



Genomic Variant Data Sharing Gains Support; Collaboration Seen as Key to Interpretation Challenge

May 02, 2016 | [Turna Ray](#)

Premium

NEW YORK (GenomeWeb) – Genomic sequencing performed as part of the BabySeq project recently revealed a puzzling variant in a newborn.

Researchers working on the project performed sequencing on the baby and identified an alteration in RYR1, a gene that codes for a protein present in skeletal muscle. Variants in this gene have been linked to congenital myopathy, a kind of muscle weakness present at birth, and RYR1 variants are also known to cause malignant hypothermia, a condition set off by anesthesia that can lead to heat stroke, infections, and even death.

But not all pathogenic variants cause both conditions. Trying to unravel how this particular variant might affect the infant's health, Heidi Rehm, director of the Laboratory for Molecular Medicine at the Partners HealthCare Personalized Medicine, turned to ClinVar, a freely available archive of genotype and phenotype relationships the NIH publicly launched three years ago.

A submission in ClinVar suggested that the variant increased the risk for malignant hypothermia. However, when Rehm's staff looked further into this submission, it turns out that the cited papers ruled out the risk for myopathy for the child, but didn't specifically link the variant to dominant malignant hypothermia.

So, she corralled the lead investigators in the BabySeq project and Les Biesecker, chief of the medical genomics and metabolic genetics branch at the National Human Genome Research Institute, who had submitted the variant to ClinVar on behalf of another project. They looked through the literature, debated about the published evidence, and consulted specialists on malignant hypothermia. The researchers couldn't pinpoint any studies where loss-of-function variants in RYR1 (which may diminish or stop normal protein function) were linked with the condition. Still, some experts thought it best to play it safe and inform the baby's doctor in the report that RYR1 gain-of-function variants (which may give rise to new kinds of protein activity) do put people at risk for malignant hypothermia.

Others disagreed. Within BabySeq, researchers are enrolling healthy and sick infants at two Boston hospitals and studying the risks and benefits of returning genomic sequencing results for childhood onset diseases to physicians. And so, researchers have to be careful not to raise unnecessary alarms, since the study is providing disease-related DNA information for not only sick babies, but also healthy ones. "We debated for quite a while," Rehm told GenomeWeb. "We debated whether to even comment on it [in the report], because then you raise issues, and people get nervous who are otherwise healthy."

After much consideration over the span of a week, Rehm and her colleagues agreed to state in the report that the baby was a carrier for a recessive RYR1 mutation for myopathy, which wouldn't affect the child, but they also explained that they did not find any evidence that this variant would put the baby at risk for malignant hypothermia. The language was deliberate. "We didn't have evidence against that fact," Rehm said. "We just appreciated that there was no evidence for that fact, and we did not want physicians jumping to their own conclusions without being informed."

After sharing and discussing the available evidence on this variant among peers, Biesecker is planning to change the condition linked to the variant in ClinVar from malignant hypothermia to myopathy.

The painstaking work required to interpret this single alteration highlights the promise as well as the challenges of public databases like ClinVar. While these databases are gaining recognition as vital resources for aggregating current knowledge on genetic variants, they should only be considered the first step in the clinical interpretation process. Labs that rely on them must not only be aware of the information gaps they contain, but must also be willing to contribute new knowledge on an iterative basis.

The discrepancy in the BabySeq project was a minor one in ClinVar, according to Rehm, but there are other far more potentially significant interpretation ambiguities in the database. "The genetic tests that we all run are useless if we're giving wrong answers, and could be harming patients," Rehm said. "We know labs are interpreting things differently, which means that at least some patients are getting the wrong answers. So, we clearly have a problem. But evaluating evidence on variants is challenging and labor intensive."

Rehm wants to crowdsource the solution. "We're each evaluating thousands of variants. If we share what we're doing, it gets each of us quicker to the answer by having a baseline evaluation already done," she said.

Outside the research setting, a few commercial labs have already tried this out. When experts from Rehm's Laboratory of Molecular Medicine, the University of Chicago, and commercial labs Ambry Genetics and GeneDx assessed more than 6,000 variants that at least two labs had submitted to ClinVar, they found their classifications agreed for 88 percent of variants. Of the approximately 724 variants for which they had conflicting classifications, the four labs over several months discussed the evidence on 232 variants and reached a consensus on 86 percent. Despite their efforts they still couldn't agree on 33 variants because of how each lab was weighing the available evidence in the literature and applying information from functional studies. These labs have yet to tackle 492 variant discrepancies within this project.

