NCBI Gene

A portal to gene-centered information from different sources

https://www.ncbi.nlm.nih.gov/gene/

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

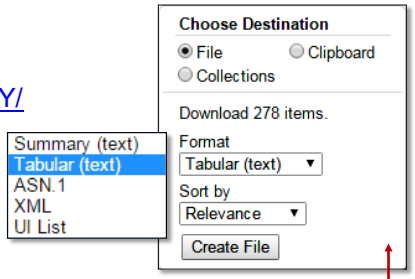
Scope and Access

The Gene database [1, 2] provides detailed information for known and predicted genes defined by nucleotide sequence or map position. Currently, Gene contains over 33 million entries and includes data from all major taxonomic groups. Each record in the database corresponds to a single gene and is derived from processing by the NCBI Reference Sequence [3, 4] and genome annotation groups [5]. You can access data from the Gene database on the web through the Gene homepage, programmatically through the Entrez Programming Utilities, or by file transfer through its FTP site.

- Gene Homepage: www.ncbi.nlm.nih.gov/gene/
- Entrez Programming Utilities: www.ncbi.nlm.nih.gov/books/NBK25501
- Gene Data Files: ftp.ncbi.nlm.nih.gov/gene/DATA/
- Gene Records in ASN.1 format: ftp.ncbi.nlm.nih.gov/gene/DATA/ASN_BINARY/

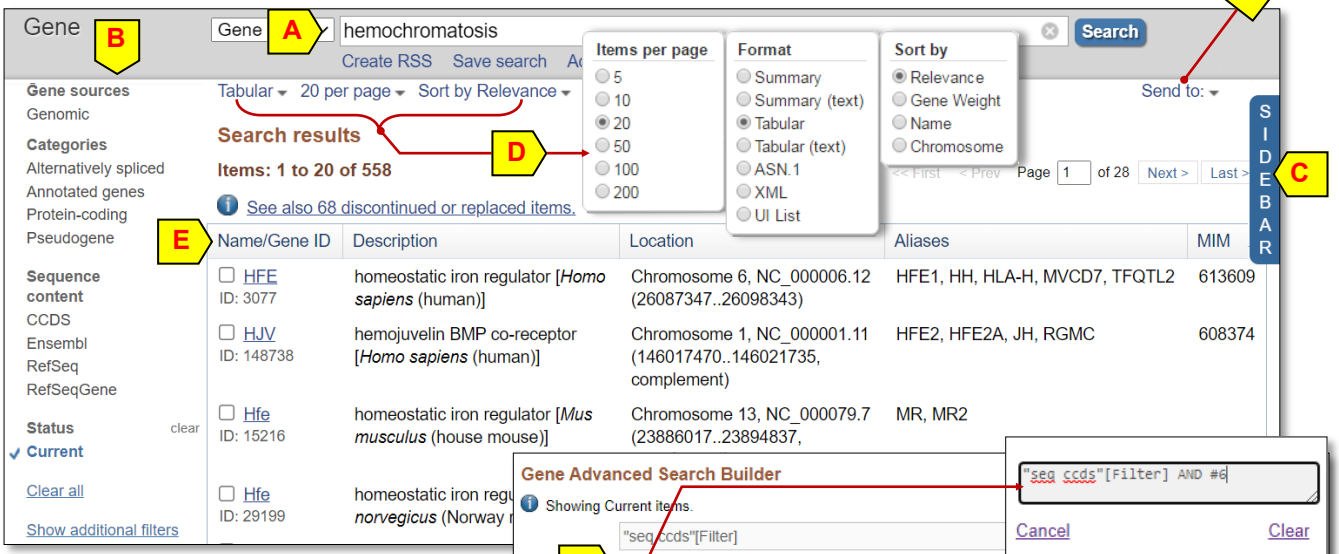
You can obtain technical details on the Gene database from the NCBI Bookshelf:

- Gene Help: www.ncbi.nlm.nih.gov/books/NBK3841/
- Gene FAQ: www.ncbi.nlm.nih.gov/books/NBK3840/



