OVERVIEW

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Databases for Neurogenetics: Introduction, Overview, and Challenges

María-Jesús Sobrido,[1,2]* Pilar Cacheiro,[3] Ángel Carracedo,[1–3] and Lars Bertram[4]

[1]*Fundación Pública Galega de Medicina Xenómica-SERGAS, Santiago de Compostela, Galicia, Spain;* [2]*Center for Network Biomedical Research on Rare Diseases (CIBERER), Institute of Health Carlos III, Majadahonda, Madrid, Spain;* [3]*Genomic Medicine Group, University of Santiago de Compostela, Galicia, Spain;* [4]*Neurospsychiatric Genetics Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany*

For the Databases in Neurogenetics Special Issue

Received 12 July 2012; accepted revised manuscript 13 July 2012.
Published online in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.22164

ABSTRACT: The importance for research and clinical utility of mutation databases, as well as the issues and difficulties entailed in their construction, is discussed within the Human Variome Project. While general principles and standards can apply to most human diseases, some specific questions arise when dealing with the nature of genetic neurological disorders. So far, publically accessible mutation databases exist for only about half of the genes causing neurogenetic disorders; and a considerable work is clearly still needed to optimize their content. The current landscape, main challenges, some potential solutions, and future perspectives on genetic databases for disorders of the nervous system are reviewed in this special issue of *Human Mutation* on neurogenetics.

Hum Mutat 33:1311–1314, 2012. © 2012 Wiley Periodicals, Inc.

KEY WORDS: Neurogenetics; LSDB; databases; HVP; genotype–phenotype

Introduction

The Human Variome Project (HVP) is an international initiative for the development of standards and actions toward public databases of genetic variation causing human disease (http://www.humanvariomeproject.org/). In addition to country-centered collections, the HVP is also fostering the organization of clinical discipline-centered working groups, one of which is the neurogenetics consortium [Haworth et al., 2011]. Neurogenetics can be defined as an intersection of the fields of genetics, neurology, and neuroscience to gain a better understanding of the molecular mechanisms involved in the function and dysfunction of the nervous system. Although genetic methodologies are having a major impact on neuroscience, there is also an increasing need of thorough, curated catalogues of the resulting knowledge, which will involve an unprecedented amount of data that, for the most part, will not find its way into traditional biomedical publications. The confluence of clinical and basic neuroscience research, high-throughput genetics analyses, innovative biostatistics tools, and information and communication technologies is producing an explosion of Web-based data in patients and control populations. These data will facilitate both research and diagnostic work; however, it is also essential that neuroscientists and clinicians learn to exploit and make appropriate use of the information becoming available.

Locus-specific databases (LSDBs) are repositories of sequence information on specific genes associated with medical conditions to help clarify the pathogenic role of a genetic variant in a given patient. The extensive use made of available LSDBs by the research and clinical communities alike demonstrates that these repositories are among the most useful sources of information when confronted by the need of assigning a disease-causing role to a given variant. However, for reliability and efficiency, LSDBs need continuing expert effort to annotate and supervise sequence variations and their associated biological effects [Samuels and Rouleau, 2011]. This special issue of *Human Mutation* (volume 33 issue 9, September 2012; http://onlinelibrary.wiley.com/doi/10.1002/humu.v33.9/issuetoc) explores the current situation of genetic databases for neurogenetics disorders and the tools used to build them. The contributing authors provide an overview of representative disease groups, the challenges, and potential sources of interpretation problems, as well as some hints of future prospects and evolution of the field.

Neurogenetics databases will have to be prepared to accommodate different types of data: single-gene variants identified in laboratories worldwide (mostly diagnostic centers), results from targeted and genome-wide association studies, and high-throughput "-omic" data (so far mostly from research laboratories), together with information on the clinical side (phenotype) and—when possible—a link to functional, cellular, and animal experimental evidence. It is also clear that the needs and challenges will be different for research-oriented repositories than for databases aimed at providing reference tools for clinical practitioners, who need reliable markers for disease diagnosis and prognosis. Clinically oriented databases should, thus, provide the highest level of validation and interpretation accuracy, while being operationally adapted to the professional profile of the expected users, who will generally not be so familiar with the language and tools of genetics research. LSDBs for neurological diseases should foster investigation of genotype–phenotype correlations and the mechanisms that govern this relationship. Databases of genetic variation causing neurological disease will provide knowledge on which to develop guidelines and update diagnostic algorithms for genetic testing in neurogenetics.

