dbSNP: Database of Short Genetic Variations

An expansive catalog of short nucleotide changes for human

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

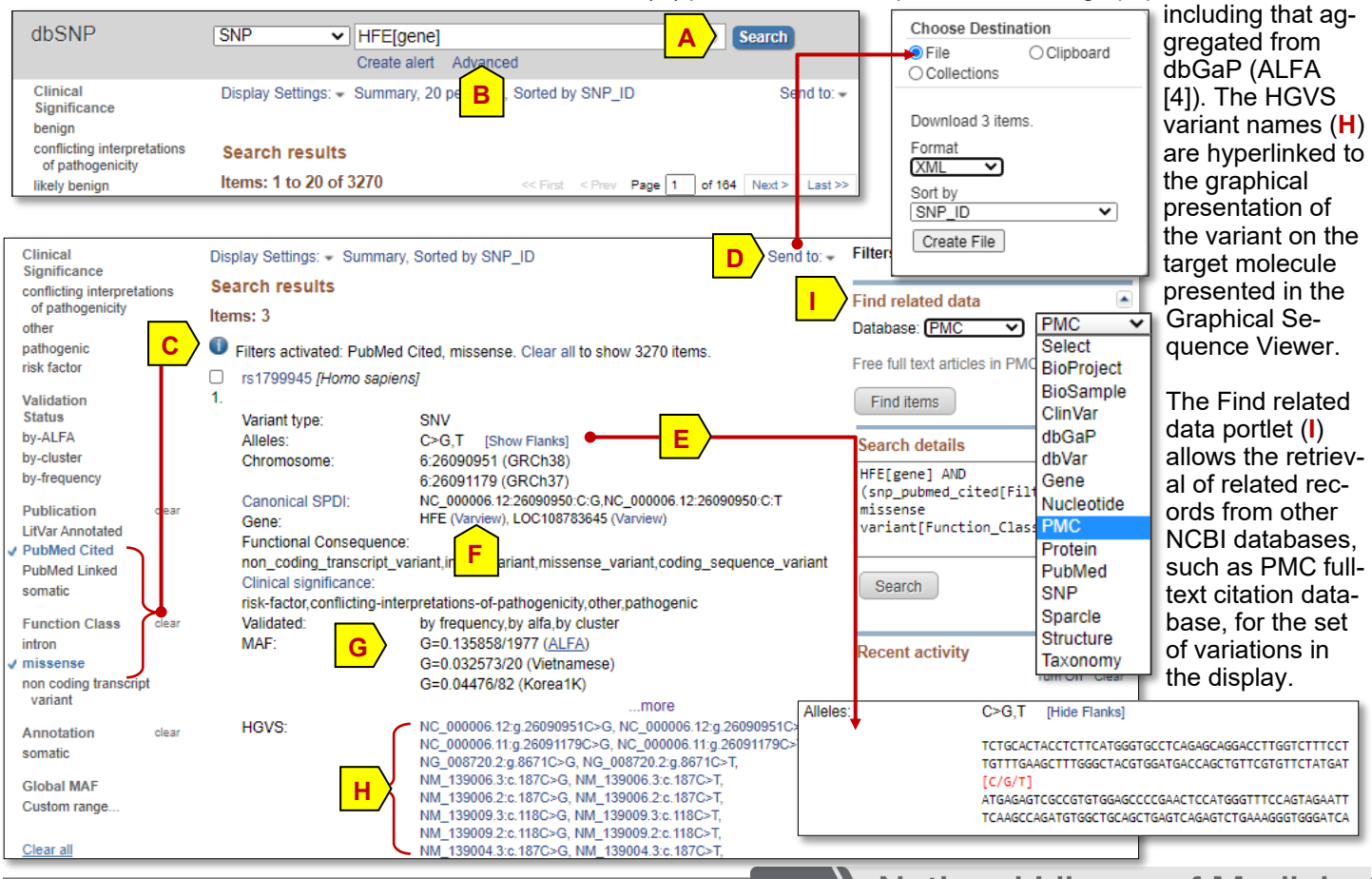
The NCBI Short Genetic Variation database (dbSNP) [1], commonly known as dbSNP, catalogs short variations in nucleotide sequences for human. These variations include single nucleotide variations, as well as insertions, deletions, and short tandem repeats less than 50 nucleotides in length. Short genetic 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 arises in somatic rather than from germline.

