

Correspondence

Smelling of roses?

Sue Povey and Hester Wain

A response to **What's in a name?** By Gregory Petsko, *Genome Biology* 2002, **3**:comment 1005.1-1005.2.

Address: HUGO Gene Nomenclature Committee, Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK.

Correspondence: Sue Povey. E-mail: nome@galton.ucl.ac.uk

Published: 31 May 2002

Genome Biology 2002, **3**(6):interactions1003.1-1003.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/6/interactions/1003>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Gregory Petsko is right, of course, in pointing out the chaos in the literature and the barriers to communication caused by free-for-all naming of gene products [1], and indeed follows on a line of broadly similar but sometimes less entertaining articles in other august journals [2-8]. A few groups (for example, [7-12]) have even tried to apply the various solutions they proposed. Here, we write about a specific part of the topic, carefully avoided by Petsko: the naming of those old-fashioned objects known as genes.

Although some of our correspondents describe in no uncertain terms our unsuitability for the job, the attempt to ensure that for each human gene there is one name and one standard abbreviation (usually known as a symbol) has occupied the Human Genome Organisation (HUGO) gene nomenclature committee [13] since 1979. There is a positive side to this endeavor. Currently we have 14,427 'approved' human gene names and symbols; these symbols are used in all the major secondary databases (LocusLink [14], Swiss-Prot [15], Genecards [16], The Genome Database (GDB) [17], Ensembl [18], and GenAtlas [19]) and are almost entirely coordinated with the symbols for equivalent genes in the mouse. You won't like every symbol (neither do we)

but they are at least all unique, and wherever humanly possible they have been settled by negotiation. The pursuit of unique standard gene symbols has been championed by *Nature Genetics* [8,20] and *Genomics* [21,22], and indeed most journals primarily concerned with human genetics do now encourage or insist upon pre-publication agreement of a unique name with the HUGO gene nomenclature committee. This can be totally confidential if required. If you believe that one gene should have one name please contact us before you publish (see [13]); if you see mistakes in our database, please tell us.

A brief inspection of many high-profile journals shows that the battle is not yet won. For example, in September 2001 the same gene was introduced in *Nature* as *Mal* [23] and in *Nature Immunology* as *TIRAP* [24], and recently a paper in *PNAS* [25] describing many defensin genes referred to *Defb19* (mouse) as the ortholog of *DEFB17* (human) and *DEFB19* (human) as the ortholog of *Defb24* (mouse). There is of course often genuine difficulty in choosing a name. In the dark ages, when there was a belief in one gene:one polypeptide chain - long before we knew that glucose-phosphate isomerase doubles

as neuroleukin [26,27] - it was decided to name genes after the function of the normal gene product. This is still the ideal naming strategy in cases for which it is applicable. At the time a gene needs a name, however, which is when someone first wants to talk about it, the information available is most often some sequence similarity to a known gene. If the best information is similarity to a fly gene, the name often refers to this, the *hedgehog* gene family being one example [28]. In fact, *Drosophila melanogaster* only has one *hedgehog* gene; *indian hedgehog*, *desert hedgehog* and *sonic hedgehog* are examples of human gene names [13,29] (belying Petsko's charge of lack of imagination, but perhaps not beyond criticism in other respects).

As more information becomes available, there is frequently discussion about changing the approved gene name, but it is impossible to encapsulate all information about a gene within its name. The most satisfactory solution is often to wait until a gene family has been defined and then for the community to propose a revised nomenclature. Some of these nomenclature problems remain unresolved for many years. One such example, the question of whether olfactory receptor genes (many of them pseudogenes) should be

named from their clustered positions on the genome or from sequence relationships [30-32], has strong protagonists on both sides but, at least so far, has been debated without personal abuse. Anyone attempting to reconcile different views of genes or gene products must be prepared for robust exchanges of a nature that one of us (S.P.) has not previously encountered in 30 years of primary research, even at its most competitive.

It is excellent that the need for a common currency in the language of genes and gene products is now recognized. Do not underestimate the task, however. And when you have explained at a meeting that rather than compete with the pharmaceutical industry in high-throughput genotyping you have decided to sort out names for all human genes, people will still ask you 'But what do you actually work on?' We may soon have a vacancy for another post-doctoral scientist in our group. Would you like to apply?

Acknowledgements

The work of the HUGO Gene Nomenclature Committee is supported by NIH contract N01-LM-9-3533 (60%) and by the UK Medical Research Council (40%).

