl ions.
- Use your search results from Nucleotide to answer the following 3 questions about SLC25A3:
 - Approximately how many sequences exist in the NCBI Nucleotide database for the mitochondrial transporter SLC25A3?
 - 30
 - 1800
 - 8000

Example 2: Mitochondrial Transporter SLC25A3

Nucleotide Nucleotide SLC25A3

Create alert Advanced

Species Summary 20 per page Sort by Default order Send to: Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree]

- Papio anubis (40)
- Macaca mulatta (33)
- Homo sapiens (23)
- Callithrix jacchus (13)
- Mus musculus (13)
- All other taxa (1706)

More...

Results by taxon

Find related data

Database:

Find items

Search details

SLC25A3[All Fields]

Items: 1 to 20 of 1828

<< First < Prev Page of 92 Next > Last >

[Ovis aries SLC25A3 \(SLC25A3\) mRNA, complete cds](#)

1. 1,377 bp linear mRNA
Accession: FJ422547.1 GI: 213688913

Search See more...

Example 2: Mitochondrial Transporter SLC25A3

- Approximately how many sequence records in your results are from the INSDC?
 - 6
 - 60
 - 700
- HINT: Find your filter options on the left of your screen.
- What is INSDC?
 - The International Nucleotide Sequence Database Collaboration (INSDC) is a long-standing foundational initiative that operates between DDBJ, EMBL-EBI and NCBI. INSDC covers the spectrum of data raw reads, though alignments and assemblies to functional annotation, enriched with contextual information relating to samples and experimental configurations.

Example 2: Mitochondrial Transporter SLC25A3

Nucleotide Nucleotide SLC25A3

Create alert Advanced

Species Summary 20 per page Sort by Default order Send to: Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree]

- Papio anubis (40)
- Macaca mulatta (33)
- Homo sapiens (23)
- Callithrix jacchus (13)
- Mus musculus (13)
- All other taxa (1706)

More...

Results by taxon

Find related data

Database:

Find items

Search details

SLC25A3[All Fields]

Search See more...

Species

- Animals (1,795)
- Plants (15)
- Fungi (4)
- Protists (6)
- Customize ...

Molecule types

- genomic DNA/RNA (840)
- mRNA (935)
- Customize ...

Source databases

- INSDC (GenBank) (682)
- RefSeq (1,144)
- Customize ...

Sequence Type

- Nucleotide (1,730)
- EST (87)
- GSS (11)

Sequence length

- Custom range...

Release date

- Custom range...

Revision date

- Custom range...

[Clear all](#)

GENE

Was this helpful?

[SLC25A3 – solute carrier family 25 member 3](#)

[Homo sapiens \(human\)](#)

Also known as: OK/SW-cl.48, PHC, PTP

Gene ID: 5250

[RefSeq transcripts](#) (3) [RefSeq proteins](#) (3) [RefSeqGene](#) (1) [PubMed](#) (160)

[Orthologs](#) [Genome Data Viewer](#) [BLAST](#) [Download](#)

RefSeq Sequences

Items: 1 to 20 of 1828

<< First < Prev Page of 92 Next > Last >>

[Ovis aries SLC25A3 \(SLC25A3\) mRNA, complete cds](#)

1. 1,377 bp linear mRNA
Accession: FJ422547.1 GI: 213688913

Example 2: Mitochondrial Transporter SLC25A3

- Approximately how many of your search results are RefSeq records?
 - 10
 - 100
 - 1200
- What is RefSeq?
 - RefSeq is short for the NCBI Reference Sequence collection, a comprehensive, integrated, non-redundant, well-annotated set of sequences, including genomic DNA, transcripts, and proteins. RefSeq sequences form a foundation for medical, functional, and diversity studies.

Example 2: Mitochondrial Transporter SLC25A3

Nucleotide Nucleotide SLC25A3

Create alert Advanced

Species Summary 20 per page Sort by Default order Send to: Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree]

- Papio anubis (40)
- Macaca mulatta (33)
- Homo sapiens (23)
- Callithrix jacchus (13)
- Mus musculus (13)
- All other taxa (1706)

More...

Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree]

- Papio anubis (40)
- Macaca mulatta (33)
- Homo sapiens (23)
- Callithrix jacchus (13)
- Mus musculus (13)
- All other taxa (1706)

More...

Find related data

Database:

Find items

Search details

SLC25A3[All Fields]

Items: 1 to 20 of 1828

<< First < Prev Page of 92 Next > Last >>

[Ovis aries SLC25A3 \(SLC25A3\) mRNA, complete cds](#)

1. 1,377 bp linear mRNA
Accession: FJ422547.1 GI: 213688913

Search See more...

Species

- Animals (1,795)
- Plants (15)
- Fungi (4)
- Protists (6)
- Customize ...

Molecule types

- genomic DNA/RNA (840)
- mRNA (935)
- Customize ...

Source databases

- INSDC (GenBank) (682)
- RefSeq (1,144)
- Customize ...

Sequence Type

- Nucleotide (1,730)
- EST (87)
- GSS (11)

Sequence length

- Custom range...

Release date

- Custom range...

Revision date

- Custom range...

[Clear all](#)

Example 2: Mitochondrial Transporter SLC25A3

- Now start a new search using the Advanced page
- Search for SLC25A3 in the Title field.
- How much of a difference is there between the results for this search and your last one?
 - There are MORE results by searching in the Title field
 - There are LESS results by searching in the Title field
 - There is not a significant difference
- Hint: Compare the following searches:
 - SLC25A3
 - SLC25A3[title]

Example 2: Mitochondrial Transporter SLC25A3

Nucleotide Nucleotide Help

Create alert Advanced

Species Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾ Filters: [Manage Filters](#)

Animals (921)

Customize ...

Molecule types

genomic DNA/RNA (13)

mRNA (885)

Customize ...

Source databases

INSDC (GenBank) (137)

RefSeq (790)

Customize ...

Sequence Type

Nucleotide (829)

EST (87)

GSS (11)

Sequence length

Custom range...

Release date

Custom range...

Revision date

Custom range...

