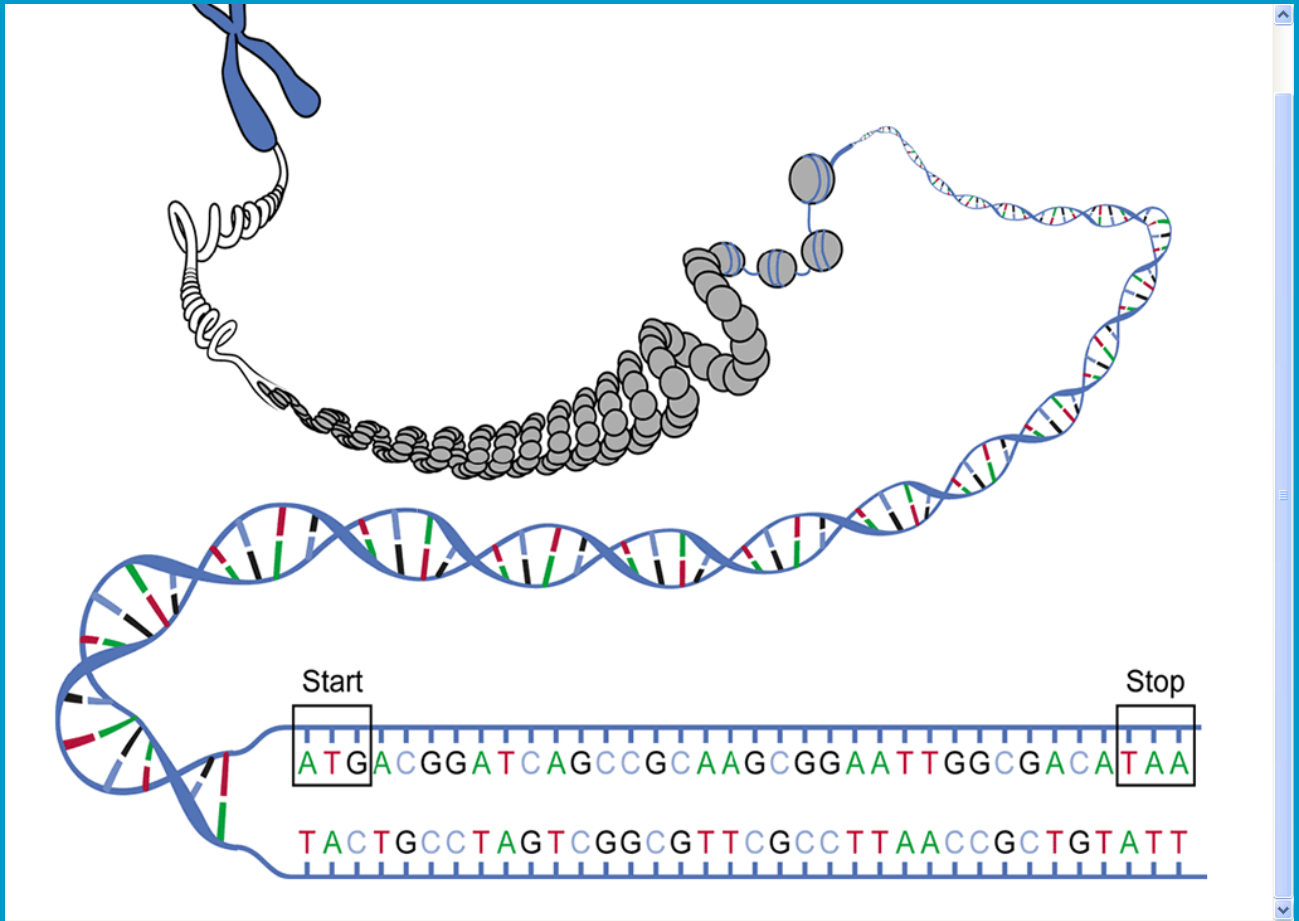
- One of the most important aspects of bioinformatics is identifying genes within a long DNA sequence.

