dbSNP: Database for Short Genetic Variations

Catalog of nucleotide changes for human and other model organisms

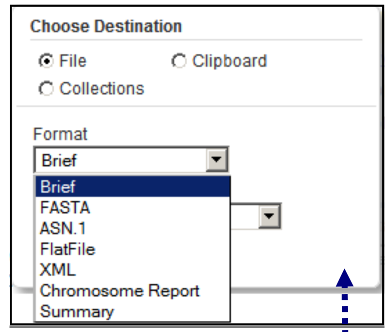
http://www.ncbi.nlm.nih.gov/snp/

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

Scope and access

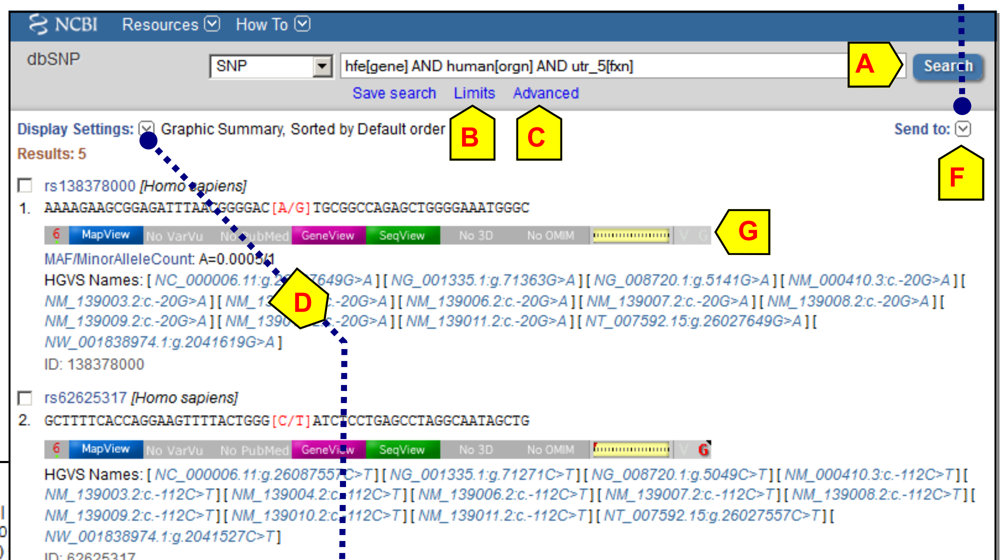
The NCBI Short Genetic Variations (SNV) database, also known as dbSNP, catalogs short variations in nucleotide sequences from a wide range of organisms. These variations include single nucleotide variations, short nucleotide insertions and deletions, short tandem repeats and microsatellites. SNVs may 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/SNPeutils.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

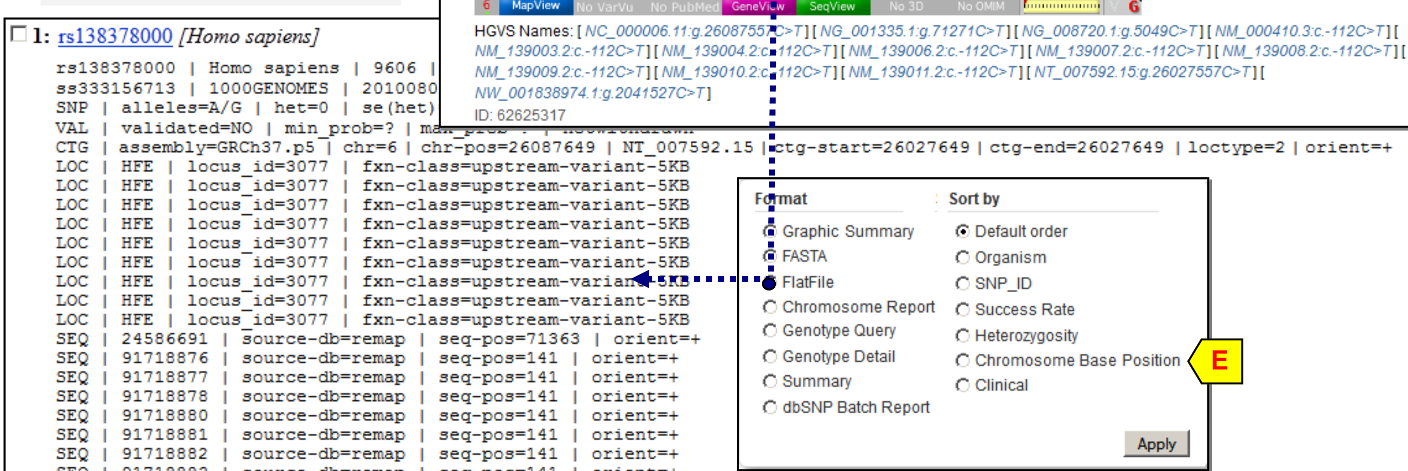


Searching for and displaying SNP records

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).



