Update on human genome completion and annotations: Gene nomenclature

Daniel W. Nebert[1]* *and Hester M. Wain*[2]

[1]Department of Environmental Health and Center for Environmental Genetics (CEG), University of Cincinnati Medical Center, Cincinnati, OH 45267–0056, USA
[2]HUGO Gene Nomenclature Committee, Department of Biology, University College London, Wolfson House, London, NW1 2HE, UK
**Correspondence to*: Tel: +1 513 558 4347; Fax: + 1 513 558 3562; E-mail: dan.nebert@uc.edu

Date received (in revised form): 31st July 2003

Abstract

Why is agreeing on one particular name for each gene important? As one genome after another becomes sequenced, it is imperative to consider the complexity of genes, genetic architecture, gene expression, gene–gene and gene–product interactions and evolutionary relatedness across species. To agree on a particular gene name not only makes one's own research easier, but will also be helpful to the present generation, as well as future generations, of graduate students and postdoctoral fellows who are about to enter genomics research.

Keywords: *human gene nomenclature, LocusLink, Genew, mouse genome database (MGD), cyclooxygenases-1 and -2, fatty acid synthase (long-chain and short-chain), NADPH-cytochrome P450 oxidoreductase*

Introduction

As mainframe computers were developed in the 1940s and personal computers designed some years later, a short-sighted decision was made, early on, to designate the date in six numbers (eg 25-07-58). It was not until some time later that computer scientists realised that, when the date changed from 31-12-99 to 01-01-00 at the turn of the century, almost all computers worldwide would 'think' the year was now 1900 instead of 2000. This 'Y2K bug' hysteria turned out not to be as disruptive as anticipated, but the important moral of the story is that we, as geneticists and genomicists, should constantly be thinking about the future. We need to plan ahead (Figure 1) with the genome, looking to the future in an orderly, organised fashion. With the amount of bioinformatics information being thrown at us at alarmingly increasing rates, the least we can do for the present generation, as well as future generations of scientists, is to agree on a systematic gene nomenclature system — not only for the human genome, but ideally for all eukaryotic and prokaryotic genomes.

Background and history

As early as the 1960s, problems with nomenclature in human genetics were recognised. The complete guidelines for human gene nomenclature were first presented at the 1979 Edinburgh Human Genome Meeting.[1] More recent updates of the guidelines have now appeared.[2–5] The HUGO Gene Nomenclature Committee (HGNC) began with one very dynamic person, Phyllis McAlpine, and, since the mid-1990s, has grown to the equivalent of five full-time editorial staff. A major goal of the HGNC is 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. The human genome sequence is now essentially complete; because this is public information, scientists worldwide expect a user-friendly standardised system for the efficient identification of genes of interest. Approved gene symbols, therefore, help to provide one of the keys to unlocking the secrets of the human genomic sequence by ensuring that there are unique gene symbols identified in all the human genome browsers and other databases (http://www.gene.ucl.ac.uk/nomenclature/).

Genes

For the purposes of nomenclature, a *gene* is defined as 'a DNA segment that is responsible for a functional gene product and/or contributes to phenotype; in the absence of demonstrated function, a gene may be characterised by sequence, transcription or homology'. The DNA segment corresponding to 'a gene' should extend from the 5′-most regulatory element to the 3′-most regulatory element flanking the actual transcribed region and controlling expression;[6] this means that one gene can overlap another, one can lie inside another and one can be

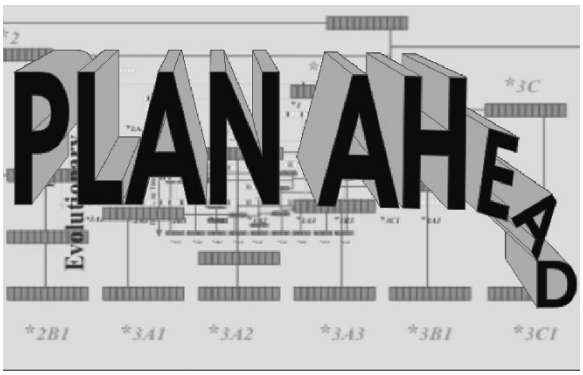

Figure 1. An illustration of what might happen if one does not plan for the future. We greatly appreciate the graphics and artwork of Dr Marian L. Miller (Department of Environmental Health, University of Cincinnati Medical Center)