[Clear all](#)

[Show additional filters](#)

Items: 1 to 20 of 927

<< First < Prev Page 1 of 47 Next > Last >

[PREDICTED: Dromiciops gliroides solute carrier family 25 member 3 \(SLC25A3\), transcript variant X2, mRNA](#)
1. [X2, mRNA](#)
1,406 bp linear mRNA
Accession: XM_043966433.1 GI: 2102757011
[BioProject](#) [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[PREDICTED: Dromiciops gliroides solute carrier family 25 member 3 \(SLC25A3\), transcript variant X1, mRNA](#)
2. [X1, mRNA](#)
1,409 bp linear mRNA
Accession: XM_043966432.1 GI: 2102757007
[BioProject](#) [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[PREDICTED: Cervus elaphus solute carrier family 25 member 3 \(SLC25A3\), transcript variant X4, mRNA](#)
3. [mRNA](#)
1,404 bp linear mRNA
Accession: XM_043880604.1 GI: 2102011317
[BioProject](#) [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[PREDICTED: Cervus elaphus solute carrier family 25 member 3 \(SLC25A3\), transcript variant X3, mRNA](#)
4. [mRNA](#)

Results by taxon

Top Organisms [Tree]

Papio anubis (39)
Macaca mulatta (24)
Homo sapiens (10)
Cebus imitator (10)
Mus musculus (10)
All other taxa (834)
[More...](#)

Find related data

Database:

Search details

SLC25A3[title]

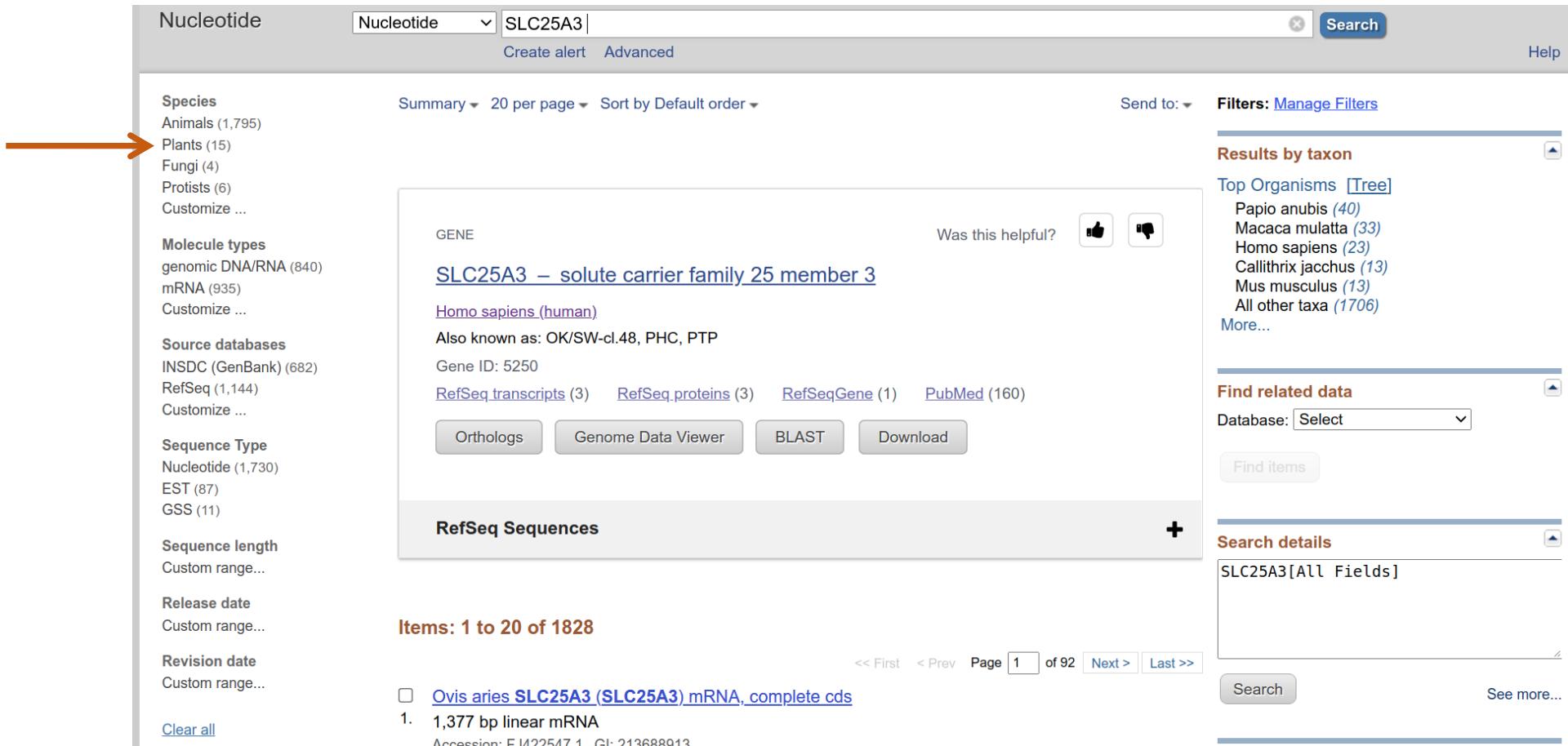
See more...

Recent activity

Example 2: Mitochondrial Transporter SLC25A3

- Return to your search results for your SLC25A3 search (WITHOUT the [title] field restriction).
- Are there any SLC25A3 sequences in plants?
 - Yes
 - No
- Hint: Check the Species filters in the left column of your results.

Example 2: Mitochondrial Transporter SLC25A3



Nucleotide Nucleotide SLC25A3

Create alert Advanced

Species Send to:

Animals (1,795)
Plants (15) →
Fungi (4)
Protists (6)
Customize ...

Molecule types
genomic DNA/RNA (840)
mRNA (935)
Customize ...

Source databases
INSDC (GenBank) (682)
RefSeq (1,144)
Customize ...

