dbSNP: Database for Short Genetic Variations

Catalog of nucleotide changes for human and other model organisms

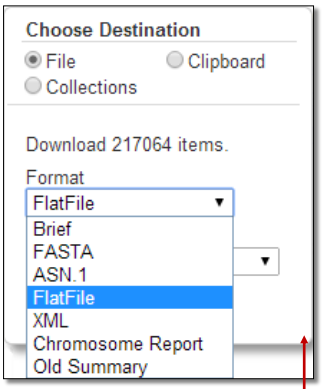
https://www.ncbi.nlm.nih.gov/snp/

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

Scope and Access

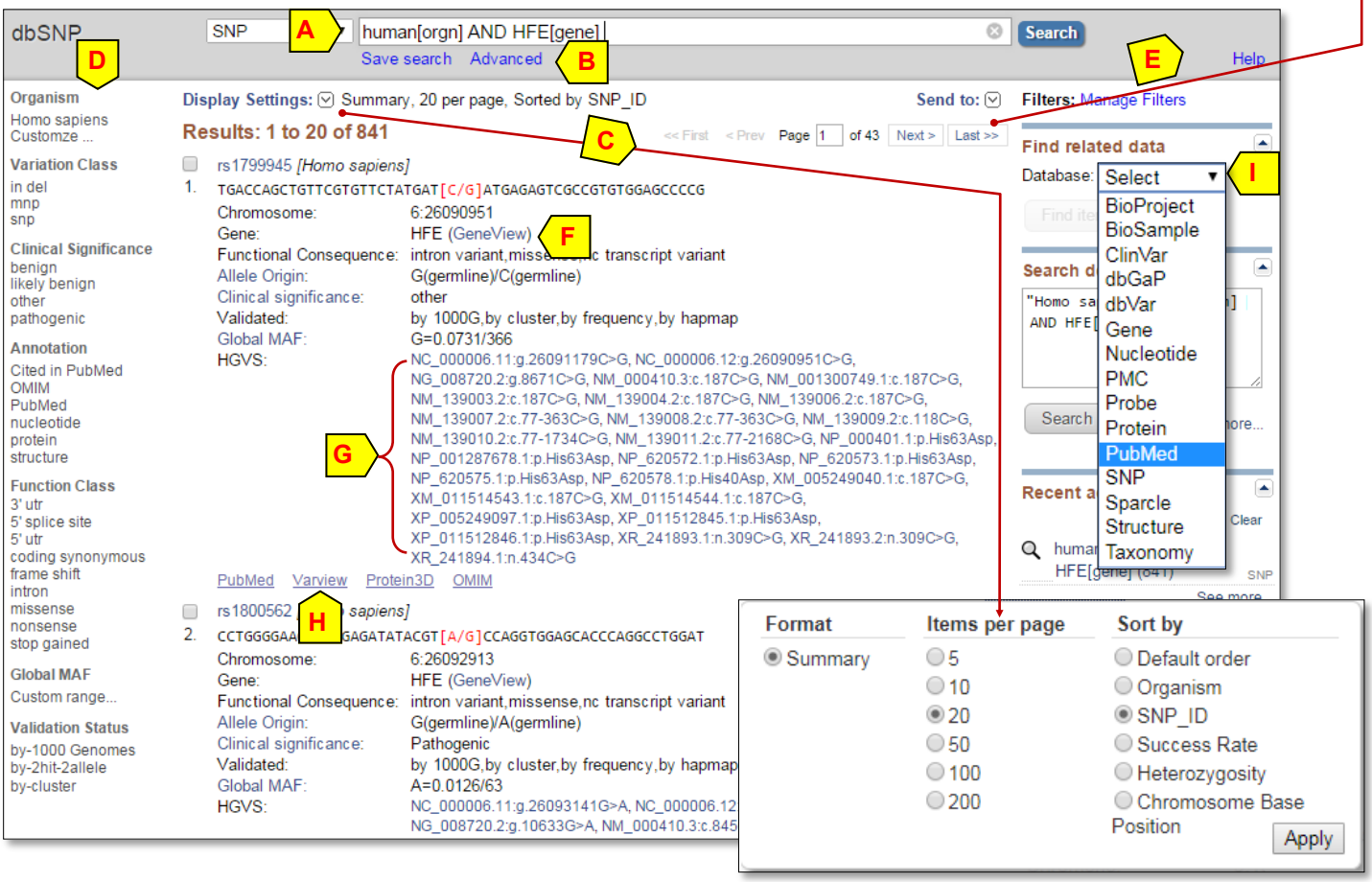
The NCBI Short Genetic Variations database (dbSNP) [1], commonly known as dbSNP, catalogs short variations in nucleotide sequences for human. These variations include single nucleotide variations, short nucleotide insertions and deletions, short tandem repeats. Short Genetic Variations may be common, thus representing true polymorphisms, or they may be rare. Some rare human entries have additional information associated with them, including disease associations from ClinVar [2], genotype information and allele origin, as some variations are somatic rather than germline events.