located on the opposite DNA strand from the other. By contrast, a genetic *locus* is not synonymous with a gene, but rather refers to a map position. There may exist several genes within one locus, or cluster, located on a chromosome. There are even guidelines for naming the small-interfering RNA (siRNA) genes, of which there are an estimated 1,000 to 2,000 throughout the human genome.[7] These DNA segments, only about 21 base pairs in length, are considered genes, since they clearly encode a functional product that can alter the phenotype by silencing their mRNA targets (ie 'RNA interference'; RNAi).

Gene families

Over the past two decades of genomics/bioinformatics, it has become clear that, once a gene evolved 1 to >3 billion years ago from a random DNA sequence, it took much less energy for gene-duplication and crossing-over events to form new genes than to create a new gene from scratch. Thus, all of the newly created genes exhibit homology to their ancestral gene — leading to the formation of modern-day gene families and superfamilies. This concept of evolutionary divergence from a common ancestor is consistent with the concept of clusters of orthologous protein groups[8] and the prediction[9] that all present-day genes are likely to be derived from a 'core' of between 7,000 and 12,000 genes that existed more than 500 million years ago; numerous functional genes that have evolved since that time generally represent duplication or cross-over events. This predicted range of the number of ancestral genes is also in agreement with a report[10] in which all *Drosophila melanogaster* proteins could be assigned to 8,065 distinct core families, about 5,000 of which are shared with *Caenorhabditis elegans*.

Homologous genes can be identified more readily if they are designated with a stem (or root) symbol. A root symbol is very much encouraged by the HGNC as the basis for a hierarchical series of genes (eg for the *ABC* family, subfamily *A*, *ABCA1*, *ABCA2*, *ABCA3*, *ABCA4*) that are either the result of evolutionary divergence of an ancient ancestral gene, or have conserved functions — via pathways, interactions or protein domains. Such a root symbol allows the easy identification of other related members in both database searches and the literature.

Homologous regions of 15–25 per cent of nucleotides or amino acids can be detected by the various alignment programs, and denote divergence from an ancestral gene. A small almost-invariant DNA motif or protein domain — functioning as an enzyme active-site, cofactor docking site or ligand-binding site — is further evidence of divergence from an ancestral gene. One of the earliest examples of this nomenclature approach for homologous genes was the cytochrome P450 (*CYP*) gene superfamily, in which it was agreed upon that approximately 40 per cent or more amino acid similarity allows two members to be placed in the same family and about 55 per cent or more similarity allows two members to be assigned to the same subfamily.[11,12] These cut-off values follow the original recommendations of Margaret Dayhoff.[13] Several dozen additional gene superfamilies and large gene families have since followed this same format.[14]

Orthologous genes

Homologues represent genes (that can be in numerous species) arising from a common ancestor. *Orthologues* are genes in different species that evolved from a common ancestral gene by speciation. By contrast, *paralogues* are genes related by duplication within a genome.[8] There is a need for consistent nomenclature between orthologous genes in different species, and this is particularly well coordinated for human and mouse genes. This is achieved through editorial work with the Mouse Genomic Nomenclature Committee ((MGNC); at The Jackson Laboratory, Bar Harbor, ME). For any new mouse or human gene symbol request, the DNA sequence is analysed by BLAST to try to identify the orthologous gene. Following this, an appropriate symbol is suggested that is unique in both species. There is a constant exchange of files between the mouse and human nomenclature committees, ensuring that even when there is no identified orthologue, consistent symbols are still reserved.

Further cross-species nomenclature coordination has also recently been encouraged by the curators of the ARK database (http://www.thearkdb.org/), who state that '... animal gene nomenclature should follow the rules for human gene nomenclature, including the use of identical symbols for homologous genes and the reservation of human symbols for yet unidentified animals' genes'. It is not the guidelines *per se* that are important, however, but the reasoning behind them. The guidelines are an aid to naming human genes, the only rule that can never be broken is that every symbol must be unique.

