HGMD® Professional 2012.3 (Release date 28th September 2012)

HGMD Introduction

Citing HGMD

If you refer to HGMD in any publication, please cite:

Rationale

Human gene mutation is a highly specific process, and this specificity has important implications for the nature, prevalence and therefore diagnosis of genetic disease. Indeed, the recognition that certain DNA sequences are hypermutable has yielded clues as to the endogenous mutational mechanisms involved and provided insights into the intricacies of the processes of DNA replication and repair (Cooper and Krawczak 1993). In practical terms, a fuller understanding of the mutational process may prove important in molecular diagnostic medicine by contributing to improvements in the design and efficacy of mutation search procedures and strategies in different genetic disorders.

The Human Gene Mutation Database (HGMD) represents an attempt to collate known (published) gene lesions responsible for human inherited disease. This database, whilst originally established for the study of mutational mechanisms in human genes (Cooper and Krawczak 1993), has now acquired a much broader utility in that it embodies an up-to-date and comprehensive reference source to the spectrum of inherited human gene lesions. Thus, HGMD provides information of practical diagnostic importance to (i) researchers and diagnosticians in human molecular genetics, (ii) physicians interested in a particular inherited condition in a given patient or family, (iii) genetic counsellors, (iv) personal genomics and NGS researchers.

Data coverage

The Human Gene Mutation Database 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).

These data comprise various types of mutation within the coding regions, splicing and regulatory regions of human nuclear genes. Somatic mutations and mutations in the mitochondrial genome are thus not included, although in the latter case, links to Mitomap are now provided. Each mutation is entered only once in order to avoid confusion between recurrent and identical-by-descent lesions.

HGMD does not usually include mutations lacking obvious phenotypic consequences although a few such variants have been included where they could conceivably have some clinical effect (e.g. albumins, butyrylcholinesterases). Many published mutation searches identify more than one genetic change in a single patient. In such cases,
the relationship between a given lesion and the clinical phenotype has not always been immediately clear, and the curators of HGMD have had to rely exclusively upon the judgements of authors, peer reviewers and journal editors. The possibility of unintentional inclusion of some lesions with little or no pathological significance can therefore not be ruled out.

HGMD includes disease-associated/functional polymorphisms. To be included, there must be a convincing association of the polymorphism with the disease/functional phenotype. For a more complete explanation of how we choose to include such polymorphisms, please read our polymorphism inclusion criteria.

**Evidence for pathological authenticity**

Pathological mutations that dramatically disrupt the structure of a given gene are self-evidently very likely to be responsible for the associated clinical phenotype. However, for other categories of lesion, pathological mutations are often difficult to distinguish from polymorphisms with little or no clinical significance, particularly if their structural or functional consequences are subtle (Cotton and Scriver, 1998)*. Evidence for their authenticity in a pathological context therefore usually comes from one or more different lines of evidence:

- Absence in normal controls.
- Novel appearance and subsequent cosegregation of the lesion and disease phenotype through the family pedigree.
- Absence of any other lesion in the gene that could be responsible for the observed clinical phenotype.
- Previous independent occurrence in an unrelated patient.
- Non-conservative amino acid substitutions are more likely to disrupt protein function.
- Location in a protein region of known structural or functional importance.
- Location in an evolutionarily conserved nucleotide sequence and/or amino acid residue.
- In vitro demonstration of reduced gene expression/mRNA splicing/activity or stability of protein product consequent to mutation.
- Demonstration that the mutant protein has the same properties in vitro as its in vivo mutant counterpart.
- Reversal of the pathological phenotype in patient/cultured cells by gene replacement.

Despite the best efforts of the HGMD curators, it may be assumed that some categories of gene lesion listed in HGMD (e.g. missense mutations, regulatory mutations, splicing mutations) are likely to include some mutations that are not
actually causative even although they have been reported as being so. In some cases the evidence for pathogenicity may be dubious, such variants can be identified by the addition of a question mark (?) to the given disease/phenotype, which indicates that some degree of uncertainty is involved.

Data collection

Data are collected by the manual and computerised screening of journals and publicly available locus specific databases (LSDBs). Where possible, data are included from the original reports; entries are referenced to ‘Mutation Updates’ and review articles if the original publication is not available. Please note that ambiguously-described mutations are not included in the database until clarification has been obtained from the authors.

Curation policy

Disease-causing mutations are entered into HGMD where the authors of the corresponding report(s) believe that the mutation(s) being reported are in some way involved in conferring the associated clinical phenotype upon the individuals concerned. Disease-associated polymorphisms (DPs) are entered into HGMD where there is evidence for a significant association with a clinical or laboratory phenotype along with additional evidence that the polymorphism is itself of likely functional relevance (e.g. missense change, alters transcription factor/miRNA binding site etc). Functional polymorphisms (FPs) are entered into HGMD where the authors have demonstrated that the polymorphism in question exerts a direct functional effect (e.g. as evidenced by a luciferase reporter gene assay). Disease-associated polymorphisms with supporting functional evidence (DFPs) must meet both of the above criteria. Finally, frameshift or truncating variants (FTVs) are variants that alter the reading frame of the encoded protein (e.g. micro-deletion, nonsense mutation), although no associated laboratory or clinical phenotype has yet been reported. FTVs are generally identified in large-scale genome/exome screening studies, and may represent either latent protein deficiency states or heterozygous carrier states for recessive disorders. Coverage of this type of variant is not yet comprehensive.

