

MutaDATABASE: a centralized and standardized DNA variation database

To the Editor:

The field of molecular diagnostics for genetic disease is entering a new era of whole-genome sequencing. The unprecedented speed and scale with which human sequence variants are being identified is presenting the diagnostics community with practical logistical problems in retrieving, collecting, sharing and depositing clinical and molecular information on genetic disease. In response to this problem, a large consortium of diagnostic testing laboratories in Europe, the United States, Australia and Asia have recently joined efforts to create a central database that provides a repository of DNA variations and allows open access to the whole community. Here, we outline the main features of this universal database, referred to as MutaDATABASE (Fig. 1; <http://www.MutaDATABASE.org>).

There are several reasons why the finding of a pathogenic DNA variation is important for an individual affected with a genetic disease and for his or her family. A molecular diagnosis may offer confirmation of a clinically or biochemically suspected disease, carrier detection, disease risk assessment, prenatal diagnosis, management guidance and/or long-term prognosis and may also facilitate genetic counseling of family members.

Of crucial importance is obtaining certainty that the DNA variation

identified in an affected individual is indeed a pathogenic variant responsible for the person's symptoms, and not simply one of the multitude of benign variations with limited or no clinical significance. Even before the advent of whole-genome sequencing, the clinical significance of many DNA variations was unclear, leaving both affected patients and their treating physicians in the dark.

One reason why this has been the case is that many DNA variations are not made public through reporting in scientific journals or submission into public databases. Even when variants are published, there is usually only limited accompanying clinical information to allow a molecular correlation. Relatively few novel variations are currently being reported in the literature, as it is difficult to get variation reports accepted in journals. Furthermore, only a minority of variations get entered into public databases because such data entry is tedious and laborious with the current platforms; submission is not demanded by journals, granting organizations or accreditation bodies (such as Clinical Laboratory Improvement Amendments (CLIA) in the United States or the International Organization for Standardization (ISO) in Europe and Australia); some testing labs may not see the added value of doing so; and a few may consider the data on variants they are paid

to find as intellectual property belonging to them and therefore decline to share the information.

The problem is far greater when it comes to submission of the clinical information necessary to make genotype-phenotype correlations. First, few diagnostic labs even receive detailed clinical information upon submission of a test sample from the physician or referral laboratory. Second, it is not commonplace for physicians to access DNA variation databases and deposit information. There are also serious issues concerning standardization of clinical information, as well as privacy concerns with depositing an individual's clinical information in a public database, even when de-identified.

Thus, even if DNA variations are entered into a database, detailed clinical information on the individual showing a given variation is usually lacking. Consequently, only a small fraction of DNA variations is currently listed in any publicly accessible database, and even for those, correlation with clinical data is very often unavailable or inadequate. Furthermore, many DNA variation classifications in the current literature are incorrect as a result of the lack of standards and nomenclature for classifying variants, the limited data sets available and the statistically nonsignificant associations that have been made to date. These incorrect classifications have