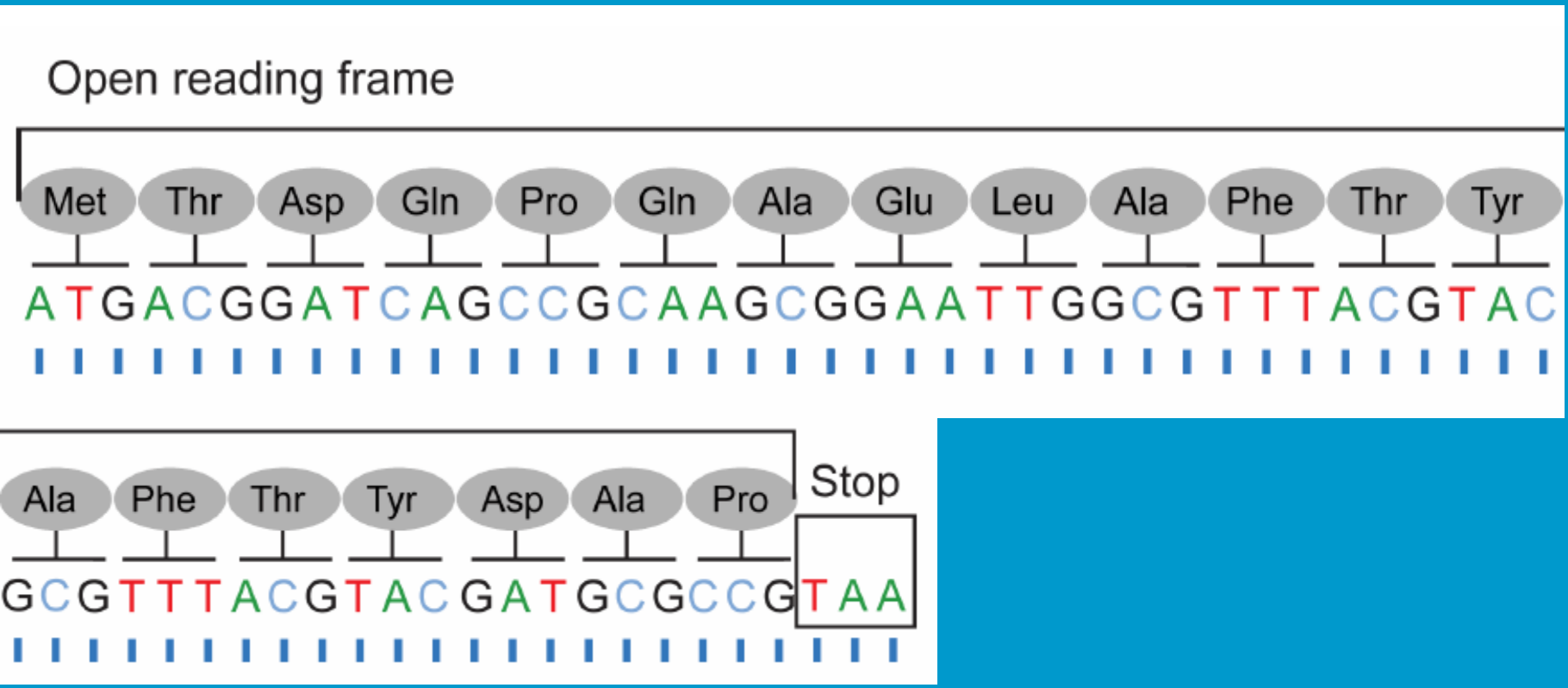
Until the development of bioinformatics, the only way to locate genes along the chromosome was to study their behavior in the organism (*in vivo*) or isolate the DNA and study it in a test tube (*in vitro*).

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.



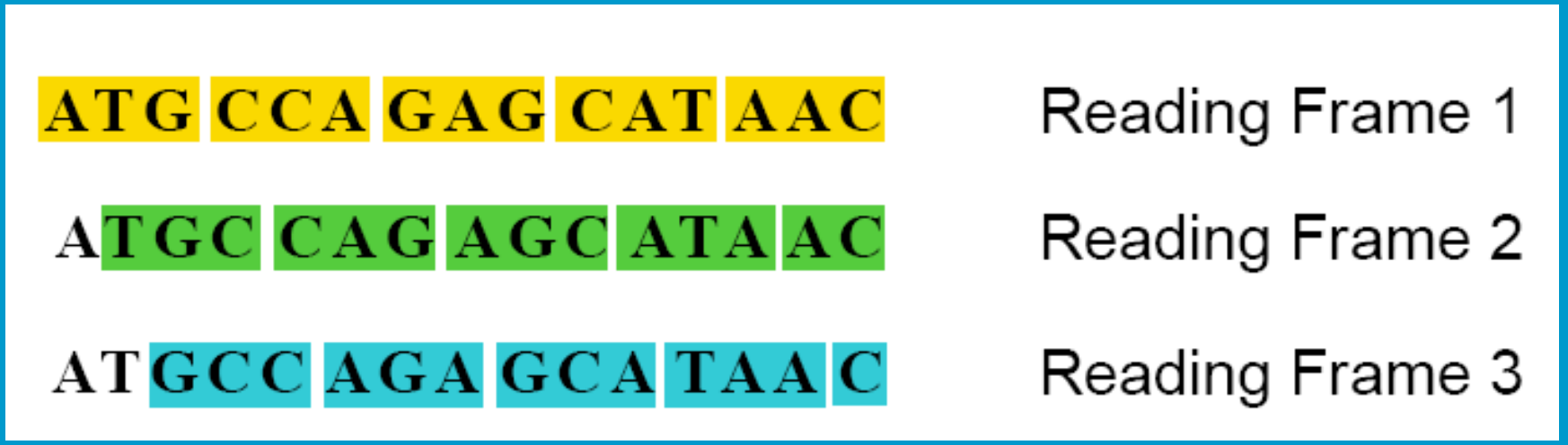
When examining an unknown DNA sequence, one indication that it may be part of a gene is the presence of an open reading frame (ORF). An ORF is any stretch of DNA that when transcribed into RNA has no stop codon

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.



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.

3. Likewise, codons for methionine do not always mark the start of a gene sequence. Methionine codons are also found within genes.