Short nucleotide variation data can be accessed through the dbSNP homepage and EUtils API:
www.ncbi.nlm.nih.gov/snp and www.ncbi.nlm.nih.gov/books/NBK25501
VCF files JSON files are available for download through FTP:
ftp.ncbi.nlm.nih.gov/snp/latest_release/
API services based on the SPDI notation system [3] is available at:
api.ncbi.nlm.nih.gov/variation/v0/
dbSNP data can also be examined under the genomic context through the 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**), or use the Advanced (**B**) page to create complex queries for more precise results. This interface now accepts SPDI notation (e.g., NC_000008.11:19953314:G:A), HGVS (e.g., NM_000237.3:c.1421C>G), and GRCh37 chromosomal position (e.g., 63499726[POSITION_GRCH37] AND 8[CHR]). More information is at: https://go.usa.gov/xGkFa.

A field-limited term **HFE[gene]** retrieves variations mapped to the HFE gene, and selecting from the preset filters in the left column refines the list to those matching the selected criteria (**C**). The Send to dialog box (**D**) allows downloading of retrieved SNPs to a local file in supported formats. The newly introduced Show Flank link (**E**) dynamically insert the short flanking sequences under the Alleles field. The VarView (**F**) link graphically presents the variant under the context of genomic annotation in the Variation Viewer. The MAF field (**G**) provides allele frequencies from large population studies, including that aggregated from dbGaP (ALFA [4]). The HGVS variant names (**H**) are hyperlinked to the graphical presentation of the variant on the target molecule presented in the Graphical Sequence Viewer.

