Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium

Laura M. Amendola,[1,16] Gail P. Jarvik,[1,16,*] Michael C. Leo,[2] Heather M. McLaughlin,[3] Yassmine Akkari,[4] Michelle D. Amaral,[5] Jonathan S. Berg,[6] Sawona Biswas,[7] Kevin M. Bowling,[5] Laura K. Conlin,[7] Greg M. Cooper,[5] Michael O. Dorschner,[8] Matthew C. Dulik,[9] Arezou A. Ghazani,[10] Rajarshi Ghosh,[11] Robert C. Green,[3,12,15] Ragan Hart,[1] Carrie Horton,[13] Jennifer J. Johnston,[14] Matthew S. Lebo,[3,12] Aleksandar Milosavljevic,[11] Jeffrey Ou,[1] Christine M. Pak,[4] Ronak Y. Patel,[11] Sumit Punj,[4] Carolyn Sue Richards,[4] Joseph Salama,[1] Natasha T. Strande,[6] Yaping Yang,[11] Sharon E. Plon,[11] Leslie G. Biesecker,[14] and Heidi L. Rehm[3,12,15,*]

Evaluating the pathogenicity of a variant is challenging given the plethora of types of genetic evidence that laboratories consider. Deciding how to weigh each type of evidence is difficult, and standards have been needed. In 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) published guidelines for the assessment of variants in genes associated with Mendelian diseases. Nine molecular diagnostic laboratories involved in the Clinical Sequencing Exploratory Research (CSER) consortium piloted these guidelines on 99 variants spanning all categories (pathogenic, likely pathogenic, uncertain significance, likely benign, and benign). Nine variants were distributed to all laboratories, and the remaining 90 were evaluated by three laboratories. The laboratories classified each variant by using both the laboratory's own method and the ACMG-AMP criteria. The agreement between the two methods used within laboratories was high (K-alpha = 0.91) with 79% concordance. However, there was only 34% concordance for either classification system across laboratories. After consensus discussions and detailed review of the ACMG-AMP criteria, concordance increased to 71%. Causes of initial discordance in ACMG-AMP classifications were identified, and recommendations on clarification and increased specification of the ACMG-AMP criteria were made. In summary, although an initial pilot of the ACMG-AMP guidelines did not lead to increased concordance in variant interpretation, comparing variant interpretations to identify differences and having a common framework to facilitate resolution of those differences were beneficial for improving agreement, allowing iterative movement toward increased reporting consistency for variants in genes associated with monogenic disease.

Introduction

The assessment of pathogenicity of genetic variation is one of the more complex and challenging tasks in the field of clinical genetics. It is now clear that enormous genetic variation exists in the human population. Most of this variation, including very rare variants, is unlikely to contribute substantively to human disease. For example, a typical genome sequence and reference genome have about 3.5 million differences, of which 0.6 million are rare or novel.[1] As such, the challenge of interpreting the clinical significance of this variation is well recognized as a barrier to furthering genomic medicine.[2,3]

We have previously reported both inconsistencies across laboratories in the classification of Mendelian-disease variants and high discordance in the use of a single simple classification system, whereby reviewers showed a bias toward overestimating pathogenicity.[4] Furthermore, recent analyses of variant classifications in ClinVar showed that for the 11% (12,895/118,169) of variants with two or more submitters, interpretations differed in 17% (2,229/12,895).[3] Inconsistency of the classification of variants across professional genetics laboratories has been reported elsewhere.[5] These data highlight the need for a more systematic and transparent approach to variant classification.