Short nucleotide variation data can be accessed via the SNP homepage and EUtils API:
www.ncbi.nlm.nih.gov/snp and www.ncbi.nlm.nih.gov/projects/SNP/SNPeutils.htm
VCF files and database bcp files are available for download through FTP and Aspera client at:
ftp.ncbi.nlm.nih.gov/snp/ and www.ncbi.nlm.nih.gov/public/?snp/organisms/
The SNP API service, SPDI [3], is available at:
api.ncbi.nlm.nih.gov/variation/v0/
SNP data can also be accessed interactively through Variation Viewer:
www.ncbi.nlm.nih.gov/variation/view/



Searching for and Displaying SNP Records

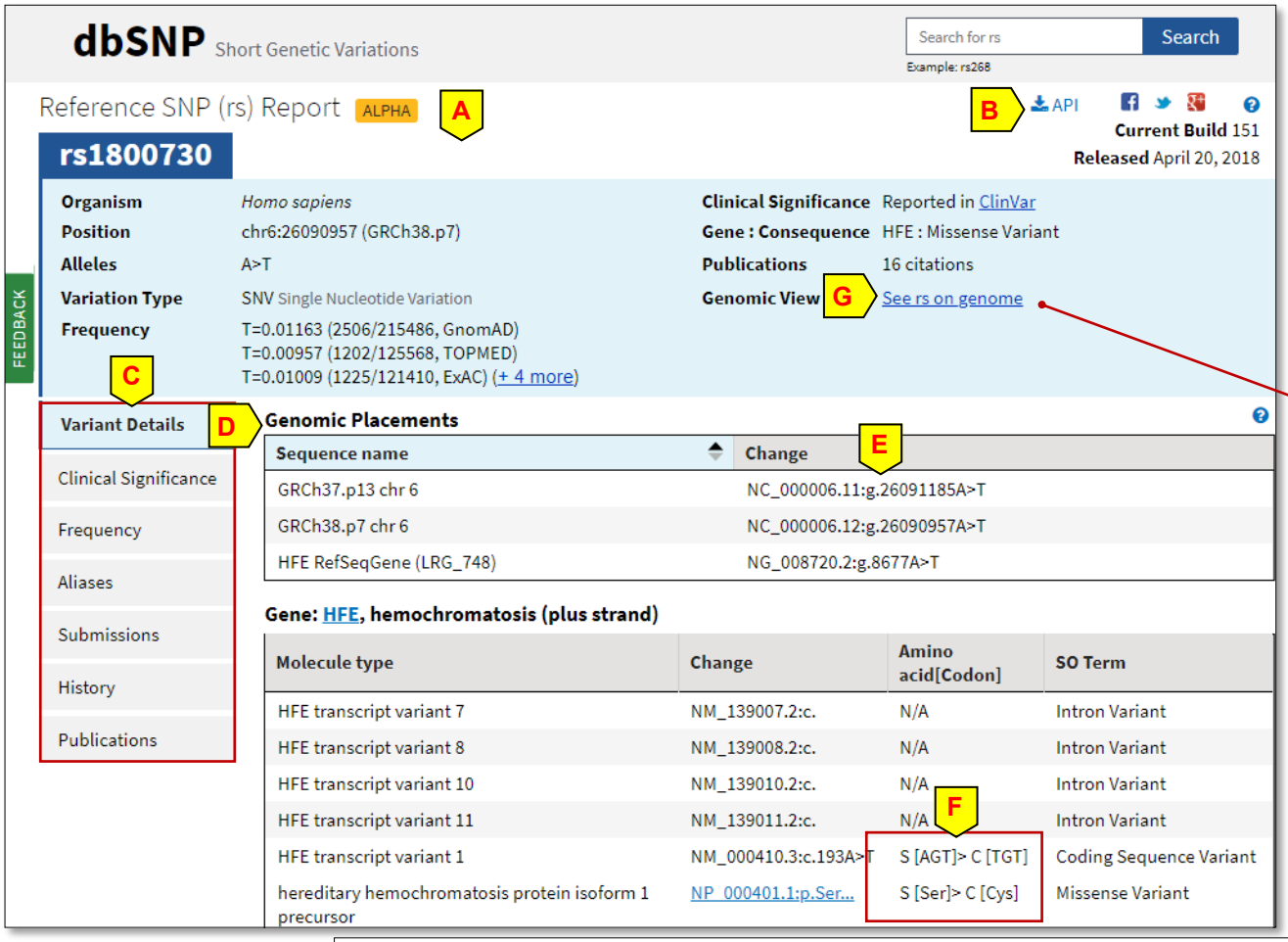
You can search for variations on the dbSNP homepage by typing a query term in the search box and clicking the Search button (**A**). You can also use the Advanced (**B**) page to create complex queries to produce more precise results. The search below, "*hfe[gene] AND human[orgn]*", retrieves variations mapped to the human HFE gene. You can use options in the Display settings popup (**C**) to change the number of records displayed or sort retrieved variations in a different order. You can further narrow down retrieved variations by selecting filters present in the left column (**D**), or save them to a local file using the Send to (**E**) option. Use links to separate displays to see gene-centric listings (GeneView, **F**), graphical presentation under the context of genome or mRNA sequences (via HGVS names, **G**), or gene-centric display in a genomic context (Varview, **H**). Using the "Find related data" portet (**I**), you can retrieve related entries from other NCBI databases for the set of variations in the display.