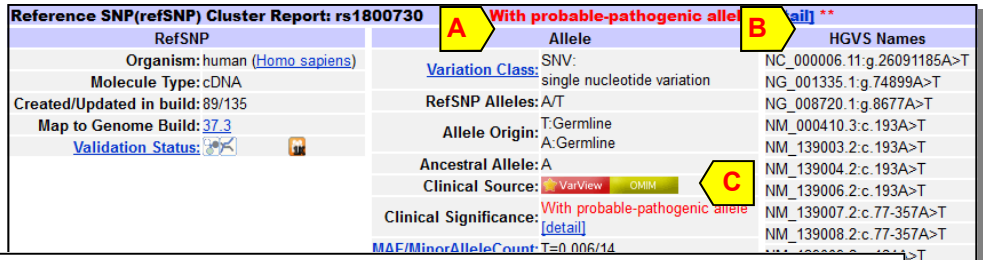
A SNP record in FlatFile format

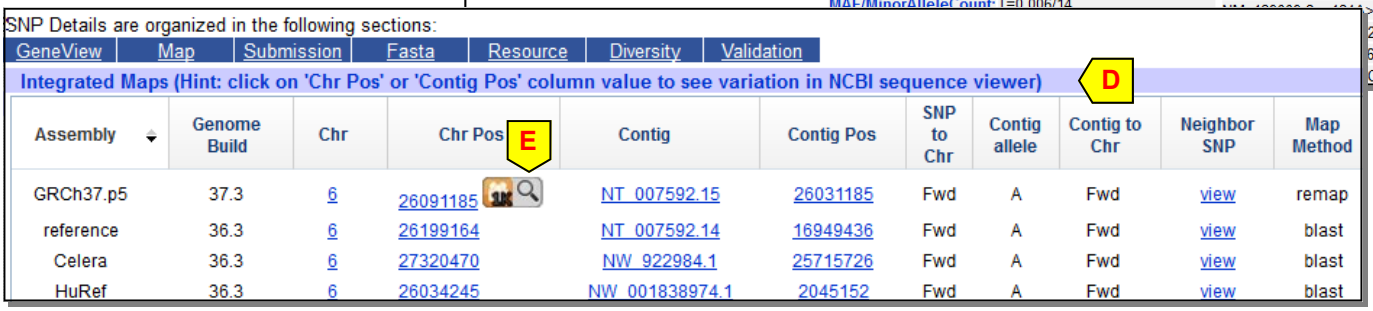


Contact: info@ncbi.nlm.nih.gov | (301) 496-2475

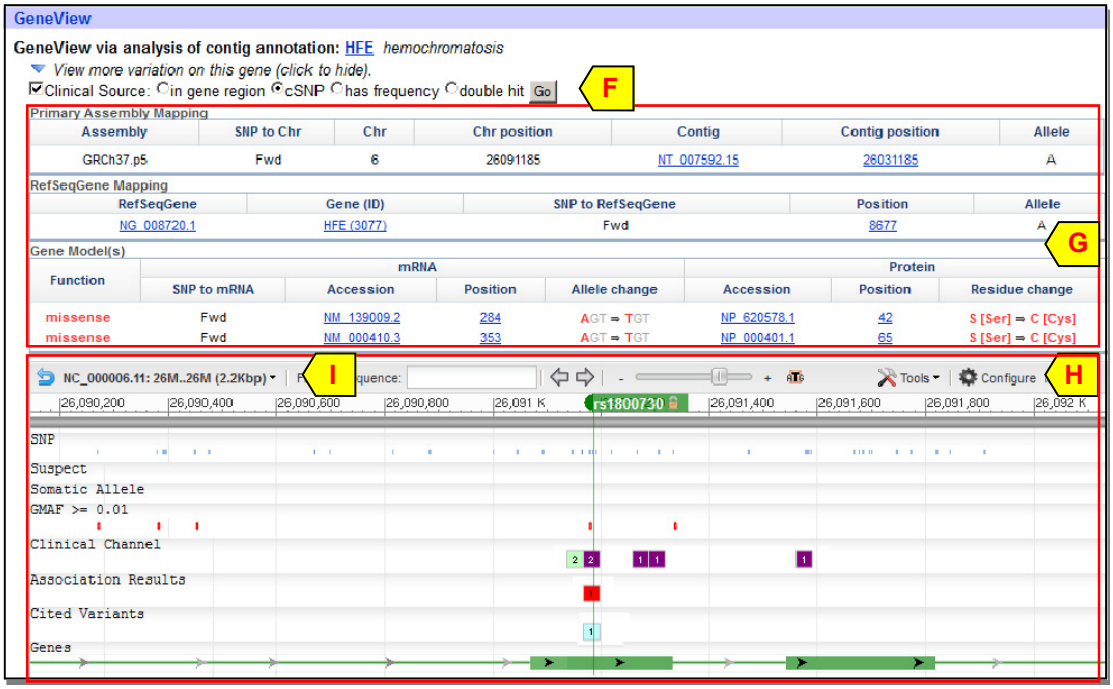
The reference SNP cluster report

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.

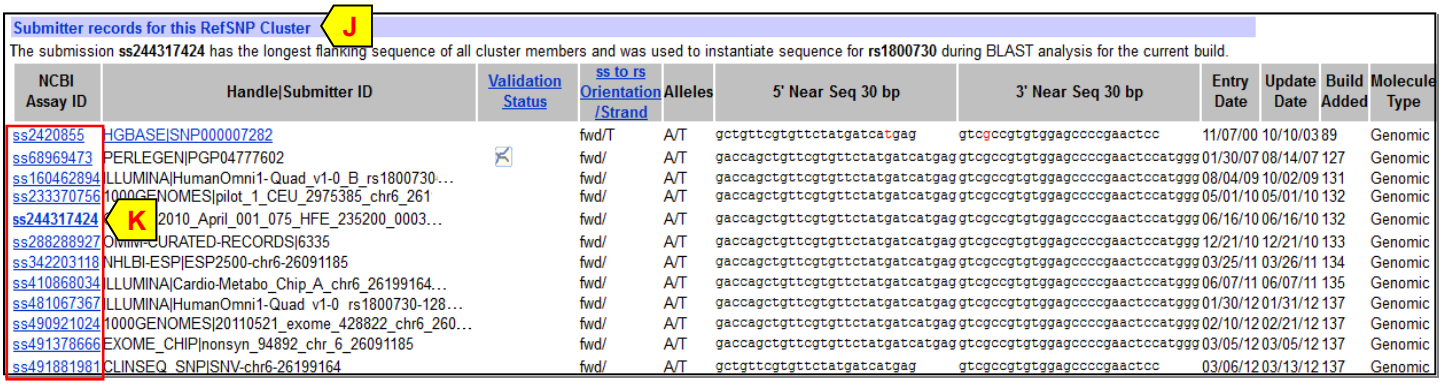




In the GeneView section (right), clicking the Go button (**F**) activates a display, known as SNP: GeneView (p.4), detailing the variations mapped to the gene. The variation mapping and the protein coding changes are summarized in the tables (**G**) below. This is followed by a graphical display of the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