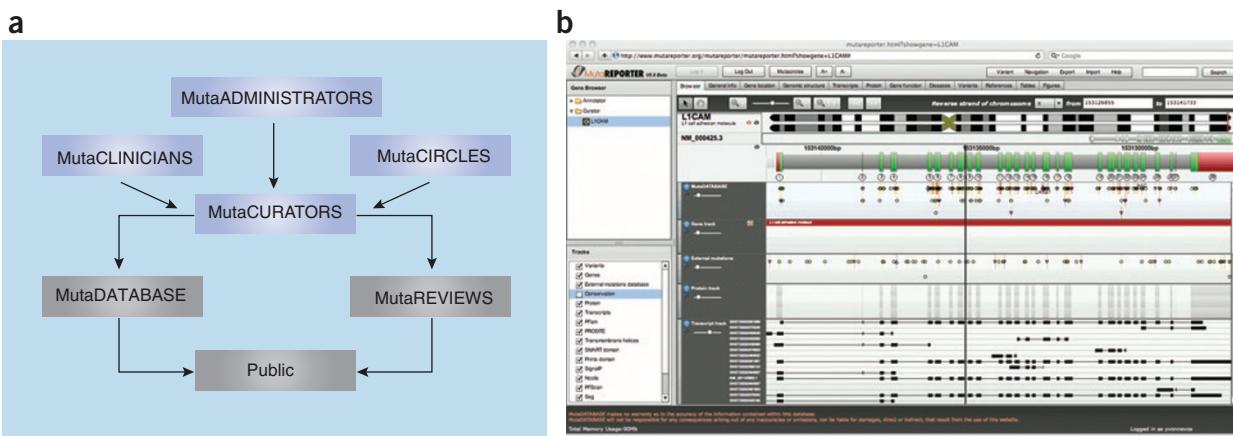


Figure 1 Organization and structure of MutaDATABASE. (a) As new variations are discovered in human genes, experts in the field ('MutaCURATORS') will collect and review all information from the literature and other gene databases, together with information entered by labs and clinicians ('MutaCLINICIANS'). Molecular and clinical information can be submitted into MutaDATABASE directly or through specifically designed 'MutaREPORTER' software, which allows users to circulate all information and questions in community groups centering around a given disease gene ('MutaCIRCLES'). Reviews ('MutaREVIEWS') on specific genes, their variants and associated phenotypes will be published online and will be freely available. (b) A screenshot of the MutaREPORTER software.



CORRESPONDENCE

the potential to cause harm to affected individuals, if they have not already done so. Last but by no means least, although many DNA variation databases exist, no freely available common platform is available wherein all disease genes are represented.

This problem is expected to increase significantly in the upcoming years, for two main reasons. First, the amount of DNA variation identified will grow exponentially in the era of whole-exome or whole-genome sequencing, which is around the corner. Currently, one or a few candidate disease genes are analyzed in each individual, revealing, in the best case, a few DNA variations. However, in upcoming years many affected patients will have their complete exome or genome analyzed through massive parallel sequencing on the currently available 'next-generation' and upcoming 'third-generation' sequencing platforms. This will reveal thousands of variants in each individual tested, and the clinical significance of each of these variants will have to be determined. Second, current molecular testing is limited to a well-selected group of patients usually with a suspected monogenic disorder. As monogenic disease is rare (with an incidence of about 1–3 per 300 births), only a limited number of individuals have been tested in the past decade. But molecular testing is now shifting to the evaluation of risk factors for common multifactorial diseases that affect the majority of the population. It is therefore anticipated that in the near future the whole genome (or exome) of many more people will be sequenced.

It is therefore clear that we need a central DNA variation database as a repository that allows open access to the whole community. Recently, several diagnostic testing laboratories from around the world have joined efforts to create the MutaDATABASE. This is a new, freely available, online database developed to contain standardized information on each human disease gene that not only will list all DNA variations identified in that gene but also aims to combine that data with clinical information for the individuals carrying the DNA variation. As shown in **Figure 1**, each single gene in the database will be curated by experts in the field ('MutaCURATORS') who collect and review all information from the literature and other gene databases as well as information entered by labs and clinicians. Molecular and clinical information can

be submitted into MutaDATABASE directly or through specifically designed software ('MutaREPORTER'). The software additionally allows its users to circulate all information and questions in community groups centering around a particular disease gene ('MutaCIRCLES'), while keeping track both of the initial submitter of a novel variant and of every observer. The whole system functions as a closed-loop, fully automated information system whereby molecular and clinical info can be both extracted and submitted, with gene-specific curators as gatekeepers reviewing all information. In addition, gene-specific reviews will be published online and become freely available. These

We need a central DNA variation database as a repository that allows open access to the whole community.

'MutaREVIEWS' will contain detailed information on each gene, DNA variations in that gene and diseases caused by those DNA variations.

Many obstacles remain before we attain the goal of a freely accessible database of variants with the clinical correlation necessary to assess their pathogenicity. The effort to create the MutaDATABASE is a first step toward resolving some of the logistical issues and is therefore an important and necessary step towards fulfilling the promise of gene testing in improving the health and counseling of our patients and their families.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Sherri Bale¹, Martijn Devisscher², Wim Van Criekinge³, Heidi L Rehm⁴, Frederik Decouttere², Robert Nussbaum⁵, Johan T Den Dunnен⁶ & Patrick Willems⁷

¹GeneDx, Gaithersburg, Maryland, USA.

²GENOHM, Ghent, Belgium. ³BIOBIX, University of Ghent, Ghent, Belgium. ⁴Harvard Medical School, Cambridge, Massachusetts, USA. ⁵Institute of Human Genetics, University of California, San Francisco, San Francisco, California, USA. ⁶Leiden University Medical Center, Leiden, The Netherlands. ⁷GENDIA, Antwerp, Belgium.

e-mail: patrick.willems@genetic-diagnostic.net