Gene products (proteins)

The standard nomenclature of proteins is beyond the scope of this review. The history of enzyme nomenclature is reviewed online at http://www.chem.qmul.ac.uk/iubmb/enzyme/history.html, and is an ongoing goal of the International Union of Pure and Applied Chemistry and the International Union of Biochemistry and Molecular Biology (IUPAC-IUBMB).[15] Examples of these abbreviations for classes of enzymes include enzyme commission (EC) numbers, for example, transaminases (EC 2.6.1), kinases (EC 2.7.1) and nucleases (EC 3.1.11-31). The names of genes coding for enzymes should be based on those recommended by the Nomenclature Committee of the IUBMB, for example, FPGS is an abbreviation of 'folylpolyglutamate synthase'. These can be found at http://ca.expasy.org/enzyme/.

Why do gene names matter?

Recently, David Botstein–one of the grandfathers of the Human Genome Project–was awarded the US$150,000 Peter Gruber Foundation Prize in Genetics during the XIXth International Congress of Genetics (July, 2003) in Melbourne, Australia. He spoke about the need for standardised gene nomenclature in his address to delegates at this meeting, stating that: 'Biologists would rather share a toothbrush than a gene's name'. Fruit-fly geneticists have always used more whimsical gene names such as 'daughterless, groucho, hedgehog, mad (mothers against decapentaplegic), plutonium and saxophone'. These quaint names, however, do have potentially serious consequences, because fly genes generally have homologues in the human and, when trying to maintain consistency with this nomenclature, functional information can be lost. In addition, explaining to the parent of a child with a variation in the 'sonic hedgehog' gene (one of three *Drosophila* orthologues in humans) may well be more complicated because of the gene's humorous name.

The symbols reflecting the fruit fly gene names are currently undergoing close scrutiny, because it has been realised that they are not all obviously unique (see http://flystocks.bio.indiana.edu/Caps.htm). Historically, a fruit fly gene symbol was capitalised if the first characterised allele was dominant. Unfortunately, this has meant that some symbols differ only in upper- versus lower-case, for example, Delta (*Dl*) versus dorsal (*dl*). This has resulted in some confusion in publications, databases and cross-species nomenclature assignments, and it has now been proposed to eliminate all capital letters in *Drosophila* gene designations (see http://flystocks.bio.indiana.edu/caps.htm).

Botstein urged a universal genetic language, in which genes would be identified by function, rather than by humorous or obtuse references to movie stars, rock music groups and the like. Such a system would allow computers to cross-reference species databases, build information about patterns of gene activity in complex networks and build hypotheses about as-yet undiscovered cellular systems. 'We are going to have to live with the computer', said Botstein. 'What we do with computers has to become more sophisticated, because the combinatorial complexities are unreasonably large. The challenge ahead is to understand how genes interact — and to present the data in such a way that ordinary people, not just statisticians, can understand them.'

Function is, of course, very important, but this is usually related to the gene product, rather than the gene. One gene may have many functions, but it does only have one location in the genome. It is the unique naming of this piece of DNA (the gene) that is needed in order to enable all of the functional information to be collated correctly, thus allowing us to move on to the more fascinating discoveries yet to be made in the genome.

Who needs official gene symbols?

For each unequivocally established human gene, a name and symbol (short-form abbreviation) are approved by the HGNC. Each symbol is unique, and each gene will have only one approved gene symbol.[5] It is important to provide a unique representation for each gene, so that colleagues can talk with one another about any particular gene sequence or family. Having one unique symbol also facilitates electronic data retrieval from publications and databases. Furthermore, it is important that each symbol preferably maintains parallel construction in, for example, different members of a gene family.

The HGNC should be contacted as quickly as possible with new members of gene families, because some symbols may be reserved in their database. Obtaining a gene symbol before publication will avoid any possible conflicts with existing symbols and will ensure that the gene is promptly recorded in the LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink/) and Genew (http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl) databases.

