

## NCBI's GenBank Quick Start An NCBI Mini-Course

GenBank® is the NIH nucleotide sequence database. It is part of the International Nucleotide Sequence Database Collaboration, with the DNA DataBank of Japan (DDBJ) and the European Molecular Biology Laboratory (EMBL). These three organizations exchange data on a daily basis that now includes sequences from about 160,000 organisms.

### The mini-course covers following topics:

- \* History and growth of GenBank
- \* Different types of entries and how to identify them from the accession numbers
- \* The format of a typical GenBank entry and understanding the features annotated on it
- \* The submission tools, BankIt and Sequin, and choosing which one to use
- \* Overview of the processing of GenBank entries
- \* Overview of the GenBank FTP directory: how to download sequences in batch
- \* BLAST databases and their equivalents in GenBank
- \* Some tips on searching GenBank efficiently

The following handout includes the screen shots of the exercise demonstrated in the workshop.

URL: <http://www.ncbi.nlm.nih.gov/Class/minicourses/quickgenbank.html>

Course developed by Medha Bhagwat (bhagwat@ncbi.nlm.nih.gov)

## Problems:

1. Retrieve entries containing the word "pannexin" from [Entrez-Nucleotide](#). The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and the Protein Data Bank. The summary of the top 20 entries is displayed by default. Currently, how many entries are there in the Entrez-nucleotides CoreNucleotide, EST and GSS databases associated with pannexin?

Change the pull down menu "Show" to a number large enough to view the summaries of all entries. Determine the source databases and types of entries such as mRNA or genomic. Consult the [Accession Number Prefixes](#) page, if necessary.

Access entries from the CoreNucleotide database. Keep the genomic entries from GenBank by using the Preview/Index page (biomol genomic [properties] AND srcdb genbank [properties]). Is GenBank a redundant database? Note multiple pannexin 1 segmented sets and individual exon sequences.

Let us use the entry AY048509 as a sample GenBank entry with feature annotations. When the entry was last updated? Was its sequence updated after release? Does it contain a complete coding region? What are the locations of the 5'UTR? What is the accession number for the protein entry?

Go back to the Entrez report. Change the source database to RefSeq, molecule type to mRNA and human as organism by using the Preview/index page. Now, RefSeq entries for human pannexin 1, pannexin 2 and pannexin 3 mRNAs are displayed. Save the sequences to a file by changing the "Display" to FASTA and "Send to" file.

Check the sequence revision history of NM\_015368. What is the most recent change? How many times has the sequence of the entry been updated? When did these updates occur?

## GenBank History

Originally built and maintained at  
Los Alamos National Laboratory (LANL)

Early 1990s, Congress awarded responsibility to NCBI

Initially, indexers scanned the literature  
and typed in the sequences

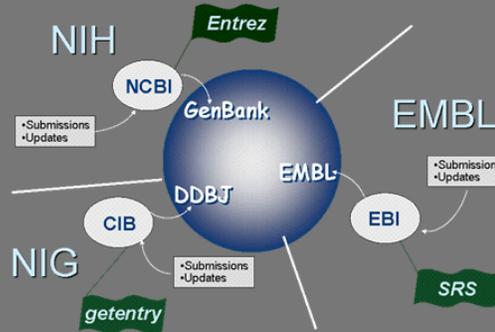
Now sequences are deposited directly by labs

Direct submissions since 1993



## International Nucleotide Sequence Database Collaboration

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



GenBank	RefSeq
Archival/repository	Curated
Redundant	Non-redundant
Submitter owner	NCBI owner
Sequenced	Combined/edited



## Types of Entries

1. Individual mRNA/Genomic
2. Sets such as Pop, Phy, Mut and environmental
3. Segmented sets
4. Expressed Sequence Tags (EST)
5. Genome Survey Sequence (GSS)
6. Sequence Tagged Site (STS)
7. Whole Genome Shotgun (WGS)
8. High Throughput Genomic (HTG)
9. High Throughput cDNA (HTC)
10. Full-Length Insert cDNA (FLIC)
11. Complete genomes
12. Third Party Annotation (TPA)



NCBI

All Databases PubMed Nucleotide Protein Genome

Search Nucleotide for pannexin Go Clear

Limits Preview/Index History Clipboard Details

The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases in these databases continues to grow at an exponential rate. As of April 2006, there are over 130 billion bases in GenBank and RefSeq alone.

