QIAGEN Knowledge Base and ClinVar: Avoiding the Knowledge Blind Spot

How QIAGEN 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 included that there was available information at the time to reclassify Christian's mutation before he was tested, and that if Athena had used that information, his diagnosis and treatment would have changed. If appropriate treatment had been administered, Christian's death could have been avoided (GenomeWeb, 2016).

This case brings to light the danger of "knowledge blind spots" in modern genetic testing. As genetic testing becomes more widespread and the knowledge base for interpretation grows dramatically, the risk of missing something critical for making the right diagnosis also increases. New literature with data on genetic variants has increased six-fold in the last 15 years, with more than 18,000 inherited disease-associated germline mutations published in 2017 alone (Figure 1). Today, more than 100 new publications are released each month. Such a deluge of medical information makes accessing the right research at the right time more challenging—and even more important.

In the race to keep pace with the exponential growth of medical genetic knowledge, public and private databases have been created. These databases aim to assist physicians in making clinical diagnoses, providing an overview of variant information, suggesting therapeutic interventions, and recommending treatments. To that end, medical knowledge-based systems perform a variety of functions, including knowledge acquisition, knowledge translation, and utilizing data from a range of sources, such as evidence-based medical literature, approved clinical practice guidelines (CPG), and drug labels. Due to this multitude of heterogeneous sources, developing a complete and consistent clinical knowledge base is a titanic undertaking.

Public disease-specific and gene-specific databases, such as the Human Gene Mutation Database (HGMD), ClinVar, and various locus-specific databases, can document functionally or clinically validated genetic variants that are pathogenic for particular diseases. The HGMD® is a comprehensive collection of germline mutations in nuclear genes associated with human inherited disease and compiled exclusively from published literature. ClinVar archives the clinical significance of variants reported ▷
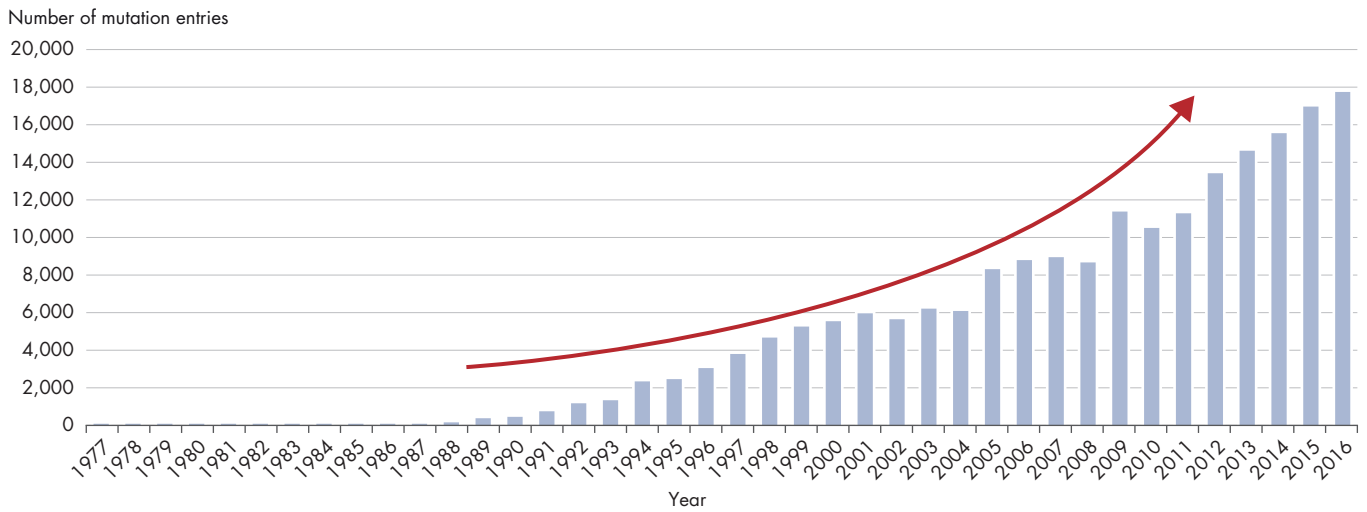
Number of mutation entries

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

directly from crowd-sourced submitters. However, these databases often contain variants that are incorrectly classified without a primary review of evidence, and they sometimes have contradictory records on the assessment of pathogenicity.