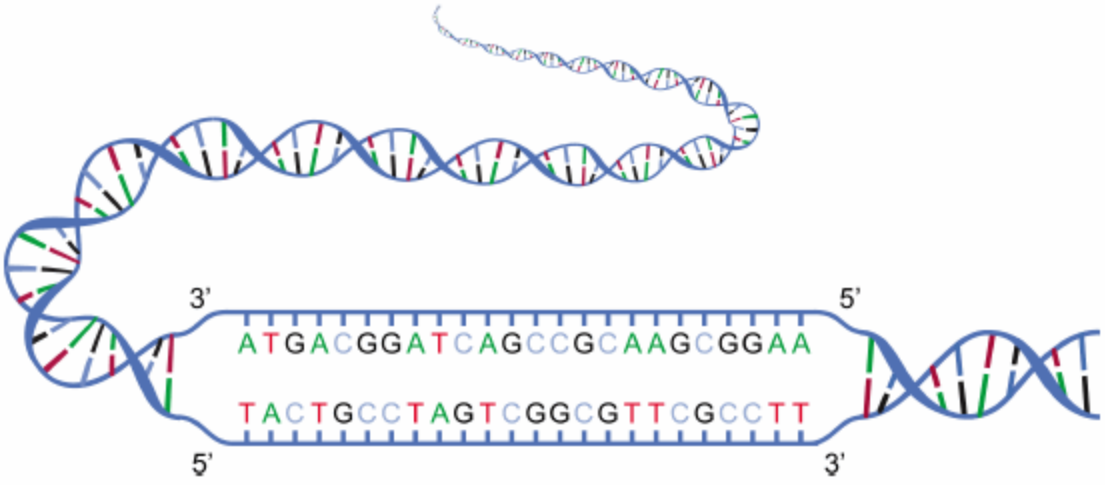
Nevertheless, searching for ORFs identifies regions of the DNA sequence that might be parts of genes.

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.

A computer program looking 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.



In a DNA strand, the end of the strand with the phosphate group is called the five prime end. The opposite end with the sugar is called the three prime end.

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.

❖The ORF encodes a protein that is similar to previously described ones.

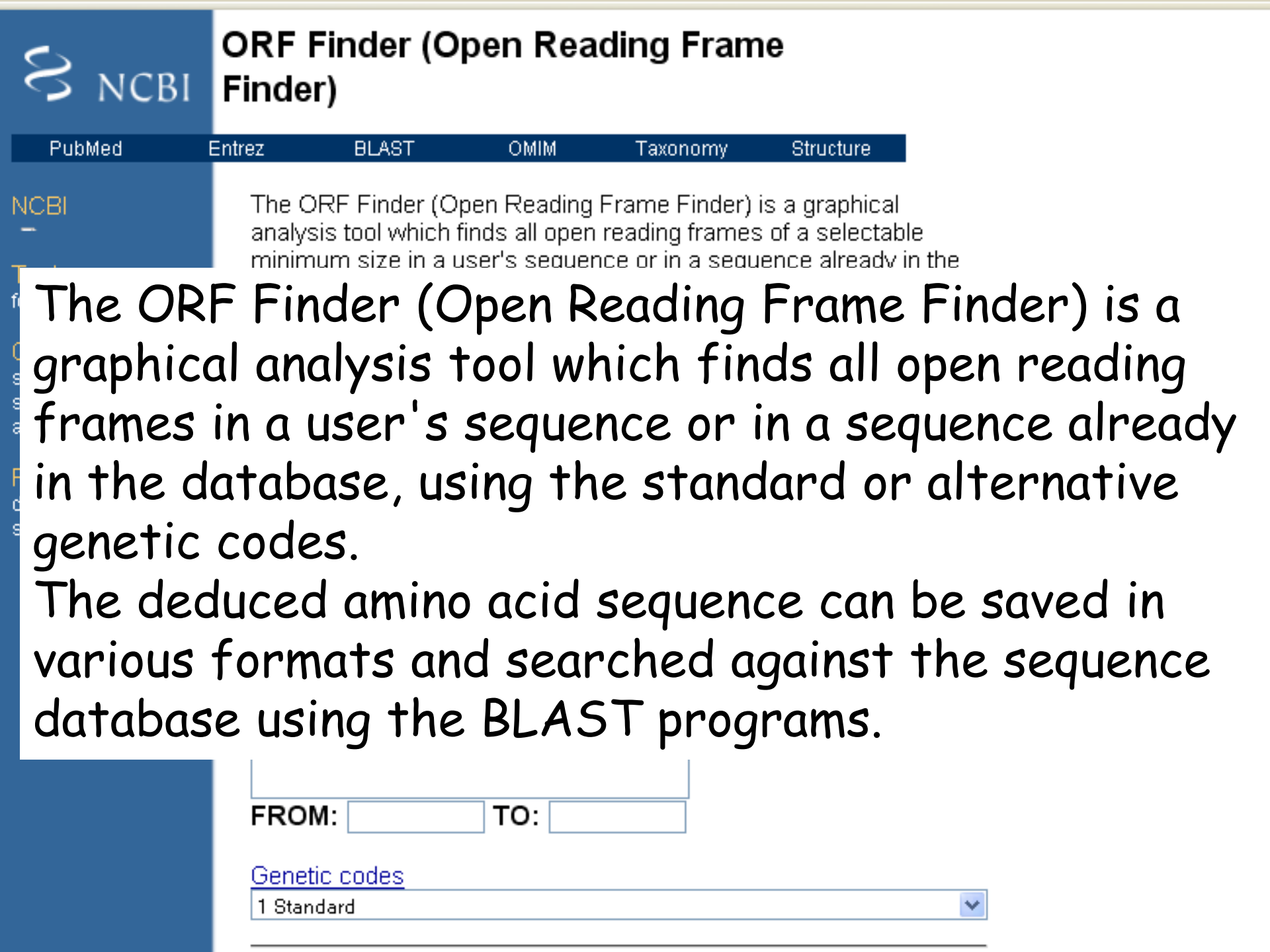
❖The ORF has a typical *GC* content, codon frequency, or oligonucleotide composition

❖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.