**Human Genome**  
Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#).

About Entrez  
Entrez Nucleotide Help | FAQ  
Entrez Tools  
Check sequence revision history  
LinkOut

NCBI

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search Nucleotide for pannexin Go Clear

My NCBI [Sign In] [Register]

About Entrez  
Entrez Nucleotide Help | FAQ  
Entrez Tools  
Check sequence revision history  
LinkOut

Found 188 nucleotide sequences

Please choose one of the following:

164	CoreNucleotide records
24	EST (Expressed Sequence Tags) records
0	GSS (Genome Survey Sequence) records

NCBI Nucleotide

Search CoreNucleotide for pannexin Go Clear Save Search

Limits Preview/Index History Clipboard Details

About Entrez Found 188 nucleotide sequences. CoreNucleotide [164] EST [24]

Display Summary Show 20 Sort by Send to

All: 164 Bacteria: 0 RefSeq: 84 mRNA: 95

Items 1 - 20 of 164 Page 1 of 9 Next

1: [BX957245](#) Reports Links  
Zebrafish DNA sequence from clone CH211-192N14 in linkage group 18 The 3' end of the gene for a novel protein similar to vertebrate megalencephalic leukoencephalopathy with subcortical cysts (MLC1) and the gene for a novel protein similar to vertebrate pannexin 2 (PANX2), complete sequence gi|51534232|emb|BX957245.8|[51534232]

2: [BC152944](#) Reports Order cDNA clone, Links  
Synthetic construct Homo sapiens clone IMAGE:100016148, MGC:184213 pannexin 3 (PANX3) mRNA, encodes complete protein gi|157170155|gb|BC152944.1|[157170155]

3: [NM\\_172729](#) Reports Links  
Mus musculus nucleotide-binding oligomerization domain containing 1 (Nod1), mRNA gi|142361399|ref|NM\_172729.2|[142361399]

4: [NM\\_138952](#) Reports Links  
Mus musculus receptor (TNFRSF)-interacting serine-threonine kinase 2 (Ripk2), mRNA gi|142347210|ref|NM\_138952.3|[142347210]

5: [NM\\_145857](#) Reports Links  
Mus musculus nucleotide-binding oligomerization domain containing 2 (Nod2), mRNA gi|83977453|ref|NM\_145857.2|[83977453]

6: [NM\\_199397](#) Reports Links  
Rattus norvegicus Pannexin 1 (Panx1), mRNA

NCBI Nucleotide

Search CoreNucleotide for pannexin Go Clear Save Search

Limits Preview/Index History Clipboard Details

About Entrez Found 188 nucleotide sequences. CoreNucleotide [164] EST [24]

Display Summary Show 20 Sort by Send to

All: 164 Bacteria: 0 RefSeq: 84 mRNA: 95

Items 1 - 20 of 164 Page 1 of 9 Next

1: [BX957245](#) Reports Links  
Zebrafish DNA sequence from clone CH211-192N14 in linkage group 18 The 3' end of the gene for a novel protein similar to vertebrate megalencephalic leukoencephalopathy with subcortical cysts (MLC1) and the gene for a novel protein similar to vertebrate pannexin 2 (PANX2), complete sequence gi|51534232|emb|BX957245.8|[51534232]

2: [BC152944](#) Reports Order cDNA clone, Links  
Synthetic construct Homo sapiens clone IMAGE:100016148, MGC:184213 pannexin 3 (PANX3) mRNA, encodes complete protein gi|157170155|gb|BC152944.1|[157170155]

3: [NM\\_172729](#) Reports Links  
Mus musculus nucleotide-binding oligomerization domain containing 1 (Nod1), mRNA gi|142361399|ref|NM\_172729.2|[142361399]

4: [NM\\_138952](#) Reports Links  
Mus musculus receptor (TNFRSF)-interacting serine-threonine kinase 2 (Ripk2), mRNA gi|142347210|ref|NM\_138952.3|[142347210]

5: [NM\\_145857](#) Reports Links  
Mus musculus nucleotide-binding oligomerization domain containing 2 (Nod2), mRNA gi|83977453|ref|NM\_145857.2|[83977453]

6: [NM\\_199397](#) Reports Links  
Rattus norvegicus Pannexin 1 (Panx1), mRNA

NCBI Nucleotide

Search CoreNucleotide for pannexin Go Clear Save Search

Limits Preview/Index History Clipboard Details

About Entrez

Entrez Nucleotide Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI (Cubby)

Related resources BLAST

Reference sequence

Found 188 nucleotide sequences. CoreNucleotide [164] EST [24]

Display Summary Show [20] Sort by Send to

All: 164 Bacteria: 0 RefSeq: 84 mRNA: 95

Items 1 - 20 of 164 Page 1 of 9 Next

1: BX957245 Reports Links  
Zebrafish DNA sequence from clone CH211-192N14 in linkage group 18 The 3' end of the gene for a novel protein similar to vertebrate megalencephalic leukoencephalopathy with subcortical cysts (MLC1) and the gene for a novel protein similar to vertebrate pannexin 2 (PANX2), complete sequence gi|51534232|emb|BX957245.8|[51534232]