For more than 20 years, QIAGEN has been building the industry's largest, most up-to-date clinical database. Unrivalled in size, design, and integrity, the QIAGEN Knowledge Base enables clinicians to find disease-causing variants faster, and with fewer false leads, by tapping into the knowledge of millions of scientific findings that have been manually curated by hundreds of MDs and PhDs. It is precisely this differentiator — QIAGEN's team of expert curators — that makes the QIAGEN Knowledge Base an invaluable tool in modern genetic testing. Unlike ClinVar, the QIAGEN Knowledge Base includes multiple layers of manual and automated quality control to maintain accuracy and consistency of phenotype and genotype representations. Researchers and clinicians who rely solely upon public databases, such as ClinVar, run the risk of misinterpreting results with missing, inadequate, inconsistent, or outdated information. To evidence the value of the QIAGEN Knowledge Base in modern genetic testing, this paper compares the QIAGEN Knowledge Base to ClinVar, analyzing the clinical breath and clinical depth of both highly-used resources.

Why Bigger is Better

Like the rest of the world, health care is becoming increasingly connected, but also increasingly complex. The sheer volume of health care data is growing at an astronomical rate: 153

exabytes (one exabyte = one billion gigabytes) were produced in 2013 and an estimated 2,314 exabytes will be produced in 2020, translating to an overall rate of increase of at least 48 percent annually (International Data Corporation (IDC), 2014). When it comes to identifying genetic disorders, misdiagnoses and delayed diagnoses are often attributed to non-specificity and heterogeneity of signs and symptoms, rarity of conditions and limited access to current publications and research. As data continues to grow, heterogeneous symptoms and rare conditions will be better understood, but only if this wealth of information can be collected, analyzed and accessed in real-time.

The QIAGEN Knowledge Base makes a seemingly impossible task, not only feasible, but inherently simple. As of January 2018, the QIAGEN Knowledge Base has analyzed more than 700,000 clinical cases and includes knowledge from over 4,000 peer-reviewed journals. Within this massive database, more than 10 million variants, including 6.2 million clinically-annotated variants from cases worldwide, provide clinically-relevant information on over 23,000 genes. Such comprehensiveness of genetic information available in the QIAGEN Knowledge Base is unmatched by both market and public competitors, making it the best available resource for clinicians when diagnosing genetic disorders.

Take ClinVar for example. ClinVar is freely available online and widely popular. However, the database has significantly fewer documented variants. As of January 2018, ClinVar reported just over 110,000 clinically-relevant variants. To

Table 1. Comparison of QIAGEN Knowledge Base to ClinVar in terms of variant interpretation from a carrier screening panel of 324 genes

Indication	Genes	Variants with QIAGEN curated bibliography	Variants with ClinVar annotations	Variants with QIAGEN content or ClinVar associations	Variants with only QIAGEN content
Carrier Screening	324	105565	49690	133400	62.75%
Inherited Cancer	114	100418	60676	138188	56.09%
Cardiovascular	92	47022	26147	61978	57.81%
ACMG 59	59	81232	46734	103553	54.87%

illustrate the end-result difference between using ClinVar and using the QIAGEN Knowledge Base, a top clinical testing laboratory in the United States conducted a study that compared both resources in terms of variant interpretation from a carrier screening panel of 324 genes. The study found that the QIAGEN Knowledge Base identified 105,565 variants, each linked to a curated bibliography, while ClinVar identified 49,690 variants. The QIAGEN Knowledge Base had 60% more unique variants than ClinVar. Similarly, the study tested inherited cancer, cardiovascular disease, and ACMG 59 panels with 114, 92, and 59 genes, respectively, and reported the QIAGEN Knowledge Base to have 50% more unique variants across all three indications (Table 1). Further, in a study comparing the bibliographies of the QIAGEN Knowledge Base and HGMD Professional (which is included in the QIAGEN Knowledge Base) for a particular variant in BRCA1, the QIAGEN Knowledge Base yielded 113 verified references, while HGMD Professional yielded seven.