Laboratories performing and reporting the results of clinical genetic testing are now tasked with considering a plethora of types of genetic evidence, some applicable to all genes and others specific to individual genes and diseases. To date, laboratories have developed their own methods of variant assessment because the prior American College of Medical Genetics and Genomics (ACMG)

[1]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA; [2]Center for Health Research, Kaiser Permanente, Portland, OR 97227, USA; [3]Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA 02139, USA; [4]Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR 97239, USA; [5]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA; [6]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; [7]Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [8]Center for Precision Diagnostics, Department of Pathology, University of Washington, Seattle, WA 98195, USA; [9]Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [10]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA; [11]Baylor College of Medicine, Houston, TX 77030, USA; [12]Brigham and Women's Hospital and Harvard Medical School, Cambridge, MA 02115, USA; [13]Clinical Diagnostics, Ambry Genetics, Aliso Viejo, CA 92656, USA; [14]Intramural Research Program, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA; [15]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[16]These authors contributed equally to this work
*Correspondence: pair@u.washington.edu (G.P.J.), hrehm@partners.org (H.L.R.)
http://dx.doi.org/10.1016/j.ajhg.2016.03.024.

variant-reporting guidelines did not address the weighting of evidence for variant classification.[6] Some laboratories assign points to types of evidence and generate a score,[7] and others define specific combinations of evidence that allow them to arrive at each classification category[8] or use a Bayesian framework to combine data types into a likelihood ratio.[9] Still others have simply relied on expert judgment of the individual body of evidence on each variant to make a decision.

Deciding how to categorize and weigh each type of evidence is challenging, and guidance has been needed. Making the task even more challenging is that the true pathogenicity is not known for most variants, and it is therefore difficult to validate approaches to variant assessment, particularly for addressing variants that have limited evidence. However, combining the collective experience of experts in the community to begin to build a more systematic and transparent approach to variant classification is essential, and this has led the ACMG and Association for Molecular Pathology (AMP) to develop a framework for evidence evaluation. The initial framework was published in early 2015[10] and focused on variants in genes associated with Mendelian disease.

The ACMG-AMP guidelines defined 28 criteria (each with an assigned code) that address evidence such as population data, case-control analyses, functional data, computational predictions, allelic data, segregation studies, and de novo observations. Each code is assigned a weight (stand-alone, very strong, strong, moderate, or supporting) and direction (benign or pathogenic), and then rules guide users to combine these evidence codes to arrive at one of five classifications: pathogenic (P), likely pathogenic (LP), variant of uncertain significance (VUS), likely benign (LB), or benign (B). In some cases, the strength of individual criteria can be modified at the discretion of the curator, and the overall classification can be modified with expert judgment. As an example, a minor allele frequency (MAF) greater than the disease prevalence but less than 5% is coded benign strong (BS1); this is considered strong evidence against pathogenicity for a highly penetrant monogenic disorder and supports a LB classification when it is combined with at least one supporting line of evidence against pathogenicity (BP1–BP6). If BS1 is combined with another strong line of evidence against pathogenicity (BS2–BS4), this supports a B classification. Conversely, a variant predicted to be null (PVS1) would be classified as LP if it is absent from population databases (PM2) or P if it is observed to be de novo with confirmed paternity and maternity (PS2). If not enough lines of evidence are invoked to classify a variant as P, LP, LB, or B, or there are valid but contradictory lines of evidence, a variant is interpreted as a VUS.

We set out to evaluate how the ACMG-AMP guidelines compare to individual laboratory approaches to variant classification and explore the variance in the use and interpretation of the pathogenicity criteria. Nine laboratories participating in the Clinical Sequencing Exploratory Research (CSER) consortium evaluated the use of the new ACMG-AMP guidelines and in-house interpretations to assess inter-laboratory concordance by either method of variant classification. Our goals were to evaluate consistency of the use of the ACMG-AMP codes and subsequent pathogenicity classification. Further, we used these criteria to analyze the basis for discordance and sought to reconcile differing implementations with an eye to guidance clarification.