2: BC152944 Reports Links  
Synthetic construct Homo sapiens clone IMAGE:100016148, MGC:184213 pannexin 3 (PANX3) mRNA, encodes complete protein gi|157170155|gb|BC152944.1|[157170155]

NCBI Nucleotide

Search CoreNucleotide for pannexin Preview Go Clear

Limits History Clipboard Details

About Entrez

Entrez Nucleotide Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI (Cubby)

Related resources BLAST

Reference sequence project

Search for Genes

Submit to GenBank

Search for full length cDNAs

Accession All Fields  
Author  
EC/IRN Number  
Feature key  
Filter  
Gene Name  
Genome Project  
Issue  
Journal  
Keyword  
Modification Date  
Organism  
Page Number  
Primary Accession  
Properties  
Protein Name  
Publication Date  
SeqID String  
Sequence Length  
Substance Name  
Text Word  
Title  
Volume

Click Preview to see only the number of search results.  
To save search indefinitely, click query # and select Save in My NCBI.  
To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

Most Recent Queries

Query	Time	Result
#88 Search pannexin	15:30:30	164
#82 Search EF535850	13:57:53	1
#79 Select 1 document(s)	13:31:23	1

Add Term(s) to Query or View Index:  
Enter a term in the text box; use the pull-down menu to specify a search field.  
Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.  
Multiple terms selected from index will be ORed; click AND to add to search.

Properties All Fields Preview Index

Click AND OR NOT to add a term to the query box

NCBI Nucleotide

Search CoreNucleotide for pannexin Preview Go Clear

Limits Preview/Index History Clipboard Details

About Entrez

Entrez Nucleotide Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI (Cubby)

Related resources BLAST

Reference sequence project

Search for Genes

Submit to GenBank

Search for full length cDNAs

Enter terms and click Preview to see only the number of search results.  
To save search indefinitely, click query # and select Save in My NCBI.  
To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

Most Recent Queries

Query	Time	Result
#88 Search pannexin	15:30:30	164
#82 Search EF535850	13:57:53	1
#79 Select 1 document(s)	13:31:23	1

Add Term(s) to Query or View Index:  
Enter a term in the text box; use the pull-down menu to specify a search field.  
Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.  
Multiple terms selected from index will be ORed; click AND to add to search.

Properties biomal Preview Index

Click AND OR NOT to add a term to the query box

biomal cma (31295)  
biomal genomic (38976307)  
biomal genomic mma (42)  
biomal mma (2483185)  
biomal other (26417)  
biomal other genetic (21907)  
biomal pre ma (1024)  
biomal ma (2728204)  
biomal rma (7869)  
biomal scma (114)

NCBI Nucleotide

Search CoreNucleotide for [pannexin AND "biomol genomic"[Properties]] Preview Go Clear

Limits Preview/Index History Clipboard Details

Most Recent Queries

Search	Time	Result
#88 Search pannexin	15:30:30	164
#82 Search EF535850	13:57:53	1
#79 Select 1 document(s)	13:31:23	1

Add Term(s) to Query or View Index:

- Enter a term in the text box, use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
- Multiple terms selected from Index will be ORed; click AND to add to search.

Properties srcdb Preview Index

Click AND OR NOT to add a term to the query box

srcdb ddbj (2348287)  
 srcdb ddbj/embl/genbank (39456301)  
 srcdb embl (3597543)  
 srcdb genbank (33500471)  
 srcdb pdb (7932)  
 srcdb refseq (2394753)  
 srcdb refseq inferred (1333)  
 srcdb refseq known (999755)  
 srcdb refseq model (433891)  
 srcdb refseq predicted (22149)

NCBI Nucleotide

Search CoreNucleotide for [pannexin AND "biomol genomic"[Properties] AND "srcdb"] Save Search

Limits Preview/Index History Clipboard Details

Found 15 nucleotide sequences. CoreNucleotide [15]

Display Summary Show 20 Sort by Send to

All: 15 Bacteria: 0 RefSeq: 0 mRNA: 0

Items 1 - 15 of 15 One page.

<input type="checkbox"/>	1: CH473993 Reports	Rattus norvegicus 181000120144526 genomic scaffold, whole genome shotgun sequence gi 74325685 gb CH473993.2 [74325685]	Links
<input type="checkbox"/>	2: CH474096 Reports	Rattus norvegicus 342000077087993 genomic scaffold, whole genome shotgun sequence gi 74325592 gb CH474096.2 [74325592]	Links
<input type="checkbox"/>	3: CH466550 Reports	Mus musculus 232000009743879 genomic scaffold, whole genome shotgun sequence gi 74140747 gb CH466550.2 [74140747]	Links
<input type="checkbox"/>	4: CH466522 Reports	Mus musculus 232000009795437 genomic scaffold, whole genome shotgun sequence gi 70980458 gb CH466522.1 [70980458]	Links
<input type="checkbox"/>	5: CH471065 Reports	Homo sapiens 211000035833829 genomic scaffold, whole genome shotgun sequence gi 71517789 gb CH471065.1 [71517789]	Links
<input type="checkbox"/>	6: CH471138 Reports	Homo sapiens 211000035835149 genomic scaffold, whole genome shotgun sequence gi 71514907 gb CH471138.1 [71514907]	Links

