

ORIGINAL ARTICLE

Clinical reassessment of post-laboratory variant call format (VCF) files

Lamia Alsubaie¹, Saeed Alturki², Ali Alothaim², Ahmed Alfares^{3*}

ABSTRACT

Background: Next-generation sequencing has been leading the genetic study of human disease for the past 10 years, generating a huge amount of sequence variant data, which are stored in variant call format (VCF) files. The aim of the study was to reassess the utility of VCF files for reanalysis.

Methodology: This is a descriptive observational study of Saudi patients with undiagnosed genetic conditions. VCF files from 20 samples were referred to the molecular laboratory by physicians for reanalysis using variant interpretation software.

Results: Seven cases ($n = 20$) have been reported differently from the outside laboratory. This accounts for almost 35% of all cases and is mainly due to the ability to gather more information about the patient's phenotype. One whole-genome sequence (WGS) case changed from inconclusive to negative. In addition, we identified variants related to the patient's phenotype in six cases; two of them were WGS and four were whole-exome sequence, all reported as negative before the reanalysis.

Conclusion: Comprehensive phenotyping of individuals is a crucial step in identifying candidate phenotype-related variants. We outline the benefit obtained from access to the patient's medical records and communication with referring physicians.

Keywords: Report, variants, classification, VCF, reassessment, genomic, genetic, counseling, NGS.

Introduction

DNA sequencing technologies are currently being developed at an incredible speed to identify genetic variants that underlie Mendelian as well as complex disorders. While the sequencing technology has shifted to become more routine, the analysis and interpretation of the genetic variants found by this technology are still challenging tasks for clinical laboratories. The variant interpretation of a personal genome, which was recently reported to differ from a reference human genome by about 3.5 million Single nucleotide polymorphisms (SNPs) and 1,000 large (>500 bp) Copy number variation (CNVs) (1), depends on our skills at associating genotype and phenotype. Previous studies have outlined the importance of many factors in identifying candidate phenotype-related variants, such as phenotyping (2), population allele frequency, and the exploration of the medical and scientific literature (3).

The American College of Medical Genetics and Genomics (ACMG) recommendations for variant interpretation and reporting in 2000 (4) [further updated in 2007 (5), 2013 (6), and 2015 (7)] enhance our ability to use genetic data generated by sequencing to diagnose genetic diseases. However, despite the availability of sequence variation, geneticists and genetic counselors still face difficulties with clinical validity in some cases.

Multiple factors that aim to highlight possible causative variants have been discussed in the literature. For instance, allele frequency databases, especially ethnically matched ones, may help to avoid the false negative and/or false positive variants that may contribute to the patient phenotype (8). A few published articles have investigated these factors and their impact on variant classification (5). Shearer et al. (9) utilized ethnic-specific differences to report pathogenic variants related to non-sensorineural hearing loss. They found that 4.2% of the reported pathogenic variants were benign (10). In addition, a previous study found that 27% of the disease mutations reported in the literature lack strong evidence favoring variant pathogenicity (11).

Correspondence to: Ahmed Alfares

*Department of Pediatrics, College of Medicine, Qassim University, Qassim, Saudi Arabia.

Email: fars@qu.edu.sa

Full list of author information is available at the end of the article.

Received: 14 October 2017 | **Accepted:** 02 December 2017

Geneticists and genetic counselors recognize that these elements change over time; hence, the classification of a novel variant should be reassessed, and its causative factors reconsidered. Genomic sequencing is expected to impact the role of genetic counselors; gathering clinical information will not be restricted to clinical genetic counselors and likewise, clinical genetic counselors will be expected to perform variant interpretation, clinical validation, and possibly other lab-oriented tasks in order to deliver genomic counseling (i.e., helping patients deal with the implications of genomic sequencing information).

Materials and Methods

We retrospectively analyzed 20 variant call format (VCF) files that were provided to us by referring physicians for further assessment because the whole-genome sequence (WGS)/whole-exome sequence (WES) test results could not explain the patient's phenotype. All patients consented to and were informed about the next-generation sequencing (NGS) and the different possible results: positive, negative, or inconclusive.