The breadth of the QIAGEN Knowledge Base is a direct result of QIAGEN's unique curation model. Each week, the content of the database is updated, with more than 4,000 articles added per month and 3,000 findings per day. This continuous refresh increases the accuracy of diagnoses. When examining genetic variants, the decisions that clinicians have to make about thorough interpretation and proper treatment affect their patient's health and medical outcomes. Having all available data and information at one's fingertips is crucial for better care.

Going to Great Depths

Variant classification is the cornerstone of clinical molecular genetic testing. The validity and utility of genetic testing require that variant classifications be evidence-based, objective, and systematic. Most public databases, including ClinVar and HGMD, document the clinical significance of genetic variants, but they are mostly provided by submitters or manually compiled from scientific literature. Since submitters or authors often differ in their assessment criteria of variant pathogenicity, the quality of entries in these databases is highly heterogeneous. As a result, it is expected that a proportion of pathogenic variants in ClinVar may be false positives that have been misclassified (MacArthur et al., 2014). Further, once data has been submitted, users seldom go back to update results in light of new cases or information.

At the 2018 American College of Medical Genetics and Genomics annual meeting, strong warnings were issued against using ClinVar exclusively for genetic interpretation because ClinVar submitters solely base their interpretation on ClinVar-provided evidence, thereby biasing the interpretation value of ClinVar. Instead, clinicians were urged to continue using ClinVar, but to do so in conjunction with other tools and databases, such as HGMD.

These observations further support the importance of QIAGEN's effort to compile a high-quality, gold-standard knowledge base that increases the accuracy of genetic testing. Unlike ClinVar, the QIAGEN Knowledge Base enables the clinical interpretation of genetic variants according to the 2017 American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) recommended guidelines for standardized interpretation of sequence variations. To better describe the causality of variants identified in genes associated with Mendelian diseases, the ACMG and AMP recommend a widely used five-tiered categorization system — pathogenic, likely pathogenic, uncertain significance, likely benign, and benign — for classifying variants. The QIAGEN Knowledge Base generates versioned and reproducible criteria, based on a meticulous search of published/unpublished clinical evidence, for each variant and helps human interpreters quickly understand the clinical significance of genetic variants.

In 2013, Peterson et al. conducted a review that compared current human variant resources and found that of the 67,555 ▷

variants described in ClinVar, only 13,465 were classified as clinically significant. At present, the majority of ClinVar's 110,000 recorded variants are not clinically significant. Approximately 30,000 variants are labelled as clinically significant, but this clinical significance is determined and defined by the submitter. In other words, ClinVar staff do not review all the submissions and most of the content cannot be applied to a clinical setting. On the other hand, with an in-house expert curation team, the QIAGEN Knowledge Base supports consistent and credible variant interpretation. Further, the QIAGEN Knowledge Base is the sole proprietor of HGMD Professional, a database that lists only the mutations that have been shown to be disease-associated. Whereas ClinVar accepts all variants, regardless of whether or not they are pathogenic or benign, the QIAGEN Knowledge Base provides sub-resources that let users quickly look up published mutations known to be associated with a particular gene or disease. This feature saves considerable time and increases the reliability of the test report.

Reliability is an essential criterion for clinical genetic testing. Reported results are used to make a prognosis, confirm a diagnosis, adjust a medication dose, make treatment decisions, and monitor therapy. Clinicians need evidence sources to be managed in a transparent, yet usable framework. The QIAGEN Knowledge Base supports clear, evidence-based variant classification. All submitted curation undergoes an extensive review process, allowing curators to improve with feedback. In addition, automatic quality control steps detect systematic errors during and after the curation process. This rigorous quality assurance procedure is a defining hallmark of QIAGEN.

