• One of the most important aspects of bioinformatics is identifying genes within a long DNA sequence.

Until the development of bioinformatics, the only way to locate genes along the chromosome was to study their behavior in the organism (in vivo) or isolate the DNA and study it in a test tube (in vitro). **Bioinformatics allows scientists to** make guesses about where genes are located simply by analyzing sequence data using a computer (in silico).

In principle, locating genes should be easy. DNA sequences that code for proteins begin with the three bases ATG that code for the amino acid methionine and they end with one or more stop codons; either TAA, TAG or TGA.



The codons between the start and stop signals code for the various amino acids of the gene product but do not include any of the three stop codons.

Open reading frame





When examining an unknown DNA sequence, one indication that it may be part of a gene is the presence of an open reading frame (ORF). An ORF is any stretch of DNA that when transcribed into RNA has no stop codon Computer programs can be used to check an unknown DNA sequence for ORFs.

The program transcribes each DNA strand into its complementary RNA sequence and then translates the RNA sequence into an amino acid sequence.

But there are problems

1. Each DNA strand can be read in three different reading frames. This means that the computer must perform six different translations for any given double-stranded DNA sequence.

ATG CCA GAG CAT AAC	Reading Frame 1
ATGC CAG AGC ATA AC	Reading Frame 2
ATGCC AGA GCA TAA C	Reading Frame 3

2. The presence of an ORF doesn't guarantee that the DNA sequence is part of a gene. We expect that, just by chance, there will be some long stretches of DNA that do not contain stop codons yet are not parts of genes. Likewise, codons for methionine do not always mark the start of a gene sequence. Methionine codons are also found within genes.

<u>Nevertheless</u>, <u>searching for ORFs</u> <u>identifies regions of the DNA</u> <u>sequence that might be parts of</u> genes. ORFs are just one feature that a computer program looks for when locating potential genes.

Genes are also characterized by specific control sequences involved in transcription and translation signals.

<u>A computer program looking</u> for ORFs must read each DNA strand in the 5' to 3' direction.

The end of the strand with the phosphate group is called the 5' end and the opposite end with the sugar is called the 3' end. The two strands run in opposite directions. That is, one strand runs in the 5' to 3' direction while the complementary strand runs in the 3' to 5' direction.



The enzymes and ribosomes that carry out protein synthesis only work in one direction.

During transcription, the mRNA is made in the 5' to 3' direction.

During translation, the mRNA is read in the 5' to 3' direction.

Why it is important identifying genes using bioinformatics

Almost half of the genes identified by the Human Genome Project have no known function.

Researchers are using bioinformatics to identify genes, establish their functions, and develop gene-based strategies for preventing, diagnosing, and treating disease. When a computer program finds a DNA sequence that satisfies all of these gene features (an ORF plus the appropriate control sequences), it identifies the sequence as likely coming from a gene.

Only testing the DNA sequence in the laboratory can prove that the gene is active in an organism however.

Prokaryotes

Archeal and bacterial genes typically comprise uninterrupted stretches of DNA between a start codon and a stop codon.

Therefore, a prokaryotic gene can be defined simply as the longest ORF for a given region of DNA. Translation of a DNA sequence in all six reading frames is a straightforward task, which can be performed using translation tools (eg ORF Finder). Of course, this approach is oversimplified and may result in a certain number of incorrect gene predictions, although the error rate is rather low.

It is always desirable to have some additional evidence that a particular ORF actually encodes a protein. Such evidence can be obtained using various methods. *<u>The ORF encodes a protein that is similar</u> to previously described ones.

*The ORF has a typical GC content, codon frequency, or oligonucleotide composition

*The ORF is preceded by a typical ribosome-binding site.

*The ORF is preceded by a typical promoter.

The most reliable of these approaches is a database search for homologs.

In several useful tools, DNA translation is seamlessly bound to the database searches.

S NCBI ORF Finder (Open Reading Frame Finder)

PubMed	Entrez	BLAST	OMIM	Taxonomy	Structure
ICBI	The C	RF Finder (O	pen Reading	Frame Finder)	is a graphical
-	analys minim	sis tool which f num size in a ι	finds all open Iser's sequer	i reading frames nce or in a sequ	s of a selectat Jence alreadv
			_	•	

The ORF Finder (Open Reading Frame Finder) is a graphical analysis tool which finds all open reading frames in a user's sequence or in a sequence already in the database, using the standard or alternative genetic codes.

The deduced amino acid sequence can be saved in various formats and searched against the sequence database using the BLAST programs.

FROM:	TO:		
<u>Genetic codes</u>			
1 Standard			~

Unicellular eukaryotes

Gene organization is more complex and poses additional problems for intron identification. In some simple eukaryotes genes are guite compact with few introns, and gene prediction can be done using the same approach as in prokaryotic genomes.

Multicellular eukaryotes

Gene organization is so complex that identification poses a major problem. Eukaryotic genes are often separated by large intergenic regions, and the genes themselves contain numerous introns, many of them long.