NCBI
ORF Finder (Open Reading Frame Finder)
PubMed
Entrez
BLAST
OMIM
Taxonomy
Structure
NCBI
The ORF Finder (Open Reading Frame Finder) is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a user's sequence or in a sequence already in the
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:
Genetic codes
1 Standard

Unicellular eukaryotes

Gene organization is more complex and poses additional problems for intron identification.
In some simple eukaryotes genes are quite 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.

Gene

A Directory of Genes

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

Gene

is a derivative database.

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.

Entrez Gene

NCBI Resources How To

Entrez Gene
Genes and mapped phenotypes

Search: Gene Limits Advanced search Help

Search Clear

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

- Submit GeneRIFs
- Submit Correction
- Statistics
- BLAST
- Genome Workbench
- Splign

Other Resources

- HomoloGene
- OMIM
- RefSeq
- RefSeqGene
- UniGene
- 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

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 human cftr gene

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

Gene ID: 1080, updated on 23-Oct-2011

Summary

Official Symbol CFTR provided by HGNC

Official Full Name cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) provided by HGNC

Primary source HGNC:1884

Locus tag tcag7.78

See related Ensembl:ENSG00000001626; HPRD:03883; MIM:602421

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 CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1

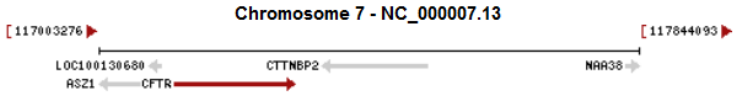
Summary This gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily. ABC proteins transport various molecules across extra- and intra-cellular membranes. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN20, White). This protein is a member 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]

Genomic context

Location : 7q31.2

Sequence : Chromosome: 7; NC_000007.13 (117120017..117308719)

See CFTR in MapViewer



Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence NC_000007 chromosome 7 reference GRCh37.p5 Primary Assembly

Go to nucleotide Graphics FASTA GenBank

Links

Order cDNA clone
BioAssay
BioAssay, by Gene target
BioProjects
BioSystems
Books
CCDS
Conserved Domains
dbVar
Full text in PMC
Genome
GEO Profiles
HomoloGene
Map Viewer
Nucleotide
OMIM
Probe
Protein
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.

Entrez Gene

Genes and mapped phenotypes

Search: Gene Limits Advanced search Help

Search Clear

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

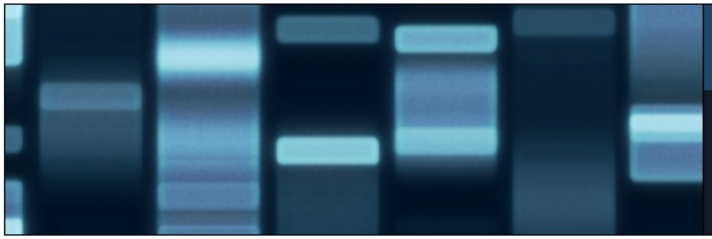
- Submit GeneRIFs
- Submit Correction
- Statistics
- BLAST
- Genome Workbench
- Splign

Other Resources

- HomoloGene
- OMIM
- RefSeq
- RefSeqGene
- UniGene
- Protein Clusters

RefSeqGene

National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA



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

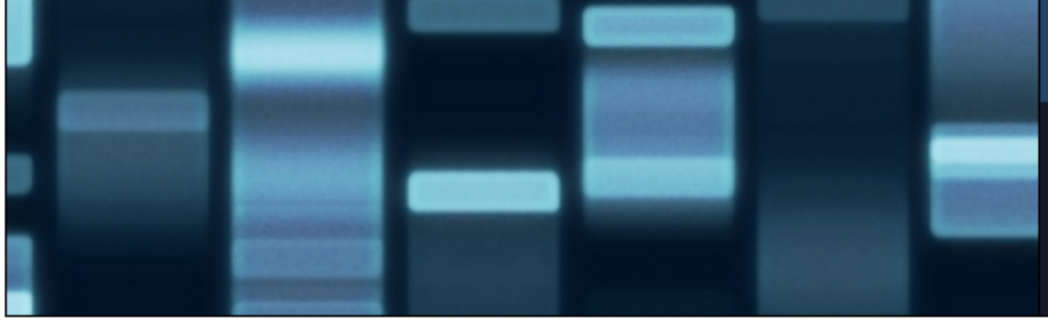
- Entrez Gene
- RefSeq
- Locus Reference Genomic (LRG)
- Variation Databases (dbSNP, dbVar)
- Online Mendelian Inheritance In Man
- Contact Us

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

1) defines genomic sequences of well-characterized genes to be used as reference standards.
2) 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

1) well-supported,
2) exist in nature,
3) 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)