*Correspondence to: María-Jesús Sobrido, Fundacion Pública Galega de Medicina Xenómica, Clinical Hospital of Santiago, level -2. Travesía da Choupana s/n 15706 Santiago de Compostela, Spain. E-mail: ssobrido@telefonica.net

Contract grant sponsor: REGENPSI network Consellería de Educación e Ordenación Universitaria, Xunta de Galicia, Spain (2009/019).

© 2012 WILEY PERIODICALS, INC.

Current Landscape of LSDBS for Neurological Diseases

About half of all human genes are expressed in the brain, while also about half of all disorders listed in the Online Mendelian Inheritance In Man (OMIM®) catalogue (http://omim.org/) and with known molecular basis are related to neurological phenotypes. According to Orphanet, one of the most widely used portals of information on rare diseases (http://orpha.net), there are around 6,000 to 8,000 rare diseases [Rath et al., 2012], including over 2,000 neurological entities, many of which have been linked to specific genes. Most neuromuscular and sensory diseases are caused by genetic factors, while for others the etiologies and heritability are not well understood. In many instances, mutations have been identified that appear to increase the risk for a disease, even though they have not been proven to be a direct cause of the disease (e.g., APOE4 in late-onset Alzheimer disease). Even in conditions where we know of specific genetic mutations (e.g., familial Parkinson's disease [PD], muscular dystrophy, and many forms of epilepsy), it has become clear in recent years that the genetics of many of these disorders is complex, entailing contribution from many genes. The same points can be made with regard to psychiatric disorders, which are among the most prevalent disorders worldwide. Indeed, advances in genetics have brought the fields of neurology and psychiatry closer together than ever before.

To provide a gross landscape of mutation databases for neurogenetics, the authors of this special issue conducted a review of LSDBs for genes involved in neurological disorders. These results were presented in the fourth Biennial Meeting of the HVP Consortium [Paris, 2012]. Briefly, we identified neurologic records using the OMIM Clinical Synopsis advanced search tool [Amberger et al., 2011]. Disorders with unknown molecular basis, as well as nondisease phenotypes, susceptibility factors, and diseases with unconfirmed mapping, were filtered out. We then searched for phenotype–gene relationships with OMIM Morbid Map and identified the approved Hugo Gene Nomenclature Committee (HGNC) symbol through the mim2gene tool. Finally, we searched the GEN2PHEN Knowledge Center LSDB listing [Webb et al., 2011; http://www.gen2phen.org/data/lsdbs] to identify mutation databases for the genes selected in the previous stage and compared the database system used, curator status, number and update of variants reported. The results were compared with the situation of neurogenetic databases in 2010 [Mitropoulou et al., 2010]. A total of 1,025 genes were identified, for 98 of which no database was listed. On the other hand, a total of 1,272 databases were available for the remaining 927 genes. For further analysis, we removed databases without variant entries, most of them without a curator, and databases with restricted access. After the previous stepwise filtering, 616 databases remained covering 471 genes, meaning that 46% of genes related to neurological conditions are covered in at least one LSDB. Of the 616 databases, 443 (72%) were built with the Leiden Open Variation Database (LOVD) [Fokkema et al.,2005], while the remaining 173 (28%) corresponded to University of Antwerp databases [Cruts et al., 2012], MUTbase [Riikonen and Vihinen, 1999], Universal Mutation Database (UMD) [Béroud et al., 2000], and other platforms, including Web-based tables. Regarding redundancy, for 363 genes (77%) only one database was detected, two databases per gene were found for 81 genes (17%), and ≥3 databases per gene for 27 genes (6%). Since 2010, the number of neurological genes covered increased from 283 to 471 (an increase of 66%). There was one individual curator in over 65% of the LSDBs for which this information was available, whereas <15% had three or more curators and <3% were curated by a larger consortium or institute. In conclusion, currently there is a freely available mutation database for about half of the genes known to cause Mendelian disorders of the nervous system, most of the neurogenetics LSDBs are built under LOVD, and there is a moderate degree of overlap between databases. Although the number of databases for neurogenetic disorders has significantly increased during the past two years, many of the recently created LSDBs have very limited content and/or little information on the curating process.