Contact: info@ncbi.nlm.nih.gov

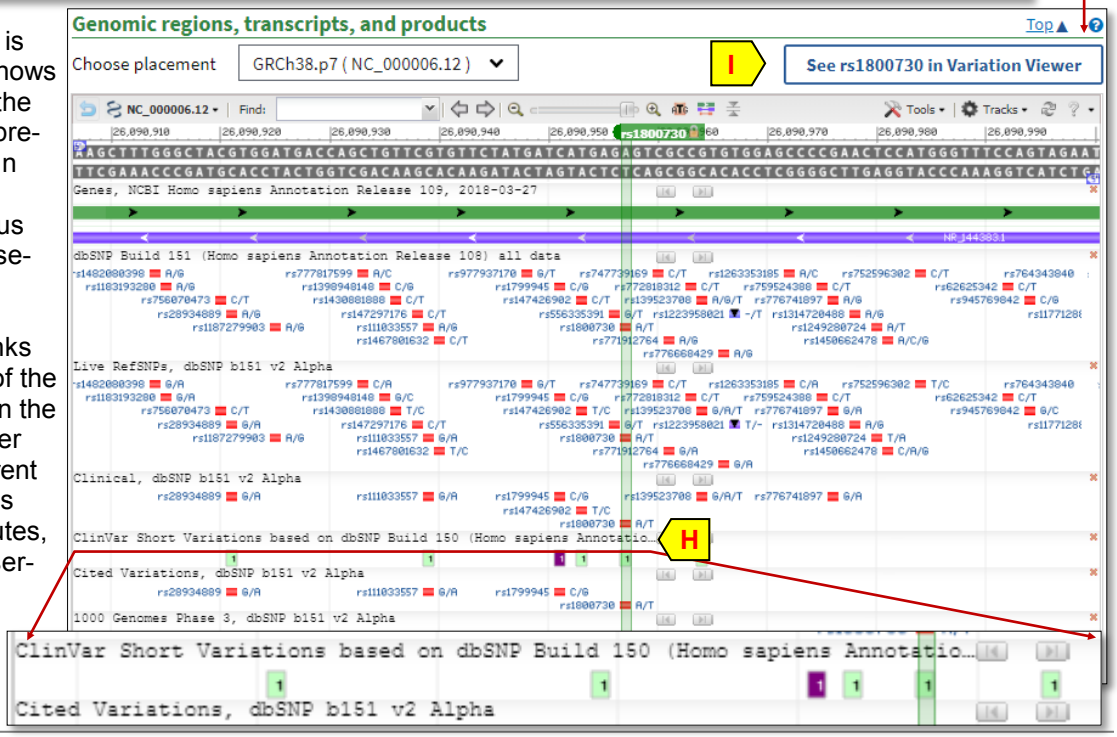
The New Reference SNP Report

The new Reference SNP Report linked from rsIDs, such as rs1800730 shown below and on p.3, shows details of a variation record. The summary section at the top (**A**) provides an overview of the variant in a self-explanatory key/value format. It reports the allele in the forward orientation of the chromosome record. The information in display is also available in JSON format through the API link at the upper right (**B**). The new report separates details of the variation into various categories (**C**) and displays them under separate tabs.



The Variant Details tab (**D**) is open by default. This tab shows the genomic placement of the variant on the current and previous genome assemblies in HGVS format (**E**), with the gene/transcript mapping plus the relevant molecular consequences (**F**) given below.

The "See rs on genome" links to the graphical panel (**G**) of the display. It puts the variant in the context of genome and other neighboring variants. Different tracks bins mapped variants according to specific attributes, such those with clinical assertions (**H**). Use the button at upper right (**I**) to go to Variation Viewer for more detailed examination.



The New Reference SNP Report (cont.)

Other tabs in the new Reference SNP Report provide category-specific information.

The **Clinical Significance** tab (**A**) lists related clinical assertions for the variant from ClinVar, with IDs linking directly to the records there.