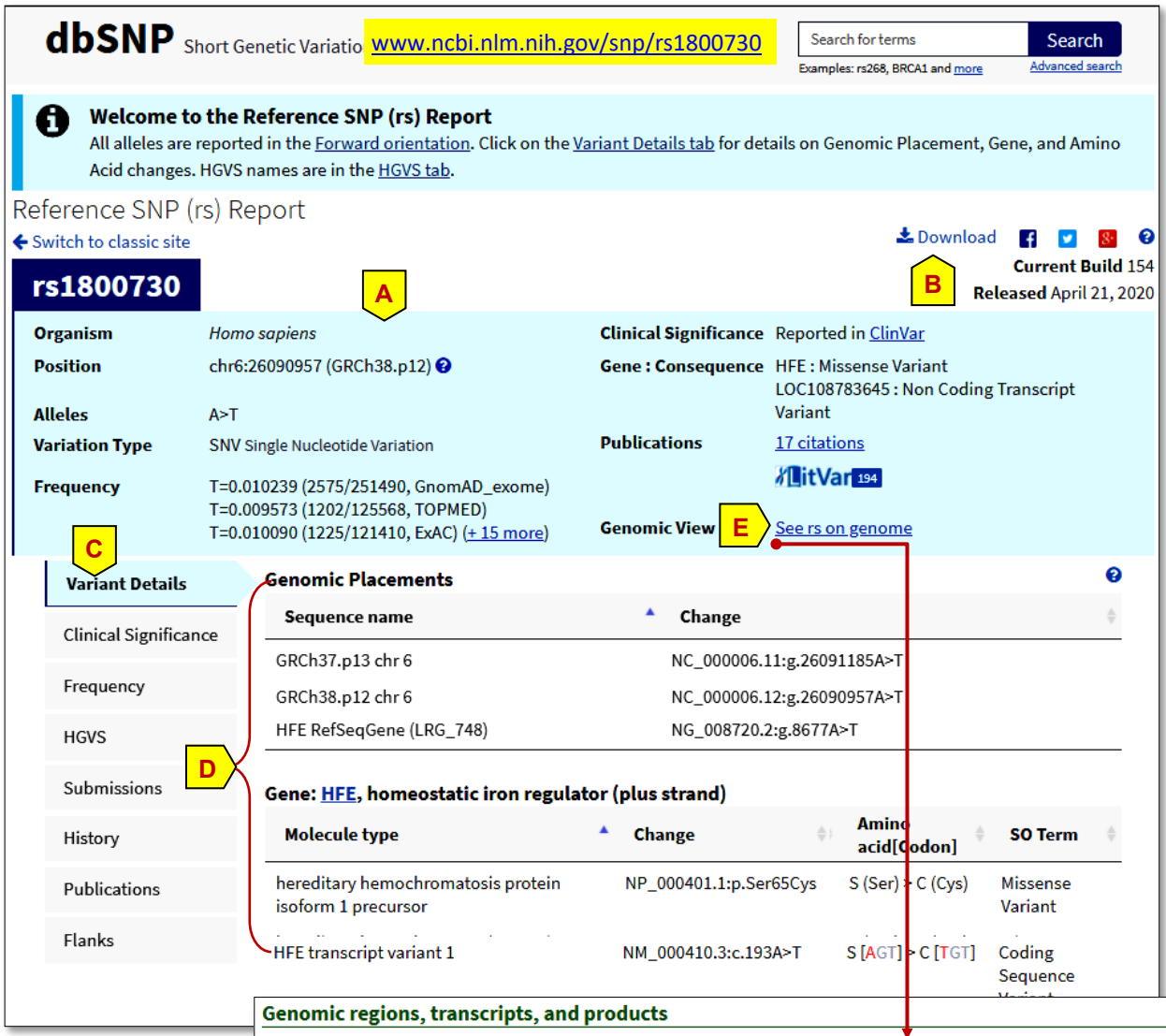


The Find related data portlet (**I**) allows the retrieval of related records from other NCBI databases, such as PMC full-text citation database, for the set of variations in the display.

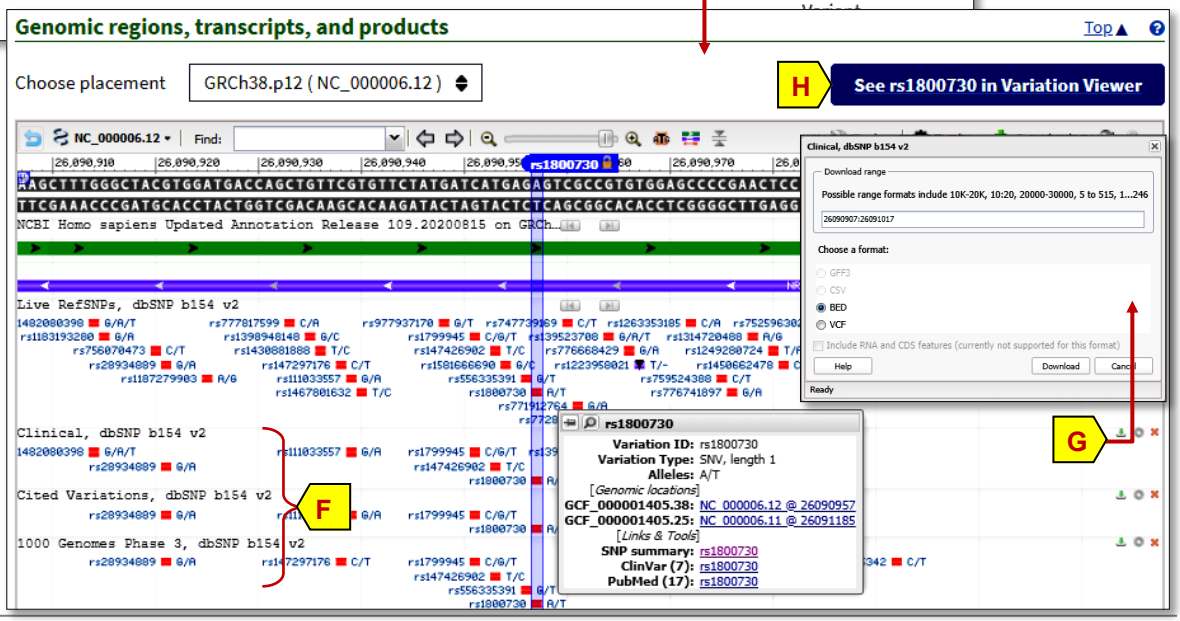
NIH National Library of Medicine
National Center for Biotechnology Information

