

HGMD and ClinVar: Avoiding the Knowledge Blind Spot

How HGMD Enables Better Care with Better Knowledge

Introduction

In February 2016, a mother filed a lawsuit relating to the death of her son Christian against Athena Diagnostics, ADI Holding Company, and Athena's parent company, Quest Diagnostics. Christian was born seemingly healthy on August 23, 2005, but several months later, experienced a series of unrelenting seizures. He had a massive battery of tests in early 2007, including the sequencing of a gene called SCN1A. Athena, which performed the genetic tests, reported that he had a variant of unknown significance (VUS) there. With no clear genetic answers, his doctors treated him for an undiagnosed mitochondrial disorder, which had minimal effect on his continuing seizures. On January 5, 2008, Christian died from a seizure.

Six years later, thinking of starting a family again, his mother wanted to get her own DNA sequenced to learn whether the disease that had affected her son could affect any future children. Again, she turned to Athena, and in addition to her own results, she requested Christian's 2007 lab report. She saw from the revised report that Athena had reclassified Christian's VUS to a disease-associated mutation, which suggested he had a form of childhood epilepsy called Dravet syndrome or severe myoclonic epilepsy of infancy. Several of the medications used to treat seizures in young children, including Christian, are toxic to children with Dravet and can increase the risk of death.

What his mother now wanted to know from Athena was when and why they reclassified the variant. As a former special education teacher, she taught herself the nuances of scientific literature and found out that the same SCN1A mutation Christian carried had been identified in an Australian family in 2006, before Christian's DNA was tested. Even more concerning was a patent document on the

SCN1A gene that listed this mutation (a change in a single amino acid in the gene) as pathogenic. When Athena refused to answer, Williams sued. Her allegations include that Athena had had enough information to reclassify Christian's mutation before he was tested, and that if they had done so, it would have changed his diagnosis and treatment such that his death from a seizure related to Dravet syndrome could have been avoided (GenomeWeb, 2016).

This case reflects the significance of "knowledge blind spots" in modern genetic testing—valuable information that goes undetected due to inadequate mining and interpretation mechanisms. In the past decade alone, the number of inherited disease-associated germline mutations published per year has doubled, with approximately 18,000 mutation entries already made in 2017 alone (Figure 1). Thus, the sheer volume of papers published every year presents a serious bottleneck for hereditary disease testing, both in terms of keeping pace with discoveries and in interpreting clinical test results with the most up-to-date research. ▷

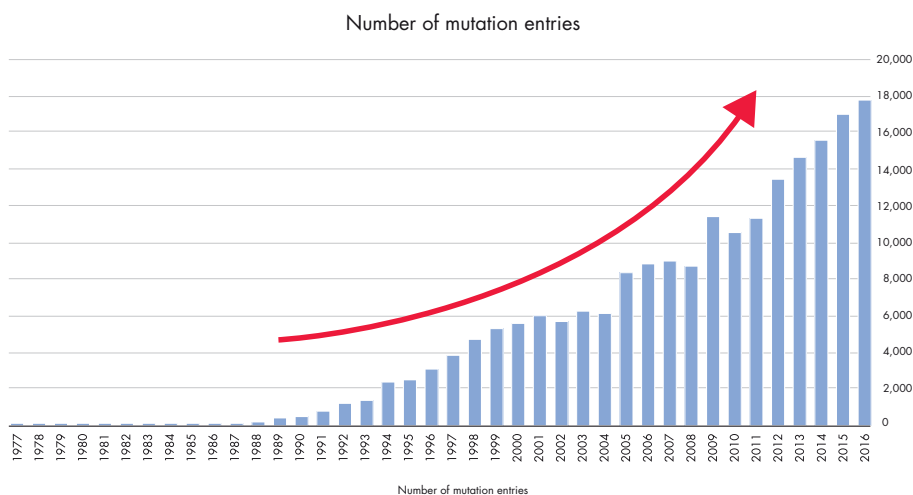


Figure 1. The number of inherited disease-associated germline mutations published per year has doubled since 2007 (within 10 years).

To this end, several variant databases have been developed and cultivated, predominantly the Human Genome Mutation Database (HGMD®) and ClinVar, both of which record disease-causing or disease-associated (pathogenic) variation. However, while both databases are extensively used in genetic testing programs, they are not equal—not even comparable.

HGMD Professional is the gold standard, industry-leading resource for comprehensive coverage of published germ-line human inherited disease mutations. Unlike any other variant database, HGMD attempts to collate all known gene lesions responsible for causing human inherited disease together with disease-associated and functional polymorphisms that have been cited in peer-reviewed literature. ClinVar, on the other hand, is a freely available archive for interpretations of both germline and somatic variants in which clinical significance is sometimes reported directly from submitters, ranging from clinical testing laboratories to research programs to the general public. In an era when new high-throughput sequencing technology affords greater specificity, sensitivity, and accuracy than ever before, so too must the accompanying tertiary analysis provide the same quality of exactitude. This paper illustrates how HGMD exceeds ClinVar in terms of clinical breadth and clinical depth and is essential to provide better care with better knowledge.