Tools

- Clinical Remap
- Genome Workbench
- Map Viewer
- RefSeqGene BLAST

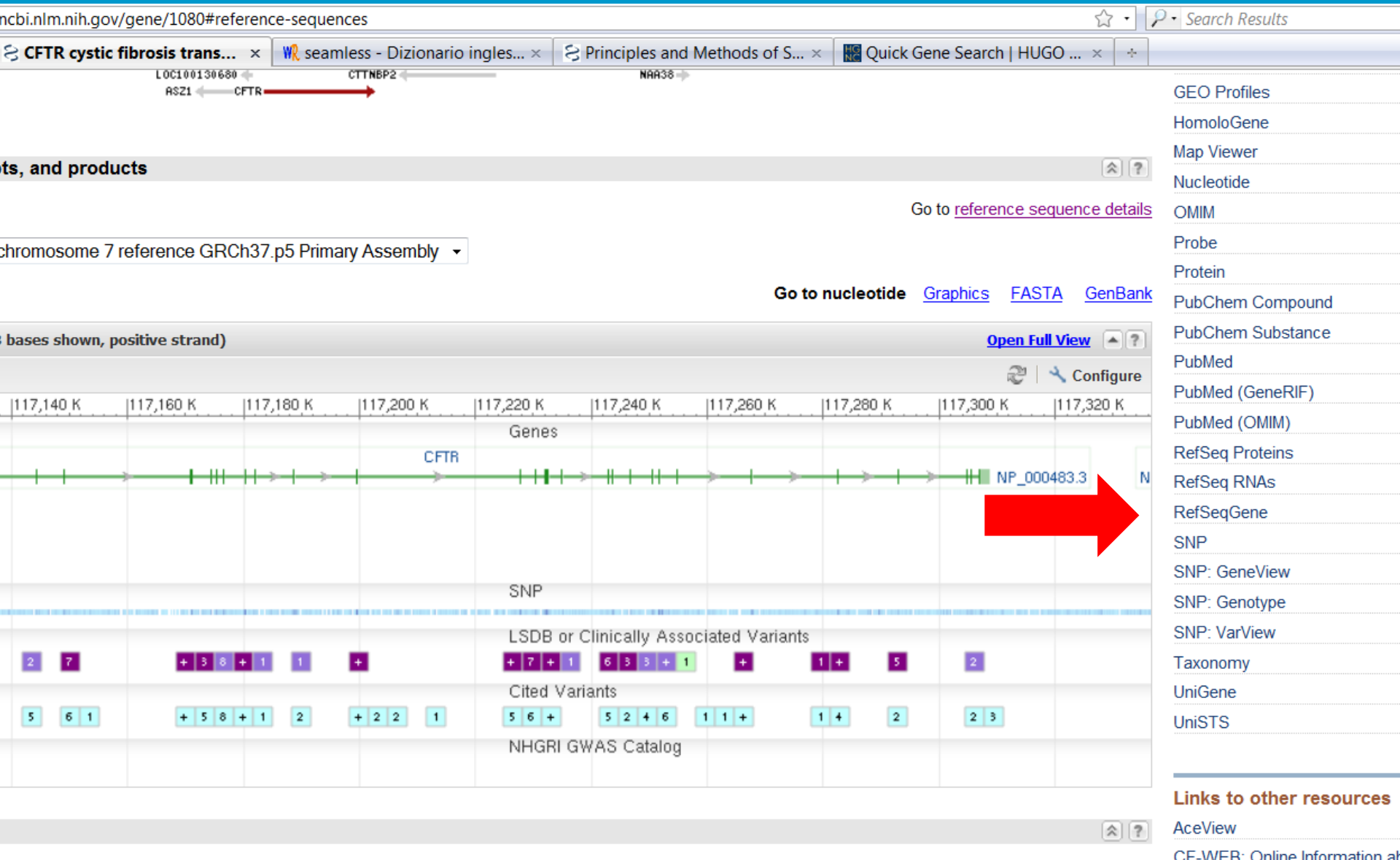
To find the list of genes for which RefSeqGene records are available.

- From RefSeqGene home page

To find RefSeqGene record

Entrez Gene
Entrez Nucleotide
OMIM.