The HGMD curators have adopted a policy of continual assessment of the curated content with respect to the mutation entries in the database. If and when additional important new information pertaining to a specific mutation entry becomes available (e.g. questionable pathogenicity, confirmed pathogenicity, additional phenotypes, population frequency, functional studies etc), then the mutation entry may be revised or recategorized. Alternatively, a comment or additional reference may be added in order to communicate this new information to users. Where new information becomes available which suggests that a given disease-causing mutation (DM) entry is likely to be of questionable pathological relevance or even a neutral polymorphism (on the basis of additional case reports, genome/population screening studies etc), it may be flagged with a question mark (DM?) or even removed from the database entirely if it turns out to have been erroneously included ab initio.

The majority of clinical phenotypes assigned to DMs in HGMD represent rare conditions that most people would consider to be “diseases”. However, it is important to note that HGMD also considers a “silent” protein deficiency or biochemical phenotype (e.g. butrylcholinesterase deficiency, reduced oxygen affinity haemoglobin etc) to be worthy of inclusion since they are potentially disease-relevant (even if they
are relatively common in the general population). Such variants may well be assigned to the DM category.

HGMD users should not assume that just because a mutation is labelled "DM", that it automatically follows that the HGMD curators are certain that the mutation is disease causing in all individuals harbouring it (i.e. that this mutation is deemed to be fully penetrant). As geneticists, we know that this is not invariably going to be the case. Indeed, it is likely that next generation sequencing programmes (such as the 1000 Genomes Project) will identify considerable numbers of "DM" mutations in apparently healthy individuals\(^1\). Such lesions should not be regarded automatically as being clinically irrelevant because it is quite possible that these mutations may be low-penetrance or late onset disease susceptibility alleles rather than neutral variants. It has always been HGMD policy to enter a variant into the database even if its pathological relevance may be questionable (while indicating this fact wherever feasible to our users), rather than run the risk of inadvertently excluding a variant that may be directly relevant to disease.

## Mutation categories

### Table 1. Summary of mutation categories in HGMD

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Genomic coordinates</th>
<th>HGVS</th>
<th>Web interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense/nonsense</td>
<td>Single base-pair substitutions in coding regions</td>
<td>YES</td>
<td>YES</td>
<td>HGMD PRO and ADVANCED</td>
</tr>
<tr>
<td>Splicing</td>
<td>Single base-pair substitutions with consequences for mRNA splicing</td>
<td>YES</td>
<td>YES</td>
<td>HGMD PRO and ADVANCED</td>
</tr>
<tr>
<td>Regulatory</td>
<td>Single base-pair substitutions causing regulatory abnormalities</td>
<td>YES</td>
<td>NO</td>
<td>HGMD PRO and ADVANCED</td>
</tr>
<tr>
<td>Small deletions</td>
<td>Micro-deletions (20 bp or less)</td>
<td>YES</td>
<td>YES</td>
<td>HGMD PRO and ADVANCED</td>
</tr>
<tr>
<td>Small insertions</td>
<td>Micro-insertions (20 bp or less)</td>
<td>YES</td>
<td>YES</td>
<td>HGMD PRO and ADVANCED</td>
</tr>
<tr>
<td>Small indels</td>
<td>Micro-indels (20 bp or less)</td>
<td>YES</td>
<td>YES</td>
<td>HGMD PRO and ADVANCED</td>
</tr>
<tr>
<td>Gross deletions</td>
<td>Deletions over 20 bp</td>
<td>NO</td>
<td>NO</td>
<td>HGMD PRO</td>
</tr>
<tr>
<td>Gross insertions</td>
<td>Insertions over 20 bp</td>
<td>NO</td>
<td>NO</td>
<td>HGMD PRO</td>
</tr>
<tr>
<td>Complex rearrangements</td>
<td>Recorded in narrative format</td>
<td>NO</td>
<td>NO</td>
<td>HGMD PRO</td>
</tr>
<tr>
<td>Repeat variations</td>
<td>Recorded in narrative format</td>
<td>NO</td>
<td>NO</td>
<td>HGMD PRO</td>
</tr>
</tbody>
</table>

All HGMD entries comprise a reference to the first literature report (primary reference) of a mutation, the associated disease state as specified in that report, the gene name, official symbol (as recommended by the HUGO Gene Nomenclature Committee; HGNC). In cases where no official gene symbol exists, a provisional symbol has been assigned by the HGMD curators, which is denoted by lower-case letters. HGMD also may provide where available additional references for a given entry.

NCBI dbSNP numbers (where identified) may also be recorded. The inclusion of a dbSNP identifier for a HGMD entry, in no way implies that the entry in question is indeed a polymorphism.
Missense/nonsense

Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet.

Splicing

Mutations with consequences for mRNA splicing are presented in brief with information specifying the relative position of the lesion with respect to a numbered intron donor or acceptor splice site. Positions given as positive integers refer to a 3' (downstream) location, negative integers refer to a 5' (upstream) location.

Regulatory

Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet. Substitutions causing regulatory abnormalities are logged in with thirty nucleotides flanking the site of mutation on both sides; the location of the mutation relative to the transcripational initiation site, initiator ATG or Termination codon is given.

Small deletions

Micro-deletions (of 20 bp or less) are presented in terms of the deleted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character ("^`). Where deletions extend outwith the coding region of the gene in question, other positional information is occasionally provided e.g. 5' UTR (5' untranslated region) or E6I6 (denotes exon 6/intron 6 boundary).