As of September 2003, there are 16,765 approved human gene symbols — meaning that the goal of naming all genes in the human genome is somewhere between one-third and perhaps more than one-half completed. Individual new symbols are requested not only by scientists but also by an increasing number of journals (eg *American Journal of Human Genetics; Animal Genetics; Annals of Human Genetics; Cytogenetic & Genomic Research; Genes, Chromosomes & Cancer; Genomics; Human Mutation; Lancet; Molecular Therapy; Nature Genetics; Radiation Research*). Publication of an article in any of these journals will not proceed until the gene under study has been officially named. This also ensures that all newly released symbols are immediately cross-indexed with other databases (eg LocusLink, RefSeq, OMIM and MGD), which increases the potential accessibility and impact of these genes in the databases.

It has been suggested that the nomenclature process could be automated, and recent publications[16] certainly indicate that this may be a viable possibility. Whereas automated assignment of gene names and symbols may well give highly systematic classifications, however, this does not always allow for the inclusion of the most useful, or indeed memorable, information.

Examples of searching for, or submitting, a gene symbol

Table I summarises the steps one is urged to take to ensure proper nomenclature of any gene. Three examples will be given here, to illustrate further how and why one should strive for a standardised gene nomenclature system. In these examples, the focus is on using the gene names as search terms, rather than comparing a DNA or protein sequence that has just been determined, by searching via BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). The three examples below comprise *genes encoding enzymes*; future updates will focus on the nomenclature of other types of gene products and DNA motifs.

Cyclooxygenase. The procedure for writing a review on prostaglandin G/H synthase-1 and -2, also known as cyclooxygenase-1 and -2, commonly nicknamed in many journals as 'COX-1' and 'COX-2' is set out below. These enzymes, which are targets of non-steroidal anti-inflammatory drugs, are pivotal in converting arachidonic acid to prostaglandins G and H, pathways that are associated with inflammatory processes, pain, rheumatoid disease, atherosclerosis, stroke, gastrointestinal tract injury and repair, oxidative stress and various cancers.[17,18] In order to determine the correct approved symbol, the first approach is to search LocusLink (for all organisms) using 'prostaglandin g synthase' or 'prostaglandin h synthase' as the full names. This will retrieve ten and 12 loci, respectively, four of which in both cases include the approved symbols for human, *PTGS1* and *PTGS2*, and the mouse and rat *Ptgs2*. Searching LocusLink with 'cyclooxygenase' will retrieve 49 hits — listed alphabetically — again, four of which include the human *PTGS1* and *PTGS2* and the mouse and rat *Ptgs2* gene records. Searching LocusLink for 'cox1', one finds three loci, which include the human *PTGS1*, the rat *Ptgs1* and the rat mitochondrial *Mt-Co1*. Searching LocusLink for 'cox2', one finds seven hits, three of which are human *PTGS2* and mouse and rat *Ptgs2*; rat mitochondrial *Mt-Co2* is also registered.

Searching Genew using 'prostaglandin g synthase' or 'prostaglandin h synthase' as the full names, however, does not retrieve any gene records. Searching Genew with 'cyclooxygenase', one can confirm that the human gene symbols are *PTGS1* and *PTGS2*, their approved names are prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase) (M59979; NM_000962) and prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase) (D28235; NM_000963), located on human chromosomes 9q32-q33.3 and 1q25.2-q25.3, respectively; aliases for *PTGS1* include COX1, PGHS-1 and PTGHS and for *PTGS2* include COX2 only. It can be seen that there is some confusion over the use of other aliases such as COX, because searching Genew for all records that begin with COX retrieves 46 records, most of which refer to the cytochrome *c* oxidase subunit genes. Thus, to use 'COX' to refer to the cyclooxygenase-1 and -2 enzymes that one is studying would not be helpful to the community, as this would only bring further confusion to the literature.