CFTR RefSeqGene



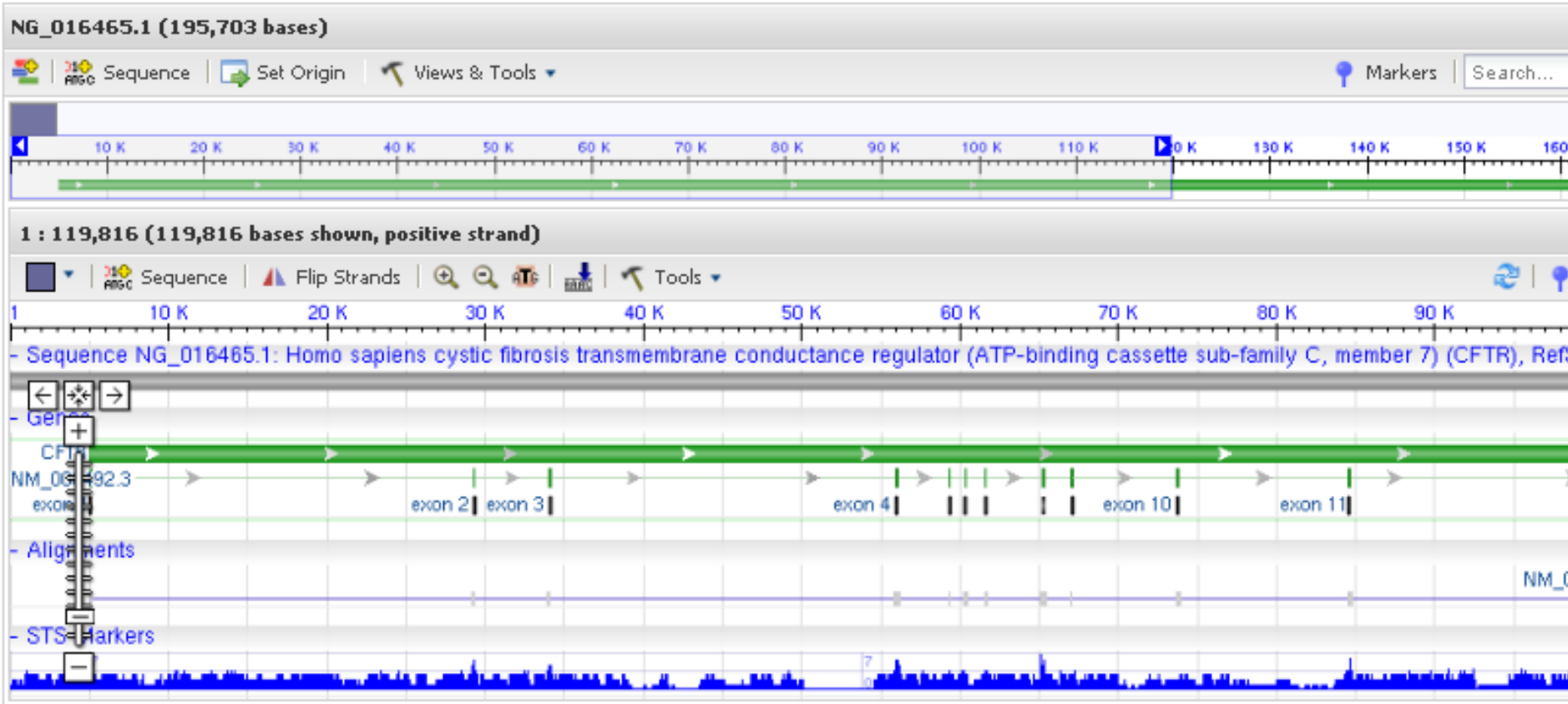
Display Settings: Graphics

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 | Views & Tools | Markers | Search...

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 | Markers | Default

4,900 5 K 5,100 5,200 5,300 5,400 5,500 5,600 5,700 5,800 5,900 6 K 6,100 6,200 6,300 6,400

ductance regulator (ATP-binding cassette sub-family C, member 7) (CFTR), RefSeqGene on chromosome 7 - Sequence NG_016465.1: Homo sapiens cystic fibrosis transmembrane co

- Genes

CFTR

NM_000492.3

exon 1

- Alignments

- STS Markers

stSG598675 REN63396 stSG598676 REN63395 ECD12042 ECD08301 ECD16416 GDB:210513 REN63393 REN63397 REN63394 REN63398 GDB:204205

Through Entrez Nucleotide by adding RefSeqGene[keyword]

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

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

Go to:

```
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 RefSeqGene 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     REFSEQ_SPAN         PRIMARY_IDENTIFIER PRIMARY_SPAN        COMP
            1-134685            AC000111.1         14670-149354
```

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: Relevance Date updated

Advanced Search: OMIM, Clinical Synopses, OMIM Gene Map Display: Toggle highlight
Search History: View, Clear

Retrieve 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

1 : * 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

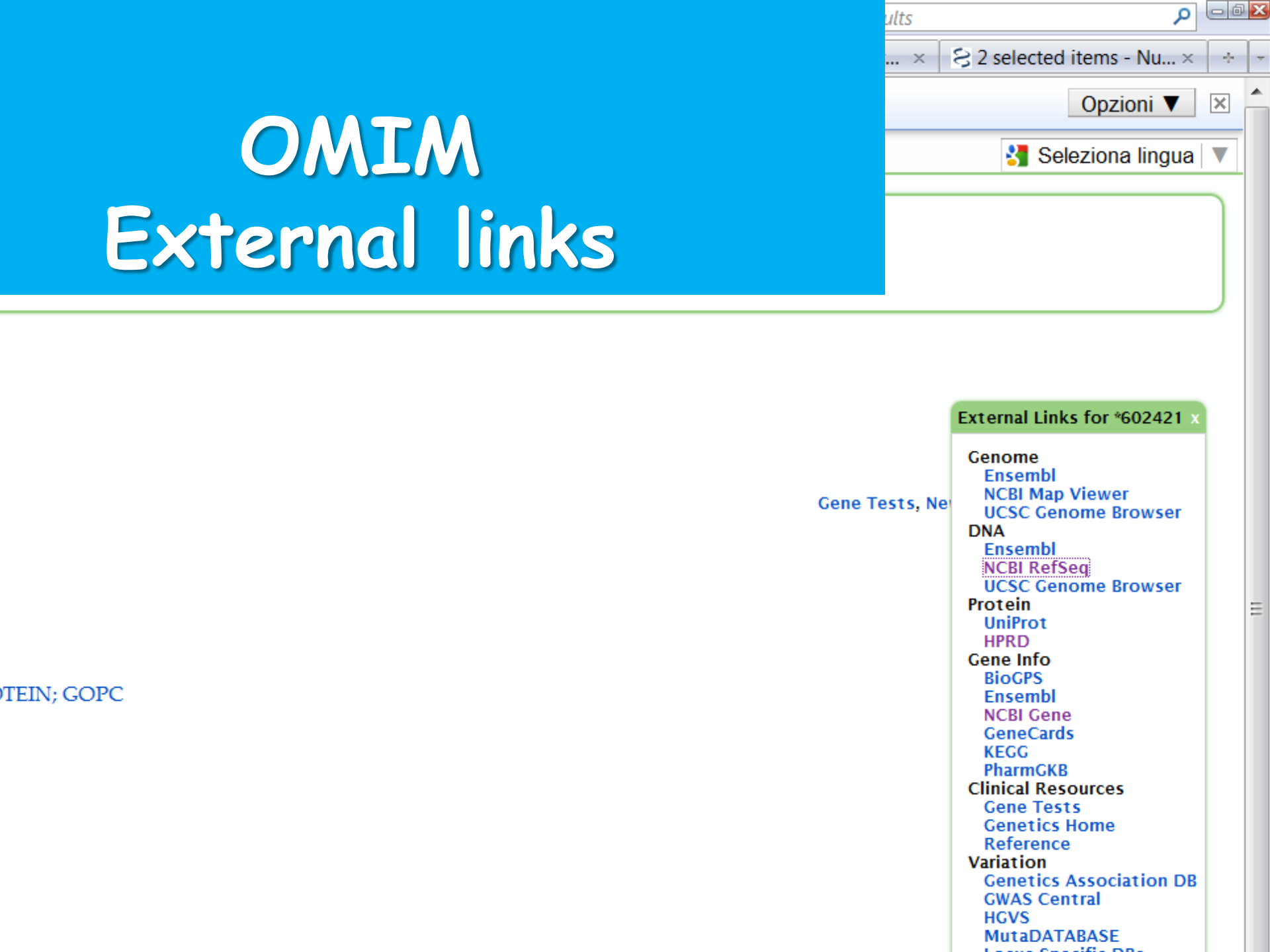
4 : * 606845. GOLGI-ASSOCIATED PDZ AND COILED-COIL DOMAINS-CONTAINING PROTEIN; GOPC
FIG/ROS1 FUSION GENE, INCLUDED
Cytogenetic location: 6q22.1 , Genomic coordinates (GRCh37): 6:117,881,431 - 117,923,704