Although the QIAGEN Knowledge Base has a significantly greater clinical depth than that of ClinVar, the database has advanced search functions to simplify navigation. The variant bibliographies contain at least 90% of the relevant references found with Google Scholar and deliver a significantly higher number of references than that of Google Scholar or manual searches. ClinVar helps users annotate variants, but the database lacks curated content from primary literature. This sticks users with the time-consuming task of searching articles to fully classify variants, which becomes especially problematic for workflows that incorporate ACMG/AMP guidelines for variant interpretation, since these guidelines require information that public sources cannot provide, such as co-segregation
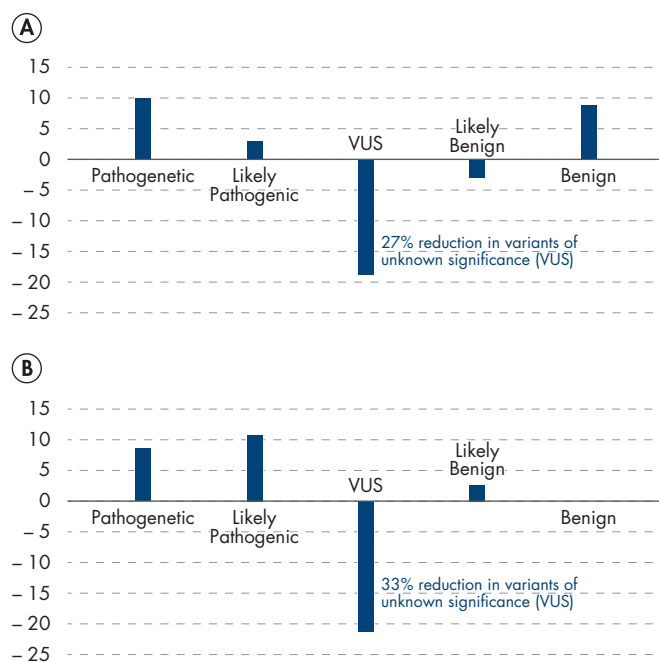

Figure 2. Variant classifications for Lynch syndrome A and cardiology disease using QCI™ Interpret B with and without QAIGEN curated primary literature. Data is represented as the difference between variant classifications with QIAGEN content minus variant classifications without QIAGEN content (i.e. public data sources only).

data, de novo status, co-occurrence with other pathogenic variants, functional study data, and real-world affected/ unaffected clinical cases. Perhaps more limiting, ClinVar does not update the clinical significance of a variant, even if new evidence changes the classification. Always current, the QIAGEN Knowledge Base continuously updates variant classifications to reflect the latest, peer-reviewed research.

To demonstrate the challenge of interpreting variants without using peer-reviewed literature, a benchmark study classified 279 randomly selected variants associated with Lynch syndrome (n=180) or cardiology diseases (n=99) with and without curated content from primary literature. The number of variants classified as having unknown significance (VUS) based on ACMG/AMP guidelines was 27-33% lower in the datasets interpreted with primary literature than in the datasets relying only on public sources (Figure 2). Further, in a 2017 evaluation conducted by Counsyl, the QIAGEN Knowledge Base reduced the median search time per 1,000 variants by more than 75% while maintaining a 98.8% concordance with their in-house manual curation efforts (Karimir et al., 2017). In addition, QIAGEN

Knowledge Base 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, the QIAGEN Knowledge Base 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 the QIAGEN Knowledge Base mitigates the occurrence of clinical knowledge "blind spots". The QIAGEN Knowledge Base is the most up-to-date, trusted, and complete collection of known, published, as well as unpublished pathogenic gene lesions responsible for human inherited disease.

The QIAGEN Knowledge Base strives to identify 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 can be a valuable resource, as long as users understand the limitations.

In today's fast-paced clinical setting, the best practice for NGS interpretation would be to use the QIAGEN Knowledge Base, as it contains multiple private and public data sources, including ClinVar. The QIAGEN Knowledge Base can lower your risk of misinterpretation, accelerate your research and improve clinical outcomes.

The QIAGEN Knowledge Base is available 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, you can deliver better care with better knowledge.

References

1. Densen, P. Challenges and Opportunities Facing Medical Education. Transactions of the American Clinical and Climatological Association. 2011;122:48-58.

2. "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things," EMC Digital Universe with Research and Analysis by IDC, April 2014.

3. Karimi K et al. Evaluation of QIAGEN Clinical Insight as a source for variant curation in a CLIA laboratory. American Society of Human Genetics 2017.

4. MacArthur D.G., Manolio T.A., Dimmock D.P., Rehm H.L., Shendure J., Abecasis G.R., Adams D.R., Altman R.B., Antonarakis S.E., Ashley E.A. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014;508:469–476

5. 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.

6. 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**.

7. 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.

Ordering and Technical Support **bioinformaticssales@qiagen.com** | Website **www.qiagenbioinformatics.com**