Challenges for Neurogenetic Databases

Many of the difficulties for neurogenetic LSDBs are not significantly different from those inherent to other fields of medicine, such as the sources of data, the challenge of expert curation and update, assessment of pathogenicity of the variants, collection and annotation of phenotype and family data, as well as ethical considerations. These and other issues are being discussed with a multidisciplinary approach within the HVP consortium and recommendations are being produced [Kaput et al., 2009; Povey et al., 2010]. However, specific characteristics of neurological disorders call for the confluence of experts in the clinical aspects, as well as in neuropathology, neurophysiology, neuroimaging, genetics, and molecular and cell biology [Haworth et al., 2011]. Some of these challenges are reviewed in more depth in this special issue of *Human Mutation*.

The list of genes and mutations associated with neurological disorders keeps expanding and genetic heterogeneity seems to be the rule. At the same time, the overlap of pathophysiological events among clinically different neurogenetic diseases is also increasingly recognized, originating a sheer complexity of genotype–phenotype relationships. Abundant examples can be found in neurology where a given gene has been associated with different clinical presentations, even within the same family [Ito, 2012; Nishioka et al., 2009]. Furthermore, often the diverse clinical profiles also reflect variability that can be demonstrated in the neurophysiologic and pathologic examination even in patients with the same mutation [Muelas et al., 2010]. Among recent discoveries further highlighting the variability of phenotypes that can be caused by a same mutation is the *C9orf72* intronic expansion associated with motor neuron, extrapyramidal, cognitive, and psychiatric manifestations [Byrne et al., 2012; Simón-Sánchez et al., 2012]. Clinically, different disorders may share related molecular mechanisms, such as mutations in gap junction proteins [Sargiannidou et al., 2010] and ion channelopathies, that can cause pain syndromes, muscle disease, cerebellar ataxia, paroxysmal dystonia, and epilepsy [Cregg et al., 2010; Crompton and Berkovic, 2009]. Also, solute carrier disorders such as *SLC2A1*-associated syndromes challenge classification of neurological disease based on the phenotype and point to a possible pathophysiological overlap between epilepsy and movement disorders [Leen et al., 2010]. The biological explanation why mutations in the same gene–gene family may cause highly variable phenotype is often unknown.

The opposite situation is also frequent in neurogenetics: an indistinguishable clinical picture can result from alterations in different genes, even with evidence of diverse pathoanatomical and pathophysiological alterations. The broad molecular heterogeneity of Charcot–Marie–Tooth disease, the spinocerebellar ataxias (SCA) or the hereditary spastic paraplegias (HSP)—a group of disorders affecting upper motor neurons—are just some examples of this. While Hersheson et al. (2012) review the broad genetic heterogeneity among inherited ataxias, Bettencourt and collaborators revise cases from the literature and from their own experience with

mutations in the SCA genes presenting with a phenotype closer to HSP than to the ataxias [Bettencourt et al., 2012]. Abel et al. (2012) further illustrate the evolving landscape of genotype–phenotype relationships in motor neuron diseases, leading to the evolution of the ALSoD database from a single-gene LSDB to a multigene and bioinformatics analysis installation. The neurodegenerative brain diseases group and neurogenetics laboratory of the University of Antwerp have developed some of the most broadly used LSDBs for Mendelian CNS disorders [Cruts et al., 2012]. The AD&FTLD and PD mutation databases described in their article currently provide thorough information on 15 genes causing Alzheimer disease (AD), frontotemporal lobar degeneration (FTLD), and PD. However, some genes can lead to a variable combination of amyotrophic lateral sclerosis, frontotemporal dementia, and PD, indicating once more that there is a need for standardized procedures and integration capabilities of the informatics tools used to build neurogenetic databases, so that efforts can be added up, cross-searches are possible and potential discrepancies and errors can be readily detected.