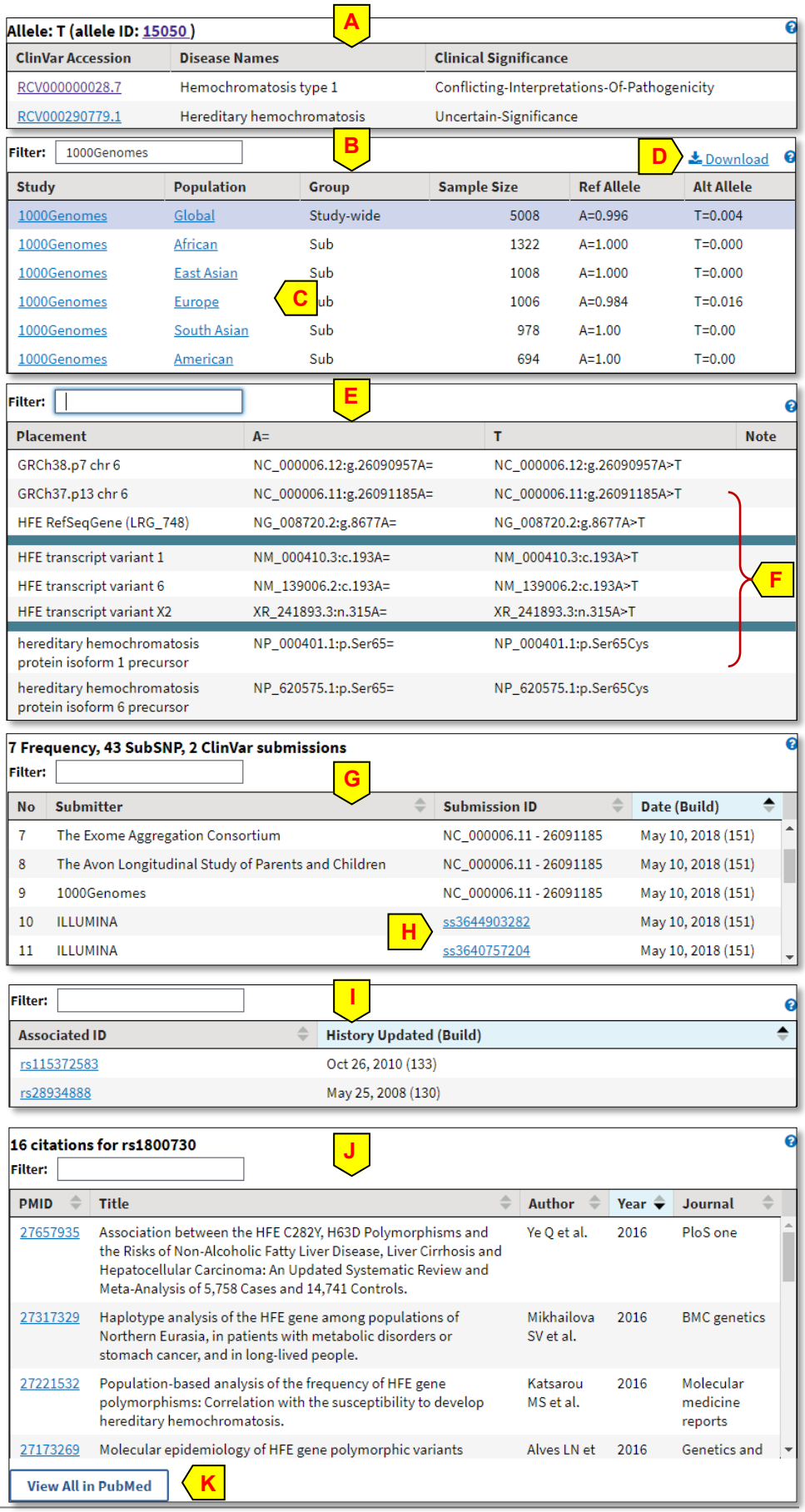
The **Frequency** tab (**B**, filtered by "1000Genomes") lists allele frequency data from major studies, such as 1000 Genomes, ExAc, and Genome Aggregation Database. It will provide frequencies for major populations (**C**) if such a frequency breakdown is submitted by these studies. You can use the "Download" link (**D**) to get the data in a tab-delimited format. This provides a way to evaluate the impact of a variant if there is no information in the **Clinical Significance** and **Publications** sections.

The **Aliases** tab (**E**) lists HGVS names that can be used to describe the same variant. These names uses different reference accessions, but they are equivalent and point to the same genomic variation. The table separates names into genomic, transcript non-coding, and transcript coding groups (**F**).

The **Submission** tab (**G**) lists submitter/batches that reported the variant. Submitters include large study projects or individual submitters. Older submission before adoption of asserted location will have ssIDs (**H**).