The coding regions compose only a minor portion of the gene.

In humans because of the clinical phenotype of the mutations we know the "correct" mRNA sequence and can identify various alternatively spliced variants as mutations.

For the majority of the human genes, multiple alternative forms are part of the regular expression pattern, and correct gene prediction ideally should

identify all of these forms.

Ideally, gene prediction should identify all exons and introns, including those in the 5'untranslated region (5'-UTR) and the 3'-UTR of the mRNA.

For practical purposes, however, it is useful to assemble at least the coding exons correctly because this allows one to deduce the protein sequence.



Since the major goal of genomic sequencing projects is to identify and characterize genes, <u>Entrez</u> <u>Gene</u> has been implemented at the NCBI to organize information about genes.



It provides information and links related to a specific gene. Each record is assigned a unique identifier, the GeneID.

Records are established for known or predicted genes, which are defined by nucleotide sequence or map position.



S NCBI Resources 🖸 How To 🖸			
Entrez Gene Genes and mapped phenotypes	Search: Gene Limits Advance	ced search Help Search Clear	
-	Canada and a state of the	Welcome to Entrez Gene	
		Entrez Gene maintains information about genes from genomes of interest to the RefSeq group.	

Using Entrez Gene	Entrez Gene Tools	Other Resources
Entrez Gene Quick Start	Submit GeneRIFs	HomoloGene
FAQ	Submit Correction	OMIM
Download/FTP	Statistics	RefSeq
Gene Mailing List	BLAST	RefSeqGene
RefSeq Mailing List	Genome Workbench	UniGene
Entrez Gene News	Splign	Protein Clusters

You are here: NCBI > Genes & Expression > Gene

GETTING STARTED	
NCBI Help Manual	
NCBI Handbook	
Training & Tutorials	

RESOURCES
Literature
DNA & RNA
Proteins
Sequence Analysis
Genes & Expression
Genomes & Maps
Domains & Structures
Genetics & Medicine
Taxonomy
Data & Software
Training & Tutorials

POPULAR
PubMed
Nucleotide
BLAST
PubMed Central
Gene
Bookshelf
Protein
OMIM
Genome
SNP
Structure

8

FEATURED		
GenBank		
Reference Sequences		
Map Viewer		
Genome Projects		
Human Genome		
Mouse Genome		
Influenza Virus		
Primer-BLAST		
Sequence Read Archive		

NCBI INFORM About NCBI Research at N NCBI Newslett

NCBI FTP Site

Gene: the central function

is to establish unique identifiers for genes that can be tracked and, in so doing, support accurate connections with the defining sequences, nomenclature and other descriptors.

Gene: A curated database

Gene can be considered a curated database because many of the contributing databases are curated. Additionally, records in Gene may

be reviewed by NCBI staff.

Gene collectcs confirmed, inferred, validated.....records

- Entrez Gene includes records for confirmed genes and for genes predicted by annotation processes.
- The evidence for a gene can be inferred from the status of the RefSeq.

For example, RefSeqs that are termed as predicted or model have less supporting evidence than those in the validated, provisional, or reviewed categories.

Gene: a guide to additional information

Entrez Gene is not comprehensive; rather, it serves as a guide to additional information in other databases. Connections are supplied to Entrez Nucleotide, Entrez Protein, and Blink, where more sequences with significant similarity can be retrieved. In addition links out to external databases support navigation to more gene-specific information.

Entrez Gene to focus on the <u>human cftr gene</u>

S NCBI Resources	♥ How To ♥	My
Gene	Gene - Search	
	Limits Advanced	
Display Settings: 🖓 Full	Report Send to: 🖓	
		Table of contents
CFTR cystic fibro	sis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) [Homo sapiens]	Summary
Gene ID: 1080, updated	on 23-Oct-2011	Genomic context
		Genomic regions, transcripts, and produ
 Summary 		Bibliography
Official Symbol	CFTR provided by HGNC	Phenotypes
Official Full Name	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) provided by <u>HGNC</u>	Interactions
Primary source	<u>HGNC:1884</u>	General gene info
Locus tag	tcag7.78	General protein info
See related	Ensembl:ENSG0000001626; HPRD:03883; MIM:602421	Reference sequences
Gene type	protein coding	Related sequences
Organism	Homo sapiens	Additional links
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	
Also known as	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1	
Summary	This gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily. ABC proteins transport various molecules across extra- and intra-cellular	Links
	membranes. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN20, White). This protein is a member of the MRP	Order cDNA clone
	subfamily that is involved in multi-drug resistance. The encoded protein functions as a chloride channel and controls the regulation of other transport pathways.	BioAssay
	Mutations in this gene are associated with the autosomal recessive disorders cystic fibrosis and congenital bilateral aplasia of the vas deferens. Alternatively spliced transcript variants have been described many of which result from mutations in this gene. Inrovided by RefSeq. Jul 20081	BioAssay, by Gene target
	transcript variants have been described, many of which result norm mutations in this gene. [provided by Nerseq, Jul 2006]	BioProjects
Genomic context		BioSystems
		Books
Location : 7q31.2		CCDS
Sequence : Chromoso	ime: 7; NC_000007.13 (11/12001711/308/19) See CETR in ManViewer	Conserved Domains
		dbVar
	Chromosome 7 - NC_000007.13	Full text in PMC
	L OCI 001 30680 CTTNBP2 N9938	Genome
	ASZ1 CFTR	GEO Profiles
		HomoloGene
· • • • • • • • • • • • • • • • • • • •		Map Viewer
Genomic regions	, transcripts, and products	Nucleotide
	Go to <u>reference sequence details</u>	OMIM
Genomic Sequence N	C 000007 chromosome 7 reference GRCh37 n5 Primary Assembly	Probe
Senome Sequence N	s_ooooy ononosono / rotoronos anonor.po r initary / soonbiy	Protein
	Go to nucleotide <u>Graphics</u> <u>FASTA</u> <u>GenBank</u>	PubChem Compound