Material and Methods

CSER is a National Human Genome Research Institute (NHGRI)- and National Cancer Institute (NCI)-funded consortium exploring the clinical use of genomic sequencing, developing best practices, and identifying obstacles to implementation. It is composed of nine clinical U-award sites focusing on all aspects of clinical sequencing, the ClinSeq project,[11] and nine R-award sites focusing on ethical, legal, and social implications. Eight of the nine clinical U-award sites and ClinSeq participated in this exercise. These included laboratories performing exome and/or genome sequencing for the following projects: BASIC[3] (Baylor College of Medicine, Houston), PediSeq (Children's Hospital of Philadelphia), CanSeq (Dana Farber Cancer Institute, Boston), HudsonAlpha Institute for Biotechnology, MedSeq (Brigham and Women's Hospital and Partners Healthcare, Boston), NextGen (Oregon Health Sciences University, Portland), NCGENES (University of North Carolina, Chapel Hill), and NEXT Medicine (University of Washington, Seattle). Eight of the nine sites were accredited by the Clinical Laboratory Improvement Amendments (CLIA).

Selection and Classification of Variants

Each site nominated 11 variants identified in their sequencing projects for this exercise. Submitted variants were single-nucleotide substitutions or small indels (<22 bp) in genes thought to be associated with Mendelian disease. Each site was instructed to provide a range of variants in each classification category with varying degrees of difficulty. Accepted classifications were B, LB, VUS, LP, and P. Each variant submission also included whether it was identified as a diagnostic result or an incidental finding. Any internal evidence that the submitting laboratory used to classify the variant—for example, the phenotype and family history of the proband or whether parental testing identified the variant as de novo—was also provided to all laboratories. Nine variants (two P, two LP, two VUS, two LB, and one B) were selected for distribution to all laboratories without the submitting laboratory's classification; half were identified as incidental findings, half were identified as diagnostic findings, and one was from a carrier screen. The remaining 90 variants were randomly distributed to at least two other laboratories, enabling classifications from at least three laboratories for each variant. Each laboratory was asked to classify the pathogenicity by applying both their internal process and then the ACMG-AMP system. They were asked to document which ACMG-AMP criteria were invoked for the ACMG-AMP classification and note whether they found the classification of each variant difficult, moderate, or easy. Time taken for categorizing the variant was requested but not consistently recorded.

Application of Automated Tool for Calculation of Overall Classification from Evidence Codes

In order to assess whether ACMG-AMP evidence codes were combined appropriately by the variant curator, we developed a

pathogenicity calculator that combines the provided codes to generate a final classification. We used this calculator to compare the calculated ACMG-AMP classification based on tabulating the evidence codes provided by the laboratory with the final ACMG-AMP classification submitted by the laboratory. We shared these data with sites for consideration during consensus discussions and manually verified the results to identify which discrepancies were due to errors by the submitting laboratory and which were due to the use of judgment in overruling the ACMG-AMP classification.

Analysis of Intra- and Inter-laboratory Concordance

Descriptive statistics summarized the intra-laboratory classification concordance between the ACMG-AMP system and the laboratory's own process and the inter-laboratory concordance both for each laboratory's own process and for the ACMG-AMP system across laboratories. Additionally, we quantified the level of agreement. To do this, we considered the five-tier classification system in the following order—B, LB, VUS, LP, and P—and defined a one-step level of disagreement to be a range of classifications from one category to the next ordered category (e.g., from VUS to LP or LP to P); the maximum level included four steps (i.e., B to P). In addition, we tracked disagreements that were more likely to lead to medical-management differences (P or LP versus any of VUS, LB, and B) and disagreements less likely to affect clinical decision making (e.g., VUS versus LB or B, or confidence differences, such as B versus LB or P versus LP). To quantify the overall level of absolute agreement on ACMG-AMP and laboratory criteria within sites and agreement between sites using ACMG-AMP or laboratory criteria, we calculated Krippendorff's alpha (K-alpha); ranging from 0 to 1, this generalized measure of absolute agreement corrects for chance responding and can handle any number of raters, scale of measurement, and missing data. Because it focuses on disagreement, it overcomes many of the weaknesses associated with other agreement measures, such as Cohen's kappa.[12–16] In general, values of 0.80 and above are considered evidence of good agreement.[14] We also calculated 95% confidence intervals (CIs) for K-alpha by using bootstrapping with 20,000 replications.[17]