Fatty acid synthases. Fatty acid synthase, one of the principal lipogenic enzymes, converts dietary calories into a storage form of energy.[19] Fatty acids themselves can also act as

Table 1. The HUGO Gene Nomenclature Committee (HGNC) checklist for deciding on a new human gene symbol

1	Establish that the phenotype is genetic (ie inherited in a Mendelian fashion) or is a cloned gene sequence.
2	Search the LocusLink, Genew and mouse genome databases (MGD) to ensure that the entry is not already there.
3	Formulate possible combinations of letters for the symbol (check guidelines at URL http://www.gene.ucl.ac.uk/nomenclature/guidelines.html).
4	Search LocusLink, Genew and MGD to see which combinations of letters/numbers are available.
5	Check ordered lists of human and mouse symbols to ensure that your proposed new symbol will not interrupt an existing 'family' of gene symbols.
6	Be certain that your proposed new symbol is not in common use elsewhere (searches of PubMed should help to identify such problems, abbreviations).
7	Check similar entries in LocusLink to ensure that parallel construction is maintained.
8	Submit the proposed gene symbol to the HGNC for confirmation (and entry into LocusLink, if approved); a form is available on Genew.

See URL http://www.gene.ucl.ac.uk/nomenclature/information/check.shtml.

signals that regulate gene expression, and fatty acid synthase is downregulated by polyunsaturated fatty acids.[20] Let us imagine you have isolated the cDNA for human liver fatty acid synthase and are considering naming your gene *FAS*. Searching LocusLink using 'fatty acid synthase', 58 loci are found — including human *FASN*, mouse and rat *Fasn*, and fruit fly *Fas*. Searching LocusLink using the symbol 'fas', produces 149 hits, which include human *FASN* and mouse *Fasn*. *FASN* is located on chromosome 17q25 and has a GenBank accession number of NM_004104; therefore, your gene already has this approved symbol. You may feel, however, that your initial choice of *FAS* is more appropriate, in which case you should contact the HGNC and argue your case as to why you believe that 'FAS' is a better symbol for this gene than *FASN*.

Let us then suppose that you have characterised genes coding for a novel cytosolic *short-chain* fatty acid synthase and a novel cytosolic *long-chain* fatty acid synthase. Searching LocusLink using the full names, you find five loci for short-chain fatty acid synthase, which include mouse and rat *Fasn*, and 12 loci for long-chain fatty acid synthase, which include human *FASN* and three 'fatty acid-coenzyme A ligases, long-chain' genes (*FACL1*, *FACL3* and *FACL4*) representing a small family. Searching LocusLink using the symbols 'fascs', 'falcs', 'facs', 'fass', 'fasc' or 'falc', you find zero hits, except for 'facs', which gives you human *FACL2* and mouse and rat *Facl2*. Searching Genew using the full names, you find zero hits referring to either of these enzymes. Searching Genew using the symbol 'fascs', 'falcs', 'facs', 'fass', 'fasc' or 'falc' will also generate zero hits. Your conclusion would be that there is a root symbol for at least four human fatty acid-coenzyme A long-chain genes (members of an evolutionarily related family), but nothing for your short-chain fatty acid synthase.

The next step would be to contact the HGNC to make certain nothing has been 'reserved', concerning the description of this gene family. Once this has been determined, you may be encouraged to get in touch with several major players in the short-chain fatty acid field, and others in the long-chain fatty acid field, and try to come to a consensus agreement (with HGNC involvement) on symbol roots for naming the gene or genes in the short-chain fatty acid synthase family. Because *FACL* is the root symbol for fatty acid long-chain synthase (or ligase), 'FACS' would be among the most reasonable and consistent roots for your fatty acid short-chain synthase gene. In LocusLink, there is also human *ECHS1*, the gene for a 'mitochondrial enoyl-coenzyme A hydratase, short-chain', which you must confirm is not the new gene you have identified. *FACS1* thus remains the most reasonable proposed name — especially if other colleagues in the field are in agreement with your suggestion.

NADPH-cytochrome P450 reductase. This enzyme transfers the first electron from NADPH to the various cytochrome P450 (CYP) monooxygenases.[21,22] But what if a review is to be written on this topic? Searching LocusLink using the full name, 'nadph cytochrome p450 oxidoreductase' (or 'reductase' without 'oxido'), there are nine and 11 hits, respectively, including human *POR*, mouse *Por* and fruit fly *Cpr*. Including a hyphen (nadph-cytochrome p450 oxidoreductase) yields only two — human *POR* and fruit fly *Cpr*. Searching LocusLink using the older name 'nadph cytochrome *c* oxidoreductase' (or 'reductase'), curiously, produces only an NADPH oxidase plus the human (*TP53*) and mouse (*Trp53*) tumour protein-53. Searching with the term 'p450 oxido-reductase', one finds human *POR* and mouse and rat *Por*, but also more than 90 hits for the *CYP* genes. Searching LocusLink with 'por', one finds four hits — human *POR*, mouse and rat *Por*, and fruit fly porcupine *Por*. The human *POR* symbol is identified in LocusLink as the 'Official Gene Symbol and Name (HGNC)'.