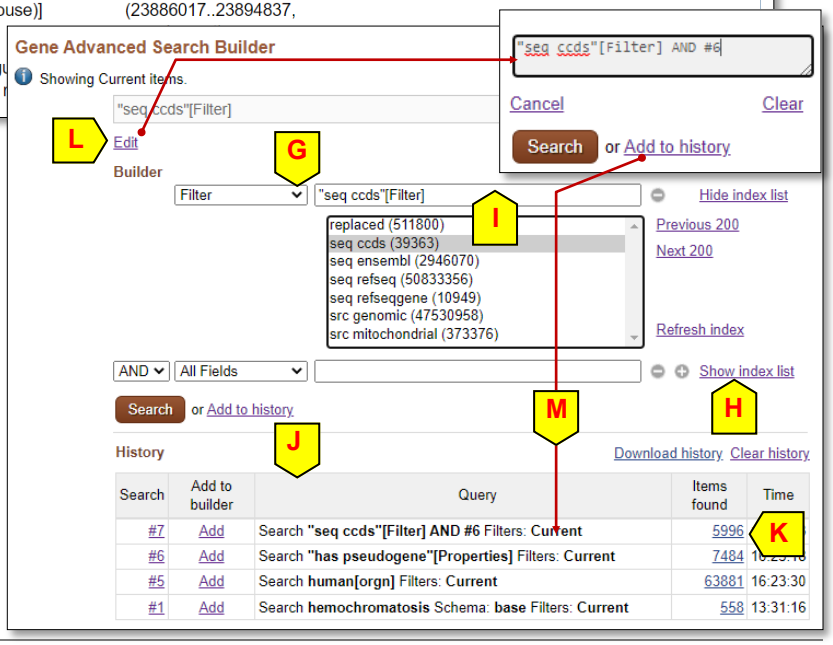
Searching Gene

Enter a set of query terms and click the Search button (**A**) to find gene records of interest. Select facet filters in the left column (**B**) to narrow down the retrieved list. Click the "SIDEBAR" (**C**) to expand and access functions listed in the right-hand column. Use the pull-down menus (**D**) at the top to change the display format or sorting order. From left to right, the default display (**E**) provides information on the official symbol and NCBI Gene ID, gene name and source organism, chromosome location, alias symbols associated with the gene, and OMIM id's if available. You can save retrieved Gene records in various formats using the "File" option available under the "Send to" link (**F**). Click the official symbol to open the full report of that Gene record.



Advanced Search Builder

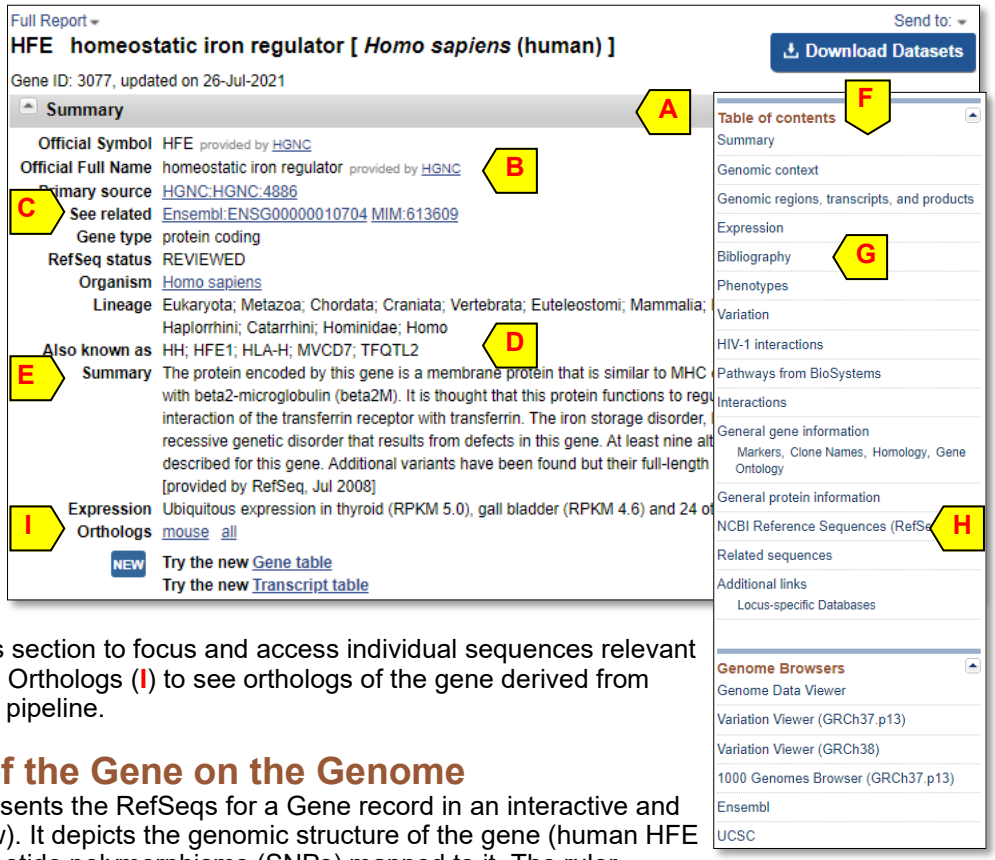
Use "Gene Advanced Search Builder" (right) to access indexed fields through a pull-down menu under the arrow (**G**), and examine terms indexed through the "Show index list" (**H**) link. Select a term to enter it in the query box (**I**) with the selected Boolean operator. Click "Add to history" (**J**) to preview and save a search to the History section below. Click the number under the Items found (**K**) column to retrieve these records. Use "Edit" (**L**) to customize the query terms. The example combines a custom term with an existing entry in the history, which was added back to the History section as a new entry (**M**). Log into your My NCBI account to permanently save a gene search result, and customize the display format using options under