Two variants were excluded from the quantitative analyses and are not represented in Figures 1A and 1B; however, they are represented in the overall concordance shown in Figures 1C and 2. One variant was a low-penetrance allele (c.3920T>A [p.Ile1307Lys] [GenBank: NM_001127510.2] in *APC* [MIM: 611731]) for which several laboratories did not assign an ACMG-AMP classification, and the other variant (c.1101+1G>T [GenBank: NM_001005463.2] in *EBF3* [MIM: 607407]) was a predicted loss-of-function variant in a gene for which there is no known association with disease. Neither of these two variants was relevant to this analysis of classifying high-penetrance variants for Mendelian conditions, for which the ACMG-AMP guidelines are intended. In addition, the two laboratories that had key personnel involved in the development of the ACMG-AMP recommendations were excluded from one study-wide sensitivity analysis to evaluate whether familiarity with the system affects concordance. Lastly, we performed a second sensitivity analysis by excluding the classifications of the submitting laboratory to determine the dependence of these results on a single laboratory and whether classification in a real case setting rather than only for the comparison study affects results.

Analysis of ACMG-AMP Lines of Evidence

We analyzed the lines of evidence used for each variant classification to identify how commonly specific evidence codes and classification rules were used across all of the variants, the overall agreement in the pattern of ACMG-AMP codes used across sites for each variant, and the consistency with which each ACMG-AMP code was used within each variant. These were determined with a frequency table, the mean of coefficient of variation (CV) values across variants with each ACMG-AMP code, and K-alpha values of ACMG-AMP codes within each variant. Descriptive statistics of how often the strength of each line of evidence was modified during variant interpretation were also calculated.

Consensus Discussions

The variants with discrepant classifications based on the ACMG-AMP guidelines were discussed via phone conferences (n = 23) or via email (n = 43). Variants were chosen for discussion by phone conference if they were interpreted by all nine laboratories or if they were discrepant by more than one category of disagreement. The laboratory that submitted each of these 23 variants presented the lines of evidence used by all laboratories and the rationale for using, not using, or altering the strength of a particular evidence code. Once all evidence was discussed, each laboratory was asked to provide a final classification. For variants for which only one laboratory was discordant for only a one-level difference, the discordant laboratory was asked to re-review their classification in light of the evidence used and classifications made by the other laboratories. The discordant laboratory then provided either a change or a decision to retain the original classification, including the rationale in both scenarios by email. During phone conferences and via email, laboratories had the opportunity to share any internal data that could have contributed to discordance.

Results

Intra-laboratory Classification Concordance between Unique Laboratory Criteria and the ACMG-AMP System

The intra-laboratory comparison of the laboratory process and the ACMG-AMP system for the 347 paired variant assessments is summarized in Figure 1A. The classifications matched for 275 of 347 (79%) variant assessments. Eleven of the 347 paired variant assessments (3.2%) differed by greater than one level. Overall, in 48 of the 72 (67%) discordant calls, the ACMG-AMP system calls were closer to VUS. Specifically, a classification of B or LB was more likely to result from using the laboratories' own criteria than from using the ACMG-AMP criteria. The K-alpha value for agreement within laboratories ranged from 0.77 to 1.00 (average = 0.91; seven of nine laboratories had K-alpha > 0.90).