Small Insertions

Micro-insertions (of 20 bp or less) are presented in terms of the inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character ("^`). Where deletions extend outwith the coding region of the gene in question, other positional information is occasionally provided e.g. 5' UTR (5' untranslated region) or E6I6 (denotes exon 6/intron 6 boundary).

Small Indels

Micro-indels (of 20 bp or less) are presented in terms of the deleted/inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character ("^`). Where deletions extend outwith the coding region of the gene in question, other positional information is occasionally provided e.g. 5' UTR (5' untranslated region) or E6I6 (denotes exon 6/intron 6 boundary).

Other mutation types
For gross deletions, gross insertions, repeat variations and complex rearrangements, information regarding the nature and location of a lesion is logged in narrative form because of the extremely variable quality of the original data reported.

**Sub-categorisation of HGMD data**

Recently, HGMD has adopted a policy of sub-categorising polymorphism (and mutation) entries. Thus, polymorphisms may now be allocated to one of three possible categories reflecting the aforementioned criteria, with one further category covering disease causing mutations:

**Disease-associated polymorphism (DP)**

A polymorphism reported to be in significant association with a disease/phenotype (p<0.05) that is assumed to be functional (e.g. as a consequence of location, evolutionary conservation, replication studies etc), although there may as yet be no direct evidence (e.g. from an expression study) of function.

**Disease-associated polymorphism with additional supporting functional evidence (DFP)**

A polymorphism reported to be in significant association with disease (p<0.05) that has evidence of being of direct functional importance (e.g. as a consequence of altered expression, mRNA studies etc).

**In vitro/laboratory or in vivo functional polymorphism (FP)**

A polymorphism reported to affect the structure, function or expression of the gene (or gene product), but with no disease association reported as yet.

**Frameshift or truncating variant (FTV)**

A polymorphic or rare variant reported in the literature (e.g. detected in the process of whole genome/exome screening) that is predicted to truncate or otherwise alter the gene product (i.e. a nonsense or frameshift variant) but with no disease association reported as yet. Please note that any variant affecting the obligate donor/acceptor splice site of a gene will not be included in this category unless there is evidence for an effect on the splicing phenotype. Variants occurring in pseudogenes will also be excluded unless evidence for a functional effect is present for both the pseudogene itself (Balakirev ES & Ayala FJ, 2003) and the variant in question.

**Disease causing mutation (DM)**

Pathological mutation reported to be disease causing in the corresponding report (i.e. all other HGMD data).
Replication studies for disease-associated polymorphic variants

The replication of disease-association studies can be a source of additional information to satisfy the inclusion criteria. If a replication study serves to support a previously tenuous genotype-phenotype correlation, then the phenotype can be ‘promoted’ from ‘association with?’ to ‘association with’ and the details of the replication study may be recorded in the comment field.

HGMD wild type and mutant alleles may be reversed

In HGMD the mutant allele is recorded as the allele responsible for the reported disease/phenotype. In some HGMD records (typically polymorphism data, where both the wild type and mutant allele may be found at high frequency), the wild type as given by RefSeq may be the same as HGMD mutant allele, this is so we preserve the relationship between reported disease/phenotype and mutant allele.
Disease-associated and/or functional polymorphisms inclusion criteria for HGMD

Aim
HGMD seeks to include DNA sequence variants that are either (i) disease-associated and of likely functional significance, or (ii) of clear functional significance even though no associated clinical phenotype may have been identified to date. The difficulty inherent in assessing published polymorphism reports describing potential disease associations has led to the adoption of a set of inclusion criteria that together describe what we consider to be a methodical and uniform approach to dealing with such variants as they appear in the literature.

Background
At present, ~55% of the polymorphic variants recorded in HGMD are 'disease-associated'. However, even in such cases where no disease association has yet been demonstrated, functional polymorphisms that alter the expression of a gene or the structure/function of the gene product are potentially very important. Although a functional polymorphism with no disease association may not have any direct and/or immediate clinical relevance, these data are potentially very valuable in terms of understanding inter-individual differences in disease susceptibility. The vast majority of polymorphic variants in HGMD are single nucleotide polymorphisms (SNPs) but a small number are of the insertion/deletion type. The polymorphic variants logged in HGMD are generally located in either gene regulatory or coding regions although it should be noted that SNPs occurring outside of these regions may nevertheless still have consequences for gene expression, splicing, transcription factor binding etc.

Definitions
The distinction between a disease-associated polymorphism and a pathological mutation is in practice often fairly arbitrary and is generally made in the context of the prevalence of the variant in the population as well as its penetrance (the frequency with which a specific genotype manifests itself as a given clinical phenotype). Variants with a minor allele frequency of ≥1% in the population being studied are, by convention, termed polymorphisms. These polymorphisms are identified in the database by the addition of terms to the clinical/laboratory phenotypic description. These additions are limited to 'association with' and 'association with?' (question marks being included to indicate that the association is judged by the HGMD curators to be somewhat tenuous).

Inclusion Criteria for Disease-Associated/Functional Polymorphisms
Polymorphic variants logged in HGMD usually fall into two discrete categories:

Disease-associated polymorphisms of functional significance (DP and DFP)
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 addition, other information (e.g. in vitro or in vivo expression/functional data, replicated association studies, epidemiological studies, evolutionary conservation data etc) should have been made available to support the contention that the polymorphism in question is itself of bona fide functional significance. Such a polymorphism (DP) could have consequences for gene expression, protein structure/function, gene splicing, etc. These supporting experimental data are required to ensure that non-causative variants (i.e. those merely in linkage disequilibrium with the actual causative variants) are not included. If the functional
data required to support the inclusion of a disease-associated variant are contained
within a subsequent article, the reference logged in HGMD will still be that which
originally reported the disease association (DFP).

