



NCBI's Database for Gene-Specific Information

Scope

The Entrez Gene database [1, 2] provides detailed information for known or predicted genes defined by nucleotide sequence or map position. Currently, Entrez Gene contains more than 3.7 million entries and includes data from all major taxonomic groups. The individual records in the database pertain to a single gene and originate from the processing and annotation of genomic sequences by the NCBI Reference Sequence [3, 4] and the Map Viewer [5] groups.

Access

To search the data in Entrez Gene via the World Wide Web:

www.ncbi.nlm.nih.gov/sites/entrez?db=gene

Entrez Gene data files, Gene References into Function (GeneRIF) data and file conversion tools are available at:

<ftp://ftp.ncbi.nih.gov/gene>

The data in Entrez Gene can also be retrieved programmatically via the Entrez Programming Utilities (eUtils):

eutils.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html

Search

Searching is as simple as entering the name of your favorite gene and pressing enter or clicking on "Go". Gene is part of NCBI's Entrez system, so you can also take advantage of functions provided under the Limits and Preview/Index tabs.

The **Limits** function can be used to restrict a search to a specific type of record in the database. Under Limits, search results can be optimized by checking one or more of the boxes under the different categories, or by limiting the search term to a field selected from the pull-down menu (A). For example, the unspecific search term "HFE" retrieves more than 30 individual gene records. However, restricting the search term to the "Gene Name" field in the pull-down menu and checking the "*Homo sapiens*" box under the "Limit by Taxonomy" section (B), will retrieve a single record for the human HFE gene (see Figure 2).

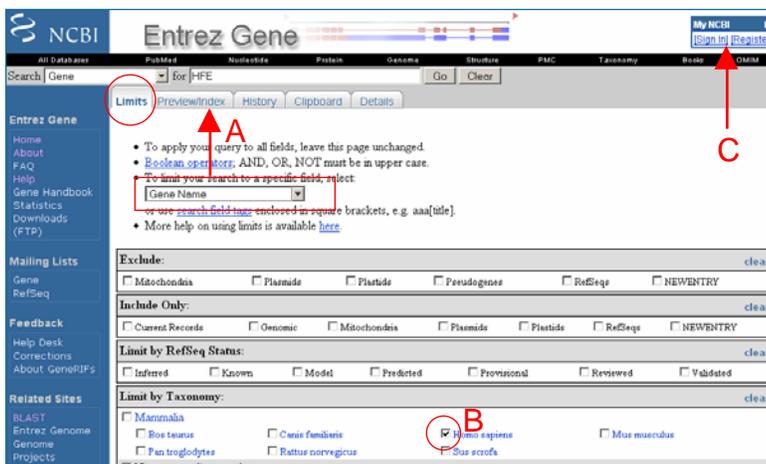


Figure 1. Entrez Gene user query interface and the Limits menu. In this search, the sample query term, HFE, is limited to the "Gene Name" field (A) and *Homo sapiens* (B). Searches can be saved temporarily and can be viewed under "History", or indefinitely from within MyNCBI (C).

The **Preview/Index** function (A) enables the construction of complex queries by restricting the search terms to specific fields combined with the Boolean operator terms "AND", "OR" or "NOT". For example, the following query:

```
Cancer [Disease/Phenotype] AND  
5 [Chromosome] AND human [Organism]
```

can be built by selecting fields under the pull-down menu and the Boolean operator term AND to retrieve all of the human, cancer-associated genes annotated on chromosome 5.

Recent queries are saved temporarily on the web can be viewed by clicking on the "History" tab. Use the **MyNCBI** tool (C) to save strategies indefinitely. For saved searches, MyNCBI also includes an option to receive e-mail alerts when new entries are released into the database.

Content

The information in a record depends on the source organism and the extent of gene-based information that is publicly available. Each Entrez Gene record contains a **Summary** section (Figure 2A) which provides a brief overview of the structure and function of the gene. For ease of navigation, the sub-topics listed in the **Table of Contents** (Figure 2B) link directly to the corresponding section within the report.