www.genenames.org



The HUGO Gene Nomenclature Committee (HGNC) has assigned unique gene symbols and names to more than 32,000 human loci, of which over 19,000 are protein coding. <u>genenames.org</u> is a curated online repository of HGNC-approved gene nomenclature and associated resources including links to genomic, proteomic and phenotypic information, as well as dedicated gene family pages.

Quick Gene Search Search symbols, keywords or IDs for: Results that ○ equal ○ begin contain Display 50 hits Search Genes

Latest News

<u>Blocking the HCN2 gene product may relieve chronic pain</u> Researchers have uncovered a new source of hope for people suffering from chronic neuropathic pain. Deleting the...

The SMARCAD1 gene has been implicated in the development of fingerprints

A study on a Swiss family with a rare condition called adermatoglyphia, which results in a lack of epidermal ridges...

New genenames.org website

Welcome to the new look and feel genenames.org website! We have made navigation of our site more intuitive by... $% \left({{{\rm{A}}_{\rm{B}}}} \right) = \left({{{\rm{A}}_{\rm{B}}}} \right) = \left({{{\rm{A}}_{\rm{B}}}} \right) = \left({{{\rm{A}}_{\rm{B}}}} \right)$

The HUGO Gene Nomenclature Committee (HGNC) has assigned unique gene symbols and names to more than 32,000 human loci, of which over 19,000 are protein coding.



Browse approved symbols by chromosome

FAQ

What is the HGNC? What is HGNC-approved nomenclature and why do we need it? Where can 1 find information about existing human gene symbols? What is a stem symbol? Where can 1 find the Nomenclature Guidelines? Do I have to use the approved symbols? How should I cite HGNC nomenclature resources? Are there nomenclature committees for other species?

More...

Search Website



wellcometrust



More...

·····		
> NCRI	Resources [v]	How To [v]
K2 INCOL		

Entrez Gene Genes and mapped

phenotypes

Limits Advanced search Help

-

Search Clear



Search: Gene

Welcome to Entrez Gene

Entrez Gene maintains information about genes from genomes of interest to the RefSeq group.

Using Entrez Gene	
Entrez Gene Quick Start	
FAQ	
Download/FTP	
Gene Mailing List	
RefSeq Mailing List	
Entrez Gene News	

Entrez Gene Tools	Other Resources	
Submit GeneRIFs	HomoloGene	
Submit Correction	OMIM	
Statistics	RefSeq	
BLAST	RefSeqGene	
Genome Workbench	UniGene	
Splign	Protein Clusters	

RefSeqGene

NCBI INFORMATION About NCBI Research at NCBI

You are here: NCBI > Genes & Expression > Gene

GETTING STARTED NCBI Help Manual NCBI Handbook

Training & Tutorials

RESOURCES Literature DNA & RNA Proteins Sequence Analysis Genes & Expression Genomes & Maps Domains & Structures Genetics & Medicine Taxonomy Data & Software Training & Tutorials Homology

Small Molecules Variation

POPULAR PubMed Nucleotide BLAST PubMed Central Gene Bookshelf Protein OMIM Genome SNP Structure

NCBI Newsletter NCBI FTP Site

Copyright | Disclaimer | Privacy | Accessibility | Contact National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA

SNCBI Resources 🕑	How To 💌		
RefSeqGene	Gene	•	Search



RefSeqGene

RefSeqGene defines genomic sequences to be used as reference standards for well-characterized genes and is part Reference Genomic (LRG) Project.