**Polymorphisms of functional significance with no reported disease association**

(PP)

If no clinical phenotype is known to be associated with a polymorphic variant, but
sufficient in vitro or in vivo expressionfunctional data1 have nevertheless been
presented to indicate functional significance, then the variant will be included in
HGMD. Typically, such data provide evidence for a direct effect on gene expression,
protein structure and/or function, gene splicing etc. These variants can thus, in a very
real sense, be considered as giving rise to a ‘deficiency’ (or occasionally a surfeit) of
a given gene transcript or protein product. Hence, the phenotype recorded in HGMD
would entail a brief descriptive of the functional effect e.g. ‘Reduced gene
expression, association with’. If, at a later date, evidence becomes available to
indicate that a disease/clinical phenotype is associated with such a polymorphism,
the disease/clinical phenotype and reference to that variant is entered into the
comment field, reflecting an additional reference and phenotype. Polymorphic
variants affecting individual drug responses, patient survival times after diagnosis,
and responses to surgical intervention, are not included in HGMD. Studies which
simply report dbSNP numbers in association with disease (e.g. from large scale
genome-wide association studies), with no additional evidence of direct functional
involvement are also not included in HGMD. Users interested in this particular
category of variation should try other databases such as the Catalogue of Published
Genome-Wide Association Studies (http://www.genome.gov/26525384/) or the
Genetic Association Database (http://geneticassociationdb.nih.gov/).

One caveat to bear in mind is that in vitro studies are not invariably accurate
indicators of the in vivo situation [see for example Cirulli ET & Goldstein DB, 2007 &
Dimas AS et al (2009)].

In some instances, the above criteria may be only partially satisfied, such that the
HGMD curators remain unconvinced as to the functional/phenotypic relevance of the
variant reported. In such cases, the polymorphism may nevertheless be included as
a result of (i) supporting information becoming available subsequent to publication of
the original (first) report, or (ii) because the associated gene/disease state was
deemed to be of sufficient importance for the variant to warrant further study. Such
variants have been ascribed the descriptor ‘association with?’ (as opposed to
‘association with’ without a question mark) to indicate that some degree of
uncertainty is involved.

**Other Categories of Variation**

**Copy number variations**

Copy number variations (CNVs) are DNA segments >1 kb in length that present with
variable numbers of copies in a given population. These variants are being reported
in the literature with an ever increasing frequency. CNVs are potentially functionally
significant and should therefore in principle be treated by HGMD in a similar manner
to any other polymorphism. Human CNVs are however already being collected by
other databases such as the Database of Genomic Variants
(http://projects.tcag.ca/variation/) and the Human Genome Structural Variation
Project (http://humanparalogy.gs.washington.edu/structuralvariation/). CNVs that are disease-associated are also being collated in databases such as DECIPHER (http://www.sanger.ac.uk/PostGenomics/decipher/), the European Cytogeneticist’s Association Register of Unbalanced Chromosome Aberrations (http://www.ecaruca.net) and the Chromosome Abnormality Database (http://www.ukcad.org.uk/cocoon/ukcad/). Whilst HGMD does not wish to replicate the excellent curatorial work of other organisations, HGMD is still interested in such variants if they meet certain criteria. HGMD will therefore include these variations if they are shown to be both of functional significance and associated with disease, and if they involve a single characterised gene that is itself clearly involved in the disease association.

**Risk haplotypes**
Reports of haplotypes associated with an increased risk of disease are not included in cases where there is no indication as to precisely which variant (or variants) within the haplotype is responsible for the disease association/functional effect. If, however, evidence is presented to support the contention that a single variant within the risk haplotype is causative and/or of functional significance to a degree which satisfies the inclusion criteria, then it would certainly be included in HGMD.

**Limitations**
The main limitation with recording disease-associated polymorphic variants of functional significance was the inclusion of only a single reference for each sequence change in HGMD. A large proportion of the papers reporting an association between a disease and a polymorphic variant do not include functional data on that variant. This problem has now been addressed with the introduction of multiple referencing.
HGMD Web Interfaces (HGMD PRO + HGMD Advanced)

HGMD PRO

Searching HGMD PRO

HGMD PRO can be searched by entering search terms and selecting the area of HGMD in which to search, and boolean searching with wildcards (*) is permitted. Alternate spelling support has recently been added. For example both “haemophilia” and “hemophilia” should now lead to the F8 and F9 genes. Please note that alternate spellings will only function for whole words.

Boolean fulltext searching

+ operator indicates that the search term must be present in each result.

e.g. +breast +cancer returns results where both breast and cancer are present.

- operator indicates that the search term must not be present in any result.

e.g. -breast +cancer returns results that contain cancer but not breast.

- operator serves as the truncation (or wildcard). Unlike the other operators, it should be appended to the word to be affected.

e.g. poly* returns results such as polyposis, polycystic and polypeptide.

" operator when used to enclose your search terms means that the search terms are treated literally (as they were typed).

e.g "hum mutat" will return results containing the exact phrase "hum mutat".

Note1: If boolean operators are not utilised, then multiple search terms are treated by a boolean search as separate entities (e.g. a search for breast cancer without quotes [“”] will return all results containing breast OR cancer.

Note2: HGMD fulltext searching in MySQL has a minimum word length of 4 characters by default. Any search terms entered with less than 4 characters may not be recognised by the HGMD fulltext search engine.