The VarSeq tool from Golden Helix (<http://goldenhelix.com>) was used for VCF file analysis and variant interpretation. Filters were used to narrow the list of candidate variants. All DNA variations were assessed through the ClinVar database (12) for clinical significance. All pathogenic and likely pathogenic variants were investigated to rule out any variant related to the study's cases. Then, all data were filtered using a chain of filters that included inheritance, variant type and variant effect [such as loss of function (LOF) or Missense], and prediction tools. Autosomal recessive inheritance and the homozygous state were at the top of the filtration chain because of a previous study by Alfares et al. (13), which found that up to 84% of positive cases from NGS in Saudi Arabia who reported consanguinity were homozygous autosomal recessive disease-causing variants. The last step before reporting was a genotype–phenotype correlation using genetic databases.

Once the candidate variants were identified, the questionable variants were investigated individually in different public genomic interpretation databases. They were then compared with an ethnically matched proprietary database. The ACMGG and Association for Molecular Pathology (AMP) criteria for variant assessment were used to classify each variant. In this classification, each criterion is weighted as very strong (PVS1), strong (PS1–4), moderate (PM1–6), or supporting (PP1–5). Note that the numbering does not convey any difference in weight (7).

This study was approved by the Institutional Research Board of King Abdullah International Medical Research Center. The data collection and analysis were conducted retrospectively at King Abdulaziz Medical City. Informed consent was obtained from all individual participants included in the study.

Results

Each case was investigated, and changes were documented at both the variant level and report level. The samples ($n = 20$) were divided into 14 WGS cases (70%) and 6 WES cases (30%) (Figure 1). All WGS cases had a previous negative result from an outside lab, except for one case in which the previous result indicated a variant of uncertain significance. Likewise, all WES cases had a previous negative result from an outside lab, except for one case in which the previous result indicated two variants; one as a variant of uncertain significance and one as likely pathogenic (Table 1).

Out of the 20 cases, we reported seven cases differently from the outside laboratory reports, which accounts for almost 35% of the total sample. Hence, variant reassessment results were concordant in 65% of the sample. The WGS case with the previous variant of uncertain significance report was reported after reanalysis as negative. The inconclusive outside lab report was regarding a variant that was possibly consistent with a genetic diagnosis of Loeys–Dietz syndrome. The patient had a heterozygous variant of uncertain significance in the *TGFBR1* gene (c.1433A > G). However, the Loeys–Dietz syndrome did not match the patient phenotype and the asymptomatic father carried the same variant. Therefore, we excluded this variant as causative of the patient's symptoms.

We identified variants that were related to the patient's phenotype in six cases: two of them were WGSs with previous negative results and four of them were WESs with previous negative results. The ACMGG variant classification was reviewed and assigned to all the reported variants. For instance, case 3 had two variants of uncertain significance in genes associated with a deafness phenotype, and the patient has deafness along with other features such as developmental delay and swallowing difficulty. The first variant was identified in a heterozygous state in the *USH1C* gene, c.1812dupC (p.Ile605fs); this variant is associated with Usher syndrome and classified as pathogenic variant according to the ACMGG-AMP criteria. It is a null variant (frameshift) in a gene where the LOF is a known mechanism of disease (PVS1), absent from population databases (PM2), and its phenotype is highly specific for a single gene disorder (PP4). In addition, the second variant was found in the heterozygous state in the *OTOGL* gene, c.3809G > C (p.Arg1270Thr). This variant is associated with autosomal recessive deafness 84B and matched the criteria for a variant of uncertain significance: it was absent from population databases (PM2), computational prediction tools found that it is possibly damaging (PP3), and the phenotype is highly specific for a single gene disorder (PP4). Case 7 is a case of chronic diarrhea, ascites, arteriovenous malformation, and infantile myofibromatosis. In this case, we identified a homozygous variant in the *LCT* gene, c.4867-4G > A, which is associated with congenital lactase deficiency (OMIM: 223000). Congenital lactase

