



dbSNP: Database for Short Genetic Variations

Catalog of nucleotide changes for human and other model organisms

<http://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 (SNV) database, also known as dbSNP, catalogs short variations in nucleotide sequences from a wide range of organisms. These variations include single nucleotide variations, short nucleotide insertions and deletions, short tandem repeats and microsatellites. SNVs may be common, thus representing true polymorphisms or they may be rare. Some of these rare human entries have additional information associated with them, including disease associations, genotype information and allele origin, as some variations are somatic rather than germline events.



Short nucleotide variation data can be accessed through the SNP home page:

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

Entrez Utilities (Eutils) can be used for programmed access:

www.ncbi.nlm.nih.gov/projects/SNP/SNPutils.htm

Database bcp files are available for download through FTP or Aspera client at:

<ftp.ncbi.nlm.nih.gov/snp/> or <http://www.ncbi.nlm.nih.gov/public/>

For help documentation, SNP database FAQs are available:

www.ncbi.nlm.nih.gov/books/NBK3848/

For information on how to find known SNPs in a sequence, refer to this handout:

ftp.ncbi.nlm.nih.gov/pub/factsheets/HowTo_Finding_SNP_by_BLAST.pdf

Searching for and displaying SNP records

Searches can be performed from the SNP homepage by typing a query term in the search box and clicking the Search button (A). The Limits (B) page has an extensive list of options that restrict search results to desired categories, while the Advanced (C) page provides a query construction function for use in creating complex queries to produce more precise results. The search below, "hfe[*gene*] AND human[*orgn*] AND utr_5[*fxn*]", retrieves variations mapped to the 5'-UTR of human HFE gene. Options in the Display setting popup can be used to show SNPs in other formats, such as FlatFile (D), or sort retrieved SNP records in a different order, such as chromosome base position (E). The retrieved variations can be saved to a local file using the Send to (F) popup. Icons in Graphic Summary (G) link to separate displays to highlight specific aspects, such as genomic placement (MapView) and gene-centric listing (GeneView).

A SNP record in FlatFile format

I: rs138378000 [Homo sapiens]

```
rs138378000 | Homo sapiens | 9606 |  
ss333156713 | 1000GENOMES | 2010080  
SNP | alleles=A/G | het=0 | se(het)  
VAL | validated=NO | min_prob=? | max_prob=? | new_variation=  
CTG | assembly=GRCh37.p5 | chr=6 | chr-pos=26087649 | NT_007592.15 | ctg-start=26027649 | ctg-end=26027649 | loctype=2 | orient=+  
LOC | HFE | locus_id=3077 | fxn-class=upstream-variant-5KB  
LOC | HFE | locus_id=3077 | fxn-class=upstream-variant-5KB  
LOC | HFE | locus_id=3077 | fxn-class=upstream-variant-5KB  
LOC | HFE | locus_id=3077 | fxn-class=upstream-variant-5KB  
LOC | HFE | locus_id=3077 | fxn-class=upstream-variant-5KB  
LOC | HFE | locus_id=3077 | fxn-class=upstream-variant-5KB  
LOC | HFE | locus_id=3077 | fxn-class=upstream-variant-5KB  
LOC | HFE | locus_id=3077 | fxn-class=upstream-variant-5KB  
SEQ | 24586691 | source-db=remap | seq-pos=71363 | orient=+  
SEQ | 91718876 | source-db=remap | seq-pos=141 | orient=+  
SEQ | 91718877 | source-db=remap | seq-pos=141 | orient=+  
SEQ | 91718878 | source-db=remap | seq-pos=141 | orient=+  
SEQ | 91718880 | source-db=remap | seq-pos=141 | orient=+  
SEQ | 91718881 | source-db=remap | seq-pos=141 | orient=+  
SEQ | 91718882 | source-db=remap | seq-pos=141 | orient=+  
SEQ | 91718883 | source-db=remap | seq-pos=141 | orient=+
```

The reference SNP cluster report

Details of a variation record are given in the Reference SNP Cluster Report (shown in sections below and on the next page). This display is linked from the rsID (rs1800730) and provides a summary of the allele (A) and mapping information in Human Genome Variation Society (HGVS) nomenclature (B). The VarView icon (C) links to a gene-centric display for clinically associated SNPs (see p.4). The detailed genome mapping information is summarized in the following table (D). The magnifying glass icon (E) links to this variation in the 1000 Genomes Browser with genotype from different populations, if available.