Comprehensive and reliable genotype–phenotype databases should accelerate the understanding of the molecular and cellular mechanisms underlying neurogenetic conditions. As this understanding evolves from the single molecular alteration to pathophysiological routes, the nosology of neurological diseases will also evolve, so that syndromes with similar molecular pathogenesis will be grouped together more so than those with similar clinical presentation. The above-mentioned issues pose a great challenge to the differential diagnosis in clinical neurology. It is increasingly clear that efforts toward genotype interpretation also need better tools for a harmonized, computer-readable description of the phenotype, including the capture of the evolving clinical picture in either treated or untreated patients. Köhler et al. (2012) review this question and illustrate how the Human Phenotype Ontology (HPO; http://www.human-phenotype-ontology.org), combined with disease classification systems such as that from Orphanet, provide a systematic approach to phenotype representation and link to other biomedical data sources. Expansion of currently available ontologies to specific clinical domains will be achievable with the coordination of bioinformatics and clinical experts [Robinson and Mundlos, 2010; Schofield and Hancock, 2012]. The phenotypic manifestations of the nervous system's correct function and dysfunction are often more apparent that those from other organ systems, in the form of abnormal movements, language, sensory, cognitive, and behavioral traits that are often subjectively measured. However, over the centuries such observations have produced a very rich, geographically and culturally variable, *non*-controlled medical vocabulary. The construction of phenotype ontologies will, therefore, be especially difficult in the realm of neurological and psychiatric disorders.

Another challenge for the development of LSDBs in neurogenetics comes from the nature of some of the mutation types frequently encountered in neurodegenerative diseases. Expansions of repetitive elements in the genome, also called dynamic mutations, underlie neurological disorders such as Huntington disease, fragile-X mental retardation, and several SCA, among others. These oligonucleotide repeats amenable to pathologic expansions vary in structure, size and location of the repetitive element within the gene. Some of the main difficulties posed by this type of mutations to build mutation databases, discussed by Martindale et al. (2012), include the definition of the reference repeat structure, accurate sizing, and naming of the alleles, establishing normal and abnormal repeat number, and mitotic and meiotic instability of the expanded allele.

Mitochondrial cytopathies represent another group of disorders often manifesting at the level of the nervous system and posing a challenge for database curators, as reviewed by Elson et al. (2012). Additional factors such as nuclear modifying genes or the mitochondrial DNA haplotype background may influence phenotypic expression of mitochondrial disease. Both the central and peripheral nervous system may be involved in many ways, complicating the diagnosis of mitochondrial disorders and calling for the concurrence of experts in different clinical areas [McFarland et al., 2010]. However, given the peculiar features of mitochondrial biology, together with the fact that there is a common pathophysiological theme—dysfunction of the cellular energy production—the maintenance of a unique, coordinated LSDB to deal with mitochondrial disorders seems desirable.

In addition to LSDBs for monogenic or oligogenic disorders, the study of genetic variants influencing the development of more common neurological conditions (e.g. PD, AD, multiple sclerosis) has been a very active fields in recent years. A model for the storage, curation, and usage of the large amount of information emanating from both targeted and genome-wide association studies are described by Lill and Bertram (2012), as exemplified in the PDGene and AlzGene databases. Both complex disease and monogenic neurogenetic databases will have to be prepared to deal with the "next-generation" sequencing (NGS) technologies, which represent a new challenge for genetic information storage and interpretation [Hersheson et al., 2012; Lill and Bertram, 2012]. In such scenario, the foreseeable difficulties for interpreting the pathogenic relevance of rare variants will only be worked out with the availability of curated LSDBs and databases of rare variants seen in healthy individuals in various ethnic backgrounds. The continuous pathogenetic frontier between Mendelian and complex neurodegenerative diseases will possibly be challenged as the scientific community moves from SNP-based association studies to the analysis of individual genomes.

