# HGMD 2022

## Frequently Asked Questions

Sample to Insight

### Is the first position in a chromosome counted as 0 or 1 for the purpose of determining genomic coordinates?

The first position in the chromosome is 1.

**Note:** Answer intended for online and download users.

### Is the risk allele for a given site-disease association available (or derivable from data) in HGMD?

The allele associated with the HGMD phenotype is available in the MUTNOMEN table (mutBASE column). The wildBASE column is the wild-type nucleotide sequence (NULL for insertions) and the mutBASE column is the mutated nucleotide sequence (NULL for deletions). When looking at the core data tables (e.g. MUTATION, PROM), the phenotype allele should be the variant allele.

**Note:** Answer intended for download users.

### Which strand are the sequences in the database taken from?

All sequence data (wild-type, mutated, and flanking bases) are given as they would appear on the strand which encodes the protein in question (i.e. the 'coding' strand). For example, in the case of CM014827, the MUTNOMEN table lists wildBASE='T' and mutBASE='C' to indicate the T>C polymorphic change described in the referenced article. The strand encoding the STX1A gene is the minus strand on the assembly, so this substitution would be equivalent to an A>G change in the assembly sequence. wildBASE and mutBASE sequence are obtained from the core mutation tables, whilst the flanking sequence is derived from the assembly. The flanking sequence is converted into the complementary sequence if the mutation was described on the non-coding strand in the reference, so as to correspond to the coding strand. Therefore, all data in HGMD are 'coding strand' data.

**Note:** Answer intended for online and download users.

### How may I find the location of mutations in a specific splice variant for genes with multiple splicing variants? Is the codon numbering system for different mutations in a gene consistent, i.e., is the numbering for different mutations all based on one splicing variant? If yes, where could I find the accession number of this mRNA splicing variant?

Codon numbering is consistent with the cDNA sequences provided (along with the NCBI accession numbers). HGMD mutations are mapped to these sequences wherever possible. Currently, there is only one splice variant sequence per gene. This splice variant is usually based on either the longest or most common mRNA isoform.

**Note:** Answer intended for online and download users.

### How does HGMD represent a simple single-base insertion?

For all insertions, the start coordinate is one less than the end coordinate. Additionally, for insertions, wildBASE in the mutnomen table is NULL and mutBASE represents the inserted bases located between the start and end coordinates. A risk allele of length one is an insertion of length one.

**Note:** Answer intended for online and download users.

## Is there is a way to get a list of SNPs for genes which are not known to be mutation-causing or associated with inherited disease?

HGMD records disease-causing mutations and disease-associated/functional polymorphisms. Neutral polymorphism data are available in other databases (e.g. dbSNP and HapMap). dbSNP data have been integrated into HGMD Professional for missense/nonsense SNPs only.

**Note:** Answer intended for online and download users.

## I have tried to open Get map under Mutation Viewer for a gene. The window could not be opened, and every time I tried to open Get map, all my windows were closed and I had to re-login?

Try installing "Java Runtime Environment version 6" locally on your computer. One other known security problem with the Mutation Viewer could be an issue of the newest Java version which introduced Internet Explorer-like security settings. The settings can be changed in Windows 7 by opening Programs -> All programs -> Java -> Configure Java. In the Security tab, you need to add the site https://portal.biobase-international.com to the site list using the Edit Site List button. Please note that the mutation viewer applet has been discontinued.

**Note:** Answer intended for online users.

## Does HGMD provide any information regarding polymorphisms found in human genes?

Disease-associated/functional polymorphisms are included in HGMD. To be included as disease-associated, a statistically significant ($p<0.05$) association between the polymorphism and a clinical phenotype must have been reported. In case no clinical phenotype is known to be associated with a polymorphic variant, but sufficient in vitro or in vivo expression/functional data have nevertheless been presented to indicate functional significance, then the variant will be included in HGMD.

**Note:** Answer intended for online and download users.

## How does HGMD define a genomic alteration: as a mutation or a polymorphism?

A polymorphism is a mutation found at a frequency of >1% in any population. However, HGMD does not necessarily distinguish between the two along such rigid lines.