Non-boolean searching

This type of search works in a different way to the boolean fulltext search. The fulltext index is not utilised, therefore there is no minimum word length in force (2 and 3 character searches will function). No operators are used, and multiple search terms will be treated as a literal phrase. Partial matches will also be returned by the search (except for the HGNC/OMIM/GDB/Entrez ID search, will will return exact matches only).

Search for the word transporter.
• Returns results where the word transporter is present. Will also return partial matches such as transporters and cotransporter.

Search for the words prostate cancer.

• Returns results where the words prostate cancer occur as a literal phrase.

Search for the gene symbol APC.

• Returns results where APC occurs, including the APC gene itself, plus partial matches e.g. APCDD1 and PROC (where APC is an alias).

Note: This search does not utilise an index, and therefore may be somewhat slower than a Boolean fulltext search.

Primary gene search

Enter search term:

The genes present in HGMD may be found by utilising seven different gene-oriented search options.

1. All fields search - Searches for your search terms in all fields listed below at once (2-7).

2. Gene symbol search - Searches HGMD for the official HUGO Gene Nomenclature Committee gene symbol. Any gene symbol aliases that have been identified are also be included in this search.
   Official symbol example: 'ABCC2'
   Gene symbol alias example: 'MRP2'

3. Gene description search - Searches HGMD for the official HUGO Gene Nomenclature Committee gene name. Any gene name aliases that have been identified are also be included in this search.
   Official description example: 'ATP-binding cassette, sub-family C (CFTR/MRP), member 2 (CMOAT)'
   Gene description alias example: 'Canalicular multispecific organic anion transporter'

4. Chromosomal location search - Searches HGMD for the chromosomal location of HGMD genes.
   Example: '10q24'

5. HGNC/OMIM/GDB/Entrez ID - Searches HGMD for the externally-assigned gene identifiers from the HUGO Gene Nomenclature Committee database, Online Mendelian Inheritance in Man, the Genome Database (legacy only) and the Entrez Gene database.
   HGNC Example: '5384'
   OMIM Example: '601107'
   GDB Example: '6089489'
   Entrez Example: '1244'
6. Disease/phenotype search - Searches HGMD for the disease/phenotype associated with reported mutations in HGMD genes. Example: 'Dubin-Johnson syndrome'

7. Gene ontology search - Searches for the ontology terms that have been assigned (by the Gene Ontology Consortium) to the genes present in HGMD. Example: 'organic anion transmembrane transporter activity', 'GO:0008514' or '0008514'

Detailed results will contain a list of gene information, diseases and external links/IDs associated with your search terms. Concise results are limited to listing gene information (symbol, description, chromosomal location) only.

To access the HGMD record for that gene, click on the gene symbol listed. Some portions of the returned text may be highlighted in green. This indicates which part of the results match your search terms.

Secondary gene search

**Symbol:**

The secondary search allows users to jump to a specific mutation data set if the gene symbol is known. This search will only function with the correct HUGO Nomenclature Committee gene symbol.

Other search options

The ability to browse HGMD genes by gene symbol (A-Z), chromosomal location (1-22, X and Y), or to view pre-queried HGMD data (a random HGMD gene entry, genes newly added for the current release, genes updated with new mutation data for the current release, genes by total number of mutations, or genes sorted by ontology term) is also available.

Mutation search

There are four ways HGMD may be searched for specific mutations.

1. Codon number search - Searches for mutations affecting a particular codon in the coding region. This search will return results from the missense/nonsense, small deletions, small insertions and small indels mutation categories. Results will contain a list of genes/diseases with links to both the gene page and specific mutations found during your search. For small deletions, small insertions and small indels, the codon number you input refers to the first affected codon, not the last whole codon as presented by "^" in the HGMD entry (i.e. - in the entry CD010589 AAAS 156 TTGCGT^GTCTttGCATGGCACC, the first affected codon is 157). Please note that this search will not pick up mutations either beginning or wholly within an intron as there will not be a first affected codon to search for. Example: `157`

NOTE: You may now restrict your search to disease causing mutations or disease-associated/functional polymorphisms. The default is to search both types. Please note that for convenience, "frameshift or truncating variant" is included under both types.
2. Accession number search - Searches HGMD for specific accession numbers (if known). Partial accession numbers will be accepted (with wildcards *). Results will contain a list of genes with links to both the gene page and specific accession number(s) found during your search.
Example: 'CM035497' or 'CM035*'

3. Search using official HGVS mutation nomenclature - Searches HGMD for mapped mutations using the official nomenclature as described by den Dunnen JT and Antonarakis SE (2001) Hum Genet 109: 121-24. This feature currently works for missense/nonsense mutations, small deletions, small insertions and small indels that have been successfully mapped. Please note that for deletions, insertions and indels, HGVS nomenclature requires that the most 3-prime affected nucleotides are specified.
Example: '298C>T' or 'R100X'
Example: '20_21delTT' or '20_21del2' or '20_21del'

Example: '104_105insAA' or '104_105ins2' or '104_105ins' and
'2195_2198dupAACA' or '2195_2198dup4' or '2195_2198dup'

Example: '385_386delAGinsGTT' or '385_386del2ins3' or '385_386delins'

4. Search using chromosomal coordinates - Searches HGMD for mapped mutations using chromosomal coordinates. This feature currently works for missense/nonsense mutations, small deletions, small insertions and small indels that have been successfully mapped to the genome (currently build 36.3/hg18). Users may enter a specific coordinate range, a single coordinate or may choose to display mutations from an entire chromosome, though this will tend to produce a large set of results.
Example: 'chr1:2327114_2333800'
Example: 'chrX:77130681_77132008'

Example: 'chr7:1942983'

Example: 'chr2'

**HGMD PRO Reference search**

There are six ways HGMD may be searched for references.