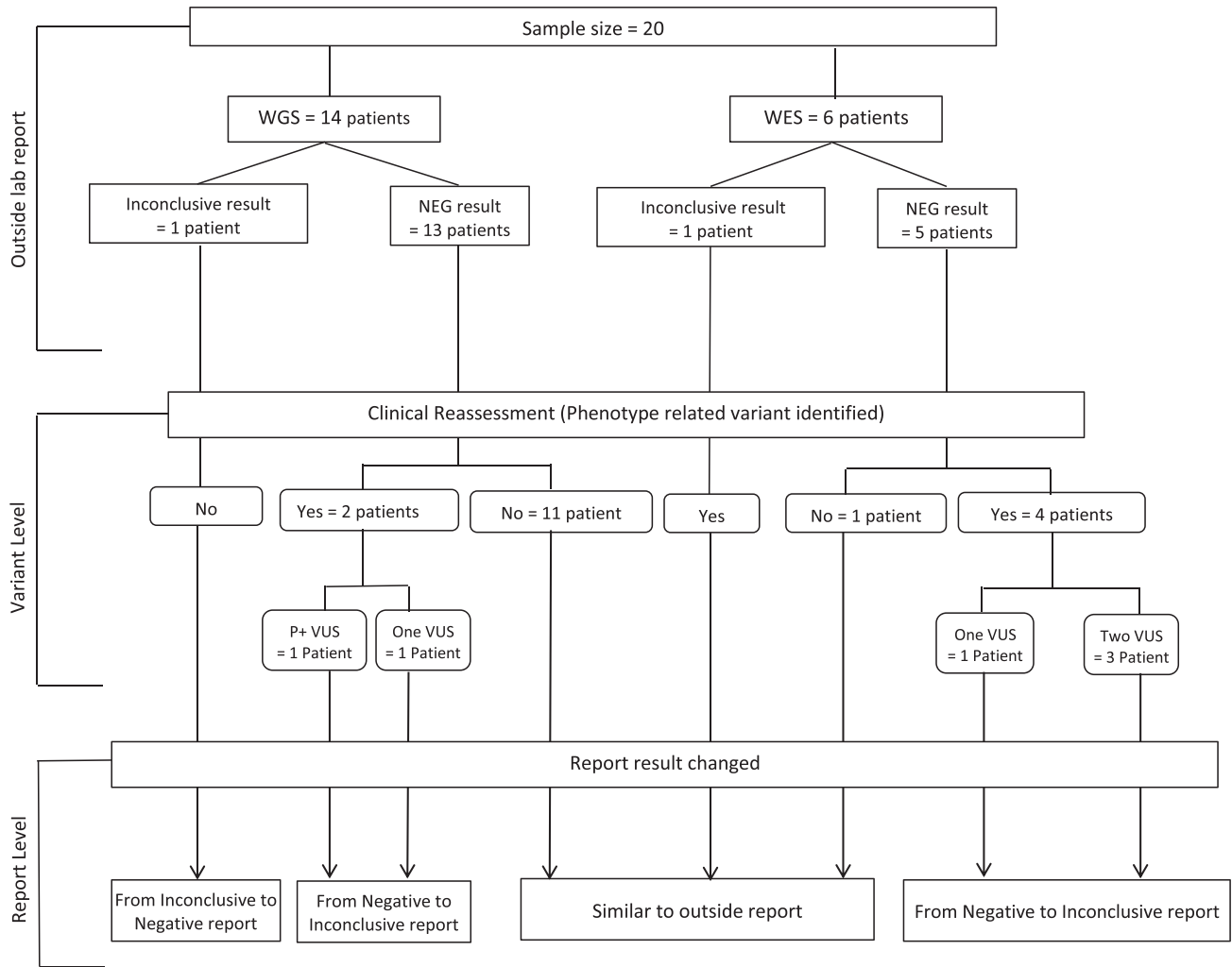


Figure 1. Flow chart showing the distribution of the cases reported by an outside lab and the reassessment process for both the variant and report levels.

deficiency is an autosomal recessive disease manifested with neonatal diarrhea, dehydration, and metabolic acidosis. The described variant is classified as a variant of uncertain significance: it is a null variant (spliced site) in a gene where the LOF is a known mechanism of disease (PVS1) and the phenotype is highly specific for a single gene disorder (PP4). The second homozygous variant addressed in this case is in the *PCCA* gene, c.802C > T (p.Arg268Cys) and is a variant of uncertain significance in the gene associated with propionic acidemia (OMIM: 606054). Our lab and the geneticist excluded the variant as a phenotype causation with several normal urine organic acid tests. However, we included this variant in the reassessment report because it has been described previously in the ClinVar database as likely to be pathogenic (SCV000343761.2) although our patient is asymptomatic (Table 1).