<input type="checkbox"/> 6:	<a href="#">CH471138</a>	Reports	Homo sapiens 211000035835149 genomic scaffold, whole genome shotgun sequence gi 71514907 gb CH471138.1 [71514907]	Links
<input type="checkbox"/> 7:	<a href="#">AF398508</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, exons 3, 4, 5, and complete cds gi 15193202 gb AF398508.1 AF398506S3[15193202]	Links
<input type="checkbox"/> 8:	<a href="#">AF398507</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, exon 2 gi 15193201 gb AF398507.1 AF398506S2[15193201]	Links
<input type="checkbox"/> 9:	<a href="#">AF398506</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, exon 1 gi 15193200 gb AF398506.1 AF398506S1[15193200]	Links
<input type="checkbox"/> 10:	<a href="#">AH010983</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, complete cds gi 15193199 gb AH010983.1 SEG_AF398506S[15193199]	Links
<input type="checkbox"/> 11:	<a href="#">AY048509</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, partial cds gi 15808666 gb AY048509.1 [15808666]	Links
<input type="checkbox"/> 12:	<a href="#">AF283663</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, exons, partial cds gi 14794510 gb AF283663.1 AF283661S3[14794510]	Links
<input type="checkbox"/> 13:	<a href="#">AF283662</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, exon gi 14794509 gb AF283662.1 AF283661S2[14794509]	Links
<input type="checkbox"/> 14:	<a href="#">AF283661</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, exon gi 14794508 gb AF283661.1 AF283661S1[14794508]	Links
<input type="checkbox"/> 15:	<a href="#">AH010945</a>	Reports	Homo sapiens pannexin 1 (PANX1) gene, partial cds gi 14794507 gb AH010945.1 SEG_AF283661S[14794507]	Links

Limits Preview/Index History

Display  Show  Send to

Range: from  to   Reverse complemented strand Features:  Refresh

---

1: [AY048509](#) Reports Homo sapiens pann...[gi:15808666]

		Reports	
<a href="#">Features</a>	<a href="#">Sequence</a>	▶ ASN.1	
		▶ XML	
LOCUS	AY048509	▶ Summary	1430 bp DNA linear PRI 29-SEP-2001
DEFINITION	Homo sa	▶ Brief	1 (PANX1) gene, partial cds.
ACCESSION	AY048509	▶ FASTA	
VERSION	AY048509	▶ TinySeq XML	56
KEYWORDS	.	▶ GenBank	
SOURCE	Homo sa	▶ INSDSeq XML	
ORGANISM	Homo sa	▶ GenBank(Full)	Chordata; Craniata; Vertebrata; Euteleostomi;
		▶ GI List	Mammalia; Archontoglires; Primates; Haplorrhini;
		▶ Graphic	Catarrhini; Homo.
REFERENCE	1 (bases 1 to 1430)	▶ <b>Revision History</b>	
AUTHORS	Baranova, A., Ivanov, D., Skoblov, M., Pestova, A., Kelmanson, I., Shagin, D., Usman, N., Lukyanov, S. and Panchin, Y.		
TITLE	Mammalian pannexin family homologous to invertebrate gap-junction proteins are differentially expressed in nervous tissue		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1430)		
AUTHORS	Ivanov, D. and Baranova, A.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-JUL-2001) Genome Analysis Laboratory, Vavilov Institute of General Genetics, Gubkina 3, Moscow 119991		
FEATURES	Location/Qualifiers		
source	1..1430		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="11"		
	/map="between D11S917 and D11S4176"		
	/clone="LANL LA1139-O20"		
<a href="#">gene</a>	606..>1430		
	/gene="PANX1"		
<a href="#">mRNA</a>	606..>1170		
	/gene="PANX1"		
	/product="pannexin 1"		
<a href="#">exon</a>	606..1170		

**Sequence Revision History**

PubMed  Nucleotide  Protein  Genome  Structure  PMC  Taxonomy  OMIM

Find (Accessions, GI numbers or *Fasta style Seqids*)

About Entrez

**Revision history for [AY048509](#)**

GI	Version	Update Date	Status
15808666	1	<a href="#">Sep 29, 2001 9:01 PM</a>	Live

