Hum Genet (2001) 109 : 678–680
DOI 10.1007/s00439-001-0615-0

NOMENCLATURE RECOMMENDATIONS

Sue Povey · Ruth Lovering · Elspeth Bruford
Mathew Wright · Michael Lush · Hester Wain

The HUGO Gene Nomenclature Committee (HGNC)

Received: 6 September 2001 / Accepted: 6 September 2001 / Published online: 24 October 2001

HGNC – past, present, future

The need for standard nomenclature in human genetics was recognised as early as the 1960s, and in 1979 full guidelines for human gene nomenclature were presented at the Edinburgh Human Genome Meeting (HGM) and subsequently published (Shows et al. 1979). The current Chair of the Human Gene Nomenclature Committee, Sue Povey, was elected at the HGM meeting in Heidelberg in 1996. Since then, under the auspices of the international Human Genome Organisations and with the acronym HGNC, we continue to strike a compromise between the convenience and simplicity required for the everyday use of human gene nomenclature and the need for adequate definition of the concepts involved. Numerical identifiers are satisfactory for computers, but when humans need to talk about a gene they prefer to use a name. Increasingly journals are requesting approved gene nomenclature before publication, although more standardisation in this respect would make a significant contribution to the annotation of the human genome (Povey et al. 1997; White et al. 1998). A recent analysis of networks of human genes from 10 million MedLine records illustrates the ingenuity currently required to extract information from the literature (Jenssen et al. 2001).

The committee has grown from a single force (Dr Phyllis J. McAlpine) to the equivalent of five professional full-time staff, and operates through the Chair with key policy advice from an International Advisory Committee (IAC, http://www.gene.ucl.ac.uk/nomenclature/IAC.shtml). We also use a team of specialist advisors who provide support on specific gene family nomenclature issues (http://www.gene.ucl.ac.uk/nomenclature/advisors.html).

S. Povey (✉) · R. Lovering · E. Bruford · M. Wright · M. Lush
H. Wain
University College London, Wolfson House,
4 Stephenson Way, London, NW1 2HE, UK
e-mail: nome@galton.ucl.ac.uk,
Tel.: +44-20-76795027, Fax: +44-20-73873496

Regular nomenclature workshops are held, frequently to coincide with the annual meeting of the American Society of Human Genetics (ASHG) and the HGM, to ensure that we are approving gene names in line with the needs of the scientific community. Guidelines for human gene nomenclature were last published in 1997 (White et al. 1977) and are also available online. New guidelines will be published in 2002 and a draft version can be inspected at http://www.gene.ucl.ac.uk/nomenclature/guidelines/draft_2001.html. For details of previous and future workshops see http://www.gene.ucl.ac.uk/nomenclature/workshops.html.

HGNC – what it does and delivers

For each known human gene locus the HGNC approves a short-form abbreviation, known as a gene symbol, and also a longer and more descriptive name (see Table 1). All approved symbols are stored in Genew, the Human Gene Nomenclature Database (http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl). Each symbol is unique, and the committee ensures that each gene locus is only given one approved gene symbol. The approved symbols are included in secondary databases (LocusLink, Ensembl,

Table 1 Summary of human gene nomenclature as of August 2001

Total of 13316 approved gene symbols and 9337 literature aliases
Human symbols are:
Unique and must not clash with mouse nomenclature
Upper case, without punctuation
A combination of Latin letters and Arabic numerals
To obtain an approved gene symbol:
Visit http://www.gene.ucl.ac.uk/nomenclature/ and complete a submission form
Or e-mail: nome@galton.ucl.ac.uk if you have several genes
You must provide the gene sequence
We will check that the gene is new to us
An approved name and symbol will be negotiated

OMIM, SWISS-PROT, HGMD, GDB, GENATLAS, GeneCards), where each symbol is unambiguously associated with a single gene. Use of approved symbols greatly increases the efficiency of electronic literature retrieval and our database does include all literature aliases for each approved symbol, which helps to clarify the identity of a gene. Frequent interaction with the mouse nomenclature committee (http://www.informatics.jax.org/mgihome/nomen/gene.shtml#genenom) ensures that for the vast majority of genes the same symbol (differing only in case) is used in the two species and usually also in other mammals.

The original aim was to give each gene a name indicating something about the function of the normal gene product and this is still done where possible, although in very many cases more is known about the sequence than its function. In preference, each symbol maintains parallel construction in different members of a gene family (functional or related by sequence). However, there is no attempt to include all known information about a gene in its name; by far the most important feature is that the nomenclature is unique.

Obtaining a name and symbol for a new gene

Individual new symbols are requested by scientists, journals and databases (e.g. RefSeq, OMIM, GDB, MGD), and groups of new symbols by those working on gene families, chromosome segments or whole chromosomes. We would like to encourage authors to submit their request for a new gene symbol via the web-based submission form (http://www.gene.ucl.ac.uk/nomenclature/submit.html), or to contact us directly by e-mail (nome@galton.ucl.ac.uk) if they have large data sets. The submission form involves entering contact details along with any other relevant information, including a suggested gene name and short-form gene symbol. We also require sequence data for each new gene submitted. The HGNC editors validate every record that is entered in this way by further analyses and database searching. As the human genome sequence analysis nears completion there is an increasing demand for the rapid approval of gene symbols. In all cases, considerable efforts are made to approve a symbol acceptable to workers in the field and, wherever possible, all interested parties are included in negotiations.