**Note:** Answer intended for online and download users.

## Two new germinal mutations recently identified in our lab are not reported in HGMD. How may I submit these to HGMD?

HGMD records disease-causing mutations published in the literature. At the moment, the only way to get these mutations into HGMD would be to publish them.

**Note:** Answer intended for online and download users.

## How frequently is HGMD updated?

HGMD has quarterly updates, released towards the end of each quarter.

**Note:** Answer intended for online and download users.

## ? Is the article/paper identified by the PMID in an ALLMUT record one of the studies that established the genetic association with the disease/phenotype?

! The ALLMUT table should contain the first (primary) literature report for a given variant, along with the associated disease/phenotype.

**Note:** Answer intended for download users.

## ? The HGMD application lists multiple articles per ALLMUT record. Where are these stored, and will all papers specified for an ALLMUT record be in agreement about the risk allele?

! Additional references are stored in the EXTRAREFS table. The risk allele will not always be the same between different literature reports (which will report different phenotypes and functional studies from different populations).

**Note:** Answer intended for download users.

## ? Do you include variants from genome-wide association studies (GWAS) papers in HGMD?

! HGMD does include SNPs from GWAS studies whenever there is evidence for a likely effect on function (which is, in fact, lacking for most GWAS studies). HGMD includes the first example of all mutations causing or associated with human inherited disease, plus disease-associated/functional polymorphisms reported in the literature. HGMD may also include additional reports for certain mutations if these reports serve to enhance the original entry (e.g. functional studies). To be included, there must be a convincing association of the polymorphism with the phenotype. These polymorphisms are currently identified in the database by an addition to the phenotypic description. These additions are limited to association — association with and increased or lower risk, depending on how the polymorphism was reported.

**Note:** Answer intended for online and download users.

## ? Which genome build are you currently using for mapping HGMD mutations?

! As of release 2015.2, we use GRCh38/hg38 by default for the genomic (chromosomal) coordinates present in HGMD. Support for GRCh37/hg19 will continue to be provided (via LiftOver) for the foreseeable future. Users are however advised to switch to using the HGMD genomic coordinates based on build 38 (GRCH38/hg38) where possible.

**Note:** Answer intended for online and download users.

## ? In the quick search part of the HGMD advanced search, what does the ranking score relate to?

! The quick search ranking score relates to the number of matches found for the query keyword(s) across the gene symbol, disease term, title of mutation report, abstract of mutation report and dbsnp identifier fields. The higher the score, the more relevant the mutation to the query keyword(s).

**Note:** Answer intended for online users.

## What is SIFT?

SIFT (Sorting Intolerant From Tolerant) predicts whether an amino acid substitution (AAS) affects protein function based on sequence homology and the physical properties of amino acids (NG et al. 2001). For disease-causing missense mutations in HGMD, around 80% are predicted to be deleterious by SIFT (Mort et al. 2010).

**References**

Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. (2001) *Genome Res* **11**:863-874.

Mort M, Evani US, Krishnan VG, Kamati KK, Baenziger PH, Bagchi A, Peters BJ, Sathyesh R, Li B, Sun Y, Xue B, Shah NH, Kann MG, Cooper DN, Radivojac P, Mooney SD. (2010) In silico functional profiling of human disease-associated and polymorphic amino acid substitutions. *Hum Mutat* **31**:335-346.

**Note:** Answer intended for online and download users.

## How do I interpret the SIFT score? (ntsub.sift_score)?

An AAS with a SIFT score of less than 0.05 is predicted to be deleterious, one with a score greater than or equal to 0.05 is predicted to be tolerated.

**Note:** Answer intended for online and download users.

## What is MutPred?

MutPred (Mutation Prediction) predicts whether an amino acid substitution (AAS) is likely to be disease-associated or neutral in humans (Li et al. 2009) — an assessment which is represented by the MutPred Score. In addition, it predicts the molecular cause of disease/deleterious AAS based upon the gain or loss of 14 different structural and functional properties, e.g. loss of a phosphorylation site. This constitutes the MutPred hypothesis.

**References**

Li B, Krishnan VG, Mort M, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* **25**:2744-2750.

**Note:** Answer intended for online and download users.

## How do I interpret the MutPred score?

The MutPred Score is the probability (expressed as a figure between 0 and 1) that an AAS is deleterious/disease-associated. A missense mutation with a MutPred score >0.5 could be considered as ‘harmful’, while a MutPred score >0.75 should be considered a high confidence ‘harmful’ prediction.