Links to view and download genomic, mRNA and protein sequences are given in the **Genomic regions, transcripts and products** section (Figure 2C). Additional sections include the **Genomic context** (i.e., the chromosomal location), a **Bibliography** section with links to the articles in PubMed which document gene function (GeneRIFs), sequence data (Reference Sequences and GenBank/DDBJ/EMBL) and general information about the gene (e.g., genetic markers, phenotypes, homologs and gene ontologies) and its encoded protein(s) (e.g., interaction data, metabolic pathways, conserved domains, etc.).

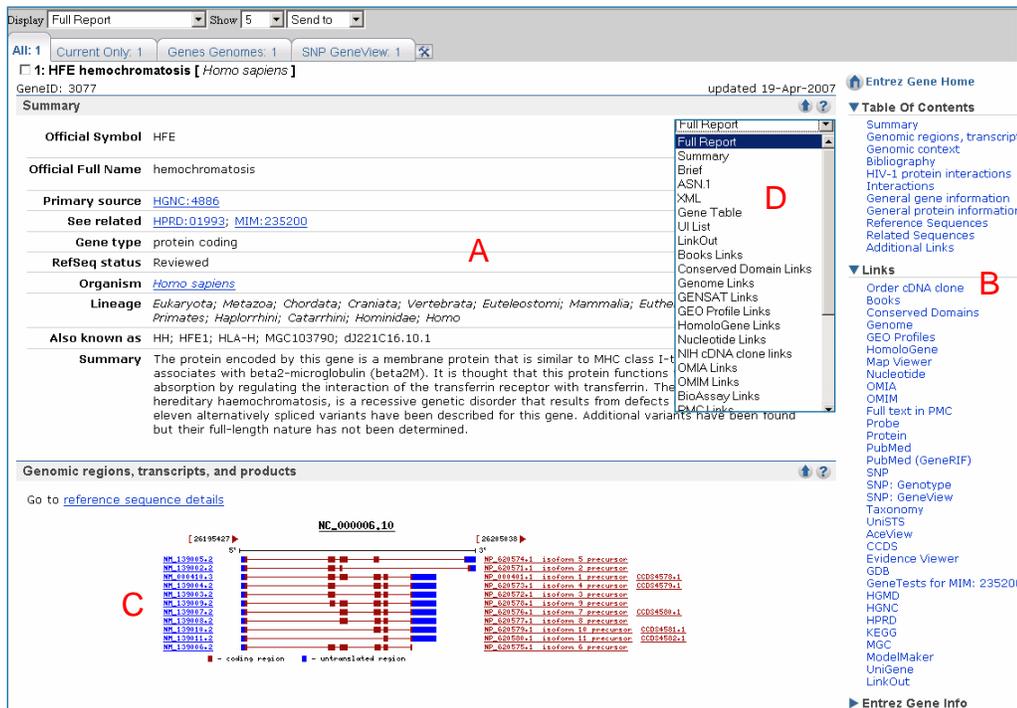


Figure 2. "Full Report" view of the Entrez Gene report for the human HFE gene. Only the top section of the full report is shown. (A) Gene "Summary" section; (B) Direct links (under "Table of Contents") to the specific elements of the Entrez Gene Report and under "Links", to other NCBI databases and selected external databases.; (C) "Genomic regions, transcripts, and proteins" section; (D) Display options.

Display Options

The default display of an Entrez Gene record is the "Full Report" view (Figure 2). Other display formats, such as XML, are also available and can be selected from the **Display** pull-down menu located at the top left of the page. A useful display option for eukaryotic protein-coding genes is the **Gene Table**.

