

Response to ‘Mutation (variation) databases and registries: a rationale for coordination of efforts’: an IRDiRC perspective

Ségolène Aymé, Sharon F. Terry and Stephen Groft

A recent Comment article (The case for locus-specific databases. *Nature Reviews Genetics* **12**, 378–379 (2011))¹ in this journal argued that there is a need to link variation and mutation data to clinical information in order to be able to interpret diagnostic information generated by sequencing approaches and therefore to serve patients and their relatives optimally. That article analysed the various initiatives and databases in place so far. The author claims that locus-specific databases (LSDBs) or disease-specific databases are superior to comprehensive variant databases in terms of the quality of variant data (accuracy and comprehensiveness), mainly owing to the motivation and expertise of the gene and disease professionals.

In a piece of Correspondence (Mutation (variation) databases and registries: a rationale for coordination of efforts. *Nature Reviews Genetics* 25 Oct 2011 (doi:10.1038/nrg3011-c1))² in response to that Comment article, the promoters of the Human Variome Project (HVP) argue that a well-funded international effort is needed to set up a federation of thousands of LSDBs. They contend that, as resources are too limited worldwide to duplicate efforts, these databases should also be patient registries.

To avoid wasteful duplications of effort, the European Commission and the US National Institutes of Health (NIH) recently launched the International Rare Diseases Research Consortium (IRDiRC). The authors of this Correspondence are members of the steering committee of the IRDiRC “whose mission is to coordinate and foster internationally

collaborative research on rare diseases” (REFS 3,4), which include all genetically determined diseases. IRDiRC has two ambitious goals for 2020: to produce diagnostic tests for most rare diseases and to develop new therapies for 200 rare diseases.

The first goal very much depends on the storage and analysis of sequence data. The IRDiRC recognizes that achieving this goal will require global coordination of effort and substantial investment. Therefore, a governing scientific council that will produce an annotated list of diseases under study is planned in order to establish priorities, facilitate collaboration and avoid duplication. It is expected that high-throughput exome or whole-genome sequencing of affected and related individuals will be the predominant approach to gene discovery by consortium researchers. Most importantly, the consortium will require extensive sharing of data and data management tools.

The IRDiRC approach at this stage is not to choose between two possible models of data storage — either many separate LSDBs or a central database — because information technologies exist to provide solutions for a global approach without a single system. WAVE, the Web Analysis of the Variome⁵, for instance, enables centralized access to existing LSDBs, aggregates genes and their variants and integrates other relevant resources without imposing anything on the original LSDBs. What matters is making common standards, interoperability and free access a prerequisite. It is obvious that the best curators are the experts on the locus or disease

and that they must be an integral part of the system that is developed. However, they can carry out curation independently of the location of the database. A responsive software team is needed to manage the database representation of genome structure and variant effects on function and phenotype; this structure has to evolve as our understanding grows.

The funding agencies will have to decide how to allocate resources among possible database models. The research community and clinical laboratories will choose the most user-friendly and useful databases for their needs. The only moral obligations for the LSDB community will be to agree on standards and to participate in this effort, which can rapidly generate new opportunities for diagnosis and disease management. It is time to implement technologically available efficiencies and to reduce competition in the realm of rare diseases. With more than 7,000 of them, there are plenty to go around.

Ségolène Aymé is at INSERM, SC11, 96 rue Didot, F-75014 Paris, France.

Sharon F. Terry is at the Genetic Alliance, 4301 Connecticut Avenue NW, Suite 404, Washington, DC 20008, USA.

Stephen Groft is at the Office of Rare Diseases, US National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892, USA.

Correspondence to Ségolène Aymé
e-mail: segolene.ayme@inserm.fr

doi:10.1038/nrg3011-c2
Published online 25 October 2011

1. Samuels, M. & Rouleau, G. The case for locus-specific databases. *Nature Rev. Genet.* **12**, 378–379 (2011).
2. Auerbach, A. D. *et al.* Mutation (variation) databases and registries: a rationale for coordination of efforts. *Nature Rev. Genet.* 25 Oct 2011 (doi:10.1038/nrg3011-c1).
3. Abbott, A. Rare-disease project has global ambitions. *Nature* **472**, 17 (2011).
4. European Commission. Statement on the international rare disease research consortium. *European Commission* [online], http://ec.europa.eu/research/health/news-07_en.html (2011).
5. Lopes, P., Dalgleish, R. & Oliveira, J. L. WAVE: web analysis of the variome. *Hum. Mutat.* **32**, 729–734 (2011).

Competing interests statement

The authors declare no competing financial interests.