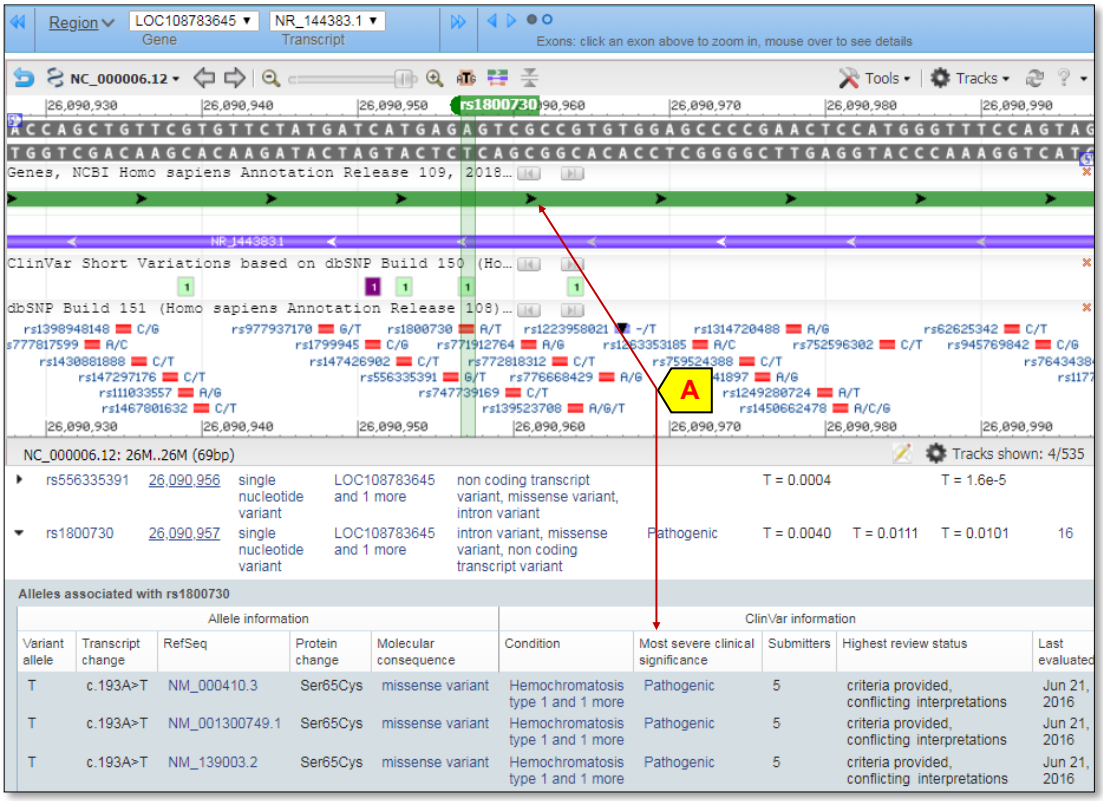
The **History** tab (**I**) tracks the change of the cluster and lists other rsIDs that has merged into this variant. In this case, variants rs115372583 and rs28934888 were determined to be duplication of rs1800730, so they were merged into a single record.

dbSNP established explicit connection between Reference SNP variant and biomedical literature citations through text-mining. The Reference SNP Report displays these connections under the **Publications** tab (**J**). You can use the "View All in PubMed" button (**K**) to retrieve the list of citations in PubMed and examine their abstracts for more information. Some of these citations may also have free full-text available from PubMed Central for detailed online reading.



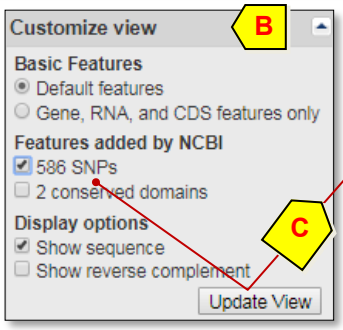
Variation Viewer

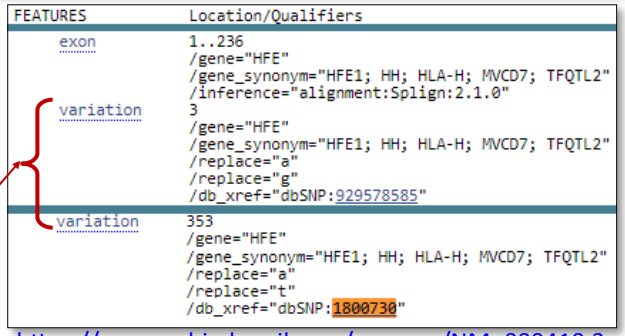
The Variation Viewer (p.2, **I**) provides an interactive display of the variant under the context of annotation of the selected genome assembly. It correlates a variation and its molecular consequences in the data table with its genomic context in the graphical display (**A**). Filters in the left hand column (not shown) are available to selectively display variants of interest. More information on this tool is available online [4 and 5]