In five out of seven changed reports, this was mainly because of the ability to gather more information about the patient's phenotype (Figure 1).

Discussion

Previous studies on the interpretation of genomic variation have proposed three elements that cause discrepancies in reports: temporal differences, internal laboratory data, and differences in allele-frequency cutoffs (2). In our study, we reported seven cases differently from the outside laboratory reports with a concordant rate (i.e., the degree of agreement) of 65%, in contrast to a 79% concordant rate achieved by a previous intra-laboratory comparison study (14).

Our study findings match up with previous research and studies (2,15) that highlight the importance of clinical data for variant classification because nearly 70% of the modified reports were identified due to the ability to gather additional patient information. The task of obtaining and evaluating relevant patient information plays a major role in the practice of a genetic counselor, despite the ignorance of this role in the literature (16).

We reported eight variants of uncertain significance (all with heterozygous status) in our population for which the

Table 1. Variant reassessment report and variant classification for the study sample.

Inheritance	NA	AR	AR	AR	AR	AD	AD	AD	AR	AR
Diagnosis	NA	Infantile hypotonia with psychomotor retardation dysmorphic facies II	Usher syndrome 1C	Autosomal recessive deafness 84B	Guanidino-acetate methyltransferase deficiency	Blepharophimosis, Hydronephrosis	Early infantile epileptic encephalopathy VI	Con-genital lactase deficiency	Propionic acidemia	AR
Internal Classification	NA	VUS	P	VUS	VUS	VUS	VUS	VUS	VUS	VUS
Zygosity	NA	Het	Het	Het	Het	Het	Het	Het	Homo	Homo
Variant result	NA	UNC80: c.3883G>C p. Glu1295Gln	USH1C: c.1812dupC p. Ile605fs,	OTOGL: c.3809G>C p. Ar-g1270Thr	GAMT: c.619T>C p. Tr-p207Arg	SOX17: c.775T>A p. Tyr259Asn	SPTAN1: c.4309C>T p. Arg-1437Cys	SPTAN1: c.5545C>T p. Arg1849Trp	LCT: c.4867-4G>A, p. Arg-268Cys	PCCA: c.802C>T, p. Arg-268Cys
NGHA Report	Negative	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
Previous NGS Report	Inconclusive	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative
Test type	WGS	WGS	WGS	WGS	WES	WES	WES	WES	WES	WES

outside lab reports were negative. This was expected as a result of access to further support clinical information. Clearly, investigating the genotype–phenotype relation is an important step for any clinician before the patient is counseled about a variant of uncertain significance. An outside lab reported a likely pathogenic homozygous variant in the *PCCA* gene but we excluded it as phenotype causation in this patient because it did not match the described phenotype. Similar studies have emphasized the availability of clinical data to help the classification of pathogenic variants.

A variant of uncertain significance is always a changeable entity that can cause a problem from a genetic counseling point of view and distress patients and their families. However, these variations should be addressed in counseling until a classification of benign or disease-related (i.e., pathogenic or likely pathogenic) status is determined. The risk of variant over-reporting for minority populations may be discussed in genomic counseling. The Arab population accounts for 423 million people, yet it is not well represented in large-scale allele databases like ESP, ExAC, gnomAD, and many others. The recently publicly available population-specific genome for the indigenous Arab population of Qatar (QTRG) can help genetic counselors reassure their patients regarding detected variants (17). Furthermore, ethnically matched genetic databases, if available, can help classify the pathogenicity of observed variants.