Accessing a more comprehensive clinical breadth

HGMD was originally established in 1996 for the scientific study of mutational mechanisms in human genes. However, it has since acquired a much broader utility as a central unified disease-oriented mutation repository utilized by human molecular geneticists, genome scientists, molecular biologists, clinicians and genetic counsellors as well as by those specializing in biopharmaceuticals, bioinformatics and personalized genomics (Stenson et al., 2014). As of October 2017, the database contained over 208,000 different lesions detected in over 8,000 different genes, with new mutation entries currently accumulating a rate exceeding 17,000 per annum. In contrast, ClinVar is young, initially released in 2013, and although it has over 100,000 different lesions detected in over 5,000 different genes, the breadth of coverage between the two databases significantly differs. As depicted in Figure 2, HGMD covers 138% more genes than ClinVar and has 187% more clinically-relevant variants than ClinVar.

In 2013, Peterson et al. conducted a review that compared current human variant resources, including HGMD and ClinVar, and found that of the 67,555 variants described in ClinVar, only 13,465 were classified as clinically significant. In com-

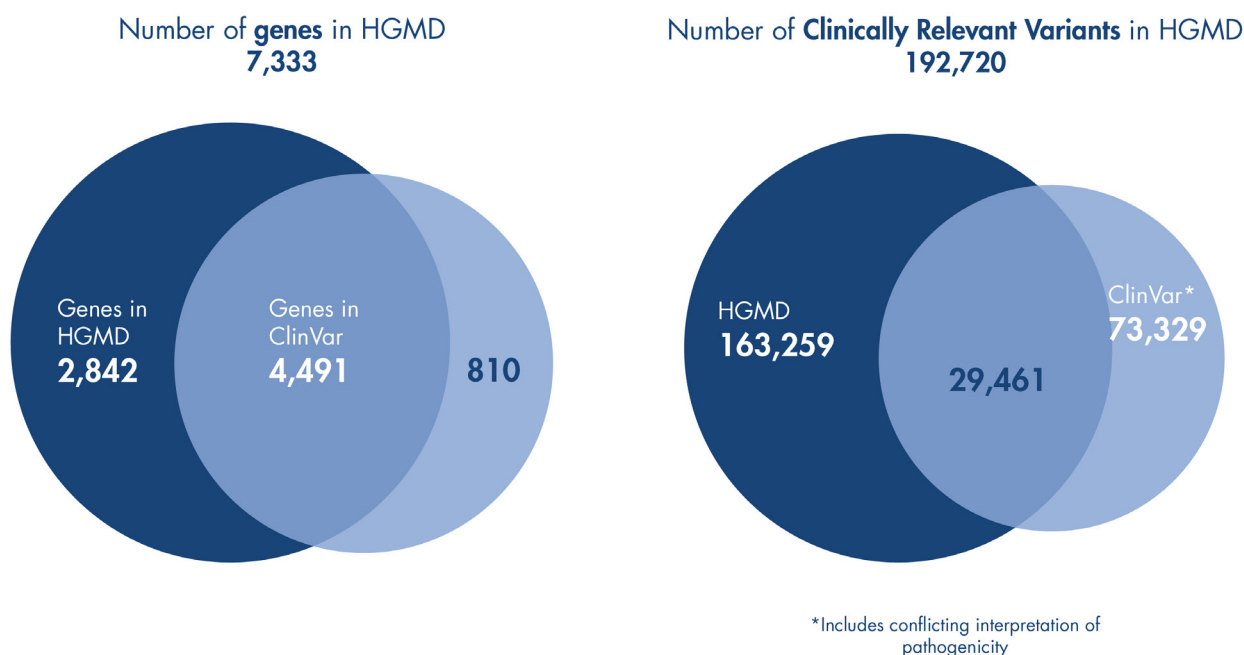


Figure 2. HGMD contains significantly more genes and clinically relevant variants than ClinVar. Data compiled as of September 2017.

Table 1. Total disease variants of each type for each resource. Overview of the total number of disease variants of each type (i.e., exonic missense variants, splice site, insertions, deletions, stop-gain, stop-loss and variants from promoter regions) contained in the ClinVar (20), HGMD (18), OMIM (19) and UniProt/Swiss-Prot (21) resources. Variants were considered only after filtering each database based on several criteria and redundant variants were removed if there were not annotated for a unique gene, wild type, mutated type and position (Peterson et al., 2013).

