Canadian Journal of Cardiology 32 (2016) 1295–1296

Editorial

Access Denied: The Controversy of Commercial Genetic Databases

Jacques Genest, MD

Institut de recherche du centre universitaire de santé McGill, Montréal, Québec, Canada

See article by Kolovic et al., pages 1297–1299 of this issue.

For the past 4 decades, DNA sequence information has increased exponentially and storage of these data in central repositories has provided considerable support for researchers and clinicians. These primary databases, however, constitute a "lower level database" that does not take into account the complexity of the proteome, network biology, phenotypes, and disease in humans.

There are a bewildering number of scientific databases for genomics, DNA sequences, proteomics, and metabolomics that are open access or require institutional (for a fee) access available for researchers. Human mutation databases are but one—albeit a very important one—of these databases.

Historical Aspects

The Human Gene Mutation Database (HGMD) is an attempt to collect and combine published gene mutations responsible for genetic disease in humans. It is maintained at the Institute of Medical Genetics in Cardiff, United Kingdom, by a group of dedicated geneticists and has been an international reference source.[1] This database includes single base-pair substitutions in the coding, regulatory, and splicing-relevant (intronic and exonic) regions of human genes, as well as insertions/deletions mutations, repeat variations, large deletions, insertions, and duplications of genes (referred to as "copy number variants") that are causally associated with a phenotype or disease. It is important to note that entries into this this database are almost always derived from the public domain in the form of peer-reviewed published reports.

The increasing complexity of capturing human disease-related mutations, to validate their causal role, to avoid duplications, to cross-link with other databases, including proteomics, metabolomic, transcriptomics, and other 'omic' databases required considerable expense of personnel, time, and financial resources, which few academic centers can provide and sustain freely. For the past few years, data from HGMD Professional have been made available to HGMD subscribers via Genome Trax (BIOBASE GmbH, Wolfenbüttel, Germany) and Alamut (Interactive Biosoftware, Rouen, France). Allowing free access to the bulk of the mutation data present in the HGMD, while generating sufficient income from its commercial distribution to support its maintenance and expansion represents a business model that should strike a balance between the competing interests of free access and ensuring long-term sustainability.

In May 2014, Qiagen (Venlo, The Netherlands), a well known biotech company, acquired the HGMD database and integrated the contents that included Genome Trax. Qiagen was founded by a team of scientists at Düsseldorf's Heinrich Heine University in 1984 and is now a world leader in molecular biology. A less up-to-date public version of the HGMD database is freely available only to registered users from academic institutions/nonprofit organizations. All commercial users are now required to purchase a license from Qiagen and the data can be used only via a license agreement with Qiagen.

The Issue

In 2003, Fu et al.[2] published a catalogue of genetic variants in monogenic dyslipidemias, the Western Database of Lipid Variants (WDLV). This catalogue was intended to help clinicians and researchers determine the potential pathogenicity of mutations identified using DNA sequencing of patients or research subjects with lipoprotein disorders. The information was published as a supplementary table in the *Canadian Journal of Cardiology*,[2] and was downloadable by individual or institutional subscribers of the journal. Not appreciated at the time was the acquisition by BIOBASE, a bioinformatics company, of the exclusive worldwide marketing of the HGMD. Thus, some of the information contained in the WDLV was considered proprietary. Under a request from Qiagen, the WDLV database was recently modified from its original form, with data considered proprietary by Qiagen removed; these new tables were then published in the place of the older more detailed tables, which were removed, so that the proprietary information is no longer available. It should be noted that the removed data were obtained before the

Received for publication March 29, 2016. Accepted March 31, 2016.

Corresponding author: Dr Jacques Genest, Institut de recherche du centre universitaire de santé McGill, 1001 boul Decarie Bloc E, Office EM12212, Montréal, Québec H4A 3J1, Canada. Tel.: +1-514-934-1934 ×34642.

E-mail: Jacques.genest@mcgill.ca

See page 1296 for disclosure information.

http://dx.doi.org/10.1016/j.cjca.2016.03.018

acquisition of the HGMD by Qiagen. The consequences of noncompliance would have been deleterious to the academic institution and, potentially to a number of clinicians and scientists.

Hegele states eloquently the need for open access to genetic databases to help clinicians and researchers sift through an increasingly complex literature.[3] It should be kept in mind that almost all disease-related genetic data are derived from physicians and scientists working with public moneys (ie, funding agencies) and adhering to the demanding peer-review system. Some laboratories and universities have altruistically endeavored to collect, collate, validate, catalogue, update, and make these data freely available. Sustainability, given the tremendous improvements and decrease in costs in DNA sequencing, is increasingly challenging for single sites or institutions.