In our study, an outside lab pointed out a possible causative variant (c.1433A > G) in the *TGFBR1* gene in one case, but a reassessment excluded this variant because the unaffected father carried the same variant. Lack of co-segregation for autosomal dominant disorders, and identifying the variant in affected and non-affected individuals, decreases the likelihood that a variant is pathogenic. In a study that compared variant interpretation between labs (18), segregation analysis was identified as the most commonly modified line of evidence. Moreover, other studies have changed variant classification based on co-segregation (10,19). Genetic segregation accompanied by the detailed phenotype information of all family members may be necessary to reach a diagnosis. It is not simply an issue of affected versus unaffected family members but every bit of clinical information is critical, especially in cases of complex disorders and/or adult-onset disorders.

The decision to clinically reassess post-laboratory VCF files was made because a geneticist felt the need to reassess these files, especially given our ability to gather more relevant information and the availability of a proprietary genetic variant database. Reassessing variants is important, especially as we accumulate additional knowledge about them. This research highlights the need for reassessment and the multiple factors that may affect variant analysis to ensure proper communication during genomic counseling. Moreover, we determined factors that have more impact on variant pathogenicity assessment in under-represented racial populations.

There are some unavoidable limitations in this study. For instance, the study was conducted on a small number of cases that were referred for reassessment by a clinician. In addition, the limited access to raw data (e.g., fastq and BAM files) constrained our ability to reassess possible phenotype-related variants, and the results rely on the testing laboratory bioinformatics pipelines alone. However, this study achieved its aim, which is to demonstrate that revisiting variants that have been reported by an outside lab can be beneficial and may result in better reporting. The discrepancies found in the results clarify the role of genetic counselors, as most genetic clinics send out their tests.

In the era of genomics, genetic counseling has become genomic counseling, and genetic counselors are required to correlate genetic variants with the patient's phenotype and clinical data. In genomic counseling, the nontraditional role of lab genetic counselor meshes with the traditional clinical role. Variant identification and interpretation expertise of a genetic counselor may be utilized, especially in a scenario that includes non-geneticist physicians, in which the genetic counselors are the primary experts. The greatest area of need is to develop a strategy for when to revisit genetic data and which data should be revisited. We hope that this study and similar studies will substantially contribute to variant interpretation guidelines and recommendations to ensure best patient care services. We emphasize in this study the benefits of effective communication between testing labs and the geneticist and/or genetic counselor to ensure proper variant classification and therefore deliver proper genomic counseling.

Conclusion

The findings of this study highlight the importance of detailed patient clinical data and clearly call for a discussion of the time and approach needed for reassessing VCF files.

Acknowledgement

None.

List of abbreviations

ACMGG	American College of Medical Genetics and Genomics
AMP	Association for Molecular Pathology
LOF	Loss of function
NGS	Next-generation sequencing
VCF	Variant call format
WES	Whole-exome sequence
WGS	Whole-genome sequence

Consent for publication

Not applicable.

Ethical approval

This study was approved by the Institutional Research Board of King Abdullah International Medical Research Center (KAIMRC). The data collection and analysis were conducted

retrospectively at King Abdulaziz Medical City. Informed consent was obtained from all individual participants included in the study.

Funding

The authors declare that there is no conflict of interests.

Declaration of conflicting interests

The authors declare that there is no conflict of interests.

Author details

Lamia Alsubaie¹, Saeed Alturki², Ali Alothaim², Ahmed Alfares³

1. King Abdullah International Medical Research Center/ King Saud bin Abdulaziz University for Health Sciences, Department of Pediatrics, Genetic Division, King Abdullah Specialized Children's Hospital, National Guard Health Affairs (NGHA), Riyadh, Saudi Arabia
2. King Abdullah International Medical Research Center/ King Saud bin Abdulaziz University for Health Sciences, Pathology and Laboratory Medicine, King Abdulaziz Medical City, NGHA, Riyadh, Saudi Arabia
3. Department of Pediatrics, College of Medicine, Qassim University, Qassim, Saudi Arabia