Using RefSeqGene

About Browse Genes with RefSeqGene Sequences See RefSeqGene Sequences in the Nucleotide Database Download Data New RefSeqGene Sequences (last 30 days)

Tools

Clinical Remap

Genome Workbench

Map Viewer

RefSeqGene BLAST

Related Sites

Entrez Gene RefSeq Locus Reference Genomic (LRG) Variation Databases (dbSNP, dbVar) Online Mendelian Inheritance In Man

Contact Us

GETTING STARTED NCBI Education NCBI Help Manual NCBI Handbook Training & Tutorials

RESOURCES Chemicals & Bioassays Data & Software DNA & RNA Domains & Structures Genes & Expression Genetics & Medicine Genomes & Maps Homology Literature Proteins Sequence Analysis

POPULAR PubMed Nucleotide BLAST PubMed Central Gene Bookshelf Protein OMIM Genome SNP Structure S

FEATURED GenBank Reference Sequences Map Viewer Genome Projects Human Genome Mouse Genome Influenza Virus Primer-BLAST Sequence Read Archive

NCBI INFORMA About NCBI Research at NCBI NCBI Newsletter NCBI FTP Site NCBI on Facebook NCBI on Twitter

NCBI on YouTube

RefSeqGene is a subset of NCBI's Reference Sequence (RefSeq) project

- defines genomic sequences of wellcharacterized genes to be used as reference standards.
 serve as a stable foundation for
 - mutations, for numbering exons and introns, and for defining the coordinates of other biologically significant variation.

Criteria for selecting RefSeqGene sequences

well-supported,
 exist in nature,
 represent a prevalent,
 'normal' allele.



RefSeqGene

RefSeqGene defines genomic sequences to be used Reference Genomic (LRG) Project.

Using RefSeqGene

About

Browse Genes with RefSeqGene Sequences

See RefSeqGene Sequences in the Nucleotide Database

Download Data

New RefSeqGene Sequences (last 30 days)

RefSeqGene BLAST

GETTING STARTED

NCBI Education

NCBI Help Manual

NCBI Handbook

Training & Tutorials

RESOURCES Chemicals & Bioassays Data & Software DNA & RNA Domains & Structures Genes & Expression POPULAR PubMed Nucleotide BLAST PubMed Central Gene 3

To find the list of genes for which RefSeqGene records are available.

From RefSeqGene home page

To find RefSeqGene record

Entrez Gene Entrez Nucleotide OMIM.

CFTR RefSeqGene

ncbi.nlm.nih.gov/gene/1080#reference-sequences		☆ ·	₽ • Search Results
S CFTR cystic fibrosis trans × 🕷 seamless - Dizion	ario ingles × 🛛 😫 Principles and Methods of S × 🛛 🏭 Quick	Gene Search HUGO × +	
L 0C100130680 CTTNBP2 ASZ1 CFTR	NAR38		GEO Profiles HomoloGene
ts, and products		* ?	Map viewer
		Go to reference seguence detail	
		Ou to reference sequence details	
hromosome 7 reference GRCh37.p5 Primary Assembly	×		Probe
	Go to nucleotide	Graphics FASTA GenBan	
			PubChem Compound
bases shown, positive strand)		<u>Open Full View</u> 🔺 ?	PubChem Substance
		🌮 🔧 Configure	PubMed
117,140,K 117,160,K 117,180,K 117,200,K	117,220 K 117,240 K 117,260 K 117,280 K	117,300 K 117,320 K	PubMed (GeneRIF)
	Genes		PubMed (OMIM)
CI	TR		RefSeq Proteins
		→ IH NP_000483.3 N	RefSeq RNAs
			RefSeqGene
			SNP
			SNP: GeneView
	SNP		SNP: Genotype
	LSDB or Clinically Associated Variants		SNP: VarView
2 7 + 3 8 + 1 1 +	+ 7 + 1 6 3 3 + 1 + 1 +	2	Taxonomy
	Cited Variants		UniGene
5 6 1 + 5 8 + 1 2 + 2 2 1	5 6 + 5 2 + 6 1 1 + 1 + 2 NHGRI GWAS Catalog	2 3	UniSTS
			Links to other resources
		☆ ?	AceView

CE-WEB: Online Information al



Display Settings: 🕑 Graphics

Homo sapiens cystic fibrosis transmembrane conductance regulator (ATP-binding cassette member 7) (CFTR), RefSeqGene on chromosome 7

NCBI Reference Sequence: NG_016465.1

GenBank FASTA

Link To This Page | Help | NG_016465.1 (195,703 bases) 🎎 Sequence 🛛 🔜 Set Origin 👎 Markers 📗 Search... 🔨 Views & Tools 🔻 >0 K 60 K 70 K 100 K 110 K 30 K 50 K 80 K 90 K 130 K 140 K 150 K 160 1:119,816 (119,816 bases shown, positive strand) 2 9 🔹 🞇 Sequence 🛛 🛝 Flip Strands 🛛 🔍 🍳 🌆 🛛 🚮 🖉 Tools 🔹 10 K 30 K 50 K 60 K 70 K 80 K 20 K 40 K 90 K Sequence NG_016465.1: Homo sapiens cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) (CFTR), Ref. < ☆ → Gei + CFIN NM_0G #92.3 exon 2 exon 3 exon 10 exon 11 exon 4 exon ... Align hents NM_0 STS-Harkers