Last but not least, all new technologies and methods—including genetic databases for public use—deserve thoughtful debate considering the ethical, social, and legal implications, and must decide whether, on balance, the "positive" uses outweigh the "negative". It is important to emphasize that we are dealing with brain disorders. As we learn more about genetic determinants, traditional distinctions between neurological and psychiatric diseases may fade or disappear altogether. Genes may influence personality, behavioral, and cognitive traits in addition to defined clinical disorders. Thus, the scientific exploration of genes that affect normal and abnormal function of the brain deserves an even more thoughtful consideration than already, and need to be linked to questions in the field of neuroethics [Illes and Bird, 2006]. The vision of the HVP is that expert-curated LSDBs coupled with standardized genetic and clinical terminology offers one efficient way to deal with these and other challenges [Samuels and Rouleau, 2011]. Because there is probably no individual or research group with equal expertise in the clinical, genetic, physiological, histopathologic, biochemical, and cell biology domains or (bio)informatic tools that are needed for a robust data synthesis and interpretation, it appears that the best framework to develop high-quality, long-standing LSDBs and other genetic databases will be within multidisciplinary networks. The main goal of the neurogenetics consortium within the HVP is exactly to reach this goal, that is to bring together experts from diverse fields in join efforts in our quest to better understand the pathogenetic forces underlying diseases of the nervous system.

References

Abel O, Powell J, Al-Chalabi A. 2012. ALSoD: a user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. Hum Mutat 33:1345–1351.

Amberger J, Bocchini C, Hamosh A. 2011. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). Hum Mutat 32:564–567.

Béroud C, Collod-Béroud G, Boileau C, Soussi T, Junien C. 2000. UMD (Universal mutation database): a generic software to build and analyze locus-specific databases. Hum Mutat 15:86–94.

Bettencourt C, Quintáns B, Ros R, Ampuero I, Yáñez Z, Pascual SI, de Yébenes JG, Sobrido MJ. 2012. Revisiting genotype-phenotype overlap in neurogenetics: triplet-repeat expansions mimicking spastic paraplegias. Hum Mutat 33:1315–1323.

Byrne S, Elamin M, Bede P, Shatunov A, Walsh C, Corr B, Heverin M, Jordan N, Kenna K, Lynch C, McLaughlin RL, Iyer PM, et al. 2012. Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. Lancet Neurol 11:232–240.

Cregg R, Momin A, Rugiero F, Wood JN, Zhao J. 2010. Pain channelopathies. J Physiol 588:1897–1904.

Crompton DE, Berkovic SF. 2009. The borderland of epilepsy: clinical and molecular features of phenomena that mimic epileptic seizures. Lancet Neurol 8:370–381.

Cruts M, Theuns J, Van Broeckhoven C. 2012. Locus-specific mutation databases for neurodegenerative brain diseases. Hum Mutat 33:1340–1344.

Elson JL, Sweeney MG, Procaccio V, Yarham JW, Salas A, Kong QP, van der Westhuizen FH, Pitceathly RD, Thorburn DR, Lott MT, Wallace DC, Taylor RW, McFarland R. 2012. Toward a mtDNA locus-specific mutation database using the LOVD platform. Hum Mutat 33:1352–1358.

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. 2011. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 32:557–563.

Haworth A, Bertram L, Carrera P, Elson JL, Braastad CD, Cox DW, Cruts M, den Dunnen JT, Farrer MJ, Fink JK, Hamed SA, Houlden H, et al. 2011. Call for participation in the neurogenetics consortium within the Human Variome Project. Neurogenetics 12:169–173.

Hersheson J, Haworth A, Houlden H. 2012. The inherited ataxias: genetic heterogeneity, mutation databases, and future directions in research and clinical diagnostics. Hum Mutat 33:1324–1332.

Illes J, Bird SJ. 2006. Neuroethics: a modern context for ethics in neuroscience. Trends Neurosci 29:511–517.

Ito D. 2012. BSCL2-related neurologic disorders/seipinopathy. In: Pagon RA, Bird TD, Dolan CR, Stephens K, Adam MP, editors. GeneReviews^{TM} [Internet]. Seattle, WA: University of Washington; 1993–2005 Dec 06 [updated 2012 Jun 07].