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

OMIM
External links



Ensembl

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.

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

Gene ID: 1080, updated on 23-Oct-2011

Summary

Official Symbol CFTR provided by HGNC

Official Full Name cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) provided by HGNC

Primary source HGNC:1884

Locus tag tcag7.78

See related Ensembl:ENSG00000001626; HPRD:03883; MIM:602421

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 CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1

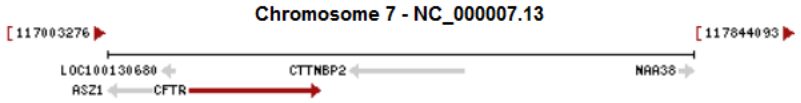
Summary This gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily. ABC proteins transport various molecules across extra- and intra-cellular membranes. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN20, White). This protein is a member 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]

Genomic context

Location : 7q31.2

Sequence : Chromosome: 7; NC_000007.13 (117120017..117308719)

See CFTR in MapViewer



Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence NC_000007 chromosome 7 reference GRCh37.p5 Primary Assembly

Go to nucleotide Graphics FASTA GenBank

Links

Order cDNA clone
BioAssay
BioAssay, by Gene ta
BioProjects
BioSystems
Books
CCDS
Conserved Domains
dbVar
Full text in PMC
Genome
GEO Profiles
HomoloGene
Map Viewer
Nucleotide
OMIM
Probe
Protein
PubChem Compoun

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 CFTR provided by HGNC

icial Full Name cystic fibrosis transme ne conductance regulator (ATP-binding cassette sub-family C, member 7) provided by HGNC

Primary source HGNC:1884

Locus tag tcag7.78

See related Ensembl:ENSG00000001626;

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 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 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]

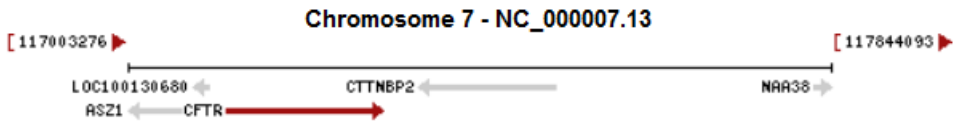
Link to ensembl

nomic context

n : 7q31.2

nce : Chromosome: 7; NC_000007.13 (117120017..117308719)

See CFTR in MapView



nomic regions, transcripts, and products

Go to reference sequence deta

Gene-based displays
- Gene summary
- Splice variants (7)
- Supporting evidence
- Sequence
- External references
- Regulation
- Comparative Genomics
  - Genomic alignments
  - Gene Tree (image)
    - Gene Tree (text)
    - Gene Tree (alignment)
  - Orthologues (51)
  - Paralogues (4)
  - Protein families (2)
- Phenotype
- Genetic Variation
  - Variation Table
  - Variation Image
  - Structural Variation
- External Data
  - Personal annotation
- ID History
  - Gene history

Configure this page
Manage your data
Export data
Bookmark this page

Gene: CFTR (ENSG00000001626)

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

Location Chromosome 7: 117,105,838-117,308,719 forward strand.

Transcripts ⊟ This gene has 7 transcripts

Show/hide columns | Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
CFTR-001	ENST00000003084	6128	ENSP00000003084	1480	Protein coding	CCDS5773
CFTR-002	ENST00000468795	682	ENSP00000419254	190	Protein coding	-
CFTR-004	ENST00000446805	575	ENSP00000417012	37	Protein coding	-
CFTR-005	ENST00000426809	4316	ENSP00000389119	1438	Protein coding	-
CFTR-201	ENST00000454343	5949	ENSP00000403677	1419	Protein coding	-
CFTR-003	ENST00000546407	222	No protein product	-	Processed transcript	-
CFTR-006	ENST00000472848	148	No protein product	-	Processed transcript	-

Transcript and Gene level displays

In Ensembl we provide displays at two levels:

- Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
- Gene views which provide displays for data associated at the gene level such as orthologues, paralogues, regulatory regions and splice variants.