Perspective

How will clinicians and scientists gain access to genetic data and to an important database such as the WDLV? First, they must go through a paid subscription to the journal containing the information (eg, in this case the *Canadian Journal of Cardiology*). For the clinician without access to either a personal or institutional subscription, access to these data is denied. An example illustrates this quandary. A clinician faced with a patient with an extreme lipoprotein disorder, say a low-density lipoprotein cholesterol level of 6.8 mmol/L, wishes to determine whether the patient has familial hypercholesterolemia and sends the patient's blood for sequencing. The results read as: mutation XXXX in the *LDLR* gene with 7 single nucleotide polymorphisms associated with an elevated low-density lipoprotein cholesterol level. The WDLV would help in determining the pathogenicity of these mutations and polymorphisms, allow the clinician to pose a firm diagnosis, and might allow the patient to access expensive medications, such as inhibitors of PCSK9.[4] It might also trigger cascade screening in the family.[5] This cost-effective case identification strategy is widely used for monogenic disorders. Presently, this access is limited to academic clinicians and researchers. Open access sites like the Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim) and Med Gen (http://www.ncbi.nlm.nih.gov/medgen) are tremendously useful, but often are difficult to navigate and the information provided might be insufficient for clinicians.

Can we blame Qiagen for limiting access to these data? Yes… and no. Anyone can retrieve human genetic information that lies in the public domain. Unfiltered, this information can be meaningless. Organizing these data is costly, time-consuming, and requires long-term funding to be sustainable. Canadian granting agencies have rarely seen the importance of such an investment. The *Canadian Journal of Cardiology* also limits access to these genetic data by means of paid subscription to the journal, at least for 1 year after publication.

The Future

The exponential growth of genetic data, driven in great part by a marked decrease in cost coupled with next-generation sequencing technology and bioinformatics support, requires novel technologies to sift through clinically relevant data. The future calls for adapting artificial intelligence technologies to define the phenotype, target DNA analysis, and interpret these data for clinicians. Even then, genetic data for relatively rare monogenic disorders, such as extreme hypertriglyceridemia, high-density lipoprotein cholesterol deficiency, and orphan lipoprotein disorders, will remain the province of specialized clinicians (except for familial hypercholesterolemia, with a prevalence of approximately 1 in 250).[6,7] The diagnosis of complex genetic traits, such as severe hypertriglyceridemia, in which gene-environment interactions work synergistically to express a phenotype,[8] will help clinicians to navigate through the ambiguities of nature vs nurture and, hopefully, help provide personalized medical treatment. For Canadian patients with familial hypercholesterolemia, DNA diagnosis is highly recommended and might guide therapy[9]; the use of the WDLV facilitates diagnosis.

Now is the time to show leadership in marshalling the resources of the Canadian Institutes of Health Research, Genome Canada, Compute Canada, and disease-based funding agencies (eg, the Heart and Stroke Foundation of Canada) to provide a service to clinicians and researchers using user-friendly queries. We can do this; for a fee…

Disclosures

The authors have no conflicts of interest to disclose.

References

1. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 2014;133:1-9.

2. Fu J, Kwok S, Sinai L, et al. Western Database of Lipid Variants (WDLV): a catalogue of genetic variants in monogenic dyslipidemias. Can J Cardiol 2013;29:934-9.

3. Kolovic M, Robinson JF, Hegele RA. Proprietary considerations in the use of cardiovascular genetic data. Can J Cardiol 2016;32:1297-9.

4. Navarese EP, Kolodziejczak M, Schulze V, et al. Effects of proprotein convertase subtilisin/kexin type 9 antibodies in adults with hypercholesterolemia: a systematic review and meta-analysis. Ann Intern Med 2015;163:40-51.

5. Nordestgaard BG, Chapman MJ, Humphries SE, et al. European Atherosclerosis Society Consensus Panel. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. Eur Heart J 2013;34:3478-90.

6. Benn M, Watts GF, Tybjærg-Hansen A, Nordestgaard BG. Mutations causative of familial hypercholesterolaemia: screening of 98 098 individuals from the Copenhagen General Population Study estimated a prevalence of 1 in 217. Eur Heart J 2016;37:1384-94.

7. de Ferranti SD, Rodday AM, Mendelson MM, et al. Prevalence of familial hypercholesterolemia in the 1999 to 2012 United States National Health and Nutrition Examination Surveys (NHANES). Circulation 2016;133:1067-72.

8. Hegele RA, Ginsberg HN, Chapman MJ, et al. European Atherosclerosis Society Consensus Panel. The polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. Lancet Diabetes Endocrinol 2014;2:655-66.

9. Genest J, Hegele RA, Bergeron J, et al. Canadian Cardiovascular Society Position statement on familial hypercholesterolemia. Can J Cardiol 2014;30:1471-81.