Kaput J, Cotton RG, Hardman L, Watson M, Al Aqeel AI, Al-Aama JY, Al-Mulla F, Alonso S, Aretz S, Auerbach AD, Bapat B, Bernstein IT, et al. 2009. Planning the human variome project: the Spain report. Hum Mutat 30:496–510.

Köhler S, Doelken SC, Rath A, Aymé S, Robinson PN. 2012. Ontological phenotype standards for neurogenetics. Hum Mutat 33:1333–1339.

Leen WG, Klepper J, Verbeek MM, Leferink M, Hofste T, van Engelen BG, Wevers RA, Arthur T, Bahi-Buisson N, Ballhausen D, Bekhof J, van Bogaert P, et al. 2010. Glucose transporter-1 deficiency syndrome: the expanding clinical and genetic spectrum of a treatable disorder. Brain 133:655–670.

Lill CM, Bertram L. 2012. Developing the "next generation" of genetic association databases for complex diseases. Hum Mutat 33:1366–1372.

Martindale JE, Seneca S, Wieczorek S, Sequeiros J. 2012. Spinocerebellar ataxias: an example of the challenges associated with genetic databases for dynamic mutations. Hum Mutat 33:1359–1365.

McFarland R, Taylor RW, Turnbull DM. 2010. A neurological perspective on mitochondrial disease. Lancet Neurol 9:829–840.

Mitropoulou C, Webb AJ, Mitropoulos K, Brookes AJ, Patrinos GP. 2010. Locus-specific database domain and data content analysis: evolution and content maturation toward clinical use. Hum Mutat 31:1109–1116.

Muelas N, Hackman P, Luque H, Garcés-Sánchez M, Azorín I, Suominen T, Sevilla T, Mayordomo F, Gómez L, Martí P, María Millán J, Udd B, Vílchez JJ. 2010. MYH7 gene tail mutation causing myopathic profiles beyond Laing distal myopathy. Neurology 75:732–741.

Nishioka K, Ross OA, Ishii K, Kachergus JM, Ishiwata K, Kitagawa M, Kono S, Obi T, Mizoguchi K, Inoue Y, Imai H, Takanashi M, Mizuno Y, Farrer MJ, Hattori N. 2009. Expanding the clinical phenotype of SNCA duplication carriers. Movement Disorders 24:1811–1819.

Povey S, Al Aqeel AI, Cambon-Thomsen A, Dalgleish R, den Dunnen JT, Firth HV, Greenblatt MS, Barash CI, Parker M, Patrinos GP, Savige J, Sobrido MJ, et al. 2010. Practical guidelines addressing ethical issues pertaining to the curation of human locus-specific variation databases (LSDBs). Hum Mutat 31:1179–1184.

Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. 2012. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. Hum Mutat 33:803–808.

Riikonen P, Vihinen M. 1999. MUTbase: maintenance and analysis of distributed mutation databases. Bioinformatics 15:852–859.

Robinson PN, Mundlos S. 2010. The human phenotype ontology. Clin Genet 77:525–534.

Samuels ME, Rouleau GA. 2011. The case for locus-specific databases. Nat Rev Genet 12:378–379.

Sargiannidou I, Markoullis K, Kleopa KA. 2010. Molecular mechanisms of gap junction mutations in myelinating cells. Histol Histopathol 25:1191–1206.

Schofield PN, Hancock JM. 2012. Integration of global resources for human genetic variation and disease. Hum Mutat 33:813–816.

Simón-Sánchez J, Dopper EG, Cohn-Hokke PE, Hukema RK, Nicolaou N, Seelaar H, de Graaf JR, de Koning I, van Schoor NM, Deeg DJ, Smits M, Raaphorst J, et al. 2012. The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. Brain 135:723–735.

Webb AJ, Thorisson GA, Brookes AJ, GEN2PHEN Consortium. 2011. An informatics project and online "Knowledge Centre" supporting modern genotype-to-phenotype research. Hum Mutat 32:543–550.