Contact: info@ncbi.nlm.nih.gov

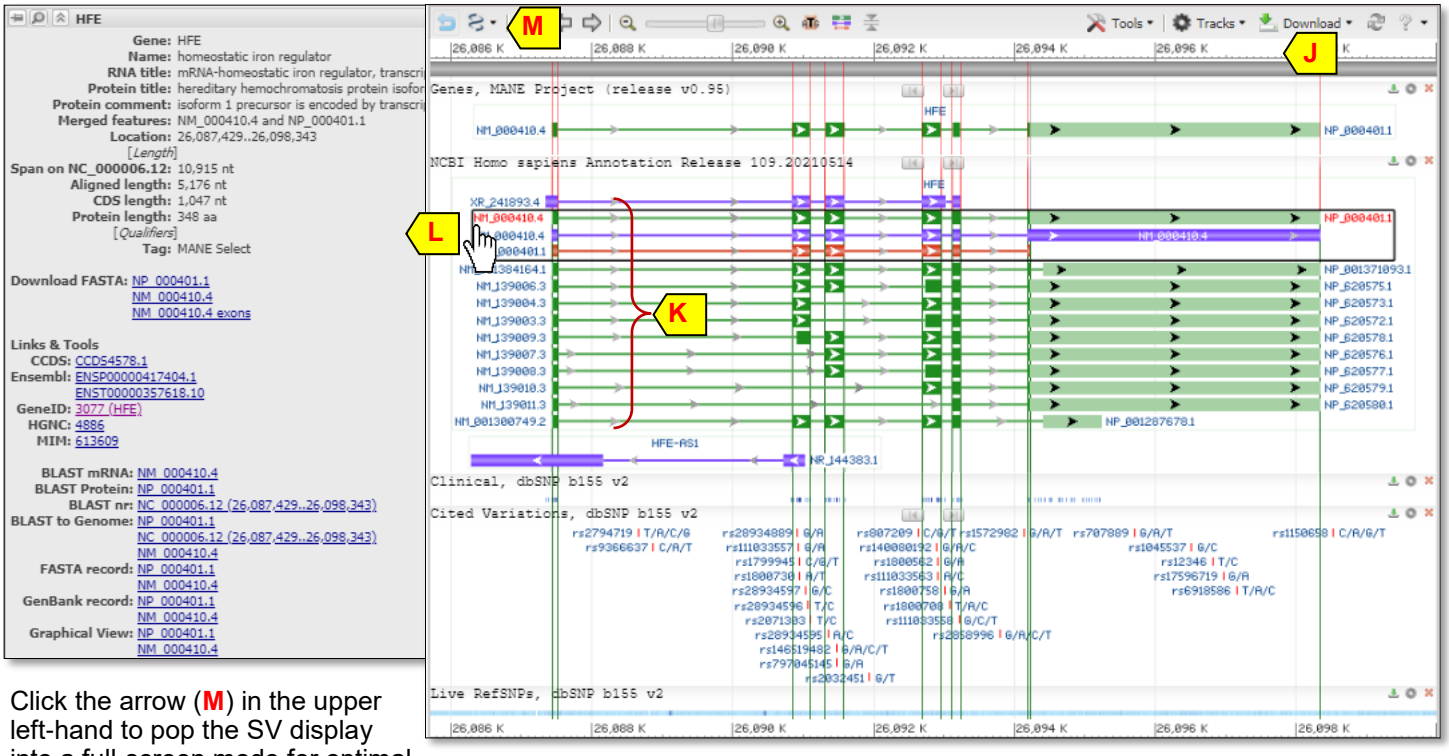
Contents of a Gene Record

Information found in a Gene record depends on publicly available data. All records, such as the well-studied human HFE gene (right), contain a Summary section (**A**) to provide an overview of the gene. The HFE gene record also contains links to corresponding records from HUGO Gene Nomenclature Committee (**B**), other external links such as OMIM [7] (**C**). It also provides a list of unofficial symbols associated with this gene (**D**) and a brief summary of the gene and its products (**E**). For ease of navigation, the Table of Contents (**F**) provides direct links to individual sections of a Gene report. For example, click Bibliography (**G**) to scroll the display to that section where publications relevant to the Gene record are listed. Similarly, click the Reference sequences (**H**) to bring