Reference SNP(refSNP) Cluster Report: rs1800730		
RefSNP	Allele	HGVS Names
Organism: human (Homo sapiens)	SNV: single nucleotide variation	NC_000006.11:g.26091185A>T
Molecule Type: cDNA	RefSNP Alleles: A/T	NG_001335.1:g.74899A>T
Created/Updated in build: 89/135	Allele Origin: T:Germline A:Germline	NG_008720.1:g.8677A>T
Map to Genome Build: 37.3	Ancestral Allele: A	NM_000410.3:c.193A>T
Validation Status:	Clinical Source: VarView OMIM	NM_139003.2:c.193A>T
	Clinical Significance: With probable-pathogenic allele [detail]	NM_139004.2:c.193A>T
	MAF/MinorAlleleCount: T=0.006/14	NM_139006.2:c.193A>T
		NM_139007.2:c.77-357A>T
		NM_139008.2:c.77-357A>T

SNP Details are organized in the following sections:

GeneView	Map	Submission	Fasta	Resource	Diversity	Validation
Integrated Maps (Hint: click on 'Chr Pos' or 'Contig Pos' column value to see variation in NCBI sequence viewer)						
Assembly	Genome Build	Chr	Chr Pos	Contig	Contig Pos	Map Method
GRCh37.p5	37.3	6	26091185	NT_007592.15	26031185	remap
reference	36.3	6	26199164	NT_007592.14	16949436	blast
Celera	36.3	6	27320470	NW_922984.1	25715726	blast
HuRef	36.3	6	26034245	NW_001838974.1	2045152	blast

In the GeneView section (right), clicking the Go button (F) activates a display, known as SNP: GeneView (p.4), detailing the variations mapped to the gene. The variation mapping and the protein coding changes are summarized in the tables (G) below. This is followed by a graphical display of the variation on the genome assembly (H) in the Sequence Viewer. The accession (I) links to a larger and more flexible display.

GeneView

GeneView via analysis of contig annotation: [HFE](#) hemochromatosis

View more variation on this gene (click to hide).

Clinical Source: ☐ in gene region ☒ cSNP ☐ has frequency ☐ double hit [Go](#)

Primary Assembly Mapping

Assembly	SNP to Chr	Chr	Chr position	Contig	Contig position	Allele
GRCh37.p5	Fwd	6	26091185	NT_007592.15	26031185	A

RefSeqGene Mapping

RefSeqGene	Gene (ID)	SNP to RefSeqGene	Position	Allele
NG_008720.1	HFE (3077)	Fwd	8677	A

Gene Model(s)

Function	mRNA				Protein		
	SNP to mRNA	Accession	Position	Allele change	Accession	Position	Residue change
missense	Fwd	NM_139009.2	284	AGT ⇒ TGT	NP_620578.1	42	S [Ser] ⇒ C [Cys]
missense	Fwd	NM_000410.3	353	AGT ⇒ TGT	NP_000401.1	65	S [Ser] ⇒ C [Cys]

WC_000006.11: 26M..26M (2.2Kbp)

Sequence:

SNP

Suspect

Somatic Allele

GMAF >= 0.01

Clinical Channel

Association Results

Cited Variants

Genes

Submitter SNP (ss), with alleles and flanking sequences (J) on which this reference SNP is based, are summarized in a table below the Sequence Viewer display. The ssIDs (K) link to submitter records providing additional details.

Submitter records for this RefSNP Cluster									
The submission ss244317424 has the longest flanking sequence of all cluster members and was used to instantiate sequence for rs1800730 during BLAST analysis for the current build.									
NCBI Assay ID	Handle Submitter ID	Validation Status	ss to rs Orientation /Strand	Alleles	5' Near Seq 30 bp	3' Near Seq 30 bp	Entry Date	Update Date	Build Added
ss2420855	HGBASE SNP000007282		fwd/T	A/T	gtcgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactcc	11/07/00	10/10/03	89
ss68969473	PERLEGEN PGP04777602		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	01/30/07	08/14/07	127
ss160462894	ILLUMINA HumanOmni1-Quad_v1-0_B_rs1800730...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	08/04/09	10/02/09	131
ss233370756	1000GENOMES pilot_1_CEU_2975385_chr6_261		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	05/01/10	05/01/10	132
ss244317424	2010_April_001_075_HFE_235200_0003...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	06/16/10	06/16/10	132
ss288288927	OMIM CURATED-RECORDS 6335		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	12/21/10	12/21/10	133
ss342203118	NHLBI ESPI ESP2500-chr6-26091185		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	03/25/11	03/26/11	134
ss410868034	ILLUMINA Cardio-Metabo_Chip_A_chr6_26199164...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	06/07/11	06/07/11	135
ss481067367	ILLUMINA HumanOmni1-Quad_v1-0_rs1800730-128...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	01/30/12	01/31/12	137
ss490921024	1000GENOMES 20110521_exome_428822_chr6_260...		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	02/10/12	02/21/12	137
ss491378666	EXOME_CHIP nonsyn_94892_chr_6_26091185		fwd/	A/T	gaccagctgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactccatggg	03/05/12	03/05/12	137
ss491881981	CLINSEQ_SNP SNV-chr6-26199164		fwd/	A/T	gtcgttcgtgttctatgatcatgag	gtcgcctgtgtggagccccgaactcc	03/06/12	03/13/12	137