References

1. Smith M. DNA sequence analysis in clinical medicine, proceeding cautiously. *Front Mol Biosci* 2017; 4:24. <https://doi.org/10.3389/fmolb.2017.00024>
2. Sobreira NL, Valle D. Lessons learned from the search for genes responsible for rare Mendelian disorders. *Mol Genet Genomic Med* 2016; 4:371–5. <https://doi.org/10.1002/mgg3.233>
3. Basel D, McCarrier J. Ending a diagnostic odyssey: family education, counseling, and response to eventual diagnosis. *Pediatr Clin North Am* 2017; 64:265–72. <https://doi.org/10.1016/j.pcl.2016.08.017>
4. Kazazian HH, Boehm CD, Seltzer WK. ACMG recommendation for standards for interpretation of sequence variations. *Genet Med* 2000; 2:302–3. <https://doi.org/10.1097/00125817-200009000-00009>
5. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hedge MR, On behalf of the ACMG Laboratory Quality Assurance Committee; et al. ACMG recommendations for standards interpretation and reporting of sequence variations: revision 2007. *Genet Med* 2008; 10(4):294–300. <https://doi.org/10.1097/GIM.0b013e31816b5cae>
6. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee; et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013; 15(9):733–47. <https://doi.org/10.1038/gim.2013.92>
7. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, On behalf of the ACMG Laboratory Quality Assurance Committee. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; 17(5):405–24. <https://doi.org/10.1038/gim.2015.30>
8. Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome Med* 2017; 9:13. <https://doi.org/10.1186/s13073-017-0403-7>
9. Shearer AE, Eppsteiner RW, Booth KT, Ephraim SS, Gurrola J, Simpson A, et al. Utilizing ethnic-specific differences in minor allele frequency to re-categorize reported pathogenic deafness variants. *Am J Hum Genet* 2014; 95(4):445–53. <https://doi.org/10.1016/j.ajhg.2014.09.001>
10. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 2011; 3(65):65ra4. <https://doi.org/10.1126/scitranslmed.3001756>
11. Berg JS. Exploring the importance of case-level clinical information for variant interpretation. *Genet Med* 2017; 19(1):3–5. <https://doi.org/10.1038/gim.2016.106>
12. ClinVar [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 1988. Accession No. NM_000282.3, Homo sapiens FYN binding protein (FYB), transcript variant 4, mRNA; [cited 2017 Apr 16]. Available from: <https://www.ncbi.nlm.nih.gov/clinvar/variation/289402>
13. Alfares A, Alfadhel M, Wani T, Alsahli S, Alluhaydan I, Almutairi F, et al. A multicenter clinical exome study in unselected cohort from a consanguineous population of Saudi Arabia demonstrated a high diagnostic yield. *Mol Genet Metab* 2017; 121:91–5. <https://doi.org/10.1016/j.ymgme.2017.04.002>
14. Green RC, Goddard K, Jarvik G, Amendola L, Appelbaum P, Berg J, et al. Clinical sequencing exploratory research consortium: accelerating evidence-based practice of genomic medicine. *Am J Hum Genet* 2016; 98(6):1051–66. <https://doi.org/10.1016/j.ajhg.2016.04.011>
15. Garber KB, Vincent LM, Alexander JJ, Bean L, Bale S, Hegde M. Reassessment of genomic sequence variation to harmonize interpretation for personalized medicine. *Am J Hum Genet* 2016; 99(5):1140–9. <https://doi.org/10.1016/j.ajhg.2016.09.015>
16. Fakhro KA, Staudt MR, Ramstetter MD, Robay A, Malek JA, Badii, R, et al. The Qatar genome: a population-specific tool for precision medicine in the Middle East. *Hum Genome Variat J* 2016; 3:16016. <https://doi.org/10.1038/hgv.2016.16>
17. Ormond KE. From genetic counseling to “genomic counseling.” *Mol Genet Genomic Med* 2013; 1(4):189–93. <https://doi.org/10.1002/mgg3.45>
18. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, et al. Performance of ACMG-AMP Variant Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet* 2016; 98:1067–76. <https://doi.org/10.1016/j.ajhg.2016.03.024>
19. Pepin MG, Murray ML, Bailey S, Leisritz-Kessler D, Schwarze U, Byers PH. The challenge of comprehensive and consistent sequence variant interpretation between clinical laboratories. *Genet Med* 2016; 18(1):20–4. <https://doi.org/10.1038/gim.2015.31>