Inter-laboratory Concordance in Classification

Considering the inter-laboratory classification for 97 variants, there was no statistically significant difference in concordance across laboratories between classifications based on laboratory criteria and those based on ACMG-AMP criteria (lab K-alpha = 0.76, 95% CI = [0.73, 0.80]; ACMG-AMP K-alpha = 0.72, 95% CI = [0.68, 0.76]). In other words, implementation of the ACMG-AMP criteria did not yield more consistent variant classification among

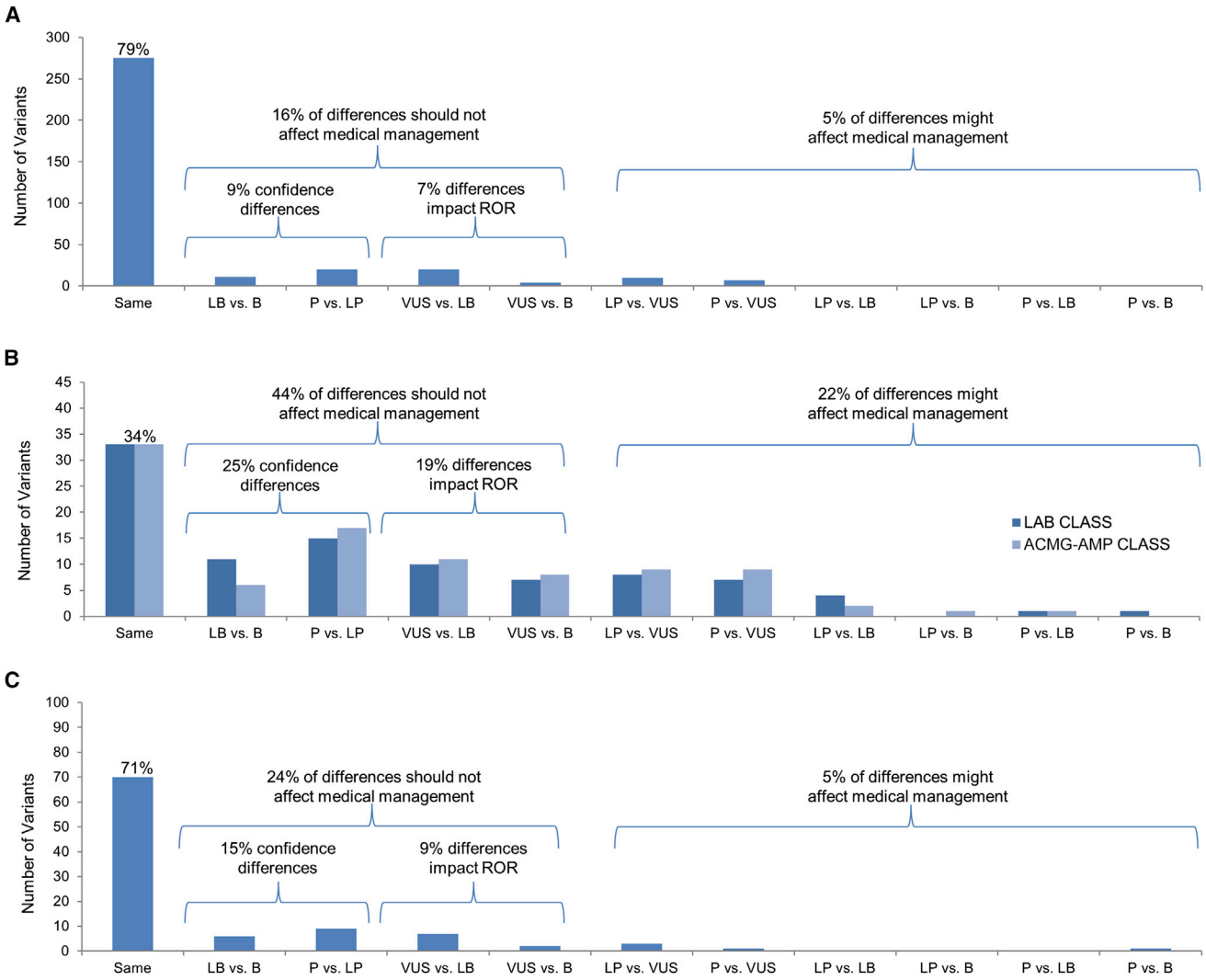

Figure 1. Distribution of Variant-Classification Comparisons according to the Extent of Differences across a Five-Tiered Classification Scheme

(A) Intra-laboratory concordance between laboratory and ACMG-AMP classification systems. This graph compares each site's use of the ACMG-AMP rules to their own laboratory classification methods.