Other Ways to Access dbSNP Data

The SNP database is fully integrated with the Entrez system, enabling the access of variation data through links present in records from other NCBI databases. For example, you can project variations mapped to a RefSeq genomic or mRNA record (with NT_, NG_, NW_ or NM_ accessions) by using the Customize view (**B**) menu in the upper right hand corner of the sequence record, simply check the SNPs checkbox and click Update View (**C**) to activate the selection.





https://www.ncbi.nlm.nih.gov/nuccore/NM_000410.3

dbSNP also integrates disease-related nucleotide variations that were reported in literature and cited in rsID format, collected by OMIM, or submitted to ClinVar. The table below is the Allelic Variant display for OMIM record 613609, which cites the rsIDs in the dbSNP column (**D**).

613609
HFE GENE; HFE
Allelic Variants (11 Selected Examples) :
D
http://omim.org/allelicVariant/613609
All ClinVar Variants

Number	Phenotype	Mutation	dbSNP	ExAC	ClinVar
.0001	HEMOCHROMATOSIS, TYPE 1 PORPHYRIA CUTANEA TARDA, SUSCEPTIBILITY TO, INCLUDED PORPHYRIA VARIEGATA, SUSCEPTIBILITY TO, INCLUDED HEMOCHROMATOSIS, JUVENILE, DIGENIC, INCLUDED ALZHEIMER DISEASE, SUSCEPTIBILITY TO, INCLUDED TRANSFERRIN SERUM LEVEL QUANTITATIVE TRAIT LOCUS 2, INCLUDED MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 7, INCLUDED	HFE, CYS282TYR	[rs1800562]	-	[RCV000210820...]
.0002	HEMOCHROMATOSIS, TYPE 1 MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 7, INCLUDED	HFE, HIS63ASP	[rs1799945]	[rs1799945]	[RCV000000027...]
.0003	HEMOCHROMATOSIS, TYPE 1	HFE, SER65CYS	[rs1800730]	-	[RCV000290779...]
.0004	HFE INTRONIC POLYMORPHISM	HFE, 5569G-A	[rs1800758]	[rs1800758]	[RCV000000031]
.0005	HFE POLYMORPHISM	HFE, VAL53MET	[rs28934889]	-	[RCV000000032]
.0006	HFE POLYMORPHISM	HFE, VAL59MET	[rs111033557]	-	[RCV000000033]
.0007	HEMOCHROMATOSIS, TYPE 1	HFE, GLN127HIS	[rs28934595]	-	[RCV000000034]
.0008	HEMOCHROMATOSIS, TYPE 1	HFE, ARG330MET	[rs111033558]	-	[RCV000000035]

References

1. The Database of Short Genetic Variation (dbSNP). Kitts A, Phan L, Ward MH, and Holmes JB. In The NCBI Handbook [Internet], 2nd ed. https://www.ncbi.nlm.nih.gov/books/NBK174586/
2. ClinVar: improving access to variant interpretations and supporting evidence. Landrum MJ, et al. Nucleic Acids Res. 2018 Jan 4;46(D1):D1062-D1067. https://www.ncbi.nlm.nih.gov/pubmed/29165669
3. New Web Services for Comparing and Grouping Sequence Variants. https://go.usa.gov/xUeKT.
4. Variation Viewer factsheet. https://ftp.ncbi.nih.gov/pub/factsheets/Factsheet_Variation_Viewer.pdf
5. Variation Viewer Online video tutorial. https://www.youtube.com/watch?v=rnWZ9MFBwUM