Accession [AY048509](#) was first seen at NCBI on Sep 29 2001 9:01 PM

Search for Genes  
 Entrez  
 LocustLink provides  
 nucleotide information for

```

TITLE      Direct Submission
JOURNAL    Submitted (27-JUL-2001) Genome Analysis Laboratory, Vavilov
           Institute of General Genetics, Gubkina 3, Moscow 119991
FEATURES   Location/Qualifiers
    source   1..1430
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="11"
             /map="between D11S917 and D11S4176"
             /clone="LANL LA1139-O20"
    gene     606..>1430
             /gene="PANX1"
    mRNA    606..>1170
             /gene="PANX1"
             /product="pannexin 1"
    exon    606..1170
             /gene="PANX1"
    CDS     990..>1170
             /number=1
             /gene="PANX1"
             /note="putative gap junction protein; contains four
             putative transmembrane domains; similar to innexin gap
             junction gene family"
             /codon_start=1
             /product="pannexin 1"
             /protein_id="AAL06604.1"
             /db_xref="GI:15808667"
             /translation="MAIAQLATEYVVFDFLLKEPTEPKFKGLRLELAVDKHMVTCIAVG
             LPLLLISLAFQEISI"

ORIGIN
1  accctgcaggc cccctctctct ctgctggctc ccggaacaag gctctgattg ggatggcaga
61 ggaagagaa acgggaccag cagcgctact caggcctcga aactccacac tcactaccgt

```

NCBI Nucleotide

Search CoreNucleotide for pannexin AND "biomol genomic"[Properties] AND "srcdb genbank"[Properties] Go Clear

Limits Preview/Index History Clipboard Details

About Entrez

- Enter terms and click Preview to see only the number of search results.
- To save search indefinitely, click query # and select Save in My NCBI.
- To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

Entrez Nucleotide Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI (Cubby)

Related resources BLAST

Reference sequence project

Search for Genes

Submit to GenBank

Search for full length cDNAs

Most Recent Queries

Search	Time	Result
#97 Select 1 document(s)	15:45:13	1
#94 Search pannexin AND "biomol genomic"[Properties] AND "srcdb genbank"[Properties]	15:42:59	15
#91 Search pannexin AND "biomol genomic"[Properties] AND "srcdb refseq"[Properties]	15:42:21	40

Add Term(s) to Query or View Index:

- Enter a term in the text box; use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
- Multiple terms selected from Index will be ORed; click AND to add to search.

Organism: human Preview Index

Click AND OR NOT to add a term to the query box

human (3026094) Up

human adenovirus 1 (54)

human adenovirus 1 isolate from a cat (3)

human adenovirus 11 (40)

human adenovirus 12 (28)

human adenovirus 14 (16)

human adenovirus 16 (19)

human adenovirus 19a (2)

human adenovirus 19p (2)

human adenovirus 2 (141) Down

NCBI Nucleotide

Search CoreNucleotide for pannexin[All Fields] AND "biomol mRNA"[Properties] Go Clear Save Search

Limits Preview/Index History Clipboard Details

Found 3 nucleotide sequences. CoreNucleotide [3]

Display Summary Show 20 Sort by Send to

All: 3 Bacteria: 0 RefSeq: 3 mRNA: 3

Items 1 - 3 of 3 One page.

1: [NM\\_052839](#) Reports Links  
Homo sapiens pannexin 2 (PANX2), mRNA  
gi|39995065|ref|NM\_052839.2| [39995065]

2: [NM\\_015368](#) Reports Order cDNA clone, Links  
Homo sapiens pannexin 1 (PANX1), mRNA  
gi|39995063|ref|NM\_015368.3| [39995063]

3: [NM\\_052959](#) Reports Links  
Homo sapiens pannexin 3 (PANX3), mRNA  
gi|39995066|ref|NM\_052959.2| [39995066]

NCBI Nucleotide

Search CoreNucleotide for Go Clear

Limits Preview/Index History Clipboard Details

Display GenBank Show 5 Send to Hide:  sequence  all but gene, CDS and mRNA features

Range: from begin to end  Reverse complemented strand Features:  SNP  STS  Exon Refresh

1: [NM\\_015368](#) Reports Links  
pann...[gi:39995063]

Comment Features Summary

LOCUS NM\_015368 Summary 2782 bp mRNA linear PRI 03-SEP-2007

DEFINITION Homo sap Brief 1 (PANX1), mRNA.