**Note:** Answer intended for online and download users.

## How do I obtain the MutPred scores from the download version? (ntsub.mutpred_score)

In the download version of HGMD Professional, a MutPred score/hypothesis has been made available (only where ntsub.mutpred_risk column is not null). For example, to download all available MutPred scores, the following SQL query could be used: SELECT mutpred_score FROM hgmd_advanced.ntsub WHERE mutpred_risk IS NOT null;

**Note:** Answer intended for download users.

## How do I interpret the MutPred hypothesis?

The MutPred hypothesis refers to the underlying structural and functional properties that the missense mutation impacts upon. The accompanying P-value indicates the assigned probability that the specified structural or functional property has been impacted upon by the mutation (a P value 0.05 or less indicates a statistically significant probability). Around 20% of missense mutations in HGMD have been assigned a MutPred hypothesis.

**Note:** Answer intended for online and download users.

## There are many common SNPs listed as mutations in HGMD. Are you going to correct these?

We are in the process of reviewing each of the variants listed in HGMD that have been found to occur at a higher frequency in normal populations than might be expected for a rare disease-causing variant. We are collaborating with the 1000 Genomes Consortium to achieve this. We are currently aware of about 700 variants assigned as Disease-Causing Mutations (DMs) in HGMD that appear with an allele frequency of greater than 1% in the 1000 Genomes Project Data.

When a variant is observed in a normal population at a higher frequency than expected, it does not necessarily mean that the variant is not a disease-causing mutation. For example, variants may be common but give rise to a (recessive) disease only in those individuals where both alleles are affected e.g. CFTR dF508. Another mechanism might involve a potentially compensating variant (allelic or non-allelic) which could be present in much of the population, but disease will occur in the absence of the compensating variant. Alternatively, some variants may be compensated for by copy number variation. Even rare, disease-causing mutations typically do not exhibit 100% penetrance for the above (and other) reasons, although there are obvious exceptions (e.g. Huntington's disease). It is therefore not unreasonable that we should expect to find some disease-causing variants in healthy individuals. Indeed, we have estimated that human genomes (from normal, apparently healthy individuals) typically contain ~100 genuine loss-of-function variants, with ~20 genes completely inactivated (MacArthur et al., 2012). To come to a conclusion about the clinical relevance of a given mutation in a particular individual therefore requires the judgment of a medical professional who can take such factors into account. Thus, mutational variants in HGMD are likely to fall into a spectrum that ranges from spurious reports (especially in the older literature) where a variant may have been found simply in association with the disease while not being the actual causative variant, to cases where we cannot tell for sure, and cases where the variant, even though common, does indeed contribute to disease causation, and hence is correctly assigned as such.

To resolve this issue is likely to be a slow iterative process, because we have to review all the supporting evidence for each variant. After review, and if so required, we shall change the status of any incorrectly ascribed DM variant to disease-associated polymorphism (DP) or assign a question mark (DM?) or remove the mutation entry entirely.

**References**

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, Albers CA, Zhang ZD, Conrad DF, Lunter G, Zheng H, Ayub Q, DePristo MA, Banks E, Hu M, Handsaker RE, Rosenfeld JA, Fromer M, Jin M, Mu XJ, Khurana E, Ye K, Kay M, Saunders GI, Suner MM, Hunt T, Barnes IH, Amid C, Carvalho-Silva DR, Bignell AH, Snow C, Yngvadottir B, Bumpstead S, Cooper DN, Xue Y, Romero IG; 1000 Genomes Project Consortium, Wang J, Li Y, Gibbs RA, McCarroll SA, Dermitzakis ET, Pritchard JK, Barrett JC, Harrow J, Hurles ME, Gerstein MB, Tyler-Smith C. (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**:823-828.

**Note:** Answer intended for online and download users.

### How many mutations in splice sites can I find in HGMD?

There are 13973 mutations with consequences for mRNA splicing currently available in HGMD release 2012.3. To find updated statistics, please use the statistics page in HGMD Professional.

**Note:** Answer intended for online and download users.

### Why doesn't the lower case "acag" representation in TACTAC^414TTAGacagAGAAGCTGGG match the c.1245_1248delCAGA position for the deletion CD982750 in SMAD4?