The SNP Report

The Reference SNP Report linked from rsIDs, such as rs1800730 shown below and on p.3, presents the available information of a dbSNP variation record. The summary section at the top (**A**) provides an overview of the variant, reports the allele in the forward orientation of the chromosome, and summary allele frequencies when available. Links to related records in other databases are listed in the right hand column. The information in display is also available in JSON format through the Download link at the upper right (**B**). That function is provided by the Variation Service API, and more information is available at: api.ncbi.nlm.nih.gov/variation/v0/#/



This SNP report separates details of the variation into various categories and lists them in the vertical tabs (**C**) below. Default "Variant Details" (**D**) lists the genomic placement in HGVS format, and gene mapping information along with protein- and transcript-level details. The "See rs# on genome" link (**E**) scrolls the display at the end, showing the variant in the context of genomic annotation and other neighboring.

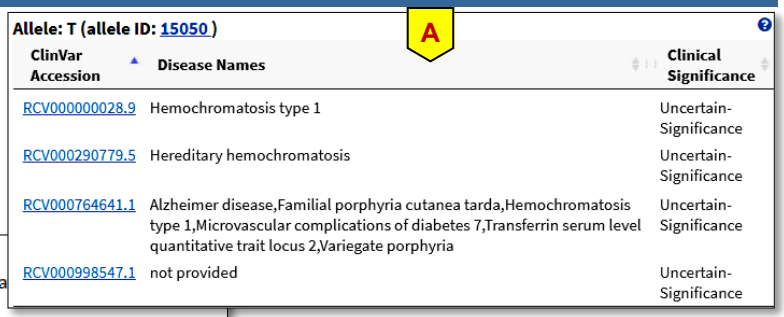


This panel groups mapped variants according to attributes , such those with corresponding ClinVar record, with literature citation, or in 1000 Genomes Phase 3 callset (**F**). Displayed variants for each track can be downloaded (**G**). Clicking the blue button (**H**) pops out the graphical presentation in a new Variation Viewer window.

Other Tabs of the SNP Report

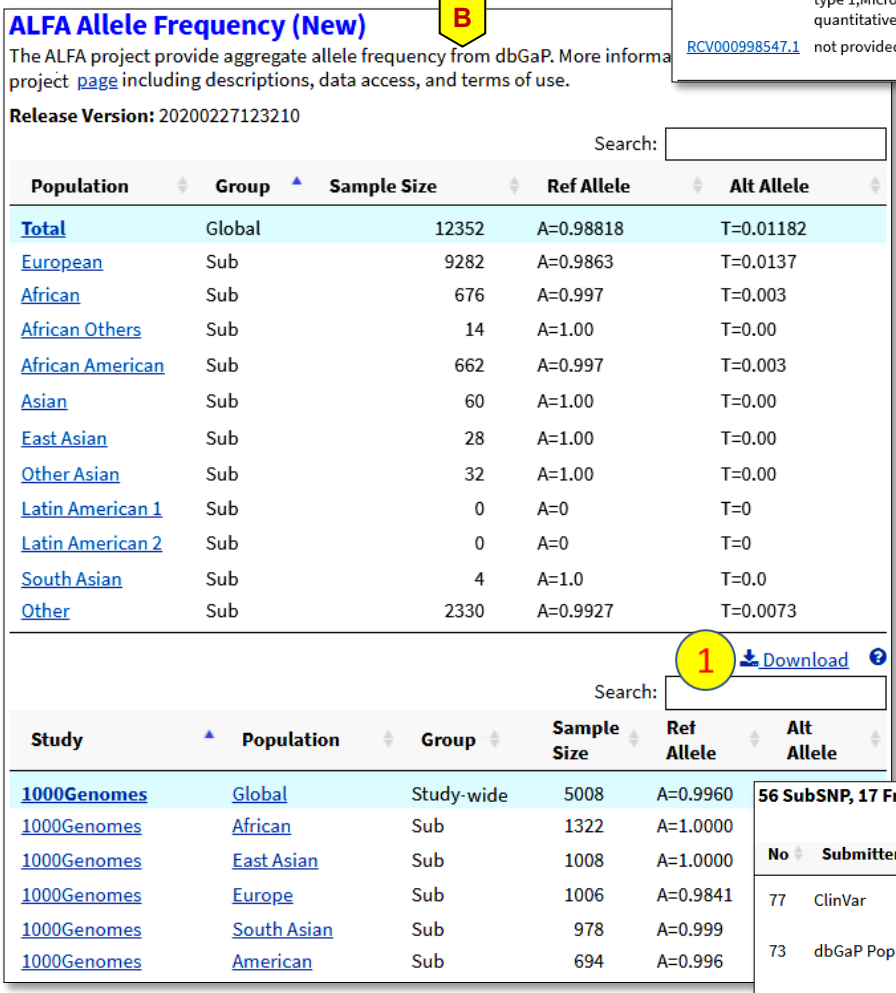
Other tabs in the SNP Report provide category-specific information.