🐸 NG_016465.1: Homo sapiens cystic fibrosis transmembrane conductance regulato	r (ATP-binding cassette sub-family C , n	nember 7) (CFTR), RefSe 🖡
<u>File M</u> odifica <u>V</u> isualizza <u>C</u> ronologia S <u>e</u> gnalibri <u>S</u> trumenti <u>A</u> iuto		
🕢 🗸 C 🗙 🏠 😒 http://www.ncbi.nlm.nih.gov/nuccore/287325315?report=graph	י ג	🕽 🔹 🐠 Ask.com
🔟 Più visitati 📄 Come iniziare 💦 Ultime notizie 📄 HotMail gratuita 📄 Personalizzazione coll 📄 R	accolta Web Slice 🏾 🏉 Siti suggeriti 📄 Windowsh	1edia
🐠 🔍 🗸 🔹 💽 Cerca 🚸 🖸 Corriere della Sera + 🥥 La	Gazzetta dello Sport 👻 🚾 YouTube 📑 Facebo	ok 🔹 🛐 Giochi 🔹 🛛 🔍 Opzior
SNG_016465.1: Homo sapiens cystic fi +		
GenBank FASTA	Link To This Page I	Help Feedback Printer-Friendly
NG_016465.1 (195,703 bases)		
警 🎇 Sequence 📑 Set Origin 🛛 🔨 Views & Tools 🔹	📍 Markers 🛛 Se	arch
1 ▶ 10 K 20 K 30 K 40 K 50 K 60 K 70 K 80 K 90 K 100 K	110 K 120 K 130 K 140 K 150 K	160 K 170 K 180 K
		· · · · · · · · · · · · · · · · · · ·
4,797 : 6,505 (1,709 bases shown, positive strand)		
📕 🔹 🎇 Sequence 🛝 Flip Strands 🔍 🔍 🏧 🚠 🔨 Tools 🔹	20 F 2020 F 2020 - 0 K - 0 102	🨂 📍 Markers Default 🔹 🔧 Op
ductance regulator (ATP-binding cassette sub-family C member 7) (CETB) RefSegGene on chromosy	ome 7 - Sequence NG 0164651: Homo sapi	ens cystic fibrosis transmembrane cr
⋲∕₹	- Genes	
	> > >	> >
exon 1	> > >	
	- Alignments	
	STS Markers	
		RE
KEW63396	REN63395 > >	ECD12042
GDB:210513 > BEN83307	BENE3304	REN63393
REN63398 > >		

8

- . .

Through Entrez Nucleotide by adding RefSeqGene[keyword]

(cftr[Gene Name] AND "Homo sapiens"[Organism]) AND RefSeqgene[Keyword]

S NCBI Resources	🖸 How To 🔽			
Nucleotide	Nucleotide	▼		Searc
		Limits	Advanced	

Display Settings: 🕑 GenBank

Homo sapiens cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) (CFTR), RefSeqGene on chromosome 7

NCBI Reference Sequence: NG_016465.1

FASTA Graphics

<u>Go to:</u> 🕑

LOCUS	NG_016465 195703 bp DNA linear PRI 13-AUG-2011					
DEFINITION	Homo sapiens cystic fibrosis transmembrane conductance regulator					
	(ATP-binding cassette sub-family C, member 7) (CFTR), RefSeqGene on					
	chromosome 7.					
ACCESSION	NG_016465					
VERSION	NG_016465.1 GI:287325315					
KEYWORDS	RefSeqGene.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;					
	Catarrhini; Hominidae; Homo.					
COMMENT	REVIEWED REFSEQ: This record has been curated by NCBI staff. The					
	reference sequence was derived from AC000111.1 and AC000061.1.					
	This sequence is a reference standard in the <u>RefSeqGene</u> project.					
	Summary: This gene encodes a member of the ATP-binding cassette					
	(ABC) transporter superfamily. ABC proteins transport various					
	molecules across extra- and intra-cellular membranes. ABC genes are					
	divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD,					
	OABP, GCN20, White). This protein is a member of the MRP subfamily					
	that is involved in multi-drug resistance. The encoded protein					
	functions as a chloride channel and controls the regulation of					
	other transport pathways. Mutations in this gene are associated					
	with the autosomal recessive disorders cystic fibrosis and					
	congenital bilateral aplasia of the vas deferens. Alternatively					
	spliced transcript variants have been described, many of which					
	result from mutations in this gene. [provided by RefSeq. Jul 2008].					
PRIMARY	REFSEO SPAN PRIMARY IDENTIFIER PRIMARY SPAN COMP					