References

- Petsko G: **What's in a name?** *Genome Biol* 2002, **3**:comment1005.1-1005.2.
- Editorial: **Obstacles of nomenclature.** *Nature* 1997, **389**:1.
- Editorial: **Wanted: a new order in protein nomenclature.** *Nature* 1999, **401**:411.
- Judson HF: **Talking about the genome.** *Nature* 2001, **409**:769.
- Pearson H: **Biology's name game.** *Nature* 2001, **411**:631-632.
- Heilbron JL: **Coming to terms.** *Nature* 2002, **415**:585.
- Williams N: **How to get databases talking the same language.** *Science* 1997, **275**:301-302.
- Editorial: **You say ptO, I say Pto.** *Nat Genet* 1998, **18**:89-90.
- Whyte BJ: **Problems of nomenclature.** *Nature* 1997, **390**:329.
- Lonsdale D: **Nomenclature regulation.** *Nature* 1998, **391**:118.
- Maltais LJ, Jackson I: **Sequencing challenge.** *Nature* 1999, **402**:347.
- White J, Wain H, Bruford E, Povey S: **Promoting a standard nomenclature for genes and proteins.** *Nature* 1999, **402**:347.
- HUGO gene nomenclature committee** [<http://www.gene.ucl.ac.uk/>]
- LocusLink** [<http://www.ncbi.nlm.nih.gov/LocusLink/>]
- Swiss-Prot** [<http://www.ebi.ac.uk/swissprot/>]
- Genecards** [<http://bioinformatics.weizmann.ac.il/cards/>]
- The Genome Database** [<http://www.gdb.org/>]
- Ensembl** [<http://www.ensembl.org/>]
- GenAtlas** [<http://www.citi2.fr/GENATLAS/>]
- White J, Maltais L, Nebert D: **Networking nomenclature.** *Nat Genet* 1998, **18**:209.
- Povey S: **Guidelines for human gene nomenclature. Community nomenclature: standardized gene symbols.** *Genomics* 2002, **79**:463-463.
- Wain HM, Lovering RC, Bruford EA, Lush MJ, Wright MW, Povey S: **Guidelines for human gene nomenclature.** *Genomics* 2002, **79**:464-470.
- Fitzgerald KA, Palsson-McDermott EM, Bowie AG, Jefferies CA, Mansell AS, Brady G, Brint E, Dunne A, Gray P, Harte MT, et al.: **Mal (MyD88-adaptor-like) is required for Toll-like receptor-4 signal transduction.** *Nature* 2001, **413**:78-83.
- Horng T, Barton GM, Medzhitov R: **TIRAP: an adapter molecule in the Toll signaling pathway.** *Nat Immunol* 2001, **2**:835-841.
- Schutte BC, Mitros JP, Bartlett JA, Walters JD, Jia HP, Welsh MJ, Casavant TL, McCray PB Jr: **Discovery of five conserved β -defensin gene clusters using a computational search strategy.** *Proc Natl Acad Sci USA* 2002, **99**:2129-2133.
- Faik P, Walker JL, Redmill AA, Morgan MJ: **Mouse glucose-6-phosphate isomerase and neuroleukin have identical 3' sequences.** *Nature* 1988, **332**:455-457.
- Chaput M, Claes V, Portetelle D, Cludts I, Cravador A, Burny A, Gras H, Tartar A: **The neurotrophic factor neuroleukin is 90% homologous with phosphohexose isomerase.** *Nature* 1988, **332**:454-455.
- Mohler J, Vani, K: **Molecular organization and embryonic expression of the hedgehog gene involved in cell-cell communication in segmental patterning of *Drosophila*.** *Development* 1992, **115**:957-971.
- Echelard Y, Epstein DJ, St-Jacques B, Shen L, Mohler J, McMahon JA, McMahon AP: **Sonic hedgehog, a member of a family of putative signalling molecules, is implicated in the regulation of CNS polarity.** *Cell* 1993, **75**:1417-1430.
- Glusman G, Bahar A, Sharon D, Pilpel Y, White J, Lancet D: **The olfactory receptor gene superfamily: data mining, classification, and nomenclature.** *Mamm Genome* 2000, **11**:1016-1023.
- Zozulya S, Echeverri F, Nguyen T: **The human olfactory receptor repertoire.** *Genome Biol* 2001, **2**:research0018.1-0018.12.
- Younger RM, Amadou C, Bethel G, Ehlers A, Lindahl KF, Forbes S, Horton R, Milne S, Mungall AJ, Trowsdale J, et al.: **Characterization of clustered MHC-linked olfactory receptor genes in human and mouse.** *Genome Res* 2001, **11**:519-530.