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



Allele: T (allele ID: 15050)

ClinVar Accession	Disease Names	Clinical Significance
RCV000000028.9	Hemochromatosis type 1	Uncertain-Significance
RCV000290779.5	Hereditary hemochromatosis	Uncertain-Significance
RCV000764641.1	Alzheimer disease,Familial porphyria cutanea tarda,Hemochromatosis type 1,Microvascular complications of diabetes 7,Transferrin serum level quantitative trait locus 2,Variegate porphyria	Uncertain-Significance
RCV000998547.1	not provided	Uncertain-Significance



ALFA Allele Frequency (New)

The ALFA project provide aggregate allele frequency from dbGaP. More informa project page including descriptions, data access, and terms of use.

Release Version: 20200227123210

Search:

Population	Group	Sample Size	Ref Allele	Alt Allele
Total	Global	12352	A=0.98818	T=0.01182
European	Sub	9282	A=0.9863	T=0.0137
African	Sub	676	A=0.997	T=0.003
African Others	Sub	14	A=1.00	T=0.00
African American	Sub	662	A=0.997	T=0.003
Asian	Sub	60	A=1.00	T=0.00
East Asian	Sub	28	A=1.00	T=0.00
Other Asian	Sub	32	A=1.00	T=0.00
Latin American 1	Sub	0	A=0	T=0
Latin American 2	Sub	0	A=0	T=0
South Asian	Sub	4	A=1.0	T=0.0
Other	Sub	2330	A=0.9927	T=0.0073

Download

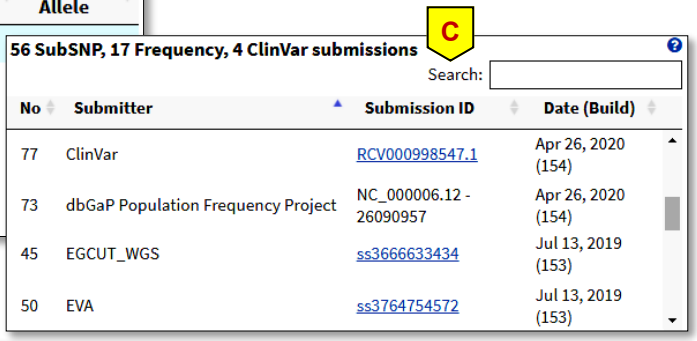
Search:

Study	Population	Group	Sample Size	Ref Allele	Alt Allele
1000Genomes	Global	Study-wide	5008	A=0.9960	
1000Genomes	African	Sub	1322	A=1.0000	
1000Genomes	East Asian	Sub	1008	A=1.0000	
1000Genomes	Europe	Sub	1006	A=0.9841	
1000Genomes	South Asian	Sub	978	A=0.999	
1000Genomes	American	Sub	694	A=0.996	

The **Frequency** tab (**B**) lists allele frequency data from major studies, such as ALFA from dbGaP samples, 1000 Genomes, ExAc, Genome Aggregation Database, etc, broken down by subpopulation if available. This provides a way to evaluate the impact of a variant if no information is available in the Clinical Significance and Publications tab. Use the "Download" link (**1**) to get the data in a tab-delimited format.