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 viewing gene level information click on the Gene tab in the menu bar at the top of the page.

Gene summary help

Name CFTR (HGNC Symbol)

Synonyms ABC35, ABCC7, CF, CFTR/MRP, dJ760C5.1, MRP7, TNR-CFTR [To view all Ensembl genes linked to the name click here.]

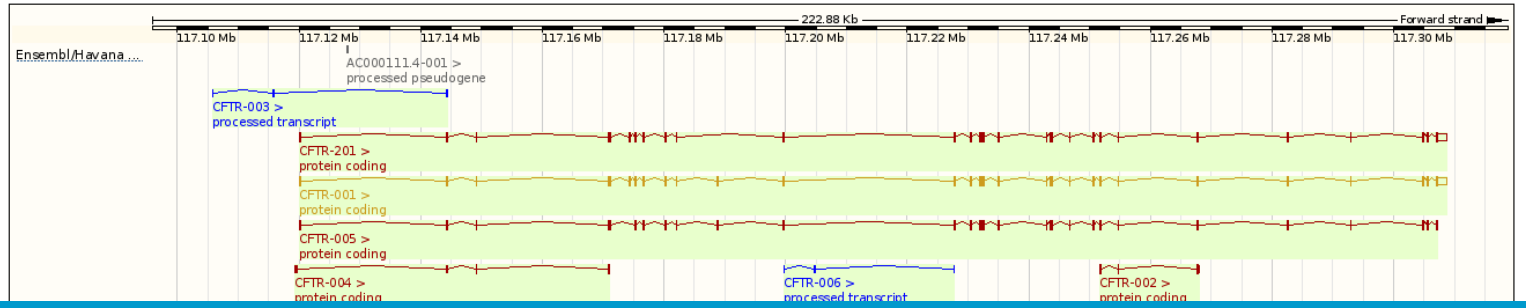
CCDS This gene is a member of the Human CCDS set: CCDS5773

Gene type Known protein coding

Prediction Method Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.

Alternative genes **This gene corresponds to the following database identifiers:**

Havana gene: OTTHUMG00000023076 (version 8) [view all locations]



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 the exon/intron structure is dispayed

Human (GRC

Transcript-based displays

- Transcript summary
- Supporting evidence (47)
- Sequence
  - **Exons (27)**
  - cDNA
  - Protein
- External References
  - General identifiers (382)
  - Oligo probes (64)
- Ontology
  - Ontology chart (42)
  - Ontology table (42)
- Genetic Variation
  - Population comparison
  - Comparison image
- Protein Information
  - Protein summary
  - Domains & features (42)
  - Variations (1379)
- External Data
  - Personal annotation
- ID History
  - Transcript history
  - Protein history

Configure this page

Manage your data

Export data

Bookmark this page

Download view as RTF

Transcript: CFTR-001 (ENST00000003084)

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

Location Chromosome 7: 117,120,017-117,308,715 forward strand.

Gene This transcript is a product of gene ENSG00000001626 - This gene has 7 transcripts

Show/hide columns | Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
CFTR-001	ENST00000003084	6128	ENSP00000003084	1480	Protein coding	CCDS5773
CFTR-002	ENST00000468795	682	ENSP00000419254	190	Protein coding	-
CFTR-004	ENST00000446805	575	ENSP00000417012	37	Protein coding	-
CFTR-005	ENST00000426809	4316	ENSP00000389119	1438	Protein coding	-
CFTR-201	ENST00000454343	5949	ENSP00000403677	1419	Protein coding	-
CFTR-003	ENST00000546407	222	No protein product	-	Processed transcript	-
CFTR-006	ENST00000472848	148	No protein product	-	Processed transcript	-

Transcript and Gene level displays

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 f
sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

Exons *help*

Show All entries | Show/hide columns

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence						.....ggaaagagcaaaaggaaggggtggtgtgcgg
1	ENSE00001343851	117,120,017	117,120,201	-	2	185	AATTGGAAGCAAATGACATCACAGCAGGTCAGAGAAAAAGG GAGTAGTAGGTCTTTGGCATTAGGAGCTTGAGCCCAGACGG GCCCGAGAGACCATGCAGAGGTCGCCTCTGGAAAAGGCCAG TTCAG
	Intron 1-2	117,120,202	117,144,306			24,105	gtgagaaggtggccaaccgagcttc.....tctgtt
2	ENSE00002474285	117,144,307	117,144,417	2	2	111	CTGGACCAGACCAATTTTGAGGAAAGGATACAGACAGCGCC CCAAATCCCTTCTGTTGATTCTGCTGACAATCTATCTGAAA
	Intron 2-3	117,144,418	117,149,087			4,670	gtatgttcatgtacattgtttagtt.....ggtccc
3	ENSE00002532105	117,149,088	117,149,196	2	0	109	AGAATGGGATAGAGAGCTGGCTTCAAAGAAAAATCCTAAAC ATGTTTTTTCTGGAGATTTATGTTCTATGGAATCTTTTTAT
	Intron 3-4	117,149,197	117,170,952			21,756	gtaaggatctcatttgtacattcat.....tttctc
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	gtaatacttccttgcacaggcccca.....atctaa
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.

Exons - Uppercase letters

- UTR (UnTranslated Region) is in purple
- Coding sequence is in black.

Flanking sequence and introns - lower case letters

- Introns are blue
- Flanking sequence upstream and downstream to the transcript is green