the NCBI Reference Sequences section to focus and access individual sequences relevant to this gene. Click the "all" link next to Orthologs (**I**) to see orthologs of the gene derived from NCBI Eukaryotic Genome Annotation pipeline.



Graphical Presentation of the Gene on the Genome

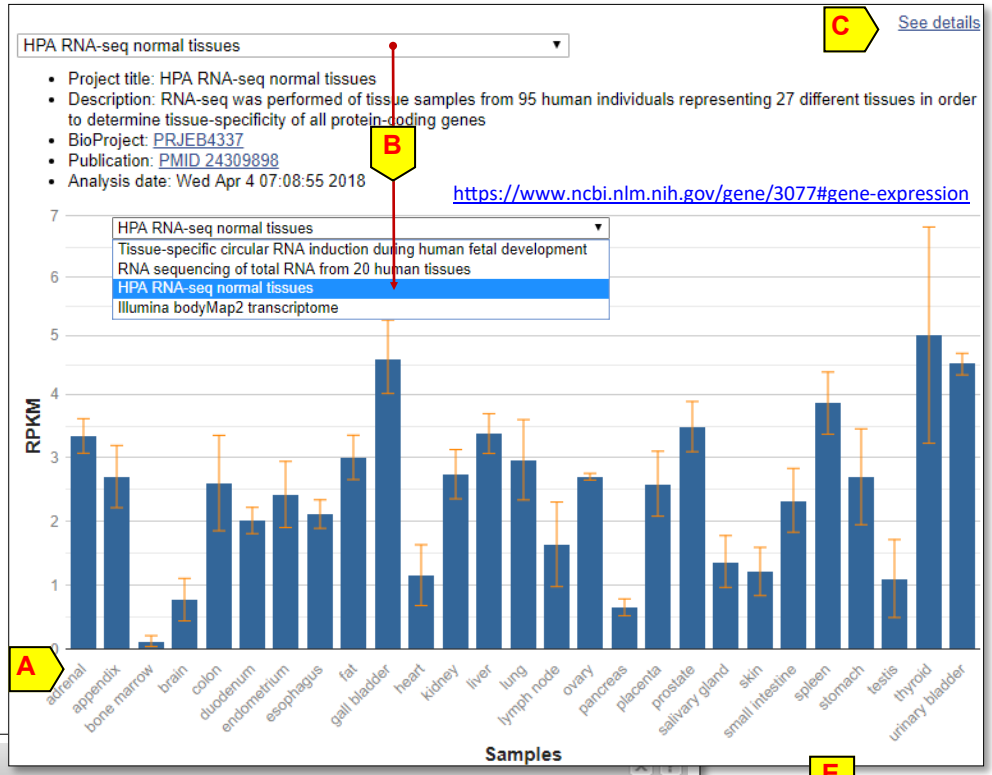
The Sequence Viewer (SV) panel presents the RefSeqs for a Gene record in an interactive and customizable graphical display (below). It depicts the genomic structure of the gene (human HFE in this case), its size, and single nucleotide polymorphisms (SNPs) mapped to it. The ruler provides the chromosome mapping information (**J**). The display lists alternatively spliced transcripts and their protein products by depicting them as bars (exons) linked by thin lines (introns, **K**). Mouseover the accession of a transcript to get the summary information in a popup (**L**).