The **HGVS** tab (not shown) contains a table of HGVS names for this variation when placed on different sequence records.

The **Submission** tab (**C**) lists equivalent submitted entries, from large projects or individual submitters. Note: only older submissions, before adoption of asserted location, have ssIDs.



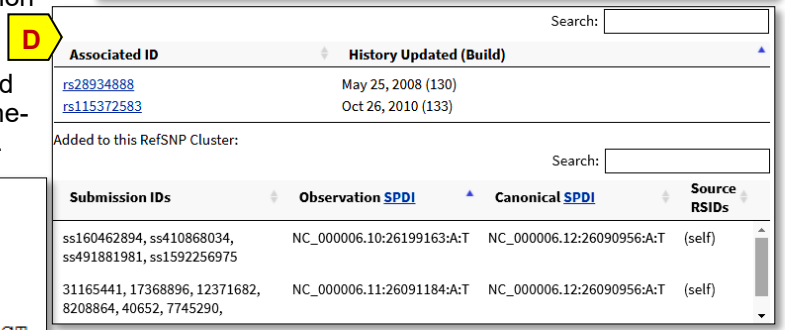
56 SubSNP, 17 Frequency, 4 ClinVar submissions

Search:

No	Submitter	Submission ID	Date (Build)
77	ClinVar	RCV000998547.1	Apr 26, 2020 (154)
73	dbGaP Population Frequency Project	NC_000006.12 - 26090957	Apr 26, 2020 (154)
45	EGCUT_WGS	ss3666633434	Jul 13, 2019 (153)
50	EVA	ss3764754572	Jul 13, 2019 (153)

The **History** tab (**D**) tracks changes of the record by listing other rsIDs that had merged with this variant, as well as submissions' observed variations and their canonical variation on current release of the genome assembly.

The **Literature** tab (not shown) listed the title of PubMed records citing this rsIDs. A button at the end allows a one-click retrieval of those records in the PubMed database.



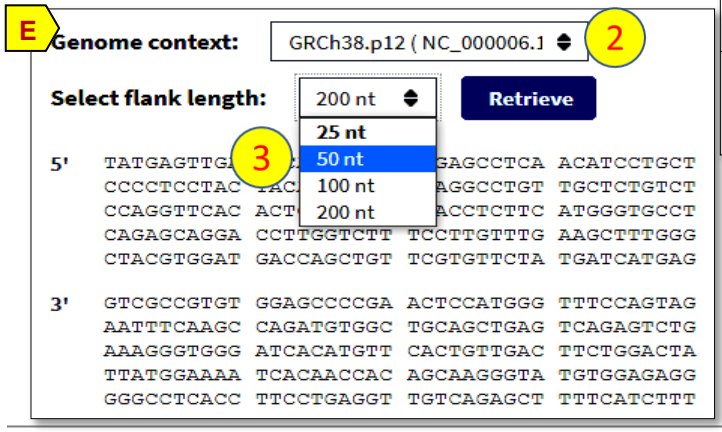
Search:

Associated ID	History Updated (Build)
rs28934888	May 25, 2008 (130)
rs115372583	Oct 26, 2010 (133)

Added to this RefSNP Cluster:

Search:

Submission IDs	Observation SPDI	Canonical SPDI	Source RSIDs
ss160462894, ss410868034, ss491881981, ss1592256975	NC_000006.10:26199163:A:T	NC_000006.12:26090956:A:T	(self)
31165441, 17368896, 12371682, 8208864, 40652, 7745290,	NC_000006.11:26091184:A:T	NC_000006.12:26090956:A:T	(self)