Deletions and insertions in HGMD are not necessarily represented at the most 3-prime (downstream) possible location of the sequence. In the specific example for SMAD4, the mutation could be delACAG or delCAGA. It is not possible to tell which at the molecular sequence level. HGVS nomenclature requires that the mutation is represented as delCAGA (most 3-prime nucleotides). That is the essential difference.

**Note:** Answer intended for online and download users.

### Does HGMD provide information about genomic polymorphisms in genes? How does the HGMD define a genomic alteration — as a mutation or as a polymorphism?

Only disease-associated/functional polymorphisms are included in HGMD. To be included as disease-associated, a statistically significant ($p<0.05$) association between the polymorphism and a clinical phenotype must have been reported. In case no clinical phenotype is known to be associated with a polymorphic variant, but sufficient in vitro or in vivo expression/functional data have nevertheless been presented to indicate functional significance, then the variant will be included in HGMD. NCBI dbSNP numbers (where identified) are also included in the comment field. A polymorphism is a mutation found at a frequency of >1% in any population.

**Note:** Answer intended for online and download users.

### How many mutation entries in HGMD have dbSNP identifiers?

38725 mutation entries in HGMD Professional 2014.1 have dbSNP identifiers. HGMD entries that have been mapped to a corresponding entry in dbSNP display the FREQ symbol where the associated dbSNP entry contains population frequency data. 9761 mutations in HGMD 2014.1 have FREQ=frequency information. To find updated statistics, please use the predefined dbSNP identifier search available on the mutation search page.

**Note:** Answer intended for online and download users.

### How many disease terms and phenotypes are reported in HGMD?

HGMD Professional 2014.1 contains currently 13779 diseases/phenotypes (conditions). The disease descriptions can be different variants of one disease as taken from the literature and is therefore sometimes redundant. The hgmd_phenbase provides disease terms from MeSH, ICD-10 and other sources mapped to HGMD disease phenotypes to provide standard disease terms and unique identifiers. 13505 ICD10 phenotypes have been mapped to the 2012.2 HGMD version, and 2086 unique MeSH terms are annotated against HGMD phenotypes. Please use the statistics page for updated information.

**Note:** Answer intended for online and download users.

## Can I find mutations associated with two (or more) phenotypes in HGMD?

Yes, if a paper reports a variant to cause two genuinely different phenotypes (a very rare occurrence) this should be reflected in the disease field (e.g. CM013504 Stargardt disease and macular degeneration). In case two or more papers describe the same lesion as responsible for different phenotypes, we will record the earliest report as the 'primary' mutation reference and add the subsequent reports as 'Additional phenotype' secondary references. In each case, the phenotype is associated with the reference in which it is described. The details of the additional phenotypes (and other information from the secondary references) are provided in the expanded mutation record (that is reached by clicking on the accession number button). Currently (2014.1) 10836 mutations have additional disease/phenotype information from secondary references.

**Note:** Answer intended for online and download users.

## How to perform a batch search in HGMD® Professional?

HGMD online provides two options for batch searches. The first option at the main search "Professional" is designed to accept a list of up to 500 variant or gene identifiers. Identifiers accepted by the batch search include dbSNP, chromosomal coordinate and HGMD accession for variants and HUGO Nomenclature Committee gene symbols and IDs, Entrez Gene IDs and OMIM IDs for genes, and Variant Call Format (VCF). At the Advanced search tool we provide "MART" has an option for batch queries which can be saved as text files. A maximum of 50 identifiers of dbSNP, EntrezGene or PubMed IDs can be loaded and searched for.

**Note:** Answer intended for online and download users.

## Why do numbers for sequences in HGMD Professional differ from numbering in (primary) references?

Mutations presented in HGMD may utilise an alternate (up-to-date) transcript numbering compared to that used in the original report (which may have been published many years ago). Please don't expect older literature/manuscript numbering to continue to match modern transcript sequences in all cases.

**Note:** Answer intended for online and download users.

## How should I reference HGMD Professional in a scientific article?

The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, Hayden M, Heywood S, Millar DS, Phillips AD, Cooper DN *Hum Genet* (2020) epub.

**Note:** Answer intended for online and download users.