Sequence Type
Nucleotide (1,730)
EST (87)
GSS (11)

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

[Clear all](#)

Results by taxon

Top Organisms [Tree]
Papio anubis (40)
Macaca mulatta (33)
Homo sapiens (23)
Callithrix jacchus (13)
Mus musculus (13)
All other taxa (1706)
[More...](#)

Find related data

Database:

Search details

SLC25A3[All Fields]

Items: 1 to 20 of 1828

<< First < Prev Page of 92 Next > Last >>

[Ovis aries SLC25A3 \(SLC25A3\) mRNA, complete cds](#)
1. 1,377 bp linear mRNA
Accession: FJ422547.1 GI: 213688913

Example 3: Human tyrosine hydroxylase

- Start a new search at the top of the screen.
- Find human tyrosine hydroxylase gene sequence records.
- What does this gene do?
 - This gene can cause Segawa syndrome and is important in the synthesis of catecholamines (which are important pharmacologic agents).
- What is the gene symbol or abbreviation for tyrosine hydroxylase?
 - The official gene symbol for tyrosine hydroxylase is TH.
 - This is included in most titles of Nucleotide records, record with TH in title
 - A place to check official gene names is the HGNC, or HUGO Gene Nomenclature Committee. (<https://www.genenames.org/>)

Example 3: Human tyrosine hydroxylase

Nucleotide Nucleotide Help

Create alert Advanced

Species Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾ Filters: Manage Filters

Animals (188)
Bacteria (1)
Customize ...

Molecule types Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾ Filters: Manage Filters

genomic DNA/RNA (93)
mRNA (125)
Customize ...

Source databases Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾ Filters: Manage Filters

INSDC (GenBank) (104)
RefSeq (125)
Customize ...

Sequence Type Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾ Filters: Manage Filters

Nucleotide (228)
EST (3)

Sequence length Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾ Filters: Manage Filters

Custom range...

Release date Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾ Filters: Manage Filters

Custom range...

Revision date Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾ Filters: Manage Filters

Custom range...

[Clear all](#)

[Show additional filters](#)

Results by taxon

Top Organisms [Tree]

Homo sapiens (117)
Rattus norvegicus (19)
Mus musculus (18)
Danio rerio (9)
Bos taurus (4)
All other taxa (64)
[More...](#)

Find related data

Database: Select

Find items

Search details

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

Items: 1 to 20 of 231

<< First < Prev Page 1 of 12 Next > >>

[Homo sapiens gene for tyrosine hydroxylase, partial cds, alternative splicing forms](#)

1. 7,854 bp linear DNA
Accession: D00269.2 GI: 62899628

Search See more...

Gene

Gene

Full Report ▾

Send to: ▾

Hide sidebar >>

TH tyrosine hydroxylase [*Homo sapiens* (human)]

Gene ID: 7054, updated on 19-Sep-2021

 Download Datasets

Summary

Official Symbol	TH provided by HGNC
Official Full Name	tyrosine hydroxylase provided by HGNC
Primary source	HGNC:HGNC:11782
See related	Ensembl:ENSG00000180176 MIM:191290
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	TYH; DYT14; DYT5b
Summary	The protein encoded by this gene is involved in the conversion of tyrosine to dopamine. It is the rate-limiting enzyme in the synthesis of catecholamines, hence plays a key role in the physiology of adrenergic neurons. Mutations in this gene have been associated with autosomal recessive Segawa syndrome. Alternatively spliced transcript variants encoding different isoforms have been noted for this gene. [provided by RefSeq, Jul 2008]
Expression	Restricted expression toward adrenal (RPKM 42.8) See more
Orthologs	mouse all

NEW

Try the new [Gene table](#)

Try the new [Transcript table](#)

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, Homology, Gene Ontology

General protein information

NCBI Reference Sequences (RefSeq)

Related sequences

Additional links

Locus-specific Databases

Gene: <https://www.ncbi.nlm.nih.gov/gene/7054>

Gene Advanced Help

Full Report Hide sidebar >>

TH tyrosine hydroxylase [*Homo sapiens* (human)]

Gene ID: 7054, updated on 19-Sep-2021

Summary

Official Symbol TH provided by HGNC
Official Full Name tyrosine hydroxylase provided by HGNC
Primary source HGNC:HGNC:11782
See related Ensembl:ENSG00000180176 MIM:191290
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 TYH; DYT14; DYT5b
Summary The protein encoded by this gene is involved in the conversion of tyrosine to dopamine. It is the rate-limiting enzyme in the synthesis of catecholamines, hence plays a key role in the physiology of adrenergic neurons. Mutations in this gene have been associated with autosomal recessive Segawa syndrome. Alternatively spliced transcript variants encoding different isoforms have been noted for this gene. [provided by RefSeq, Jul 2008]
Expression Restricted expression toward adrenal (RPKM 42.8) [See more](#)
Orthologs [mouse](#) [all](#)
NEW Try the new [Gene table](#)
Try the new [Transcript table](#)

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, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links
 - Locus-specific Databases

Gene: <https://www.ncbi.nlm.nih.gov/gene/7054>

Genomic context

Location: 11p15.5 See TH in [Genome Data Viewer](#)

Exon count: 14

Annotation release	Status	Assembly	Chr	Location
109.20210514	current	GRCh38.p13 (GCF_000001405.39)	11	NC_000011.10 (2163929..2174081, complement)
105.20201022	previous assembly	GRCh37.p13 (GCF_000001405.25)	11	NC_000011.9 (2185159..2193045, complement)

Genomic regions, transcripts, and products

Genomic Sequence: NC_000011.10 Chromosome 11 Reference GRCh38.p13 Primary Assembly

Go to reference sequence details Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

Find: 2,174 K 2,172 K 2,170 K 2,168 K 2,166 K 2,164 K

Genes, MANE Project (release v0.95)