Genome context: GRCh38.p12 (NC_000006.1

Select flank length: 200 nt

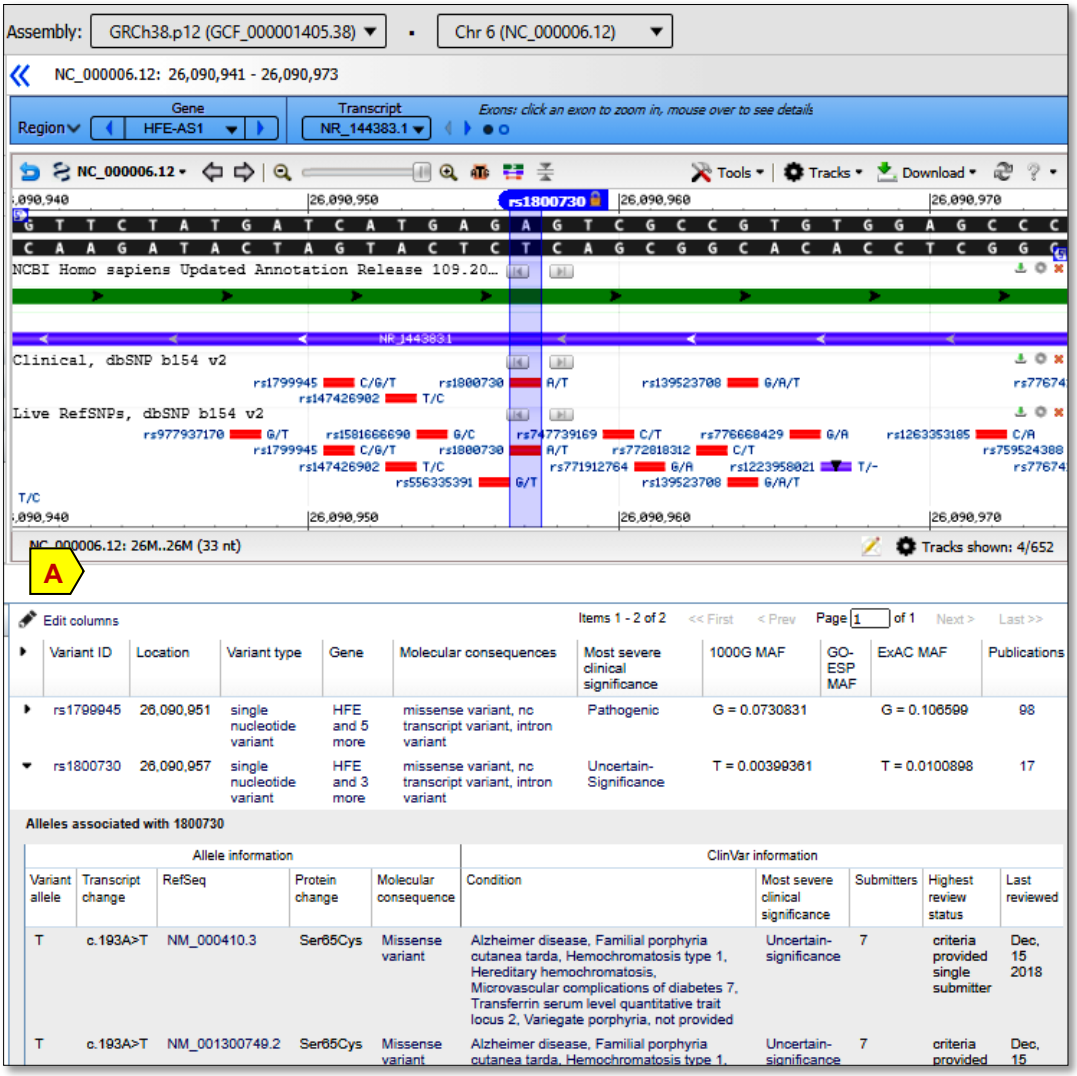
25 nt
50 nt
100 nt
200 nt

Retrieve

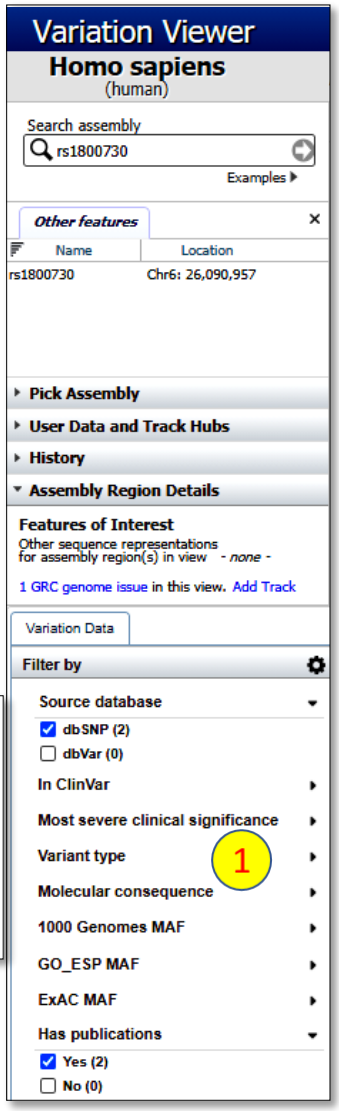
```
5'  TATGAGTTG   GAGCCTCA ACATCCTGCT
    CCCCTCCTAC TAC AGGCCTGT TGCTCTGTCT
    CCAGGTTCAC ACT ACCTCTTC ATGGGTGCCT
    CAGAGCAGGA CCTTGGTCTT TCCTTGTTTG AAGCTTTGGG
    CTACGTGGAT GACCAGCTGT TCGTGTTCTA TGATCATGAG

3'  GTCGCCGTGT GGAGCCCCGA ACTCCATGGG TTTCCAGTAG
    AATTTCAAGC CAGATGTGGC TGCAGCTGAG TCAGAGTCTG
    AAAGGGTGGG ATCACATGTT CACTGTTGAC TTCTGGACTA
    TTATGGAAAA TCACAACCAC AGCAAGGGTA TGTGGAGAGG
    GGGCCTCACC TTCCTGAGGT TGTCAGAGCT TTTCATCTTT
```

The **Flank** tab (**E**) provides access to genomic sequences flanking the reported SNP allele. The source genomic sequences is set to the current genome build (**2**), with GRCh37.p13 (hg19 equivalent) and NG RefSeqGene as other options. The length can be customized using the options in the pull-down menu (**3**) with default set to 25 nucleotides.

Variation Viewer

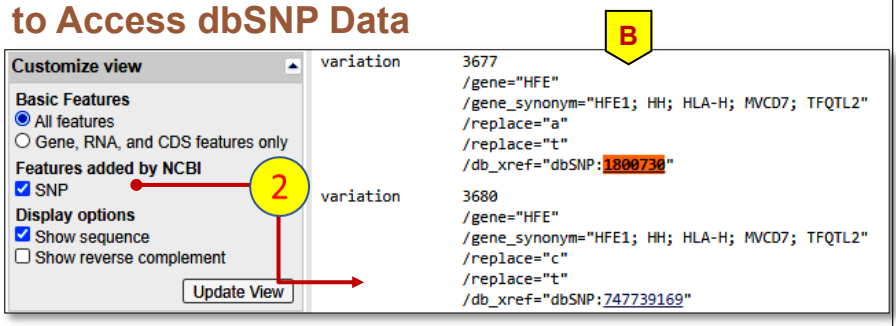


The Variation Viewer provides an interactive display of the variant under the context of annotation of the selected genome assembly. It correlates a variation and its molecular consequences in the data table with its genomic context in the graphical display (**A**). Selecting filters in the left column (**1**) updates the variants table to those fit the selected criteria. More information on this tool is available online [5, 6].



Other Ways to Access dbSNP Data

dbSNP is integrated with other databases. This allows access of variation data through links. For example, variations mapped to a RefSeq records (with NT_, NG_, NW_ or NM_ accessions) by using the SNP checkbox in the Customize view (**B**) menu of the sequence record and the SNPs checkbox. Clicking Update View (**2**) to activates the selection.



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. 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. www.ncbi.nlm.nih.gov/pubmed/29165669
3. New Web Services for Comparing and Grouping Sequence Variants. go.usa.gov/xUeKT.
4. ALFA: Allele Frequency Aggregator. www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/
5. Variation Viewer factsheet. ftp.ncbi.nih.gov/pub/factsheets/Factsheet_Variation_Viewer.pdf
6. Variation Viewer Online video tutorial. www.youtube.com/watch?v=rnWZ9MFBwUM