Name	ClinVar	OMIM	UniProt/Swiss-Prot	HGMD Public	HGMD Professional
Exonic Missense Variants	10,420	10,967	22,912	44,427	60,058
Splice Site Variants	332	896	N/A	9,455	12,665
Insertions or Deletions	264	1,666	317	13,396	19,089
Stop-gain Variants	2,344	2,521	N/A	11,110	14,972
Stop-loss Variants	32	19	N/A	32	74
Variants from Promoter Regions or UTRs	73	N/A	N/A	1,753	2,663
Total	13,465	16,069	23,229	80,173	109,521

parison, HGMD contained 109,521 clinically significant disease-associated mutations at the time (Table 1). Further, there are approximately 2,800 genes only found in HGMD with annotated variants, and 230 of these genes contain more than 50 additional annotated variants. In contrast, ClinVar has only 810 annotated genes unique to the database. Such a vast difference in the number of reported clinically significant variants exposes the limitations of using databases that have incomplete coverage. HGMD includes only mutations that have been shown to be disease-associated, while ClinVar includes variants regardless of whether or not they have been shown to be disease-associated. While ClinVar does provide access to all variants that have been observed, these observations may not be supported by peer-reviewed literature. Furthermore, as more knowledge of certain variants has accumulated, ClinVar submitters often do not update the submitted variant’s clinical significance.

Gaining a higher depth of coverage

Not only does HGMD surpass ClinVar in terms of clinical breadth with information on more genes and more variants, HGMD also provides up-to-date mutation data with a quarterly release cycle, which is essential for validating the novelty of newly found mutations. HGMD currently contains mutation entries obtained from over 57,000 primary and 29,000 additional (supplementary) peer-reviewed literature reports pub-

lished in more than 2,600 different journals. Of >10,000 identified articles screened for mutation data during 2012, 35% contained novel mutation data and 29% contained additional useful information (e.g. in vitro functional data or further clinical or phenotypic information). For ~4 % of all the missense/nonsense mutations reported in the literature during 2012, it was necessary for the HGMD curators to contact the original authors to obtain correction and/or clarification of the nature or precise location of the mutations in question. However, only half of the mutations that required author contact were satisfactorily resolved by these means, leading to their inclusion in HGMD; the ~2 % of unresolved missense/nonsense mutations will not be entered into HGMD unless or until the nature or precise location of the mutation(s) in question is determined to the satisfaction of the HGMD curators (Stenson et al., 2014). This rigorous quality assurance procedure is a defining hallmark of HGMD. Only disease-causing mutations (DM) that have been demonstrated to be involved in conferring the associated clinical phenotype upon the concerned individuals are included in HGMD. ClinVar includes conflicting interpretations of pathogenicity, which can contribute to misdiagnosis.

Although HGMD has a significantly greater clinical depth than that of ClinVar, the database is easily navigated through an advanced search application that can quickly look up individual mutations, as well as identify published mutations

associated with a particular gene or disease. For example, a clinician looking up cystic fibrosis can retrieve 3,415 relevant articles from PubMed (the “old way”), then take approximately 284 hours to read all articles and deduce a total of 1,383 mutations (calculation based on a conservative estimate of 5 minutes per article). With HGMD, the same search would take 5 minutes and yield the same total of 1,383 mutations. In addition, HGMD users can tailor their queries with specific criteria, including functional profile, amino-acid change, nucleotide substitution, size, sequence, and much more, affording greater possibilities and specificity. Therefore, HGMD saves considerable time and money by providing an alternative to tedious and labor-intensive literature searches.

Conclusion

In the case of Christian, it is impossible to determine if the outcome would have been different had the clinicians at Athena Diagnostics been able to locate the Australian paper that identified the same SCN1A mutation one year earlier. However, what can be certain is that accessing such a comprehensive and timely resource as HGMD mitigates the occurrence of clinical knowledge “blind spots”. HGMD Professional is the most up-to-date, reputable, and complete collection of known and published pathogenic gene lesions responsible for human inherited disease.

HGMD identifies every published article that describes a germline mutation and assesses whether or not the mutation has been convincingly demonstrated to be associated with a specific disease or phenotype. ClinVar, although not reviewed and regulated, is seeded with records based on allelic variants described by clinical laboratories and research groups with differing quality standards. Therefore, the best practice would be to combine HGMD with ClinVar, saving clinicians time and money while having peace of mind that their interpretation and reporting is of the highest quality, accuracy, and comprehensiveness.

HGMD is available as a free public version with restricted content and limited search options for solely academic use, and as a fully functional HGMD Professional version. The free public version of HGMD is approximately two to three years out of date compared to HGMD Professional, making the latter option of choice resource for genetic testing, variant filtering and annotation.

HGMD Professional is licensed exclusively through QIAGEN, the industry leader in bioinformatics. QIAGEN’s hereditary disease solutions are widely adopted, with unsurpassed comprehensiveness of validated content and accuracy. With QIAGEN and HGMD, you can deliver better care with better knowledge.

References

- Peterson TA, Doughty E, Kann MG. Towards precision medicine: advances in computational approaches for analysis of human variants. *J Mol Biol.* 2013;425:4047–4063.
- Ray, T. Mother’s Negligence Suit Against Quest’s Athena Could Broadly Impact Genetic Testing Labs. *GenomeWeb.* 2016; May 14, 2016. <https://www.genomeweb.com/molecular-diagnostics/mothers-negligence-suit-against-quests-athena-could-broadly-impact-genetic>
- Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. 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.

For up-to-date licensing information and product-specific disclaimers, see the respective user manual. QIAGEN user manuals are available at www.qiagenbioinformatics.com or can be requested from QIAGEN Technical Services or your local distributor.