The reference SNP cluster report (cont.)

Sequences flanking the variation are given in the [FASTA Sequence](#) section of the Reference SNP Cluster Report. Here the 5'-(**A**) and 3'-(**B**) flanking sequences and the allele (**C**), from the exemplar submitter record are used to represent the SNP record. If available, additional links will be given in the [NCBI Resource Links](#) section (**D**). Genotypes and allele frequencies information for various populations from different studies are summarized in the [Population Diversity](#) table at the end of the record (**E**). More genotype details can be retrieved using the [Genotype Detail](#) link (**F**).

Fasta sequence (Legend)

```
>gn[dbSNP]rs1800730[allelePos=251][totalLen=501][taxid=9606][snpclass=1][alleles='A/T'][mol=Genomic][build=137]
CAGGACTGCA ACTCACCCTT CACAAAATGA GGACCAGACA CAGCTGATGG TATGAGTTGA
TGCAGGTGGT TGGAGCCTCA ACATCCTGCT CCCCTCCTAC TACACATGGT TAAGGCCTGT
TGCTCTGTCT CCAGGTTTCA ACTCTCTGCA CTACCTCTTC ATGGGTGCTT CAGAGCAGGA
CCTTGGTCTT TCCTTGTITG AAGCTTTGGG CTACGTGGAT GACCAGCTGT TCGTGTCTTA
TGAATGAG
W
GTCGCTCTCT GGAGCCCGCA ACTCCATGGG TTTCAGTAG AATTCAAGC CAGATGTGGC
TGCAGCTGAG TCAGAGTCTG AAAGGGTGGG ATCACATGTT CACTGTGGAC TTCTGGACTA
TTATGGAAAA TCACAACCAC AGCAAGGGTA TGTGGAGAGG GGGCCTCACC TTCTGAGGT
TGTCAAGGCT TTTTCATCTT TCATGCATCT TGAAGGAAAC AGCTGGAAGT CTGAGGTCTT
GTGGGAGCAG
```

NCBI Resource Links

Submitter-Referenced	dbSNP Blast Analysis	UniGene Cluster ID	3D structure mapping	OMIM
GenBank Z92910	233325	NP_000401 NP_620572 NP_620573	NP_620575 NP_620578	235200.0003 613609.0003

Population Diversity

ss#	Population	Sample Ascertainment	Individual Group	Chrom. Sample Cnt.	Source	Genotype Detail				Alleles	
						A/A	A/T	T/T	HWP	A	T
ss233370756	pilot_1_CEU_low_coverage_panel			120	AF					0.967	0.033
ss342203118	ESP_Cohort_Populations			4552	GF	0.978	0.022		0.655	0.989	0.011
ss491881981	CSAqilent			1323	GF	0.983	0.017			0.992	0.009

Other ways to access data from dbSNP

The SNP database is fully integrated with the Entrez system enabling the access of variation data from records in other NCBI databases through built-in links if such data are available. For example, the [SNP: GeneView](#) link (**G**) found in the [Related information](#) section of the gene record displays variations mapped to a gene of interest. Variations mapped to a segment of the RefSeq genomic record or to an mRNA entry, with NT_, NG_, NW_ or NM_ accessions, can be viewed through [Customize view](#) (**H**) menu in the upper right hand corner of the sequence record. Checking the SNPs checkbox and clicking the [Update View](#) will add annotation of mapped variations (**I**) to the feature display.