Click the arrow (**M**) in the upper left-hand to pop the SV display into a full-screen mode for optimal results and more functionality. Features available in this mode allow customizing the display, importing data in various formats, viewing and downloading subsequences, as well as selecting a region to BLAST and bringing the alignment results back into the display as new tracks. See SV document linked off the full-screen mode for more details.

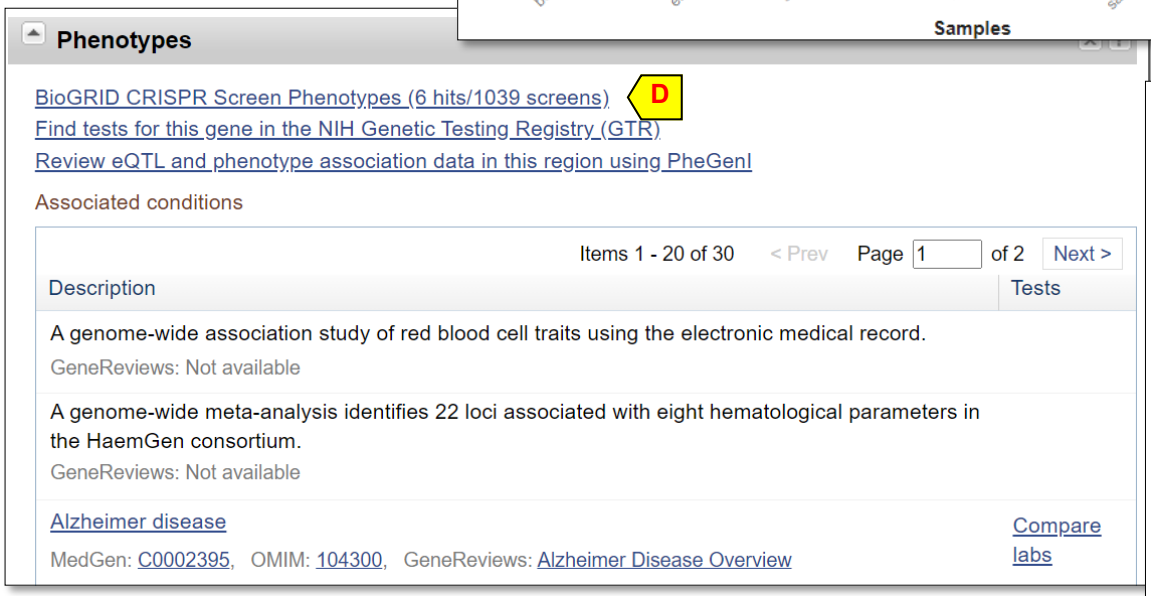
Expression

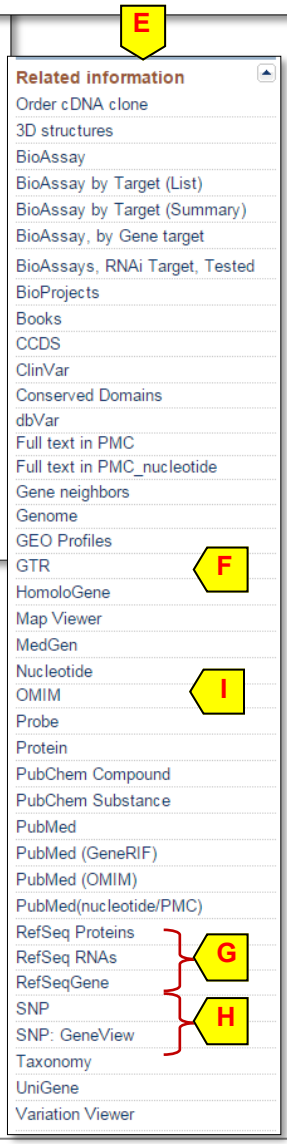
This newly added section sums up level of gene expression across various tissues (**A**), with information derived from alignment of Body Map 2 RNA-seq data (PRJNA144517). You can use the pull-down menu to select source data from different studies (**B**). Click "See details" link (**C**) to see this report in a new window, along with the underlying data table and detailed counts from individual samples.