NM_000410.3		length: 2222 bp, number of exons: 6	
NP_000401.1		length: 349 aa, number of exons: 6	
EXON	Coding EXON	INTRON	
coords	length	coords	length
62 - 297	236 bp	222 - 297	76 bp
3622 - 3885	264 bp	3622 - 3885	264 bp
4095 - 4370	276 bp	4095 - 4370	276 bp
5466 - 5741	276 bp	5466 - 5741	276 bp
5900 - 6013	114 bp	5900 - 6013	114 bp
6967 - 8022	1056 bp	6967 - 7007	41 bp

Figure 3. "Gene table" view for the human HFE gene. This view shows the coordinates of each exon and intron in one of the HFE transcript variants. The sequence of this variant is represented by the Reference Sequence NM_000410.3.

Download The entire Gene database can be downloaded from <ftp.ncbi.nih.gov/gene/DATA> (updated daily)

Help Provides access to detailed help, including how to use the eUtils for programmatic access to the data in Entrez Gene.[10]

Feedback On-line forms for submitting comments and suggestions to NCBI staff regarding the data in Entrez Gene, include corrections and for submitting GeneRIFs.

Subscriptions To subscribe to the Reference Sequence, Entrez Gene and Map Viewer Announce lists, use this root URL :

RefSeq: www.ncbi.nlm.nih.gov/mailman/listinfo/refseq-announce

Gene: www.ncbi.nlm.nih.gov/mailman/listinfo/gene-announce

Map Viewer: www.ncbi.nlm.nih.gov/mailman/listinfo/mapview-announce

[8] Joanna S. Amberger Donna Maglott and Ada Hamosh. Online mendelian inheritance in man (OMIM): A directory of human genes and genetic disorders. In: The NCBI Handbook. National Center for Biotechnology Information, 2002. www.ncbi.nlm.nih.gov/books/.

[9] Donna Maglott. Using Entrez Gene Help: Integrated Access to Genes of Genomes in the Reference Sequence Collection. In: The NCBI Help Manual 2005. www.ncbi.nlm.nih.gov/books/.

Links

A powerful feature of the Entrez Gene database is its integration with genome-related resources at NCBI or elsewhere. Figure 2B shows the Links menu for the human HFE gene record. Several of the links incorporate NCBI resources, such as Map Viewer [5], dbSNP [6], HomoloGene, and UniGene [7]. Information on genetic disorders is accessible through the Online Mendelian Inheritance in Man (OMIM) database [8]. Links to external resources, such as KEGG and Reactome, provide information on metabolic pathways. Organism-specific links, such as the one labeled "HGNC" in the Summary section of the HFE report (Figure 2A), link to the Human Genome Gene Nomenclature Committee, where more information about the official name of the gene, gene aliases and standards for the naming of genes and gene families, can be found.

References:

- [1] D Maglott, J Ostell, K D Pruitt, and T Tatusova. Entrez gene: gene-centered information at NCBI. Nucleic Acids Res, 35 (Database issue):D26-31, Jan 2007.
- [2] Kim Pruitt Donna Maglott and Tatiana Tatusova. Entrez gene: A directory of genes. In: The NCBI Handbook. National Center for Biotechnology Information, 2005. www.ncbi.nlm.nih.gov/books/.
- [3] K D Pruitt, T Tatusova, and D R Maglott. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res, 35 (Database issue):D61-65, Jan 2007.
- [4] Tatiana Tatusova Kim D. Pruitt and James M. Ostell. The reference sequence (RefSeq) project. In: The NCBI Handbook. National Center for Biotechnology Information, 2003. www.ncbi.nlm.nih.gov/books/.
- [5] Susan M. Dombrowski and Donna Maglott. Using the Map Viewer to explore genomes. In: The NCBI Handbook. National Center for Biotechnology Information, 2003. www.ncbi.nlm.nih.gov/books/.
- [6] Adrienne Kitts and Stephen Sherry. The single nucleotide polymorphism database (dbSNP) of nucleotide sequence variation. In: The NCBI Handbook. National Center for Biotechnology Information, 2003. www.ncbi.nlm.nih.gov/books/.
- [7] Lukas Wagner Joan U. Pontius and Gregory D. Schuler. Unigene: A unified view of the transcriptome. In: The NCBI Handbook. National Center for Biotechnology Information, 2003. www.ncbi.nlm.nih.gov/books/.