If anything, this exercise demonstrates the hard work that lies ahead for the genetic testing industry,

given there are around 20,000 genes and innumerable variants. Researchers and labs around the world will have to powwow over many more uncertain variants that crop up in patients' reports. And still, as the work of the four labs shows, they may not agree whether a variant is benign, pathogenic, or of uncertain classification based on the evidence available at the time.

As such, there are labs that maintain that public data repositories, such as ClinVar, are not ready to be used in clinical decision making. "The use of a database that by peer-reviewed publication has been shown to have a 17 percent discordance — meaning that if you pull out the results 100 times, 17 of those times you're going to have a different result based on the lab — no lab should be using any of that information in any clinical decision because it's not meant for that," said Johnathan Lancaster, chief medical officer at Myriad Genetic Laboratories, which is not submitting to ClinVar.

Lancaster was referring to a [report](#) Rehm penned a year ago in the *New England Journal of Medicine*, where she discussed how ClinGen, another NIH-supported project for which she's a lead investigator, was establishing expert working groups to develop standards and resolve the discrepancies in ClinVar. To illustrate the critical need for this work, she noted that in ClinVar, of the 12,895 clinical interpretations that were submitted by multiple labs at the time, 2,229, or 17 percent, had conflicting interpretations.

Across the healthcare sector, within regulatory bodies, among editors of peer-reviewed journals, and even among insurers, there is growing recognition of the public's right to be aware of the gaps in genomics knowledge, as much as the advances. "There's the recognition now that wasn't there years ago that we're sometimes classifying variants differently," Rehm said. "There's the recognition that this is really challenging work and we need to work together to do it better."

Toward transparency

Rehm is leading the data-sharing charge when it comes to genomic variants. She is working to get accrediting organizations like the College of American Pathologists (CAP) to require that labs submit their variant interpretations to open databases as a way to check quality. She wants journals to ensure that authors deposit variant data into public repositories before publishing their papers. She has urged insurers to mandate that labs share data on variants before granting coverage. At US Food and Drug Administration meetings, she has recommended that the agency regulate tests from labs that don't share variant interpretations as conferring higher risk to public health.

“ We've really set ourselves up where we're lacking the expertise, we're inundated with genetic variation, and what's being impacted is the quality. ”

Her urgings come during a period of rapid growth in the genetic testing industry. According to healthcare technology firm NextGxDx, there are more than 60,000 commercially available genetic tests in the US, accounting for more than half of the lab testing market. Taking advantage of dropping sequencing costs, labs are launching next-generation sequencing tests that can analyze dozens, if not hundreds, of genes at once. NextGxDx has tracked more than 18,800 NGS products in the genetic testing market, of which around 2,000 assess multiple genes.

A decade ago, genetic testing was limited to a handful of niche labs around the country with expertise in specific diseases. But availability of NGS changed the market dynamics.

"Given the volume of genetic testing happening now and the size of these tests [in terms of the number of genes], there's just an intense amount of variation that we're trying to evaluate without many resources," Rehm said. "And there is pressure to provide cheaper and cheaper testing and to cut costs. So, we've really set ourselves up where we're lacking the expertise, we're inundated with genetic variation, and what's being impacted is the quality."