Customize view

Basic Features

- ☒ Default features
- ☐ Gene, RNA, and CDS features only

Features added by NCBI

- ☒ 65 SNPs

Display options

- ☒ Show sequence
- ☐ Show reverse complement

[Update View](#)

exon

1..236

/gene="HFE"

/gene_synonym="HFE1; HH; HLA-H; MVCD7; TFQTL2"

/inference="alignment:Splign"

/number=1

34

/gene="HFE"

/gene_synonym="HFE1; HH; HLA-H; MVCD7; TFQTL2"

/replace="c"

/replace="g"

/db_xref="dbSNP:62625316"

49

/gene="HFE"

/gene_synonym="HFE1; HH; HLA-H; MVCD7; TFQTL2"

/replace="c"

/replace="t"

/db_xref="dbSNP:62625317"

Disease related nucleotide variations, reported in literature and collected by OMIM, are also added to dbSNP through an NCBI curation effort. The rsIDs are used to reference these variations when applicable. The table below is the [Allelic Variant](#) (**J**) display for OMIM record 613609, which cites the rsIDs in the dbSNP column (**K**).

***613609**

HFE GENE; HFE

Allelic Variants (Selected Examples):

Number	Phenotype	Mutation	dbSNP
.0001	HEMOCHROMATOSIS 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	-
.0002	HEMOCHROMATOSIS MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 7, INCLUDED	HFE, HIS63ASP	[rs1799945]
.0003	HEMOCHROMATOSIS	HFE, SER65CYS	[rs1800730]
.0004	HFE INTRONIC POLYMORPHISM	HFE, 5569G-A	-
.0005	HFE POLYMORPHISM	HFE, VAL53MET	[rs28934889]
.0006	HFE POLYMORPHISM	HFE, VAL59MET	[rs111033557]
.0007	HEMOCHROMATOSIS	HFE, GLN127HIS	[rs28934595]
.0008	HEMOCHROMATOSIS	HFE, ARG330MET	[rs111033558]
.0009	HEMOCHROMATOSIS	HFE, ILE105THR	[rs28934596]
.0010	HEMOCHROMATOSIS	HFE, GLY93ARG	[rs28934597]

Related information

- [Order cDNA clone](#)
- [3D structures](#)
- [BioAssay, by Gene target](#)
- [BioProjects](#)
- [Books](#)
- [CCDS](#)
- [Conserved Domains](#)
- [dbVar](#)
- [Full text in PMC](#)
- [Genome](#)
- [GEO Profiles](#)
- [GTR](#)
- [HomoloGene](#)
- [Map Viewer](#)
- [Nucleotide](#)
- [OMIM](#)
- [Probe](#)
- [Protein](#)
- [PubChem Compound](#)
- [PubChem Substance](#)
- [PubMed](#)
- [PubMed \(GeneRIF\)](#)
- [PubMed \(OMIM\)](#)
- [RefSeq Proteins](#)
- [RefSeq RNAs](#)
- [RefSeqGene](#)
- [SNP](#)
- [SNP: GeneView](#)
- [SNP: Genotype](#)
- [SNP: VarView](#)
- [Taxonomy](#)
- [UniGene](#)
- [UniSTS](#)

The SNP:GeneView display

The SNP:GeneView display tabulates variations mapped to splice variants of a particular gene. The top of the page lists all annotated splice variants (A). The splice variant, for which nucleotide variations are shown, is highlighted in yellow (B). The default setting displays only the non-clinical coding variations (C). Changing this to include clinical variations and those "in gene region" (D) displays all the variations in the table below (E). The table arranges these variants according to their sequential orders on the genome (F) and color-coded by their function: white for "in gene region" (G), orange for UTR (H), red for non-synonymous (I), blue for frame-shift (not shown), and yellow for intronic (K). For non-synonymous variations with 3D protein structure information, link to the structure record will be provided in the 3D column (L). If available, the global minor allele frequency computed from 1000 genomes data will be provided in the MAF column (M).

The VarView display

The VarView icon (N) at the top of a detailed report links to a display showing clinically associated variations for the gene, which are submitted by Locus Specific Database (LSDB) and other research groups. Here, variations are summarized in a table listing the changes (in HGVS format) at the genomic, transcript and protein levels (O, P and Q, respectively) as well as corresponding entries in OMIM (R). Details of the selected variant (S) are shown in the panel below (T). Related literature citations are summarized under the Literature tab (U) with entries linked to PubMed.