OMIM

< 🕞 🗸 🏠 👔 http://www.omim.org/search?index=entry&start=1&limit=10&search=cftr&sort=score+desc%2C+prefix_sort+desc
🛿 Ensembl genome br × 🖉 Genome Annotation × 🍿 pipeline - Dizionario × 😒 Homo sapiens cystic × 🖉 Genome Annotation × 👔 OMIM Entry Search
Google traduttore Visualizza questa pagina in: italiano 🔹 Traduci Disattiva per: inglese
Home About Statistics + Downloads Help + External Links Copyright Contact Us
Cftr Search Sort by: Image: Search Image: Display: Toggle highlight Advanced Search: ONIM. Clinical Synopses, ONIM Gene Map Display: Toggle highlight Date space of the second secon
Search History: View, Clear Retneve corresponding: gene map clinical synopses
Search: 'cftr' Results: 1 - 10 of 81 Show all 1 2 3 4 5 6 7 8 9 Next Last
 * 602421. CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR; CFTR Cytogenetic location: 7q31.2, Genomic coordinates (GRCh37): 7:117,120,016 - 117,308,718
2: # 219700. CYSTIC FIBROSIS; CF Cytogenetic locations: 7q31.2, 19q13.2
3: * 603831. PDZ DOMAIN-CONTAINING 1; PDZK1 Cytogenetic location: 1q21.1, Genomic coordinates (GRCh37): 1:145,727,665 - 145,764,206
 * 606845. GOLGI-ASSOCIATED PDZ AND COILED-COIL DOMAINS-CONTAINING PROTEIN; GOPC FIG/ROS1 FUSION GENE, INCLUDED Cytogenetic location: 6q22.1, Genomic coordinates (GRCh37): 6:117,881,431 - 117,923,704
5: # 277180. VAS DEFERENS, CONGENITAL BILATERAL APLASIA OF; CBAVD

- Cytogenetic location: 7q31.2
- 6: # 167800. PANCREATITIS, HEREDITARY; PCTT PANCREATITIS, CHRONIC, SUSCEPTIBILITY TO, INCLUDED Cytogenetic locations: 1p36.21, 5q32, 7q31.2, 7q34
- 7: # 211400. BRONCHIECTASIS WITH OR WITHOUT ELEVATED SWEAT CHLORIDE 1; BESC1 Cytogenetic locations: 7q31.2, 16p12.2

- 0 🔀 م 😪 2 selected items - Nu... × ... × r): Opzioni V × 🚼 Seleziona lingua 🔻

ults

OMIM External links

External Links for *602421

Genome Ensembl **NCBI Map Viewer** Gene Tests, Ne UCSC Genome Browser DNA Ensembl NCBI RefSeq UCSC Genome Browser Protein UniProt HPRD Gene Info BioGPS Ensembl NCBI Gene GeneCards KEGG PharmGKB Clinical Resources Gene Tests **Genetics Home** Reference Variation **Genetics Association DB** GWAS Central HGVS MutaDATABASE

Ξ

TEIN; GOPC





Ensembl is a joint project between the EBI (European Bioinformatics Institute) and the Wellcome Trust Sanger Institute that annotates chordate genomes (i.e. vertebrates and closely related invertebrates with a notochord).

Gene sets from model organisms such as yeast and fly are also imported for comparative analysis.

The Ensembl Annotation



Protein-coding genes are automatically annotated using Ensembl's genebuild pipeline. All transcripts are based on mRNA and proteins in public scientific databases.

The Ensembl gene set also includes automatically-annotated pseudogenes and non-coding RNAs.

S NCBI Resources	✓ How To ♥	
Gene	Cana	
Ocho	Gene v	
	Limits Advanced	
Display Settings: 🕑 Full	Report Send to: 🕑	
		Table of contents
CFTR cystic fibro	sis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) [Homo sapiens]	Summary
Gene ID: 1080, updated o	1 23-Oct-2011	Genomic context
		Genomic regions, tra
Summary		Bibliography
Official Symbol	CFTR provided by <u>HGNC</u>	Phenotypes
Official Full Name	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) provided by HGNC	Interactions
Primary source	HGNC:1884	General gene info
Locus tag	icag/./8 Ensembl:ENSG0000001626: HPPD:03982: MM:602421	General protein info
Gene type	protein coding	Reference sequence
RefSeq status	REVIEWED	Related sequences
Organism	Homo sapiens	Additional links
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	
Also known as	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1	Links
Summary	This gene encodes a member of the ATP-binding cassette (ABC) transporter superramily. ABC proteins transport various molecules across extra- and intra-cellular membranes. ABC genes are divided into seven distinct subfamilies (ABC). MDR/TAP, MRP, ALD, OABP, GCN20, White). This protein is a member of the MRP.	Order cDNA clone
	subfamily that is involved in multi-drug resistance. The encoded protein functions as a chloride channel and controls the regulation of other transport pathways.	BioAssay
	Mutations in this gene are associated with the autosomal recessive disorders cystic fibrosis and congenital bilateral aplasia of the vas deferens. Alternatively spliced	BioAssay, by Gene ta
	transcript variants have been described, many of which result from mutations in this gene. [provided by RefSeq, Jul 2008]	BioProjects
		BioSystems
 Genomic context 		Books
Location : 7q31.2		CCDS
Sequence : Chromoso	ne: 7; NC_000007.13 (117120017117308719)	Conserved Domains
	See CFTR in MapViewer	dbVar
	Chromosome 7 - NC_000007.13	Full text in PMC
		Genome
	ASZ1 CFTR	GEO Profiles
		HomoloGene
· · ·		Map Viewer
 Genomic regions, 	transcripts, and products	Nucleotide
	Go to reference sequence details	OMIM
Genomic Sequence	000007 chromosome 7 reference GRCb37 p5 Primary Assembly	Probe
Continue Sequence MC		Protein
	Go to nucleotide Graphics FASTA GenBank	PubChem Compoun