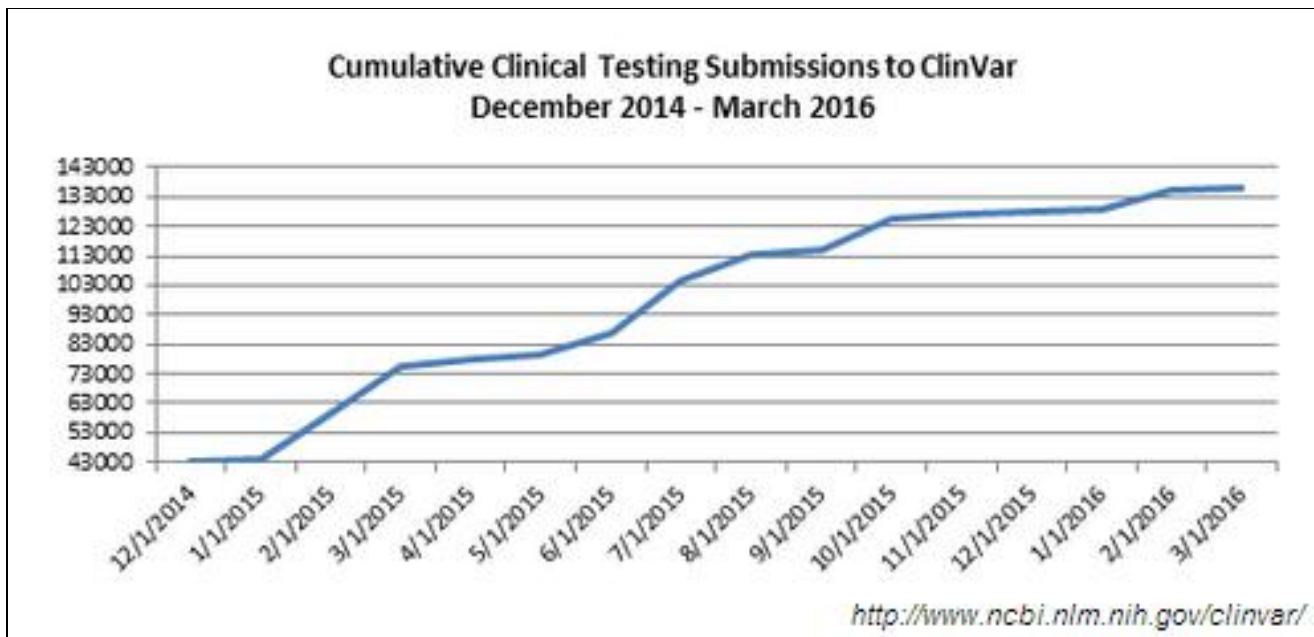
Citing these market changes, the FDA has said it wants to regulate all lab tests. Simultaneously, the agency has been working with commercial and academic labs to come up with new ways to regulate NGS panels, and variant databases [are a key consideration](#) in that effort. Two years ago, the FDA granted marketing clearance for Illumina's 139-variant cystic fibrosis assay using Johns Hopkins University's CFTR2 database to establish the validity of the markers. Since then, the agency [has said it is open](#) to approving or clearing other NGS panel tests using well-curated databases.

"We believe that aggregating information about variants can be highly useful for more rapid understanding of variant meaning, and may support regulatory authorization of tests when the data are managed and interpreted in a manner that applies good quality processes," FDA spokesperson Lyndsay Meyer told GenomeWeb.

The agency further said it encourages labs and researchers to share data about variants. "In many cases, this will allow aggregation of data to support well-informed interpretations and conclusions about the effects of variants on health," Meyer said.

Rehm also reached out to CAP two years ago to see if it would require labs to submit to ClinVar and add this to its lab quality checklist. Because ClinVar was still a new endeavor back then, Rehm didn't make much headway with CAP, but many more labs are now submitting to ClinVar.

From December 2014 to March 2016, ClinVar received more than 92,000 submissions, averaging around 6,000 submissions per month. Commercial labs, such as GeneDx, Invitae, and Ambry Genetics are among its top contributors. The two largest reference labs in the country, Quest Diagnostics and Laboratory Corporation of America, are also submitting. As of March, there were close to 180,000 submitted records, and eight of the top 10 submitters are clinical labs that have deposited about half those records.



Armed with the latest statistics, Rehm recently sent a formal letter to CAP in an effort to revive discussions. CAP accredits thousands of labs in the US and abroad. According to the organization's 2015 [annual report](#), a significant portion of its 2015 revenues — \$172.9 million of \$186.4 million — come from lab improvement programs, including funds generated from accreditation work. If an entity with CAP's influence began requiring that labs share variant data as part of the accreditation process, it would be a watershed moment. "It would show labs that the professional community feels this is an important issue, and make it easier to get labs to adopt," Rehm said.

Meanwhile, insurers are also beginning to recognize the importance of sharing data on variants. "Insurers are starting to ask questions of labs as they contemplate whether to reimburse them," Rehm said.

The molecular diagnostics industry has faced a difficult reimbursement environment in recent years. Labs have complained that the payment they're receiving from insurers is well below the list price and often below the cost of performing such tests. Insurers have countered that many of the tests marketed by labs, particularly sequencing panels that test for many genes, lack evidence showing they are medically necessary.

At least one national insurer, Aetna, is requiring that all newly contracted labs submit their variant interpretations to ClinVar as a condition of participation in its network for BRCA1 and BRCA2 genetic testing. "It is part of a quality review of the laboratory as a condition of contracting," Joanne Armstrong, Aetna's senior medical director, told GenomeWeb.