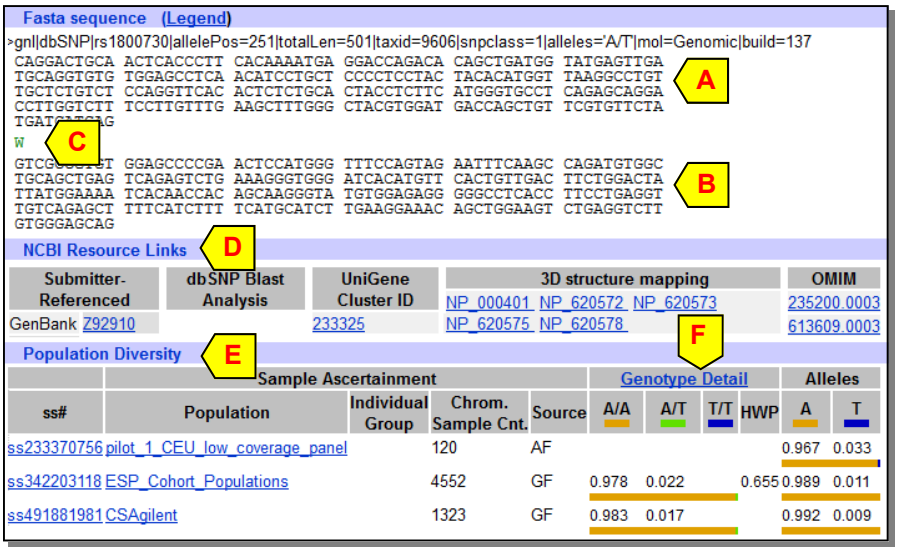


Submitter SNP (ss), with alleles and flanking sequences (**J**) on which this reference SNP is based, are summarized in a table below the Sequence Viewer display. The ssIDs (**K**) link to submitter records providing additional details.



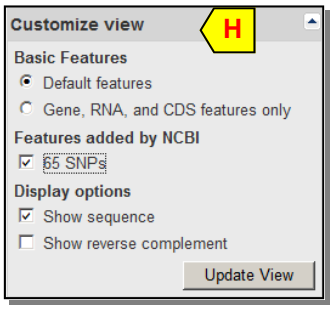
The reference SNP cluster report (cont.)

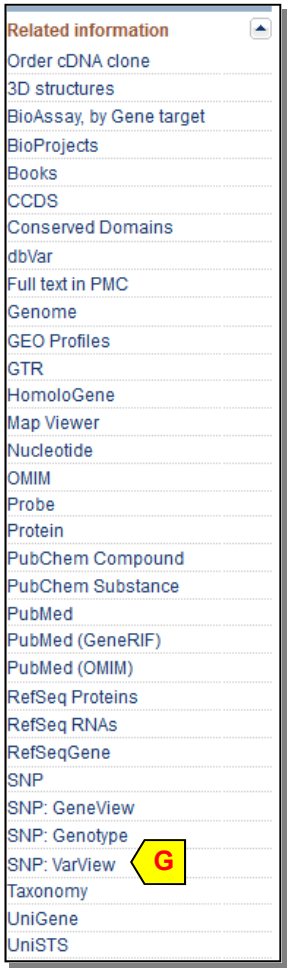
Sequences flanking the variation are given in the FASTA Sequence section of the Reference SNP Cluster Report. Here the 5'- (**A**) and 3'-(**B**) flanking sequences and the allele (**C**), from the exemplar submitter record are used to represent the SNP record. If available, additional links will be 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**).



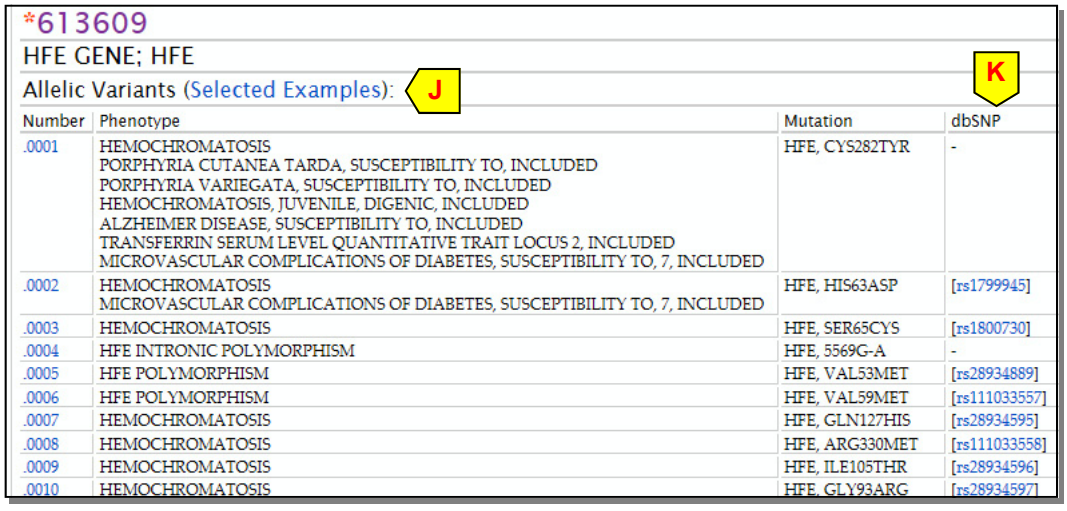
Other ways to access data from dbSNP

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.