Although open discussion is preferable, many pre-publication findings are extremely sensitive. Therefore, it is possible to obtain an approved symbol in complete confidence; this, however, rules out broader discussion with other scientists in the field. Specialist advisors will only be consulted with the express permission of the submitter. See http://www.gene.ucl.ac.uk/nomenclature/information/confidentiality.shtml for the different levels of confidentiality offered.

Some problems

There is no shortage of problems to discuss in gene nomenclature. Two have been chosen for brief mention here.

Definition of a gene

From the nomenclature point of view, a gene is currently defined as a DNA segment that contributes to phenotype/function. In the absence of demonstrated function a gene may be characterised by sequence, transcription or homology.

In general, alternate transcripts are not regarded as separate genes. However, in a few cases, where very different products share a single exon and the community working in this field would like the products to be regarded as coming from separate genes, separate approved symbols have been given; examples include the protocadherins and the UDPG glucuronosyl-transferases. Antisense transcripts and transcripts deriving from within an intron of another gene may also receive separate designations, e.g. IGF2AS and COPG2IT1. There is also at least one example, CDKN2A, of shared exons read in different frames. In this case, the locus currently has one approved gene symbol, although, as it has two distinct transcripts, p16(INK4a) and p14(ARF), it could also be regarded as two genes. However, it must be emphasised that these examples are rare and each one is considered separately.

Pseudogenes present a particular problem of definition. Although the assumption is that a pseudogene is non-functional and therefore does not qualify as a gene, it can be useful to know of their existence as they can interfere with mutation screening in a functional gene. Furthermore, there are genes (e.g. ADAM1) that occur as a pseudogene in human, but as a functional gene in mouse (see http://www.people.virginia.edu/~jag6n/Table_of_the_ADAMs.html). Therefore, the current policy is to assign a pseudogene the next number in the relevant gene symbol series. The pseudogenes are distinguished from the genes by the presence of “pseudogene” in the gene description and a terminal ‘P’, which can be removed if necessary, e.g. OR5B12P “olfactory receptor, family 5, subfamily B, member 12, pseudogene”.

Stability of names

It frequently happens that the first attribute of a gene, which is recognised and used as the basis of a name, is ultimately found not to be an essential aspect of the function of the gene product. In general, unless the original name is misleading, single gene names are not altered to accommodate new information, since stability of names is very desirable. However, the overall goal is to produce a nomenclature that is useful, and for many gene families the community working on these genes has initiated and agreed a new, more rational, nomenclature. Two such examples

are the protease inhibitors (now with the root symbol SERPIN) and the very large superfamily of ABC transporters. In both of these cases there was pressure from the biochemical community for sweeping change to every symbol, but some existing symbols had become so well known among geneticists that HGNC decided to retain a few of these with the new nomenclature as an alias, at least temporarily. Thus CFTR (cystic fibrosis transmembrane regulator) has been retained, although in the new ABC classification it is the alias of ABCC7. Among the serpins the name AGT [angiotensinogen serine (or cysteine) proteinase inhibitor] has been retained, although it has an alias of SERPINA8. This is, of course, not without controversy. Further examples of families can be seen on the family resources page at http://www.gene.ucl.ac.uk/nomenclature/genefamily.shtml.

Another controversial topic is the handling of a gene symbol that describes a phenotype following the cloning of the causative gene. The logical next step is to replace the phenotype-derived symbol with one describing the normal function of the protein. The name of the disease is often included in the new full descriptive name of the gene, to facilitate retrieval, and the disease symbol added as an alias. More frequently, however, at the moment of publication all that is known about the gene function is that it prevents the disease in question. Many authors feel that to change the symbol to some invented protein name is only a temporary, and not particularly informative, solution, and prefer to keep the phenotype nomenclature until more is known.

Acknowledgements HGNC is jointly funded by the UK Medical Research Council (40%) and the US National Institutes of Health N01-LM-9-3533 (60%).

References

Jenssen TK, Laegreid A, Komorowski J, Hovig E (2001) A literature network of human genes for high-throughput analysis of gene expression. Nat Genet 28:21–8

Povey S, White J, Nahmias J, Wain H (1997) Problems of nomenclature. Nature 390:329

Shows TB, Alper CA, Bootsma D, Dorf M, Douglas T, Huisman T, Kit S, Klinger HP, Kozak C, Lalley PA, Lindsley D, McAlpine PJ, McDougall JK, Meera Khan P, Meisler M, Morton NE, Opitz JM, Partridge CW, Payne R, Roderick TH, Rubinstein P, Ruddle FH, Shaw M, Spranger JW, Weiss K (1979). International system for human gene nomenclature (1979) ISGN. Cytogenet Cell Genet 25:96–116

White J, Maltais L, Nebert D (1998) Networking nomenclature. Nat Genet 18:209

White JA, McAlpine PJ, Antonarakis S, Cann H, Eppig JT, Frazer K, Frezal J, Lancet D, Nahmias J, Pearson P, Peters J, Scott A, Scott H, Spurr N, Talbot C Jr, Povey S (1997) Guidelines for human gene nomenclature. HUGO Nomenclature Committee. Genomics 45:468–71