-	Oene	•		Search	
		Limits	Advanced		

Settings: 🕑 Full Report

Send to:

cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) [Homo sapiens]

: 1080, updated on 23-Oct-2011

mmary)(
Official Symbol icial Full Name Primary source	CFTR provided by <u>HGNC</u> cystic fibrosis transmer to be conductance regulator (ATP-binding cassette sub-family C, member 7) provided by <u>HGNC</u> <u>HGNC:1884</u>	
See related Gene type RefSeq status Organism Lineage	Ensembl:ENSG0000001626; ! protein coding REVIEWED Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	
Also known as Summary	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1 This gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily. ABC proteins transport various molecules across extra- and intra-cellular membranes. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN20, White). This protein is a member of the MRP subfamily that is involved in multi-drug resistance. The encoded protein functions as a chloride channel and controls the regulation of other transport pathways. Mutations in this gene are associated with the autosomal recessive disorders cystic fibrosis and congenital bilateral aplasia of the vas deferens. Alternatively spliced transcript variants have been described, many of which result from mutations in this gene. [provided by RefSeq, Jul 2008]	
nomic context	t)(
n : 7q31.2 nce : Chromoso	ome: 7; NC_000007.13 (117120017117308719) See CFTR in MapVi	ev
	Chromosome 7 - NC_000007.13 [117003276] LOC100130680	
		100

		Login · Registe							
<i>e</i> Ensembl	BLAST/BLAT BioMart Tools Downloads Help & Documentation Blog Mirrors	🛃 • 🛛 🔍							
Human (GRCh37) V Location	on: 7:117.105.838-117.308.719 Gene: CFTR								
Gene-based displays	Correr CETD (ENSCO000001626)								
- Gene summary									
- Splice variants (7)	Description cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) [Source:HGNC Symbol:Acc:1884]								
- Supporting evidence	Location Chromosome 7: 117.105.838-117.308.719 forward strand.								
 External references 	Transcripts This gene has 7 transcripts								
- Regulation									
Genomic alignments	Show/bide columns								
⊟· Gene Tree (image)									
Gene Tree (text)	Name Transcript to Lengun (p) Protein to Lengun (a) Dotype CCDS CCS (CCS)								
- Orthologues (51)	CETR-002 ENST0000048206 682 ENST0000040504 1400 Protein coding CCD5/15								
- Paralogues (4)	CETRODAL ENSTRODOMARSES 575 ENSPRODUCTIVE 37 Protein coding								
 Protein families (2) Phenotype 	CFTR.005 ENST00000426809 4316 ENSP0000380119 1438 Protein coding								
Genetic Variation	CFTR-S01 ENST0000454333 5949 ENSP0000403677 1419 Protein coding								
- Variation Table	CFTR-003 ENST00000546407 222 No protein product - Processed transcript -								
Structural Variation	CFTR-006 ENST00000472848 148 No protein product - Processed transcript -								
External Data									
Personal annotation Fild History									
Gene history	in Ensembli we provide displays at two levels:								
Configure this page	 Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation. 								
Configure this page	 Gene views which provide displays for data associated at the gene level such as orthologues, paralogues, regulatory regions and splice variants. 								
💼 Manage your data	This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the menu	at the left hand side of the page. To return to							
Export data	viewing gene level information click on the Gene tab in the menu bar at the top of the page.								
🙀 Bookmark this page	Gene summary help								
	Name CETE (HONC Symbol)								
	Synonyms ABC59, ABCC7, CF, CF HZ/MRP, QJ/SUC5.1, MRP7, INR-CF IR [To view all Ensembligenes linked to the name <u>click here.]</u> CCDS This gene is a member of the Human CCDS set: <u>CCDS5773</u> Gene type Known protein coding Prediction Method Annotation for this gene is one includes both automatic annotation from Ensembli and <u>Havana</u> manual curation, see <u>article</u> . Alternative gene This gene corresponds to the following database identifiers: Havana gene: OTTH UNC0000023076 (varsion 8) biew all locations].								
	222.88 kb	Forward strand							
		17.26 MB 117.28 MB 117.30 MB							
	processed pseudogene								
	CFTR-003 >								
	processed transcript								
	CFTR-201 >								
	Crik+uui > protein coding								
	protein coding								
	CFTR-004 > CFTR-006 > CFTR-006 > CFTR-006 > crtra- protein coding	> ading							

An Ensembl gene (with a unique ENSG... ID) includes any spliced transcripts (ENST...) with overlapping coding sequence.