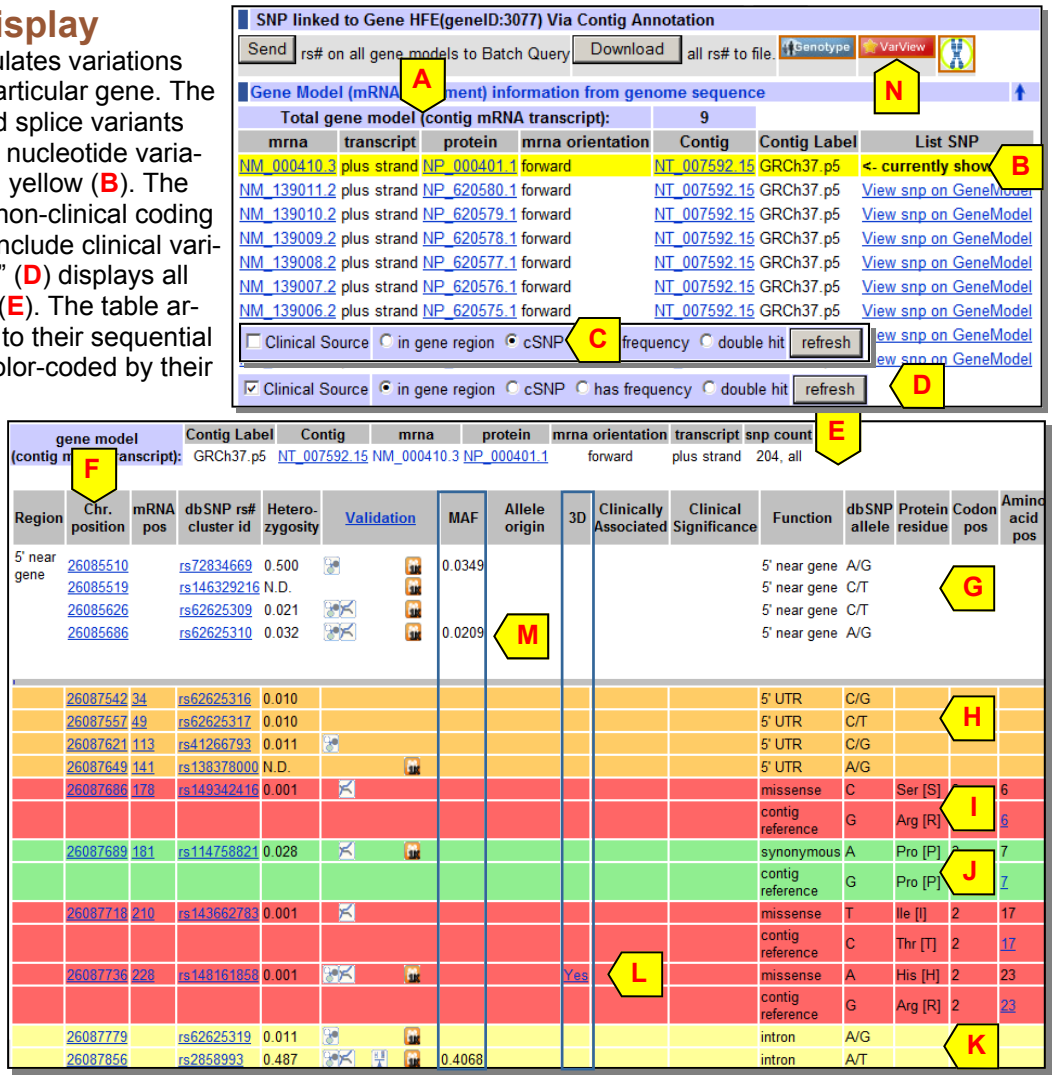


Disease related nucleotide variations, reported in literature and collected by OMIM, are also added to dbSNP through an NCBI curation effort. The rsIDs are used to reference these variations when applicable. The table below is the Allelic Variant (**J**) display for OMIM record 613609, which cites the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



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**), green for synonymous (**J**), 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.