Phenotype Information

The "Phenotypes" section summarizes available information from OMIM, NHGRI, and PubMed as a list of links. The Phenotype and Genotype Integrator [8] (PheGenI, **D**) link provides a user-friendly interface to a collection of genotype/ phenotype data from available GWA studies.





Related Records

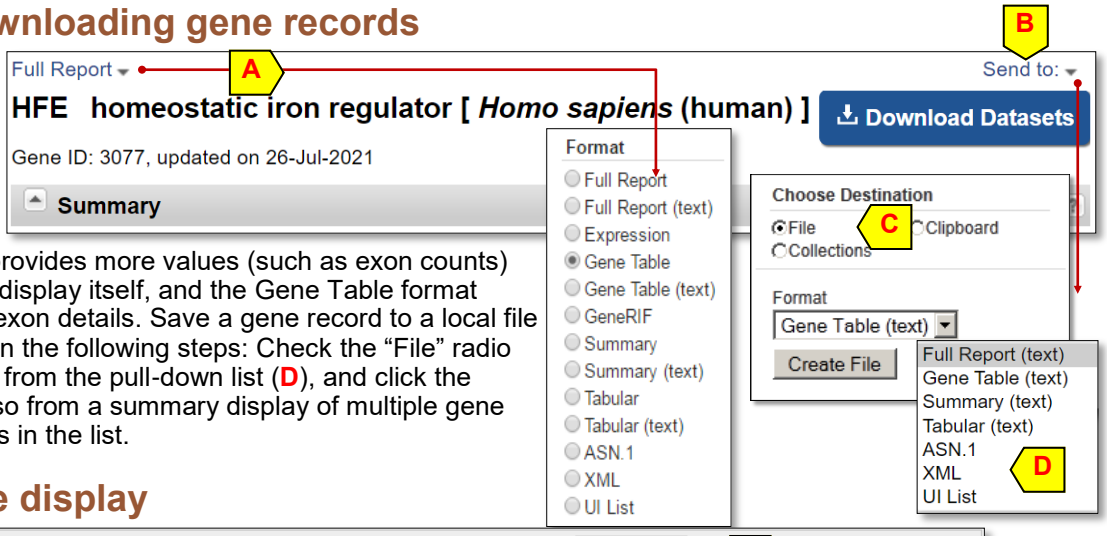
A powerful feature of the Gene database is its integration with other resources at NCBI and elsewhere. The information is provided through links under the "Related information" section (**E**) in the right hand column. For the human HFE gene, this section contains genetic testing information from GTR (**F**), protein, RNA and genomic Reference Sequence records for the gene (**G**), various subsets of SNPs for the gene in dbSNP [6] (**H**), and Information on genetic disorders from/through OMIM (**I**).

For genes involved in known biologic pathways, links to BioSystems and outside resource like Kyoto Encyclopedia of Genes and Genomes (KEGG) are also available. A gene record may contain a "Pathways from PubChem" section to provide additional information on the pathways the gene products are involved in. Certain records also have links to databases outside NCBI to provide disease information or other details.