Searching Genew using the full names, 'nadph cytochrome p450 oxidoreductase' (or 'reductase'), 'nadph cytochrome c oxidoreductase' (or 'reductase'), 'p450 oxidoreductase' (or 'reductase'), 'cytochrome c oxidoreductase' (or 'reductase'), or 'p450 (cytochrome) oxidoreductase' (or 'reductase') however, retrieves no data, although searching Genew with 'por', one finds a 'hit' for the gene named 'P450 (cytochrome) oxido-reductase' located on human chromosome 7q11.2 with an alias of 'CYPOR'. This shows that Genew is missing some relevant aliases, because the full-name query 'P450 (cytochrome) oxidoreductase' does not lead you to *POR* as the gene name, whereas the 'por' symbol does lead you to the full name. By contrast, starting your search with LocusLink sends you directly to the human *POR* and rodent *Por* genes. This minor glitch in Genew should be reported to the HGNC as soon as possible.

Designation of genus and species

It might be noted that LocusLink includes a shorthand abbreviation of the genus and species for each gene. For designating genus and species, HGNC guidelines suggest using the five-letter Swiss-Prot codes found at http://www.expasy.ch/cgi-bin/speclist. Thus, *Bos taurus* (cow) is 'BOVIN' and *Bacillus thuringiensis* is 'BACUT'. These codes are for use in publications only and should not be incorporated as part of the gene symbol. The five-letter species designation is added as a prefix to the gene symbol in parentheses. For example, HUMAN signifies *Homo sapiens* and MOUSE signifies *Mus musculus*, for example the human genes (HUMAN)*ABCA1* and (HUMAN)*SLC13A2* are orthologues of the mouse genes (MOUSE)*Abca1* and (MOUSE)*Slc13a2*.

Conclusions

For each and every gene, one is strongly advised to check the websites cited in this article and write a peer-reviewed paper or invited review with the correct gene nomenclature. Human

gene symbols usually consist of a combination of upper-case letters and Arabic numerals (eg *ABCC13*). Superscripts, subscripts and Greek symbols cannot be used; it is recommended that gene names should not have more than six letters/numbers, and that the gene name should not start with a number. Mouse gene symbols (eg *Abcc13*) usually have only the first letter capitalised. It is recommended that genes and allele symbols are underlined in the manuscript and italicised in print; protein symbols should be represented in standard upper-case fonts. Italics need not be used in gene catalogues. To distinguish between messenger (mRNA), genomic DNA (gDNA) and complementary (cDNA), the relevant prefix should be written in parentheses adjoining the italicised symbol — (mRNA)*RBP1*, (gDNA)*RBP1*, (cDNA)*RBP1* (http://www.gene.ucl.ac.uk/nomenclature/guidelines.html). Using approved gene nomenclature greatly enhances the ability of scientists to communicate and correctly associate information from publications and databases. With increasing numbers of sequenced genomes comes increased interpretation and annotation, and this will be very much simplified for geneticists and genomicists alike if everyone uses approved gene symbols.

Acknowledgments

The writing of this article was funded, in part, by NIH grant P30 ES06096 (D.W.N.).