NCBI Homo sapiens Annotation Release 109.20210514

XM_011520335.2 NM_000360.4 TH NP_000351.2

XP_011518637.1 NM_000360.4 NP_000351.2

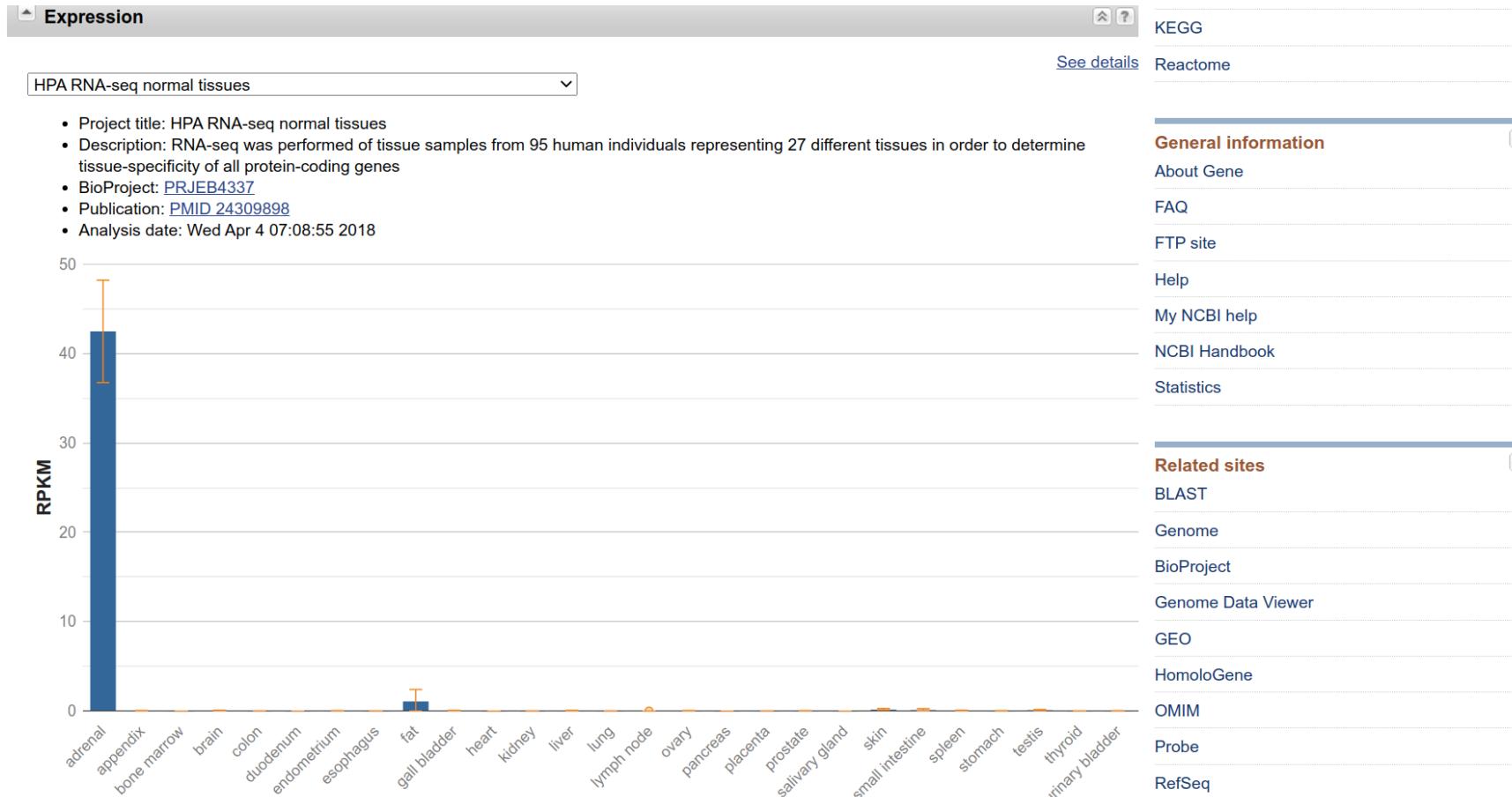
Genome Browsers

- [Genome Data Viewer](#)
- [Variation Viewer \(GRCh37.p13\)](#)
- [Variation Viewer \(GRCh38\)](#)
- [1000 Genomes Browser \(GRCh37.p13\)](#)
- [Ensembl](#)
- [UCSC](#)

Related information

- [Order cDNA clone](#)
- [3D structures](#)
- [BioAssay by Target \(List\)](#)
- [BioAssay by Target \(Summary\)](#)
- [BioAssay, by Gene target](#)
- [BioAssays, RNAi Target, Active](#)
- [BioAssays, RNAi Target, Tested](#)
- [BioProjects](#)
- [BioSystems](#)
- [Books](#)
- [CCDS](#)
- [ClinVar](#)
- [Conserved Domains](#)

Gene: <https://www.ncbi.nlm.nih.gov/gene/7054>



Gene: <https://www.ncbi.nlm.nih.gov/gene/7054>

Phenotypes

[BioGRID CRISPR Screen Phenotypes \(6 hits/1042 screens\)](#)
[Find tests for this gene in the NIH Genetic Testing Registry \(GTR\)](#)
[Review eQTL and phenotype association data in this region using PheGenI](#)

Associated conditions

Description	Tests
A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease. GeneReviews: Not available	
Autosomal recessive DOPA responsive dystonia MedGen: C2673535 , OMIM: 605407 , GeneReviews: Tyrosine Hydroxylase Deficiency	Compare labs
Genome-wide association scan for variants associated with early-onset prostate cancer. GeneReviews: Not available	
Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. GeneReviews: Not available	
Meta-analysis of genome-wide association studies in african americans provides insights into the genetic architecture of type 2 diabetes. GeneReviews: Not available	

Gene: <https://www.ncbi.nlm.nih.gov/gene/>

Gene

Gene

Advanced

Search

Help

Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene

- [Gene Quick Start](#)
- [FAQ](#)
- [Download/FTP](#)
- [RefSeq Mailing List](#)
- [Gene News](#)
- [Factsheet](#)

Gene Tools

- [Submit GeneRIFs](#)
- [Submit Correction](#)
- [Statistics](#)
- [BLAST](#)
- [Genome Workbench](#)
- [Splign](#)

Other Resources

- [OMIM](#)
- [RefSeq](#)
- [RefSeqGene](#)
- [Protein Clusters](#)

Example 3: Human tyrosine hydroxylase

- Look at the following accession records, all of which are in your search results for the tyrosine hydroxylase (TH) gene:
 - L15440.1
 - X05290.1
 - M17589.1
- How are these records different from each other? Look at their sizes, their sources, any descriptive notes, and their feature tables.
- HINT: These three records are in your search results, but for easier comparison search Nucleotide for: L15440.1 OR X05290.1 OR M17589.1

Example 3: Human tyrosine hydroxylase

Nucleotide Nucleotide

Create alert Advanced Send to:

Species Summary ▾ Sort by Default order ▾ Analyze these sequences

Animals (3) **Molecule types** **Filters:** [Manage Filters](#)

Customize ... genomic DNA/RNA (1) **Find related data**

Items: 3 Database: [Select](#)

Send to:

Species **Sequence Type** **Search details**

Animals (3) Nucleotide (3) [L15440.1\[All Fields\]](#) OR
Customize ... Nucleotide (3) [X05290.1\[All Fields\]](#) OR

Molecule types Human tyrosine hydroxylase type 4 mRNA, complete cds [M17589.1\[All Fields\]](#)

genomic DNA/RNA (1) 1. 1,921 bp linear mRNA

mRNA (2) 2. 1,921 bp linear mRNA

Customize ... 3. 1,816 bp linear mRNA

Source databases **Release date** **Recent activity**

INSDC (GenBank) (3) Human tyrosine hydroxylase type 4 mRNA, complete cds Turn Off Clear

Customize ... Human mRNA for tyrosine hydroxylase (HTH-1)

Sequence Type Human mRNA for tyrosine hydroxylase (HTH-1)

Nucleotide (3) Human mRNA for tyrosine hydroxylase (HTH-1)

Sequence length Human mRNA for tyrosine hydroxylase (HTH-1)

Custom range... Human mRNA for tyrosine hydroxylase (HTH-1)

Release date Human mRNA for tyrosine hydroxylase (HTH-1)

Custom range... Human mRNA for tyrosine hydroxylase (HTH-1)

Revision date Human mRNA for tyrosine hydroxylase (HTH-1)

Custom range... Human mRNA for tyrosine hydroxylase (HTH-1)

Clear all Human mRNA for tyrosine hydroxylase (HTH-1)

Show additional filters Human mRNA for tyrosine hydroxylase (HTH-1)

Summary ▾ Sort by Default order ▾ Send to:

Example 3: Human tyrosine hydroxylase

Each of the records you looked at included a sequence that codes for the TH gene.

L15440.1 : The record title says it all. This records a **human DNA sequence** that includes several genes, not just TH. And it only has the 3' end of the TH gene. See the top of the feature table, and you can see that base pairs 1 through 1482 are part of the TH gene.

X05290 : This records an **mRNA sequence** that encodes a **single protein product**, identified as "tyrosine hydroxylase (HTH-1)". Note that it includes only about 1,800 base pairs. In the CDS portion of the feature table, note that the protein product is called HTH-1.

M17589.1 : The feature table on this record includes only **one gene** and **one CDS region**. Note that the literature reference says that "alternative RNA splicing produces four kinds of mRNA from a single gene" and the annotation in the feature table says that this one is **type 4**.

Extra tip: to find records with **variation features**, use the Advanced Search function,

The screenshot shows the NCBI Advanced Search results page. A search term "variation" is entered in the "Feature key" search field. Below the search bar, a dropdown menu shows the selected term "variation". A blue-highlighted link "variation (6395224)" leads to the main results page, which displays 28 entries under "xref". Navigation links "Previous 200" and "Next 200" are visible at the bottom of the results list. A "Hide index list" link is located in the top right corner.

Example 4: tRNA

Search for AB009835. What is this sequence molecular type? Source of this sequence?

Nucleotide Nucleotide ▾ AB009835 Search Help

GenBank ▾ Send to: ▾ Change region shown

Drosophila melanogaster mitochondrial tRNA-Lys, complete sequence

GenBank: AB009835.1 Fasta Graphics

Go to: ▾

LOCUS AB009835 74 bp tRNA linear INV 22-NOV-2014
DEFINITION Drosophila melanogaster mitochondrial tRNA-Lys, complete sequence.
ACCESSION AB009835
VERSION AB009835.1
KEYWORDS .
SOURCE mitochondrion Drosophila melanogaster (fruit fly)
ORGANISM [Drosophila melanogaster](#)
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora.
REFERENCE 1
AUTHORS Tomita,K., Ueda,T., Ishiwa,S., Crain,P.F., McCloskey,J.A. and
Watanabe,K.
TITLE Codon reading patterns in Drosophila melanogaster mitochondria
based on their tRNA sequences: a unique wobble rule in animal
mitochondria
JOURNAL Nucleic Acids Res. 27 (21), 4291-4297 (1999)
PUBMED [10518623](#)
REFERENCE 2 (bases 1 to 74)
AUTHORS Tomita,K.
TITLE Direct Submission

Send to: ▾ Change region shown

Customize view

Analyze this sequence

Run BLAST
Pick Primers
Highlight Sequence Features
Find in this Sequence

Related information

PubMed
Taxonomy
Full text in PMC

LinkOut to external resources

FlyBase [FlyBase]

FEATURES source Location/Qualifiers

tRNA 1..74 /organism="Drosophila melanogaster"
modified_base 9 /organelle="mitochondrion"
modified_base 25 /mol_type="tRNA"
modified_base 34 /db_xref="taxon:7227"
modified_base 41 1..74 /product="tRNA-Lys"
modified_base 42 9 /mod_base=m1g
modified_base 34 25 /mod_base=p
modified_base 41 34 /mod_base=t6a
modified_base 42 41 /mod_base=p
modified_base 63 42 /mod_base=dhu
modified_base 64 63 /mod_base=p
modified_base 64 64 /mod_base=p

ORIGIN

//

1 cattagatga ctgaaagcaa gtactggtct cttaaaccat ttaatagtaa attagcactt
61 acttctaatg acca