1. All fields search - Searches for your search terms in all fields listed below (2-6).

2. First author search - Searches HGMD for the first author (surname) of the mutation references found in HGMD.
Example: 'Edwards'

3. PubMed journal search - Searches HGMD for the PubMed journal title abbreviations associated with the mutation references found in HGMD. Full journal names may also be used.
Example: 'Am J Hum Genet' or 'American journal of human genetics'

4. PubMed ID search - Searches HGMD for PubMed IDs associated with the mutation references found in HGMD.
Example: '9042910'

5. Publication year search - Searches HGMD for the specific years in which the mutation references found in HGMD were published.
Example: '1997'

6. HGMD gene search - Searches HGMD for the gene symbols associated with the mutation references found in HGMD.
Example: 'TCOF1'

Results will contain a list of genes and references associated with your search terms. To access the HGMD record for that gene, click on the gene symbol listed. You may also view the PubMed record (if available) associated with each search result.

**F.A.Q. search**

The HGMD F.A.Q. may be searched for keywords, or may be listed in its entirety if required. Boolean operators may be used (see above) and as with HGMD searching, there is a minimum word length in force, usually of 4 characters.

Problems ??

**General remarks**

Firstly, always make sure you are searching with the correct option selected. You cannot use a gene symbol search to find a disease. If you are getting error messages you need to alter your search strategy. If you are getting too many "gene not found" errors when searching for gene symbols, you should try searching alternate fields instead. For example, CD95 is the old symbol for FAS. It will not appear in a gene symbol search, but it will appear in a gene description or alias search. If you are not getting the results you require when using a gene description or disease/phenotype search, you can try to narrow your search. For example, entering "cancer" as a disease/phenotype search term will produce too many results. Narrowing it to something like "gastric cancer" may produce the desired results. Note also that there is a minimum word length of 4 characters for HGMD searching, so entering any term with less than that will not produce any results.

**Have I got the right browser ?**

The HGMD web pages should work correctly with any modern browser. These include the most popular browsers such as Netscape Navigator, NCSA Mosaic, Microsoft Internet Explorer, Mozilla, FireFox and Opera. Text-only browsers such as Lynx can also operate HGMD correctly although some pages will look odd. Whilst Netscape is used in Cardiff, the pages have not been enhanced specifically for Netscape, and should work well with any other program that supports tables and forms. Almost all browsers support forms properly, and although those browsers unable to handle tables will look somewhat untidy, they will perform all searching duties correctly.

We are working toward making our code compliant with HTML v4.01. If your browser is not capable of this, it will only mean that the presentation on your screen is not identical to ours. The results generated from search requests should not be affected.
Can I bookmark any page?

With browsers such as Netscape or NCSA Mosaic it is possible to 'bookmark' individual pages from HGMD, be it a single gene page or even this help page. We do not recommend this though, as not only are you likely to miss important notices placed on the HGMD home page, but we also reserve the right to change the internal structure of the database as required.
HGMD Advanced Search documentation

Search types

The HGMD Advanced Search has four different interfaces:

1. Substitutions Plus (What about the old “Substitution”, which can be selected from the left hand list?)
2. Micro-lesions
3. Quick search
4. Gene based Search

Figure A1. Access to HGMD Advanced Search interfaces
1. Substitutions PLUS

Overview

The Substitutions PLUS search includes all HGMD single base substitutions (missense, nonsense, regulatory and splicing) to be interrogated via one interface.

The mutation types used by this search have been mapped onto genomic reference sequences (hg18 & hg19) to allow flexible data-mining and motif-searching for transcription factors, conserved DNA domains or user-defined motifs.

Figure A2. Overview of Substitutions PLUS Interface

How do I search?

Nucleotide Substitutions can be mined using one or more of the following categories:
- Motif Search
- Base Substitution
- Amino Acid Substitution
- Splicing
- Regulatory
- Other
Motif Search

Using the drop-down box predefined motifs can be selected or a user-defined motif can be typed into the text field to the side of the drop-down box.

The user-defined motif can be defined by using regular expressions. Some characteristics of extended regular expressions are:

- '.' matches any single character.

A character class '[...]' matches any character within the brackets. For example, '[abc]' matches ‘a’, ‘b’, or ‘c’. To name a range of characters, use a dash. '[a-z]' matches any letter, whereas '[0-9]' matches any digit.

**Example 1**
The regular expression `[CA][CA]AGGTAGGTAA` would match to the 5’ splice site consensus sequence.

**Example 2**
The regular expression `[GC]ATG` would match to both GATG and CATG.

**Example 3**
The regular expression `[AG][CT][AG]` would match the sequence RYR.

Note - the motif search feature is not case sensitive and single letter codes are not supported.

If the check-box Created is selected then all substitutions that create this motif will be returned.

If the check-box Abolished is selected then all substitutions that disrupt this motif will be returned.

The Include or Exclude radio buttons allow the selection to be included or excluded from the search results respectively.

### Base Substitution

Using the drop-down menus the wild type and/or the mutant nucleotides can be selected.

### Amino Acid Substitution

Using the drop-down menus the wild type and/or the mutant amino acids can be selected.

