

A database of locus-specific databases

To the Editor:

Complete and accurate information on genetic mutations and their effects on patients is essential for proper genetic healthcare. This realization led a group of prominent human geneticists to propose a federation of Locus-Specific Database (LSDB) curators as the best mode of collecting and curating accurate lists of mutations¹. This subsequently led to the formation of the Mutation Database Initiative (MDI)² under the auspices of the Human Genome Organisation (HUGO); MDI later became a society now known as the Human Genome Variation Society (HGVS)^{3,4}.

Key activities aimed at collecting mutations were initiated, including encouraging collection of information (the first step in creating a database) by inviting reviews of mutations in genes for the journal *Human Mutation*, creating guidelines for nomenclature of mutations^{5–8}, initiating quality control of LSDB content^{9,10} and specifying the minimum content of LSDBs^{11,12}. More recently, the content of 100 representative LSDBs was published, leading to further recommendations for content¹³ and a recommended form, published by members of the initiative, for submitting mutations to LSDBs (ref. 14 and <http://www.hgvs.org/entry.html>). As a result of this activity, the number of LSDBs grew to 83 by 2002. More recently, customized software has been made available to assist new curators (for example, LOVD¹⁵ and UMD¹⁶).

Documentation of LSDBs as an aid to research and clinical care began, and a listing was posted on the HUGO/MDI website (now the HGVS website; <http://www.hgvs.org>) in early 1998 containing 209 databases; this listing was later published¹⁷. The listing has grown over the years, making it increasingly difficult to maintain; thus, a new database of LSDBs was created as a relational database on a MySQL database platform (<http://www.mysql.com>) to make curation of these sites easier. In January 2006, a program was initiated to update and add unlisted LSDBs. Dead links were investigated, and curators were contacted to create new links. This process led to the permanent deletion of four LSDBs and the addition of 176 more LSDBs

from various sources, 75 of which were from the Retina International Scientific Newsletter Mutation Databases (see below) and 72 from the IMT Bioinformatic Groups Mutation Databases (see below). The latter two sets could perhaps be called aggregated databases. With Retina International, it appears that the databases are derived directly from the literature. The latest listing (9 March 2007) now includes 672 LSDBs and is likely to grow (<http://www.hgvs.org/dblist/glsdb.html>). This number represents 32% of genes in which at least one mutation has been reported (according to the Human Gene Mutation Database (HGMD); 2,056 genes, as of 9 March 2007).

Beyond the information displayed on the HGVS website, the LSDB database includes gene-specific links to outside databases to aid in curation (EMBL's Ensembl (<http://www.ensembl.org>), the HUGO Gene Nomenclature Committee's gene nomenclature database (<http://www.gene.ucl.ac.uk/nomenclature>) and NCBi's Entrez Gene (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>)).

LSDBs are important because (i) curation by experts on the genes under consideration is hugely superior to that which can be given by experts on databases that collect mutations on all genes, such as OMIM¹⁸ and HGMD¹⁹, and (ii) curators are generally able to collect unpublished mutations from their laboratories, collaborators and abstracts, and unpublished mutations can represent as much as 50% of the total²⁰.

These LSDBs are extremely useful—indeed, vital—to research and proper healthcare, but there is usually little or no funding for this activity. This has led to a number of databases not being updated recently or even being withdrawn. A mechanism clearly needs to be found to prevent this loss of data and data collection.

If there are any LSDBs in existence that do not appear on the list, the authors would be pleased to hear of them.

*Ourania Horaitis*¹, *C Conover Talbot, Jr*²,
*Manyphong Phommarinh*³, *Kate M Phillips*¹ &
Richard G H Cotton^{1,4}

¹Genomic Disorders Research Centre, St. Vincent's Hospital Melbourne, Melbourne, Victoria 3053, Australia. ²Institute of Genetic Medicine, The Johns Hopkins School of Medicine, Baltimore, Maryland, USA. ³Montreal Children's Hospital, Medical Genetics Division - Biochemical Genetics Unit, McGill University Health Centre, Montreal, Quebec, Canada. ⁴Department of Medicine, The University of Melbourne, Melbourne, Victoria 3010, Australia.

e-mail: cotton@unimelb.edu.au

ACKNOWLEDGMENTS

This work was supported by the Australian National Health and Medical Research Council, the March of Dimes and Réseau de médecine génétique appliquée du Fonds de la Recherche en Santé du Québec.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

1. Scriver, C.R., Cotton, R.G.H., Antonarakis, S. & McKusick, V.A. *Genome Digest* **4**, 12–15 (1997).
2. Cotton, R.G., McKusick, V. & Scriver, C.R. *Science* **279**, 10–11 (1998).
3. Cotton, R.G. & Horaitis, O. *Pharmacogenomics J.* **2**, 16–19 (2002).
4. Cotton, R.G.H. & Horaitis, O. in *Nature Encyclopedia of the Human Genome* Vol. 3 (ed. Cooper, D.N.) 361–362 (Nature Publishing Group, London, 2003).
5. Beaudet, A.L. & Tsui, L.C. *Hum. Mutat.* **2**, 245–248 (1993).
6. Antonarakis, S.E. *Hum. Mutat.* **11**, 1–3 (1998).
7. den Dunnen, J.T. & Paalman, M.H. *Hum. Mutat.* **22**, 181–182 (2003).
8. den Dunnen, J.T. & Antonarakis, S.E. *Hum. Mutat.* **15**, 7–12 (2000).
9. Cotton, R.G. & Horaitis, O. *Hum. Mutat.* **15**, 16–21 (2000).
10. Cotton, R.G. & Scriver, C.R. *Hum. Mutat.* **12**, 1–3 (1998).
11. Scriver, C.R., Nowacki, P.M. & Levaslaiho, H. *Hum. Mutat.* **13**, 344–350 (1999).
12. Scriver, C.R., Nowacki, P.M. & Levaslaiho, H. *Hum. Mutat.* **15**, 13–15 (2000).
13. Claustres, M., Horaitis, O., Vanevski, M. & Cotton, R.G. *Genome Res.* **12**, 680–688 (2002).
14. Horaitis, O. & Cotton, R.G.H. in *Current Protocols in Human Genetics* Vol. 2, 7.11.1–7.11.12 (Wiley-Liss, New York, 2003).
15. Fokkema, I.F., den Dunnen, J.T. & Taschner, P.E. *Hum. Mutat.* **26**, 63–68 (2005).
16. Beroud, C. *et al.* *Hum. Mutat.* **26**, 184–191 (2005).
17. Horaitis, O., Scriver, C.R. & Cotton, R.G.H. in *The Metabolic and Molecular Bases of Inherited Disease* (McGraw-Hill, New York, 2001).
18. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. & McKusick, V.A. *Nucleic Acids Res.* **33**, D514–D517 (2005).
19. Stenson, P.D. *et al.* *Hum. Mutat.* **21**, 577–581 (2003).
20. Cotton, R.G. *Hum. Mutat.* **15**, 4–6 (2000).