References

1. Shows, T.B., Alper, C.A., Bootsma, D. *et al.* (1979), 'International system for human gene nomenclature 1979 (ISGN 1979)', *Cytogenet. Cell Genet.* Vol. 25, pp. 96–116.
2. Shows, T.B., McAlpine, P.J., Boucheix, C. *et al.* (1987), 'Guidelines for human gene nomenclature. An international system for human gene nomenclature (ISGN, 1987)', *Cytogenet. Cell Genet.* Vol. 46, pp. 11–28.
3. McAlpine, P.J. (1995), 'International system for human gene nomenclature', *Trends Genet.* Vol. 11, pp. 39–42.
4. White, J.A., McAlpine, P.J., Antonarakis, S. *et al.* (1997), 'Guidelines for human gene nomenclature. HUGO Nomenclature Committee', *Genomics* Vol. 45, pp. 468–471.
5. Wain, H.M., Bruford, E.A., Lovering, R.C., *et al.* (2002), 'Guidelines for human gene nomenclature', *Genomics* Vol. 79, pp. 464–470.
6. Nebert, D.W. (2002), 'Proposal for an allele nomenclature system based on the evolutionary divergence of haplotypes', *Hum. Mutat.* Vol. 20, pp. 463–472.
7. Scherr, M., Morgan, M.A. and Eder, M. (2003), 'Gene silencing mediated by small interfering RNAs in mammalian cells', *Curr. Med. Chem.* Vol. 10, pp. 245–256.
8. Tatusov, R.L., Koonin, E.V. and Lipman, D.J. (1997), 'A genomic perspective on protein families', *Science* Vol. 278, pp. 631–637.
9. White, J.A., Maltais, L.J. and Nebert, D.W. (1998), 'An increasingly urgent need for standardized gene nomenclature', *Nat. Genet.* Vol. 18(3), p. 209. [Internet]. Available from http://www.genetics.nature.com/nomen/nomen_article.html.
10. Rubin, G.M., Yandell, M.D., Wortman, J.R. *et al.* (2000), 'Comparative genomics of the eukaryotes', *Science* Vol. 287, pp. 2204–2215.
11. Nebert, D.W., Adesnik, M., Coon, M.J. *et al.* (1987), 'The P450 gene superfamily. Recommended nomenclature', *DNA* Vol. 6, pp. 1–11.
12. Nelson, D.R., Koymans, L., Kamataki, T. *et al.* (1996), 'Cytochrome P450 superfamily: update on new sequences, gene mapping, accession numbers, and nomenclature', *Pharmacogenetics* Vol. 6, pp. 1–42.
13. Dayhoff, M.O. (1987), Atlas of Protein Sequence and Structure (Natl. Biomed. Res. Foundation, Washington DC) Vol. 5/3, pp. 351–352.
14. Nuclear Receptors Nomenclature Committee (1999), 'A unified nomenclature system for the nuclear receptor superfamily', *Cell* Vol. 97, pp. 161–163.
15. Anon (1999), 'IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and Nomenclature Committee of IUBMB (NC-IUBMB), newsletter', *Eur. J. Biochem.* Vol. 264, pp. 607–609.
16. Kasukawa, T., Furuno, M., Nikaido, I. *et al.* (2003), 'Development and evaluation of an automated annotation pipeline and cDNA annotation system', *Genome Res.* Vol. 13, pp. 1542–1551.
17. Koki, A., Khan, N.K., Woerner, B.M. *et al.* (2002), 'Cyclooxygenase-2 in human pathological disease', *Adv. Exp. Med. Biol.* Vol. 507, pp. 177–184.
18. Whelton, A. (2002), 'COX-2-specific inhibitors and the kidney: Effect on hypertension and oedema', *J. Hypertens.* Vol. 20(Suppl. 6), pp. S31–S35.
19. Schweizer, M., Roder, K., Zhang, L. and Wolf, S.S. (2002), 'Transcription factors acting on the promoter of the rat fatty acid synthase gene', *Biochem. Soc. Trans.* Vol. 30, pp. 1070–1072.
20. Duplus, E. and Forest, C. (2002), 'Is there a single mechanism for fatty acid regulation of gene transcription?', *Biochem. Pharmacol.* Vol. 64, pp. 893–901.
21. Sheweita, S.A. (2000), 'Drug-metabolizing enzymes: mechanisms and functions', *Curr. Drug Metab.* Vol. 1, pp. 107–132.
22. Hlavica, P. and Lewis, D.F. (2001), 'Allosteric phenomena in cytochrome P450-catalyzed monooxygenations', *Eur. J. Biochem.* Vol. 268, pp. 4817–4832.