### Splicing

Mutations that have been shown to affect splicing can be searched for by:

- entering an intron number into the ivs text box.
- The site of the substitution can be selected (either donor or acceptor splice site).
The location relative to the splice site can also be selected.

**Other**

Select the field to search from the list. Enter the search term in the text field. If the 'Fuzzy Search' check box is selected then wildcards will be added to the beginning and end of the search term.

Combining of the “other” option with the categories listed above will narrow the results to e.g to a certain disease/phenotype or gene.

**Filters**

Advanced Search Results can be filtered by mutation type:
- Missense/nonsense
- Regulatory
- Splicing

Advanced Search Results can also be filtered by variant class:
- Disease-causing mutations
- Disease-associated polymorphisms
- Functional polymorphisms

**Output Fields:**
1. Mutation type e.g. regulatory
2. Variant class e.g DM,
3. HGMD id
4. dbSNP
5. Functional profiling result
6. Disease/Phenotype
7. Gene symbol
8. Entrez gene id
9. link to PROTEOME locus report
10. HGVS nomenclature
11. mutation description
12. Genomic coordinates (hg19 default)
13. MutPred
14. SIFT
15. Genomic sequence context
16. Primary Reference
17. Mutation viewer link
**Functional Profiling of Mutation Data**

HGMD is currently in the process of annotating variants with both in vitro/in vivo and in silico data to help in the ascertainment of the molecular mechanism underlying the functional effect of a given mutation.

When this process has been completed, HGMD aims to annotate mutation data for over 40 different types of functional site including exonic splice enhancers (ESE), post-translational modification sites and numerous transcription factor binding sites (TFBS) etc.

Please see the following paper for more information:

Access to the functional profiling search tools is via the ‘Substitutions Plus' Search.

**An example functional profiling search:**

Search for HGMD regulatory variants disrupting transcription factor binding sites (TFBS) from TRANSFAC. Select ‘TFBS' from functional profiling drop down menu (Figure A10).

**Figure A10.** Drop down functional profiling menu from Substitutions Plus.

Functional sites disrupted by the mutation in question are shown in the functional Profile column (Figure A11, highlighted by red box)
Clicking on the relevant functional site shows the functional profiling mutation report (Figure A12) which displays both *in vitro/in vivo* and *in silico* evidence (where available) for any functional site disruption.

**Figure A12.** Functional profiling mutation report

---

**Ability to download results (Figure A3):**

1. Results downloaded as a tab delimited file
2. Genome browser track using hg18 coordinates
3. Genome browser track using hg19 coordinates

**Figure A3.** Downloading results from a Advanced Search query

```
<table>
<thead>
<tr>
<th>Mutation type</th>
<th>Variant class</th>
<th>HGMD_ID</th>
<th>dbsnp</th>
<th>Functional Profile</th>
<th>Disease/Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulatory</td>
<td>FP</td>
<td>CR097904</td>
<td>rs327053</td>
<td></td>
<td>nc-RNA levels, association with ?</td>
</tr>
<tr>
<td>Missense</td>
<td>DP</td>
<td>CM980001</td>
<td>rs6660</td>
<td></td>
<td>Alzheimer disease, association with</td>
</tr>
</tbody>
</table>
```

Query returned 76015 mutations from 3747 different genes.
2. Micro-Lesions Search < 21bp (Insertions, Indels and Deletions)

This category comprises all micro-insertions, micro-indels and micro-deletions. This allows the data-mining to be performed on a number of fields including insertion size, deletion size and user-defined motifs. Search results are available to download, using the same method as the Substitutions Plus search.

The micro-lesions search is accessed by clicking 'Micro-lesions' from the Advanced Search menu (Figure M1, highlighted by red circle).

Figure M1. Micro-lesions search page

Output fields include:

Mutation type, HGMD ID, dbSNP, Disease/Phenotype, Gene, HGVS, Genomic coordinates, Sequence context, codon, Deleted bases, Inserted bases, Nucleotide, Reference
3. Quick Search (Search in both Nucleotide substitutions and Micro-lesions)

With the Quick Search Function, selected fields in different tables (Nucleotide substitutions and Micro-lesions) can be searched at the same time.

This search also queries the title of the original mutation report for key words. This search option is especially useful to get a quick overview of the information available in different tables, e.g. for a certain disease.

The Quick Search Results are ranked by relevance and are assigned a Ranking rating.

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).

**Figure Q1.** HGMD Advanced Quick Search

![HGMD Advanced Quick Search](image)

**Welcome to HGMD Professional version 2011.2 (24/6/2011)**

To start a search, select one of the tables below

or browse disease genes by chromosomal location

or enter your Quick Search query here: [Japanese]

![Chromosomal Location](image)

**Quick Search examples**

**Example 1**

Enter 'Japanese' as the search term (Figure Q1, highlighted with red circle). This returns all mutations linked to a Mutation Report with 'Japanese' in the title and/or 'Japanese' in the Gene or Disease fields.

**Example 2**

Enter 'stroke candidate gene' as the search term. This returns all mutations linked to a Mutation Report with 'stroke candidate gene' in the title and/or 'stroke candidate gene' in the Gene or Disease fields.
Genome-interpretation and NGS based analysis using the Advanced Search

The Advanced Search provides a collection of utilities to assist in the analysis of NGS data:
1. Results from the Advanced Search can be exported into a custom genome trax to allow importing into other software tools such as the UCSC genome browser and Genome Trax.
2. Functional profiling of mutation data to identify the underlying molecular mechanism including the mapping of HGMD variants to TFBS from Transfac.
3. Providing annotations from SIFT (Sorting Intolerant From Tolerant).
4. Providing annotations from MutPred.