The market for BRCA testing for hereditary breast and ovarian cancer became hyper-competitive after the US Supreme Court invalidated several patent claims held by Myriad Genetics. Before the 2013 landmark [decision](#), which deemed isolated gene sequences patent ineligible, Myriad held a monopoly over the BRCA testing space for nearly two decades. Now, according to NextGxDx's estimates as of February, there are 335 commercially available tests that assess BRCA1/2 genes individually or as part

of a panel.

BRCA1 and BRCA2 are highly susceptible to mutation and many of the variants are exceedingly rare, making it challenging to determine their links to cancer. Since Myriad has been testing these genes the longest, it has detected many of these rare markers in patients and has collected information on 17,000 BRCA variants in its proprietary database, which contains 46,000 variants in total. Many competing labs, however, are interpreting what these variants mean for patients' health with far less data.

"The evidence available to determine whether a variant is classified as deleterious or not deleterious is often not published [or] publicly available and thus can vary across laboratories," Armstrong said over email, adding that making variant classification information public for researchers and others to use serves patients interests.

Currently, Aetna's requirement to share BRCA variant classifications in ClinVar only applies to labs that it signed contracts with after Jan. 1, 2015. Labs that inked agreements before this date wouldn't include this provision, "but we are working to include it," Aetna officials told GenomeWeb.

Finally, peer-reviewed journals, particularly genomics-focused publications, have historically encouraged authors to share research data, including variants. Richard Cotton, a pioneer in developing methods for detecting genetic mutations and longtime editor-in-chief of the journal *Human Mutation*, was among the early champions of creating a freely accessible central resource for cataloging all variation data. Cotton, who passed away last year, chaired a meeting in 1994 in Montreal, where top geneticists from around the world gathered to discuss the need to systematically collect this information.

But this led to the development of gene-specific databases, in which collection and curation was fragmented. There was another meeting in June 2006 in Melbourne, Australia, in which attendees, again with Cotton's leadership, recognized that researchers needed to be able to search a "complete database," and launched the [Human Variome Project](#) aiming to advance nomenclature standards and build the infrastructure for sharing variant data. It was important to the meeting participants even then that variant data not be siloed, that diagnostic labs share interpretations, and that peer-reviewed journals encourage such activity.

Given this history, *Human Mutation* has [required](#) for some time that authors submit all variants featured in an article into databases before it is accepted for publication — and many journals now require or recommend the same.

Two years ago, the *European Journal of Medical Genetics* began requiring that authors publishing exome sequencing and clinical data from a single patient also submit this information to the Wellcome Trust Sanger Institute's DECIPHER database. Cold Spring Harbor's recently launched journal *Molecular Case Studies* [instructs authors](#) to submit data on interpreted variants into ClinVar and, for rare disease cases, to deposit information on candidate genes into another project, called Matchmaker Exchange. A spokesperson for the *New England Journal of Medicine* told GenomeWeb that currently the journal encourages authors to submit variants to public repositories by publication time, and it is in the process of "implementing mechanisms" to make this a requirement.

Genetics in Medicine has expected authors to share variant data since at least 2008, around the time the American College of Medical Genetics and Genomics published its 2007 standards for variant interpretation. ACMG updated these [guidelines](#) in 2015, further refining the types of evidence labs and researchers should consider and how to weigh the evidence when classifying a variant into one of five categories — pathogenic, likely pathogenic, benign, likely benign, and uncertain. The authors of these guidelines, which include Rehm, also encouraged labs to work with clinicians and others in industry to resolve interpretation differences and submit to ClinVar "to aid in the continued understanding of the impact of human variation."

As the official journal of the ACMG, last year *GIM* instituted a [mandatory checklist](#) for authors "to ensure good reporting standards and improve the reproducibility of published results." Among other things, the list asks authors to provide accession codes for data they've submitted to public repositories, including protein, DNA, and RNA sequences.

Reviewers look over this checklist during the article revision phase and follow up with the authors. If they are unwilling to provide information that shows they've deposited variant information in a freely accessible repository, then the article will not be published. It hasn't come to that at *GIM*.

"In practice, we have found our reviewers and editors are very well aware of these requirements [on variant submissions] for authors," Jan Higgins, managing editor at *GIM*, told GenomeWeb. For a few recent manuscripts, reviewers even wanted "clear indication" from the authors that they had placed variant data in a public database. "Those assurances were received from the authors," she said.

Evolving truth

“ There is no true truth for most variants. Most interpretations are expert opinions based on an evaluation of evidence.