SNP linked to Gene HFE(geneID:3077) Via Contig Annotation

Send rs# on all gene models to Batch Query Download all rs# to file

Gene Model (mRNA) information from genome sequence

Total gene model (contig mRNA transcript): 9

mRNA	transcript	protein	mRNA orientation	Contig	Contig Label	List SNP
NM_000410.3	plus strand	NP_000401.1	forward	NT_007592.15	GRCh37.p5	<- currently shown
NM_139011.2	plus strand	NP_620580.1	forward	NT_007592.15	GRCh37.p5	View snp on GeneModel
NM_139010.2	plus strand	NP_620579.1	forward	NT_007592.15	GRCh37.p5	View snp on GeneModel
NM_139009.2	plus strand	NP_620578.1	forward	NT_007592.15	GRCh37.p5	View snp on GeneModel
NM_139008.2	plus strand	NP_620577.1	forward	NT_007592.15	GRCh37.p5	View snp on GeneModel
NM_139007.2	plus strand	NP_620576.1	forward	NT_007592.15	GRCh37.p5	View snp on GeneModel
NM_139006.2	plus strand	NP_620575.1	forward	NT_007592.15	GRCh37.p5	View snp on GeneModel

☐ Clinical Source ☐ in gene region ☒ cSNP ☐ frequency ☐ double hit refresh

☒ Clinical Source ☒ in gene region ☐ cSNP ☐ has frequency ☐ double hit refresh

Region	Chr. position	mRNA pos	dbSNP rs#	cluster id	Heterozygosity	Validation	MAF	Allele origin	3D	Clinically Associated	Clinical Significance	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos
5' near gene	26085510		rs72834669		0.500		0.0349					5' near gene	A/G			
	26085519		rs146329216		N.D.							5' near gene	C/T			
	26085626		rs62625309		0.021							5' near gene	C/T			
	26085686		rs62625310		0.032		0.0209					5' near gene	A/G			
	26087542	34	rs62625316		0.010							5' UTR	C/G			
	26087557	49	rs62625317		0.010							5' UTR	C/T			
	26087621	113	rs41266793		0.011							5' UTR	C/G			
	26087649	141	rs138378000		N.D.							5' UTR	A/G			
	26087686	178	rs149342416		0.001							missense	C	Ser [S]	6	
												contig reference	G	Arg [R]	6	
	26087689	181	rs114758821		0.028							synonymous	A	Pro [P]	7	
												contig reference	G	Pro [P]	7	
	26087718	210	rs143662783		0.001							missense	T	Ile [I]	2	17
												contig reference	C	Thr [T]	2	17
	26087736	228	rs148161858		0.001				Yes			missense	A	His [H]	2	23
												contig reference	G	Arg [R]	2	23
	26087779		rs62625319		0.011							intron	A/G			
	26087856		rs2858993		0.487		0.4068					intron	A/T			

HFE Variation Viewer [Download report](#) (27113 bytes)

[\[expose gene summary\]](#)

Observed Variants [O](#) [P](#) [Q](#) [R](#) results 1 - 10 of 10

Var Class	Genomic	Transcript	Protein	Clinical interpretation	Dep	Obs	Freq	PubMed	OMIM	AI Var	rs id
SNC	g.8641G>A	c.157G>A	p.Val53Met		4	4	Q		235200.0005, 61		rs28934889
SNC	g.8659G>A	c.175G>A	p.Val59Met		2	2			235200.0006, 61		rs111033557
SNC	g.8671C>G	c.187C>G	p.His63Asp		21	21	Q	9	235200.0002, 61		rs1799945
SNC	g.8677A>T	c.193A>T	p.Ser65Cys		8	8	Q		235200.0003, 61		rs1800730
SNC	g.8761G>C	c.277G>C	p.Gly93Arg		3	3			235200.0010, 61		rs28934597

Variant Details [Submission Details](#) [Literature](#)

Citations in PubMed: 9 [\[view in PubMed\]](#)

- Sebastiani P, et al. A hierarchical and modular approach to the discovery of robust associations in genome-wide association studies from pooled DNA samples. BMC Genet. 2008 Jan 14; 9:6.
- Simoni M, et al. Functional genetic polymorphisms and female reproductive disorders: Part I: Polycystic ovary syndrome and ovarian response. Hum Reprod Update. 2008 Sep-Oct; 14(5):459-84.
- Hopkins MR, et al. Variants in iron metabolism genes predict higher blood lead levels in young children. Environ Health Perspect. 2008 Sep; 116(9):1261-6.

21 Submissions [T](#)

Method	Submitted by	Submitter Link	Details
Other	Functional Genomics Research Center Korea Research Institute of Bioscience and Biotechnology	[more]	Q
Unknown	Division of Genome Analysis, Research Center for Genetic Information Medical Institute of Bioregulation, Kyushu University	[more] QH03725	Q