<https://www.ncbi.nlm.nih.gov/nuccore/AB009835>

Example 4: tRNA

Predict the folding of AB009835 tRNA

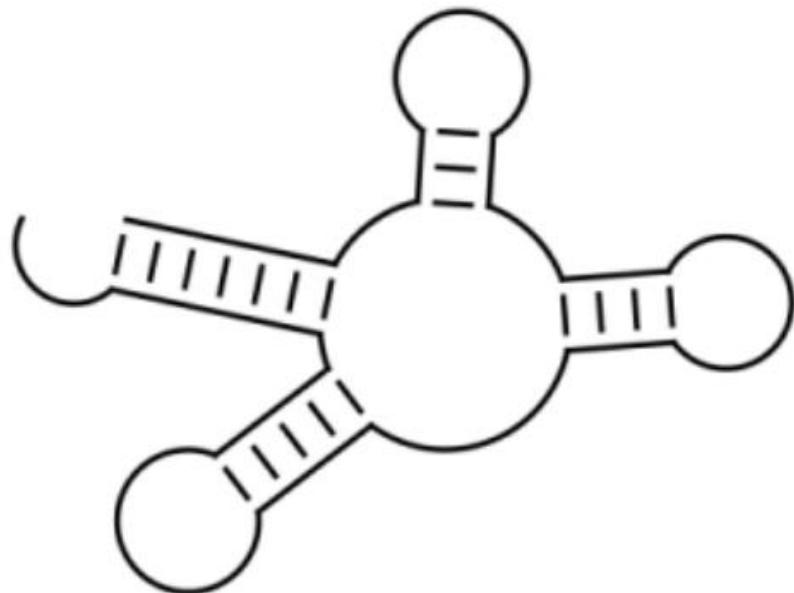


Figure 1: Secondary structure of a tRNA molecule.

FEATURES	Location/Qualifiers
source	1..74 /organism="Drosophila melanogaster" /organelle="mitochondrion" /mol_type="tRNA" /db_xref="taxon:7227"
tRNA	1..74 /product="tRNA-Lys"
modified_base	9 /mod_base=m1g
modified_base	25 /mod_base=p
modified_base	34 /mod_base=t6a
modified_base	41 /mod_base=p
modified_base	42 /mod_base=dhu
modified_base	63 /mod_base=p
modified_base	64 /mod_base=p
ORIGIN	1 cattagatga ctgaaagcaa gtactggtct cttaaaccat ttaatagtaa attagcactt 61 acttctaatg acca
	//



Search

HOME
RESEARCH
PUBLICATIONS
SERVICES +
DOWNLOADS
GROUP

SPOT-RNA: RNA Secondary Structure Prediction using an Ensemble of Two-dimensional Deep Neural Networks and Transfer Learning.

2020-01-29 by Jaswinder Singh

Our resources are limited. If you wish to run several batches per day please make use of the downloadable package or contact the administrator directly.

Submit

E-mail address (optional):	<input type="text"/>
Target (optional):	<input type="text"/>
Input your RNA Sequences: <input type="button" value="CLEAR"/>	Maximum: 2000 nts, Only one RNA sequence at a time
>Example sequence 4wj4_B UCCGCGAUAGCUCAGUCGGUAGAGCAAAUGACUGUUAAUCAUUGGGUCCUGGU UCGAGU CCAGGUCGCGGAGCCA	

Search

HOME
RESEARCH
PUBLICATIONS
SERVICES +
DOWNLOADS
GROUP

Note:

To run SPOT-RNA for long sequences or batch of sequences, please use standalone version of SPOT-RNA mentioned below.

Standalone Program to Run Locally:

[SPOT-RNA-local](#)

Datasets:

bpRNA: Initial Learning (Training TR0, validation VL0, and test TS0) [Dropbox](#) or [Nihao Cloud](#)

PDB: Transfer Learning (Training TR1, validation VL1, and test TS1) [Dropbox](#) or [Nihao Cloud](#)

Reference:

Singh, J., Hanson, J., Paliwal, K., Zhou, Y. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. Nat Commun 10, 5407 (2019) <https://doi.org/10.1038/s41467-019-13395-9>

RNA, server

RNA, RNA Secondary Structure Prediction, Pseudoknot bps, Non-Canonical bps, Base-triples

Example 5: Human PKD1

Search with Human PKD1 Gene.

Gene Gene Homo sapiens PKD1 Search Help

Create RSS Save search Advanced

Gene sources Tabular 20 per page Sort by Relevance Send to:

Genomic Categories Sequence content Status Clear all Show additional filters

Hide sidebar >>

Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree]

- Homo sapiens (102)
- Cellulosimicrobium cellulans (4)
- Bacteroides nordii CL02T12C05 (3)
- Necator americanus (1)
- Fonsecaea monophora (1)
- All other taxa (11)

[More...](#)

Find related data

Database:

Search results

Items: 1 to 20 of 122 << First < Prev Page of 7 [Next >](#) [Last >>](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> PKD1	polycystin 1, transient	Chromosome 16,	PBP, PC1, Pc-1, TRPP1	601313

See more...

Example 5: Human PKD1

How many exons of Human PKD1?

Genomic context

Location: 16p13.3 **See PKD1 in [Genome Data Viewer](#)**

Exon count: 51

Annotation release	Status	Assembly	Chr	Location
109.20210514	current	GRCh38.p13 (GCF_000001405.39)	16	NC_000016.10 (2088708..2135898, complement)
105.20201022	previous assembly	GRCh37.p13 (GCF_000001405.25)	16	NC_000016.9 (2138709..2185899, complement)

Chromosome 16 - NC_000016.10

[2026902 ► [2154165 ►

SLC9A3R2 → TSC2 → MIR6511B1 ← MIR3180-5 ← MIR4516 → RAB26 ←

NTHL1 ← LOC105371049 → MIR1225 ← PKD1 ←

Example 5: Human PKD1

How many transcripts of Human PKD1?

 Summary

Official Symbol PKD1 provided by [HGNC](#)

Official Full Name polycystin 1, transient receptor potential channel interacting provided by [HGNC](#)

Primary source [HGNC:HGNC:9008](#)

See related [Ensembl:ENSG00000008710](#) [MIM:601313](#)

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 PBP; PC1; Pc-1; TRPP1

Summary This gene encodes a member of the polycystin protein family. The encoded glycoprotein contains a large N-terminal extracellular region, multiple transmembrane domains and a cytoplasmic C-tail. It is an integral membrane protein that functions as a regulator of calcium permeable cation channels and intracellular calcium homoeostasis. It is also involved in cell-cell/matrix interactions and may modulate G-protein-coupled signal-transduction pathways. It plays a role in renal tubular development, and mutations in this gene cause autosomal dominant polycystic kidney disease type 1 (ADPKD). ADPKD is characterized by the growth of fluid-filled cysts that replace normal renal tissue and result in end-stage renal failure. Splice variants encoding different isoforms have been noted for this gene. Also, six pseudogenes, closely linked in a known duplicated region on chromosome 16p, have been described. [provided by RefSeq, Oct 2008]

Expression Ubiquitous expression in endometrium (RPKM 11.1), fat (RPKM 9.6) and 25 other tissues [See more](#)

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

NEW Try the new [Gene table](#)
Try the new [Transcript table](#)

Example 5: Human PKD1

How many transcripts of Human PKD1?