In a January *Nature Methods* [editorial](#), Rehm is quoted saying, "There is no true truth for most variants. Most interpretations are expert opinions based on an evaluation of evidence." ClinVar, she hopes, can become a dependable framework for tracking the evolving truth on variants.

”

As of April 25, more than 181,000 ClinVar records [have been deposited](#) by over 500 submitters. Among these, there are 127,000 unique variants with interpretations.

A tool called [Variant Explorer](#) provides users a snapshot of the discrepancies in ClinVar. Based on information submitted as of March 2016, there are more than 2,400 "medically significant" differences in classifications across 138 labs. Although approximately 500 submitters have deposited data into ClinVar, they have done so on only a few variants, usually in newly discovered genes that other labs haven't submitted on.

Rehm explained that medically significant differences are when one lab deems a variant likely pathogenic or pathogenic and another lab says it's of unknown significance, likely benign, or benign. The number of overall discrepancies exceed 8,700 if one factors in all the times a lab calls a variant likely benign or benign and another lab says it's of uncertain significance. But, such differences are less likely to impact medical practice, Rehm explained. ClinVar doesn't represent a variant's classification as

conflicting if the difference is in the level of confidence (likely pathogenic versus pathogenic or likely benign versus benign).

The very purpose of ClinVar, say those involved in the effort, is to shed light on these differences, so labs can work together to resolve them. To help labs do just that, the NIH in 2013 launched ClinGen, a project to establish an authoritative source of information on genes and variants that can be used in research and to personalize care. ClinGen's expert panels review data in ClinVar and submit their own interpretations. Groups working on these two programs have come up with a four-star system as a way to communicate to users how extensively a variant has been reviewed.

Around 59,500 ClinVar records currently have one star, which means that the submitting lab has provided the criteria by which they arrived at a particular clinical determination for a variant. There are approximately 8,300 two-star records submitted by multiple labs, with underlying criteria, and no conflicting classifications. More than 3,600 records carry three stars, meaning they have been reviewed and agreed upon by expert panels comprising researchers and representatives from labs with extensive experience in a particular disease. Johns Hopkins University's CFTR2 database, which the FDA used to clear Illumina's cystic fibrosis NGS panel, is an example of an expert panel. Finally, 23 records have associated practice guidelines and therefore have four stars.

But because ClinVar is very much a work in progress and the areas of disagreement between submitters are openly available for anyone to explore, some labs have refused to participate in the effort. Myriad's Lancaster described the company's decision to stop contributing to public variant databases as they currently exist as a moral and ethical stance.


Older public variant repositories have a reputation for lacking standard interpretation criteria and for not being kept up-to-date. There is published data that such databases are "very, very, very subpar," Lancaster said. "It's incomprehensible to me that someone would rely on those sorts of muddy, murky pools of water to make clinical decisions."

A few years ago, when researchers [re-evaluated 239 variants](#) classified as pathogenic in the Human Gene Mutation Database, they found that only 7.5 percent of the variants were actually pathogenic.

Researchers from Myriad [published a study](#) last year in which they evaluated more than 2,000 BRCA1/2 variant classifications across five public databases, including ClinVar. For the 116 variants that were in all the databases and deemed pathogenic in at least one, the authors reported that pathogenic classifications agreed for only four variants.

Instead of submitting to these databases, Myriad is contributing to advancing science through the more well-established peer-reviewed publication process, according to Lancaster. "Myriad does and has shared data," he said, and noted that Myriad was a founding member of the Breast Cancer Information Core database and has been one of its biggest contributors. However, Myriad stopped contributing to BIC in 2004, due to the lack of standards and poor quality.

“ There is no debate within the scientific community about what truly represents peer-reviewed publication.

Myriad is committed to sharing data "in a responsible way," Lancaster said, by continuing to publish in the peer-reviewed literature "not just specific variants and how they're classified, but on the science of variant classification." Myriad estimates that it has shared its classifications for more than 8,000 variants either through publications or databases when it was still submitting to them. 

The firm has published extensively on the analytical and clinical validation of its tests, as well as on its variant classification methodology. Recently, Myriad joined top cancer centers and other commercial labs in advancing PROMPT, a registry that collects information from patients who have been tested on multi-gene panels for cancer risk. In 2014, Myriad garnered FDA approval for BRACAnalysis (its flagship test for nearly two decades) as a companion diagnostic. As part of that process, the [agency evaluated](#) Myriad's variant classification process and a subset of variants from its proprietary database. In this way, Lancaster asserted that Myriad has been subject to peer review and even the FDA's scrutiny much more so than other labs in the genetic testing industry.