Option 1. Exporting HGMD data to other applications (e.g. Genome Trax or UCSC) by creating Custom Genome Browser Track

Results from the HGMD Advanced Search can be downloaded as custom genome browser tracks, which can then be viewed in the UCSC Genome Browser or imported in Genome Trax. Please follow Step by step guide below.

Figure A4. An Advanced Search result been viewed on the UCSC genome browser.

UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)

Step 1. Creating your HGMD Custom Genome Browser Track
- Use the Advanced Search to generate your search query.
- Click submit and navigate to results page
- Click download results as genome browser track (Figure A5)
- Save custom genome browser track to your computer (Hint: remember where you save it)
Step 2 importing your HGMD Custom Genome Browser Track into an external program (e.g. UCSC genome browser or Genome Trax)

- Navigate to UCSC genome browser gateway (http://genome.ucsc.edu/cgi-bin/hgGateway)
- Click manage/add custom tracks (Figure A6)

Step 3. finding the Custom Genome Browser track to upload.
- Click Browse (Figure A7) and upload previously saved browser track
- Click submit
Figure A7. Uploading the Custom Browser Track

Step 4
Your track should now be imported.
Click go to genome browser (Figure A8)

Figure A8.
Step 5

- View your custom track in the genome browser (Figure A9).
- Note: Large custom tracks may take a while to load.

**Figure A9.** Viewing your HGMD custom genome browser track.

Once upload to the UCSC, these HGMD custom annotation tracks are viewable only on the machine from which they were uploaded and are automatically discarded 48 hours after the last time they are accessed. The HGMD custom annotation tracks are for the sole use of HGMD Professional subscribers only.
MutationMart: batch mode search for HGMD

HGMD can be queried in a batch mode using three different types of identifier:

1. dbSNP identifier e.g. rs1800072
2. PubMed identifier e.g. 20981092
3. Entrez gene identifier e.g. 1080

Multiple identifiers (up to 500) can be queried at once, place each identifier on a new line.

MutationMart Results can optionally be downloaded to a tab delimited text file.
Tools used to predict harmful missense mutations (SIFT and MutPred)

HGMD missense mutations have been annotated with two different tools, which identify harmful missense mutations. MutPred also makes predictions about the underlying molecular mechanism disrupted as well, e.g. loss of phosphorylation site.

SIFT (Sorting Intolerant From Tolerant)

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 (DM) missense mutations in HGMD around 80% are predicted to be deleterious by SIFT. 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. For more information please refer to the SIFT website (http://sift.jcvi.org/www/SIFT_help.html).

MutPred, predicting deleterious missense mutations and underlying molecular mechanisms disrupted

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', whereas a MutPred score > 0.75 should be considered a high confidence 'harmful' prediction.

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 indicates a statistically significant probability). Around 20% of missense mutations in HGMD have been assigned a MutPred hypothesis. For more information please refer to the MutPred website (http://mutpred.mutdb.org/about.html).
Help page for Mutation Viewer

The Mutation Viewer depicts coding region mutations superimposed on the cDNA sequence of a gene. The Mutation Viewer is a Java Applet and requires at least Java version 1.5 installed. Note: splicing mutations are not supported by the Mutation Viewer.

The wild-type cDNA sequence is represented in black whilst the mutated nucleotides are shown in different colours.

- Nucleotide Substitutions are shown in RED.
- Deletions are shown in BLUE.
- Insertions are shown in MAGENTA.
- Indels are shown in GREEN.

The amino acid sequence annotation of the wild-type cDNA can be switched on and off with the **Show Amino Acids** button.

**Figure V1.** CFTR gene as displayed by the Mutation Viewer

**Mutation Viewer examples**

**Nucleotide Substitution**

*Example* Leu-Arg CTT-CGT codon 51.

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Codon</th>
<th>Wild-type</th>
<th>Mutated</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C</td>
<td>GGACTTG</td>
<td>GGACCTG</td>
</tr>
<tr>
<td>G</td>
<td>T</td>
<td>GGACTTG</td>
<td>GGACCTG</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>GGACTTG</td>
<td>GGACCTG</td>
</tr>
</tbody>
</table>

**Figure V1.** CFTR gene as displayed by the Mutation Viewer

**Mutation Viewer examples**

**Nucleotide Substitution**

*Example* Leu-Arg CTT-CGT codon 51.
Insertion
*Example* CCTGAAA^GATagatATTAATTTCA ^codon 94

- TCCTGAAAGATATTAATTTCAAGATA
- AGAT
- TCCTGAAAGATATTAATTTCAAGATA
- TCCTGAAAGATAGAATTTCAAA

Deletion
*Example* CGATT^GAAGaagCAATGGAA codon 141

- GAA6AA6CAATGGAA
- AAS
- GAA6AA6CAATGGAA
- GAA6CAATGGAA

Indel
*Example* ATTT^CCCTgggctGTA codon 59 Insert A

- ATTTCCCTGGGCTGTA
- A
- ATTTCCCTGGGCTGTA
- ATTTCCAGTA

Copyright Notice:

The Human Gene Mutation Database constitutes the intellectual property of Cardiff University. Any unauthorised copying, storage or distribution of this material without written permission from the curators would lead to copyright infringement with possible ensuing litigation. Copyright © Cardiff University 2011. All Rights Reserved.
REFERENCES


Cooper and Krawczak 1993

Cotton and Scriver, 1998