The screenshot shows the NCBI Datasets interface. On the left, there's a sidebar with links to various datasets like NCBI Datasets, Homepage, Genomes, Genes, and others. The main content area is titled "PKD1 Transcripts" (BETA) and displays information about the gene PKD1 in Homo sapiens. It says there are 10 transcripts. Below this, there's a "SWITCH TO GENE VIEW" link. At the top right, there's a search bar with "Search NCBI" and a "SEARCH" button. Below the search bar, there are buttons for "EDIT", "DOWNLOAD", and "SELECT COLUMNS". A table follows, listing 10 transcripts for the gene PKD1. The columns are: Gene ID, Symbol, Transcript, Length (nt), Protein, Length (aa), Protein name, Isoform, and Organism. The data is as follows:

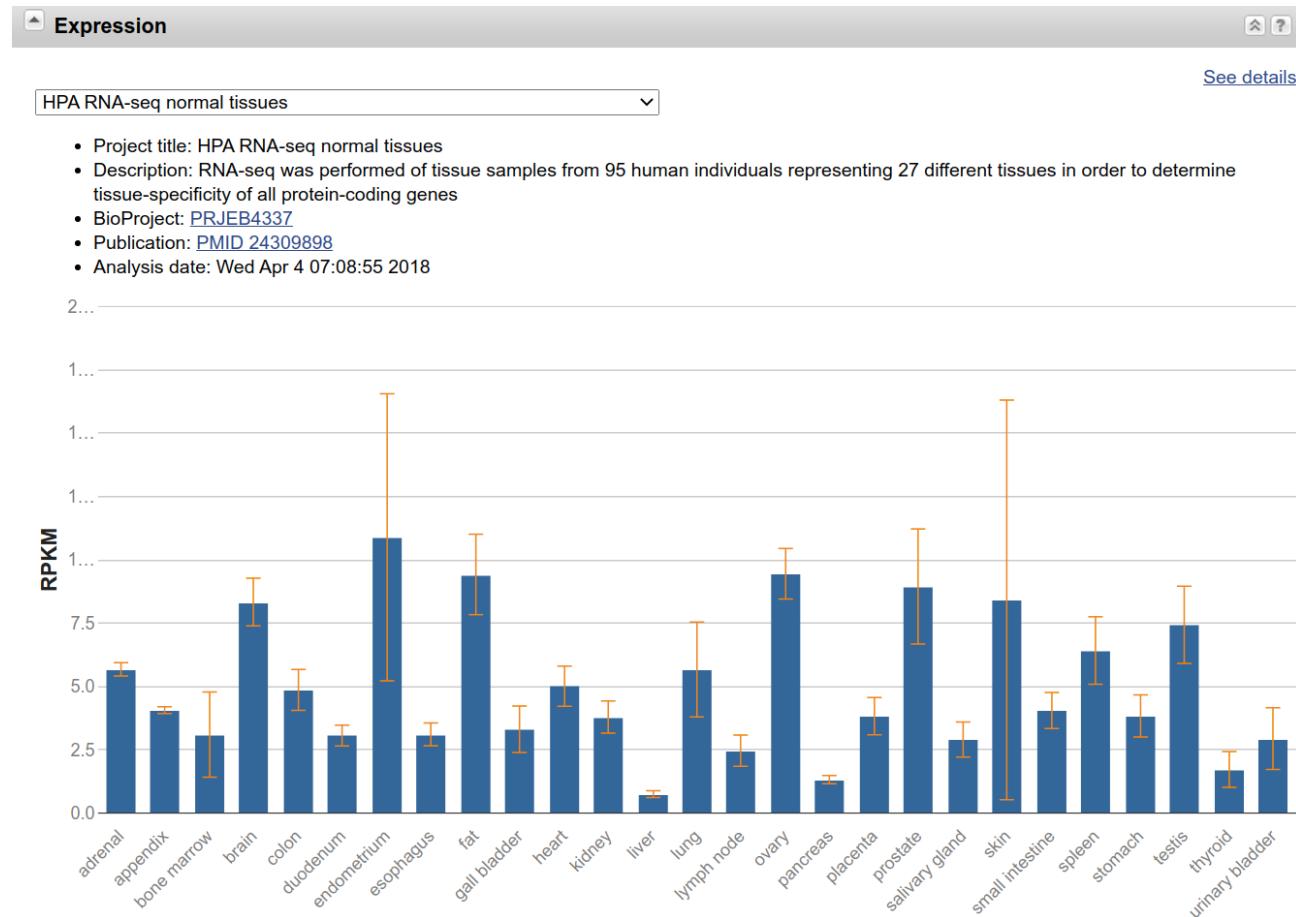
	Gene ID	Symbol	Transcript	Length (nt)	Protein	Length (aa)	Protein name	Isoform	Organism
<input type="checkbox"/>	5310	PKD1	NM_000296.4	14137	NP_000287.4	4302	polycystin-1	2 precursor	Homo sapiens
<input type="checkbox"/>	5310	PKD1	NM_001009944.3	14140	NP_001009944.3	4303	polycystin-1	1 precursor	Homo sapiens
<input type="checkbox"/>	5310	PKD1	XM_024450299.1	13994	XP_024306067.1	4319	polycystin-1	X4	Homo sapiens
<input type="checkbox"/>	5310	PKD1	XM_024450298.1	14066	XP_024306066.1	4343	polycystin-1	X1	Homo sapiens
<input type="checkbox"/>	5310	PKD1	XM_024450300.1	14052	XP_024306068.1	4273	polycystin-1	X5	Homo sapiens
<input type="checkbox"/>	5310	PKD1	XM_011522529.2	14163	XP_011520831.1	4320	polycystin-1	X3	Homo sapiens
<input type="checkbox"/>	5310	PKD1	XM_011522528.3	14166	XP_011520830.1	4321	polycystin-1	X2	Homo sapiens
<input type="checkbox"/>	5310	PKD1	XM_024450301.1	12283	XP_024306069.1	3635	polycystin-1	X6	Homo sapiens
<input type="checkbox"/>	5310	PKD1	XM_005255370.3	11727	XP_005255427.1	3288	polycystin-1	X8	Homo sapiens
<input type="checkbox"/>	5310	PKD1	XM_011522537.2	11098	XP_011520839.1	3329	polycystin-1	X7	Homo sapiens

At the bottom, there are buttons for "Rows per page" (set to 25), "Showing transcripts for 1-1 of 1 Gene", and navigation arrows.