ACCESSION NM\_015368 FASTA

VERSION NM\_015368 TinySeq XML 63

KEYWORDS . GenBank

SOURCE Homo sap INSDSeq XML

ORGANISM Homo sap GenBank(Full)  
Eukaryota  
Mammalia  
Carnivora  
Primates  
Hominidae  
Homo

REFERENCE 1 (bases 1 to 2782)  
AUTHORS Locovei,S., Scemes,E., Qiu,F., Spray,D.C. and Dahl,G.  
TITLE Pannexin1 is part of the pore forming unit of the P2X(7) receptor death complex  
JOURNAL FEBS Lett. 581 (3), 483-488 (2007)  
PUBMED 17240370  
REMARK GeneRIF: pannexin1 appears to be the molecular substrate for the permeabilization pore (or death receptor channel) recruited into the P2X(7)R signaling complex

REFERENCE 2 (bases 1 to 2782)  
AUTHORS Pelegri,F. and Surprenant,A.  
TITLE Pannexin-1 mediates large pore formation and interleukin-1beta release by the ATP-gated P2X7 receptor  
JOURNAL EMBO J. 25 (21), 5071-5082 (2006)  
PUBMED 17036048  
REMARK GeneRIF: pannexin-1 is required for processing of caspase-1 and release of mature IL-1beta induced by P2X(7) receptor activation. GeneRIF: Panx1 in the plasma membrane of the macrophage couples to the purinergic P2X7 receptor and permeabilize the macrophage membrane. This signaling is required for the processing and release

**Sequence Revision History**

difference between I and II as

**Revision history for [NM\\_015368](#)**

GI	Version	Update Date	Status	I	II
39995063	3	<a href="#">Sep 3 2007 3:02 AM</a>	Live	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Jun 26 2007 10:53 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Jun 22 2007 11:28 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Mar 25 2007 11:40 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Jan 14 2007 12:11 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Dec 3 2006 10:40 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Nov 17 2006 11:49 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Nov 3 2006 1:07 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Oct 29 2006 10:26 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Oct 15 2006 10:59 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Aug 6 2006 2:43 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Jan 29 2006 11:11 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Sep 24 2005 6:44 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Jul 8 2005 1:43 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Apr 23 2005 9:43 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Mar 2 2005 1:02 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Oct 27 2004 9:40 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Aug 23 2004 3:34 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Jan 27 2004 9:22 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Dec 23 2003 12:52 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
39995063	3	<a href="#">Dec 17 2003 5:18 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
29837657	2	<a href="#">Oct 5 2003 3:40 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
29837657	2	<a href="#">Sep 7 2003 4:39 PM</a>	Dead	<input type="radio"/>	<input type="radio"/>
29837657	2	<a href="#">May 7 2003 4:24 AM</a>	Dead	<input type="radio"/>	<input type="radio"/>

**Sequence Revision History**

NCBI

PubMed Nucleotide Protein Genome Structure PPMC Taxonomy OMIM Books

Find (Accessions, GI numbers or Fasta style Seqids)

About Entrez

Entrez

Search for Genes

Help/FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

My NCBI (Cubby)

Related resources

BLAST

Reference sequence project

Entrez Gene

Clusters of orthologous groups

Show difference in  format

GI	Version	Update Date
39995063	3	Sep 3 2007 3:02 AM
39995063	3	Jun 26 2007 10:53 PM

LOCUS NM\_015368 2782 bp mRNA linear PRI 03-SEP-2007  
 ELOCUS NM\_015368 2782 bp mRNA linear PRI 26-JUN-2007  
 DEFINITION Homo sapiens pannexin 1 (PANX1), mRNA.  
 ACCESSION NM\_015368  
 VERSION NM\_015368.3 GI:39995063  
 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 2782)  
 [ SKIP: unchanged/identical feature/sequence/publication information is not displayed ]

PUBMED 17240370  
 REMARK GeneRIF: pannexin1 appears to be the molecular substrate for the permeabilization pore (or death receptor channel) recruited into the P2X(7)R signaling complex

REFERENCE 2 (bases 1 to 2782)  
 AUTHORS Pelegrin, P. and Surprenant, A.  
 TITLE Pannexin-1 mediates large pore formation and interleukin-1beta release by the ATP-gated P2X7 receptor  
 JOURNAL EMBO J. 25 (21), 5071-5082 (2006)  
 PUBMED 17036048

REMARK GeneRIF: pannexin-1 is required for processing of caspase-1 and release of mature IL-1beta induced by P2X(7) receptor activation.  
 GeneRIF: Panx1 in the plasma membrane of the macrophage couples to GeneRIF: Panx1 in the plasma membrane of the macrophage couples to the purinergic P2X7 receptor and permeabilize the macrophage membrane, this signaling is required for the processing and release of interleukin-1beta in response to P2X7 receptor activation.  
 REMARK GeneRIF: pannexin-1 is required for processing of caspase-1 and release of mature IL-1beta induced by P2X(7) receptor activation.