"We are far more interested in advancing the field of variant classification, so that ultimately, hopefully one day, every human gene in the genome will have been subject to the same ongoing scrutiny that BRCA1 and BRCA2 have been subject to, so there is no longer any need for variant classification, because all of that information will have been catalogued," Lancaster said.

But if labs have to funnel every new, rare, and previously unseen variant through the peer-reviewed publication process, then journals have a data tsunami coming their way. This is why supporters of ClinVar believe a public repository is needed and liken the ClinGen-supported process to collaboratively resolve discrepancies as a form of peer review.

"As an editor of a clinical genetics journal, I'm all for the peer-review process for manuscripts," said James Evans, editor-in-chief of *Genetics in Medicine*. "But it's utterly naïve to think that peer review of that sort will be up to the task of adjudicating the almost infinite number of variants that exist in the global human genome. The only way forward is by pooling data in widely accessible, expertly curated databases that allow for the entire community to arrive at consensus. Efforts like ClinVar and ClinGen are precisely what the field needs."

Lancaster likened sharing data in public variant repositories to dumping data in Wikipedia. And while it may be that labs are working to resolve discrepancies in ClinVar, he said that as a physician he would not make medical decisions for patients based on information in a database that is, by definition, a work in progress.

"At the end of the day, what it always comes back to is what's best for patients today with the available tools we have," said Ron Rogers, Myriad's executive VP of corporate communications. "Our view is that the public databases are just not good enough to make life-changing decisions based on them."

"There is no debate within the scientific community about what truly represents peer-reviewed publication," Lancaster added. "If they want to use those databases and improve them with time and effort to the point where they are robust enough to be used clinically, go ahead, do that, publish the literature that shows that they are ready for prime time."

The journals, of course, aren't pristine either, and the peer-review process isn't airtight against errors. Rehm noted that 35 percent of medically significant differences in ClinVar are the result of information taken from a compendium of genes and phenotypes, called the Online Mendelian Inheritance in Man, which draws variant data directly from the published literature.

At the same time, Rehm doesn't deny that variant interpretation is a challenge for the genetic testing industry and that inconsistencies in classification can harm patients. Nevertheless, she believes ClinVar can be a valuable resource for regulators, insurers, doctors, and patients who have to make critical decisions amid evolving and imperfect science.

Although Myriad is not submitting to ClinVar, genetic counselors and doctors GenomeWeb spoke to said they trust the firm's variant interpretations compared to newer labs in the space, particularly given Myriad's lengthy experience testing BRCA1/2 genes. Frederica Lofquist, a physician with Pacific Women's OB/GYN Medical Group in San Francisco, said she uses Myriad for hereditary cancer testing precisely because the firm has a "shocking amount of data" and is transparent about its variant classification process. Lofquist, who has consulted for Myriad, said she doesn't have a good handle on how other labs are classifying variants. "Myriad does have the advantage in that they've been doing this the longest and [have] the most experience," Lofquist said, though she added, "This is a very developing field."

Rehm pointed out that labs, including Myriad, are using NGS to assess a large number of genes, and for many of these there is much less data than there is on BRCA1/2. "It is possible that there is a correlation between labs doing a poor job in interpretation and not wanting to publicly share their interpretations," she said, adding that doctors would be better off going to a lab that's willing to put its variant classification up for public scrutiny.

Ultimately, she believes ClinVar is improving the quality of variant interpretations in the genetic testing space. ClinGen is developing an interface that expert groups can use to more easily interpret variants and resolve conflicting interpretations in ClinVar. There are ClinGen working groups focused specifically on variants related to heart disease, hereditary cancer, metabolic disease, and other conditions. Bioinformaticians also are developing approaches to make variant classification work more streamlined. And there are efforts to encourage patients to share their own genetic test results and phenotypic information into a registry called GenomeConnect.

In April, there were between 16,000 and 18,000 daily hits on ClinVar. Doctors and genetic counselors are starting to check variant classifications in test results against ClinVar and calling labs for explanations if there are interpretation conflicts. In turn, labs are downloading ClinVar's monthly discrepancy reports to identify variants that need greater review in their internal databases. "It is the process of transparency by submitting to ClinVar, performing comparisons between labs, and putting your variant interpretation out there that is helping make this happen," Rehm said.

This is the first article in a series exploring efforts to improve the quality of variant interpretations in genomic testing.