<https://www.ncbi.nlm.nih.gov/gene/5310>

Example 5: Human PKD1

Is Human PKD1 expressed in all the tissues?



Example 5: Human PKD1

How many pseudogenes of Human PKD1?

Summary This gene encodes a member of the polycystin protein family. The encoded glycoprotein contains a large N-terminal extracellular region, multiple transmembrane domains and a cytoplasmic C-tail. It is an integral membrane protein that functions as a regulator of calcium permeable cation channels and intracellular calcium homoeostasis. It is also involved in cell-cell/matrix interactions and may modulate G-protein-coupled signal-transduction pathways. It plays a role in renal tubular development, and mutations in this gene cause autosomal dominant polycystic kidney disease type 1 (ADPKD1). ADPKD1 is characterized by the growth of fluid-filled cysts that replace normal renal tissue and result in end-stage renal failure. Splice variants encoding different isoforms have been noted for this gene. Also, six **pseudogenes**, closely linked in a known duplicated region on chromosome 16p, have been described. [provided by RefSeq, Oct 2008]

Expression Ubiquitous expression in endometrium (RPKM 11.1), fat (RPKM 9.6) and 25 other tissues [See more](#)

Example 5: Human PKD1

How many pseudogenes of Human PKD1?

General gene information

- [Markers](#)
- [Related pseudogene\(s\)](#)
10 found [Review record\(s\) in Gene](#)
- [Homology](#)
[Homologs of the PKD1 gene](#): The PKD1 gene is conserved in dog, cow, mouse, rat, chicken, zebrafish, and frog.
[Orthologs from Annotation Pipeline](#): 399 organisms have orthologs with human gene PKD1
[Orthologs](#)
- [Gene Ontology](#) [Provided by GOA](#)

Function	Evidence Code	Pubs
enables Wnt-activated receptor activity	IDA	PubMed
contributes to calcium channel activity	IDA	PubMed
enables calcium channel activity	ISS	
enables carbohydrate binding	IEA	
enables protein binding	IPI	PubMed
enables protein domain specific binding	IPI	PubMed
enables protein kinase binding	IPI	PubMed

Example 5: Human PKD1

How many pseudogenes of Human PKD1?

Gene Gene related_functional_gene_5310[group] Search Create RSS Save search Advanced Help

Gene sources: Tabular ▾ 20 per page ▾ Sort by Relevance ▾ Send to: ▾ Hide sidebar >>

Categories: Genomic

Items: 10 Filters: Manage Filters

Showing Current items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> PKD1P1 ID: 339044	polycystin 1, transient receptor potential channel interacting pseudogene 1 [<i>Homo sapiens</i> (human)]	Chromosome 16, NC_000016.10 (16317609..16350608)	HG1, NPIP	
<input type="checkbox"/> PKD1P6 ID: 353511	polycystin 1, transient receptor potential channel interacting pseudogene 6 [<i>Homo sapiens</i> (human)]	Chromosome 16, NC_000016.10 (15125139..15154873, complement)	HG6	
<input type="checkbox"/> PKD1P6-NPIPP1 ID: 105369154	PKD1P6-NPIPP1 readthrough [<i>Homo sapiens</i> (human)]	Chromosome 16, NC_000016.10 (15104723..15131601, complement)		

Find related data: Database: Select

Find items

Search details: related_functional_gene_5310[group] AND alive[prop]

Search See more...

Recent activity: Turn Off Clear

related_functional_gene_5310[group] AND (alive[prop]) (10) Gene

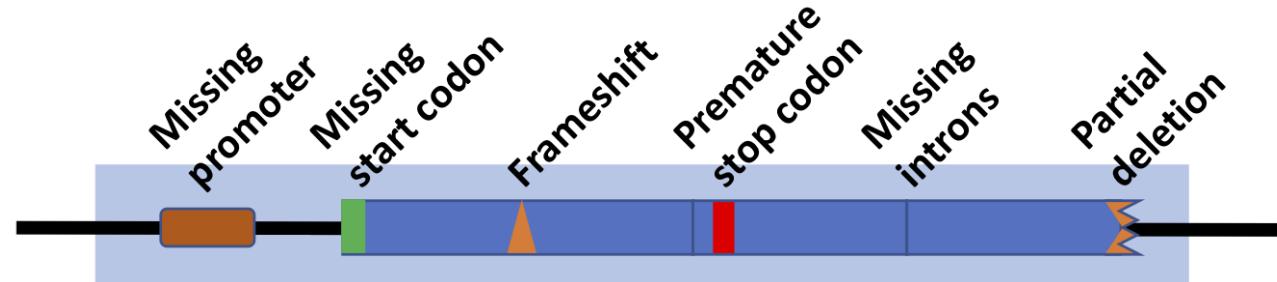
https://www.ncbi.nlm.nih.gov/gene/?Term=related_functional_gene_5310%5Bgroup%5D

Example 5: Human PKD1

What is Pseudogenes?

- Pseudogenes are nonfunctional segments of DNA that resemble functional genes.
- Most arise as superfluous copies of functional genes, either directly by DNA duplication or indirectly by reverse transcription of an mRNA transcript.
- Pseudogenes are usually identified when genome sequence analysis finds gene-like sequences that lack regulatory sequences needed for transcription or translation, or whose coding sequences are obviously defective due to frameshifts or premature stop codons.

Common defects of pseudogenes:



Summary: Nucleotide and Gene

- What do we need the Nucleotide/sequence for?
- Which information does the Gene database give us?