**Submit to GenBank**

NCBI

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

SITE MAP  
Guide to NCBI resources

Accession numbers  
For manuscript citation

BankIt

Sequin

SequinMacroSend  
Upload .sqn files directly

TBL2ASN  
Command line

▶ **Submitting Sequence Data to GenBank**

▶ **Submit now!!**

[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

▶ [GenBank](#)

## Submission Procedure

Individual mRNA/Genomic	<a href="#">BankIt</a>
Sets	Sequin
EST	batch-sub
GSS	batch-sub
STS	batch-sub
WGS	tbl2asn/ftp
HTG	fa2htgs/ftp
HTC	tbl2asn
FLIC	tbl2asn/ftp
Complete genomes	tbl2asn
TPA	<a href="#">BankIt/Sequin</a>

## Accession Numbers

### [GenBank/DDBJ/EMBL](#)

Nucleotide: 1 letter + 5 numbers

OR 2 letters + 6 numbers (AF123456)

Protein: 3 letters + 5 numerals

WGS: 4 letters + 2 numbers

### [RefSeq](#)

2 letters + underscore + 6 numbers (NM\_123456)

## Processing of Direct Submissions

Entries are checked for:

- Biological validity
- Vector contamination
- Publication Status
- Formatting/spelling

Completed sequences sent to submitters for review

5 days to review/update

Can hold until published

Released records available for searching in [Entrez](#) and BLAST within few days

Release to ftp every 2 months and daily updates



 **NCBI GenBank Overview**

[PubMed](#) [Entrez](#) [BLAST](#) [OMIM](#) [Books](#) [Taxonomy](#) [Structure](#)

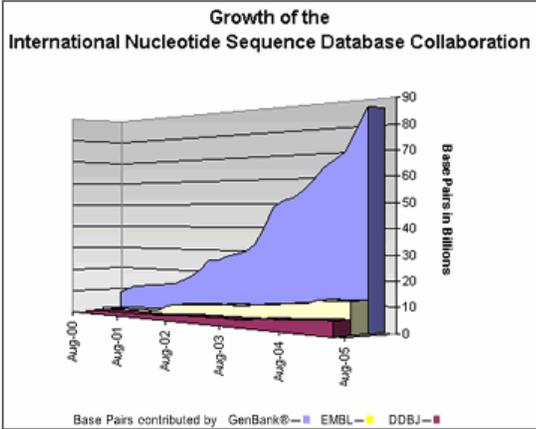
Search  for

[NCBI](#)  
[SITE MAP](#)  
[Submit to GenBank](#)  
[Updates](#)  
[Search GenBank](#)  
[Entrez Nucleotide](#)  
[BLAST](#)

### International sequence databases exceed 100 gigabases

In August 2005, the INSDC announced the DNA sequence database exceeded 100 gigabases. GenBank is proud of its contributions toward this milestone. We thank all the scientists who have worked through the submission process at GenBank and made their sequence data available to the world. See the related [press release](#).

#### Growth of the International Nucleotide Sequence Database Collaboration



Date	GenBank (Billions)	EMBL (Billions)	DDBJ (Billions)	Total (Billions)
Aug-00	~1	~1	~1	~3
Aug-01	~5	~2	~2	~9
Aug-02	~15	~5	~3	~23
Aug-03	~35	~10	~5	~50
Aug-04	~60	~15	~8	~83
Aug-05	~80	~20	~10	~110

## ► What is GenBank?

GenBank<sup>®</sup> is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research 2007 Jan ;35\(Database issue\):D21-5](#)). There are approximately 65,369,091,950 bases in 61,132,599 sequence records in the traditional GenBank divisions and 80,369,977,826 bases in 17,960,667 sequence records in the WGS division as of August 2006.

The complete [release notes](#) for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

An example of a GenBank [record](#) may be viewed for a *Saccharomyces cerevisiae* gene.

```
GBREL.TXT      Genetic Sequence Data Bank
                August 15 2007

                NCBI-GenBank Flat File Release 161.0

                Distribution Release Notes

76146236 loci, 79525559650 bases, from 76146236 reported sequences

This document describes the format and content of the flat files that
comprise releases of the GenBank nucleotide sequence database. If you
have any questions or comments about GenBank or this document, please
contact NCBI via email at info@ncbi.nlm.nih.gov or:
```

```
GenBank
National Center for Biotechnology Information
National Library of Medicine, 38A, 8N805
8600 Rockville Pike
Bethesda, MD 20894
USA
Phone: (301) 496-2475
Fax: (301) 480-9241
```

```
GenBank releases do not include sequence records that originate from
third-parties (TPA) or from NCBI's Reference Sequence (RefSeq) project.
Rather, GenBank is the archival/primary resource which those other
efforts draw upon. For information about TPA and RefSeq, please visit:
```

```
http://www.ncbi.nlm.nih.gov/Genbank/TPA.html
http://www.ncbi.nlm.nih.gov/RefSeq
```

### =====

#### TABLE OF CONTENTS

### =====

1. INTRODUCTION
  - 1.1 Release 161.0
  - 1.2 Cutoff Date
  - 1.3 Important Changes in Release 161.0
  - 1.4 Upcoming Changes
  - 1.5 Request for Direct Submission of Sequence Data
  - 1.6 Organization of This Document
2. ORGANIZATION OF DATA FILES