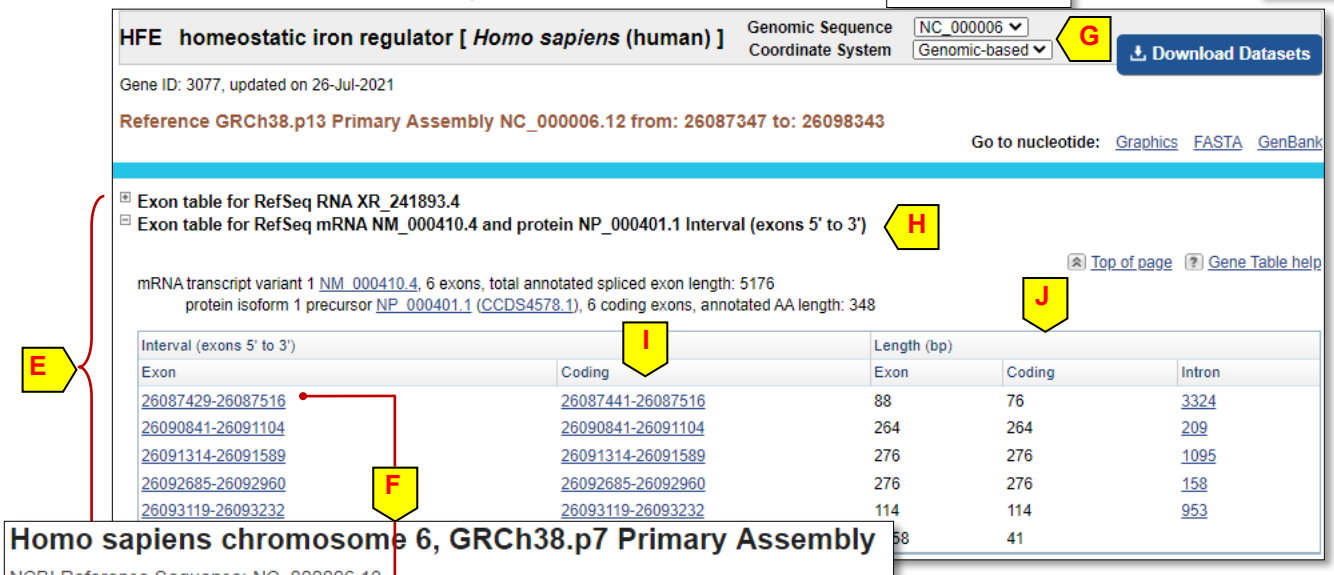
The NCBI genome annotation process also analyzes and collects homologs for available genes and makes them available as the Ortholog link in the Summary section (p. 2).

Displaying and downloading gene records

Use the pull-down menu (**A**) to change the display format of a Gene record. The browser will automatically switch the display format selected from the pull-down menu. The "Tabular (text)" option provides more values (such as exon counts) than visible in the "Tabular" display itself, and the Gene Table format provides transcript-specific exon details. Save a gene record to a local file using the "Send to" link (**B**) in the following steps: Check the "File" radio button (**C**), select file format from the pull-down list (**D**), and click the "Create File" button. Doing so from a summary display of multiple gene records saves all the records in the list.



Special Gene Table display



With genome annotation, the "Gene Table" format will be available to display genomic coordinates of exons and introns for all annotated transcript variants of the gene (**E**). Clicking a set of coordinates (**F**) retrieves the subsequence for that element in FASTA format. Default coordinates are based on the highest level reference assembly for that genome, such as the reference assembly for the chromosome. If needed, you can also select to use coordinates on other RefSeq records through options in the pull-down menus (**G**). One of the splicing variants for human HFE displayed under "Gene Table" format is NM_000410.3 encoding a protein product NP_000401.1 (**H**). This variant has six exons, all of which are coding (**I**). Columns under the Length (bp) heading (**J**) contain the exon lengths and the length of the intervening introns.

References

1. Maglott D, et al. Gene: a gene-centered information resource at NCBI. Nucleic Acids Res. 2014. pii: gku1055. [Epub ahead of print]
2. Maglott D, et al. Gene, The NCBI Handbook, 2nd Ed. NCBI, 2013. www.ncbi.nlm.nih.gov/books/NBK169435/
3. Pruitt K, et.al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 2014. 42(Database issue):D756-63.
4. Pruitt K, et al. Chapter 18, The NCBI Handbook. NCBI, 2003. www.ncbi.nlm.nih.gov/books/NBK21091/
5. Thibaud-Nissen, et. al. Eukaryotic Genome Annotation Pipeline, The NCBI Handbook, 2nd Ed. www.ncbi.nlm.nih.gov/books/NBK169439/
6. Kitts A, Sherry S. Chapter 5, The NCBI Handbook. NCBI, 2003. www.ncbi.nlm.nih.gov/books/NBK21088/
7. Amberger JS, et al. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2014 Nov 26. pii: gku1205. [Epub ahead of print]
8. Gene Help Manual www.ncbi.nlm.nih.gov/books/NBK3841/
9. MyNCBI help manual. https://www.ncbi.nlm.nih.gov/books/NBK3842/