Transcripts from the Ensembl genebuild, the Havana/Vega set and the Consensus Coding Sequence (CCDS) set may all be clustered into the same gene.

Transcripts that belong to the same gene ID may differ in splice events, exons, and can give rise to very different proteins (isoforms) arising from alternative splicing.

Selecting a specific transcript et the exon/intron structure is dispayed

21-

Transcript: CFTR-001 (ENST0000003084)

cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) [Source:HGNC Symbol;Acc:1884] Chromosome 7: 117.120.017-117.308.715 forward strand

This transcript is a product of gene ENSG0000001626 - This gene has 7 transcripts

General identifiers (382)	Show/hide columns Filter							
intology	Name 🍦	Transcript ID 🍦	Length (bp)	Protein ID	Length (aa) 🍦	Biotype 🍦	CCDS 🍦	
Ontology chart (42)	CFTR-001	ENST0000003084	6128	ENSP0000003084	1480	Protein coding	CCDS5773	
Ontology table (42)	CFTR-002	ENST0000468795	682	ENSP00000419254	190	Protein coding	-	
Population comparison	CFTR-004	ENST0000446805	575	ENSP00000417012	37	Protein coding	-	
Comparison image	CFTR-005	ENST0000426809	4316	ENSP00000389119	1438	Protein coding	-	
Protein Information	CFTR-201	ENST0000454343	5949	ENSP00000403677	1419	Protein coding	-	
Domains & features (42)	CFTR-003	ENST0000546407	222	No protein product	-	Processed transcript	-	
Variations (1379)	CFTR-006	ENST0000472848	148	No protein product	-	Processed transcript	-	
vtornal Data								

Transcript and Gene level displays

Transcript-based displays

Ontology chart (42) Ontology table (42) Genetic Variation

 Variations (1379) External Data

Personal annotation

Transcript history

Protein history 🎤 Configure this page

Population comparison Comparison image Protein Information Protein summary

Description

Location

Gene 🖃

 Transcript summary Supporting evidence (47)

> Exons (27) **cDNA**

⊡ Sequence

Protein External References General identifiers (382) Oligo probes (64)

⊡ Ontology

⊡ ID History

Views in Ensembl are separated into gene based views and transcript based views according to which level the information is more appropriately associated with. This view is a transcript level view. To sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page

Exons help

/lanage your data								
Export data	Show All 💌 entries					Show/hide columns		
	No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
Bookmark this page		5' upstream sequence					U	
Download view as RTF	1	<u>ENSE00001343851</u>	117,120,017	117,120,201		2	185	AATTGGAAGCAAATGACATCACAGCAGGTCAGAGAAAAAG GAGTAGTAGGTCTTTGGCATTAGGAGCTTGAGCCCAGACGG GCCCGAGAGACCATGCAGAGGTCGCCTCTGGAAAAGGCCAG TTCAG
		Intron 1-2	117,120,202	117,144,306			24,105	gtgagaaggtggccaaccgagcttctctgtt
	2	ENSE00002474285	117,144,307	117,144,417	2	2	111	CTGGACCAGACCAATTTTGAGGAAAGGATACAGACAGCGCC CCAAATCCCTTCTGTTGATTCTGCTGACAATCTATCTGAAA
		Intron 2-3	117,144,418	117,149,087			4,670	gtatgttcatgtacattgtttagttggtccc
	3	ENSE00002532105	117,149,088	117,149,196	2	0	109	AGAATGGGATAGAGAGCTGGCTTCAAAGAAAAATCCTAAA ATGTTTTTTCTGGAGATTTATGTTCTATGGAATCTTTTAI
		Intron 3-4	117,149,197	117,170,952			21,756	gtaaggateteatttgtacatteattttete
	4	ENSE00000718611	117,170,953	117,171,168	0	0	216	GAAGTCACCAAAGCAGTACAGCCTCTCTTACTGGGAAGAAT GATAACAAGGAGGAACGCTCTATCGCGATTTATCTAGGCAT ATTGTGAGGACACTGCTCCTACACCCAGCCATTTTTGGCCT ATGAGAATAGCTATGTTTAGTTTGATTTATAAGAAG
		Intron 4-5	117,171,169	117,174,329			3,161	gtaatacttccttgcacaggccccaatctaa
	5	ENSE00000718620	117 174 330	117 174 419	0	0	90	ACTTTAAAGCTGTCAAGCCGTGTTCTAGATAAAATAAGTAT

CTTTCCAACAACCTGAACAAATTTGATGAA

Exons, introns and flanking sequence are shown for one transcript (ENST...) in the 5' to 3' direction, regardless of whether it is a forward or reverse-stranded gene.

<u>Exons - Uppercase letters</u>

- •UTR (UnTranslated Region) is in purple
- •Coding sequence is in black.
- <u>Flanking sequence and introns –</u> <u>lower case letters</u>
- Introns are blue
 Flanking sequence upstream and downstream to the transcript is green