## GenBank Database Divisions

BCT	Bacterial sequences
PRI	Primate sequences
ROD	Rodent sequences
MAM	Other mammalian sequences
VRT	Other vertebrate sequences
INV	Invertebrate sequences
PLN	Plant and Fungal sequences
VRL	Viral sequences
PHG	Phage sequences
SYN	Synthetic sequences
UNA	Unannotated sequences

BLAST database: nr, month except for  
environmental samples

## GenBank Database Divisions

Division		BLAST database
EST	Expressed Sequence Tags	<a href="#">dbest</a> , month
STS	Sequence Tagged Sites	<a href="#">dbsts</a> , month
GSS	Genome Survey Sequences	<a href="#">dbgss</a> , month
HTG	High Throughput Genomic sequences Phase 0, 1 and 2	<a href="#">htgs</a> , month
PAT	Patent sequences	patent, month
HTC	High Throughput <a href="#">cDNA</a>	nr, month
ENV	Environmental sampling sequences	<a href="#">env</a> , month

**Problem 2:** Retrieve entries containing the word "FOXP2" from [Entrez-Nucleotide](#). The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The summary of top 20 entries is displayed by default. Currently, how many entries are there in the Entrez-nucleotides CoreNucleotide, EST and GSS databases associated with FOXP2?

Change the pull down menu "Show" to a number large enough to view the summaries of all entries. Determine the source databases and types of entries such as mRNA or genomic. Consult the [Accession Number Prefixes](#) page, if necessary.

Access entries from the CoreNucleotide database. Keep the genomic entries from GenBank by using the Preview/Index page (biomol genomic [properties] AND srcdb genbank [properties]). Is GenBank a redundant database? Note the multiple FOXP2 exon sequence entries. Are they duplicate entries?

Let us use the entry AF515032 as a sample GenBank entry with feature annotations. When the entry was last updated? Was its sequence updated after release? Does it contain a complete coding region? What is the accession number for the protein entry?

Go back to the Entrez report. Change the source database to RefSeq, molecule type to mRNA and human as organism by using the Preview/index page. Note that some of the entries are not for the FOXP2 gene. Restrict to the FOXP2 gene using the Preview/index page and the gene name field from the "All Fields" pull down menu. How many human FOXP2 transcript variants have been annotated? Save the sequences to a file by changing the "Display" to FASTA and "Send to" file.

Check the sequence revision history of NM\_148900. What is the most recent change? How many times has the sequence of the entry been updated? When did these updates occur?

**Problem 3:** Retrieve entries containing the word "bacteriophage" from [Entrez-Nucleotide](#). The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. Currently, how many entries are there in the Entrez-nucleotides CoreNucleotide, EST and GSS databases associated with bacteriophage?

Click on the CoreNucleotide link. The summary of top 20 entries is displayed by default. Select '500' from the Show pull-down menu to see a variety of different genbank entries. Determine the source databases and types of entries such as mRNA or genomic. Consult the [Accession Number Prefixes](#) page, if necessary.

Next, keep the GenBank entries containing the terms bacteriophage and complete sequence in the title of the record by using the Preview/Index page (bacteriophage[Title] AND "complete sequence"[Title] AND "srcdb genbank"[Properties]). Are all of these complete bacteriophage genome sequences?

Let us use the entry U32222 as a sample GenBank entry with feature annotations. Which bacteriophage sequence does it represent? When was the entry last updated? Was its sequence updated after the first release? What are the different ways to determine this? Access the Revision History page. Notice the 3 other accession numbers on the upper right side of the page. What do they

represent? Return to the GenBank entry.

Count the number of entries under the Reference field. How many of these are actual PubMed entries? What are the other types of References? Now scroll to the Features table. This GenBank record is a well-annotated phage genome entry. Note the different types of features are annotated here. Navigate to the variation feature at position 3980. What is the nucleotide sequence in this entry at position 3980? What is it replaced with in the Vam38 mutant? What is the amino acid change caused due to this nucleotide change? What is the translation table (genetic code) used for the translation of the bacteriophage sequence? Access that [translation table](#). What are the codons for glutamine using this translation table? Find the codon that is present in the DNA sequence of this bacteriophage (displayed at the bottom of the record) for the wild type amino acid (glutamine). Determine the variant codon in the Vam38 mutant. Which gene includes this variation? Click on the corresponding CDS feature. What sequence is now displayed? Navigate back to the full GenBank record for U32222. Display the entry in the "Graph" format to display the graphical representation of the annotation of various features on the entry. Finally, save the sequence to a file by changing the "Display" to FASTA and "Send to" file.