(B) Inter-laboratory concordance of 97 variants. This graph compares the same calls, based on either the ACMG-AMP rules or the site's rules, between laboratories.

(C) Inter-laboratory concordance after consensus efforts. This graph shows a final comparison of calls between sites after consensus-building efforts.

these laboratories. All laboratories reviewing the variant (three to nine) agreed for 33 (34%) when they used either the ACMG-AMP system or their own criteria. No significant difference was found in inter-laboratory concordance when the two laboratories that contributed to the ACMG-AMP classification recommendations were removed from the analysis (K-alpha lab = 0.77, 95% CI = [0.73, 0.80]; K-alpha ACMG-AMP = 0.70, 95% CI = [0.66, 0.74]) or when the site that submitted the variant classifications was removed from the analysis (K-alpha lab = 0.76, 95% CI = [0.71, 0.80]; K-alpha ACMG-AMP = 0.75, 95% CI = [0.71, 0.78]). The distribution of types of disagreement among laboratories using each method is shown in Figure 1B. A total of 43/194 (22%) classifications had category differences that are more likely to influence medical decision making (P or LP versus VUS, LB, or B), the majority of which (33) were P or LP versus VUS. An additional 36 classifications (19%) involved differences between VUS and LB or B, which could have an impact on results reported by the laboratory given that many laboratories do not report LB or B results and that reporting VUS results could result in a more lengthy disclosure process and uncertainty of follow-up. The remaining 25% of variant classifications were differences in the confidence of calls (P versus LP or B versus LB), which are unlikely to have an impact on clinical care.

Consensus Discussions

The interpretation of 33/99 (34%) variants was identical across all sites that used the ACMG-AMP guidelines. After

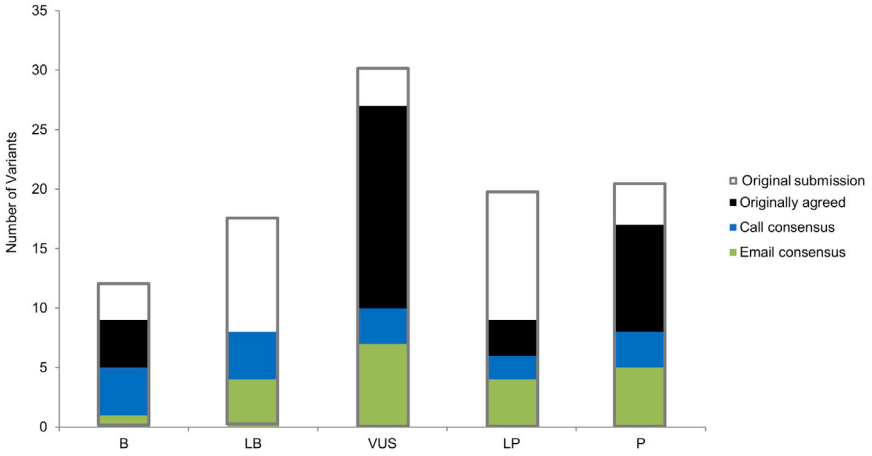

Figure 2. Distribution of 99 Variants Submitted for Assessment
Gray outlines illustrate the distribution of variant classifications submitted for assessment. Green bars indicate those calls that were agreed upon after initial review, blue bars indicate those calls agreed upon after email exchange, and black bars indicate those calls agreed upon after discussion on conference calls.

either emails or conference calls among the reporting laboratories, consensus on variant classifications based on the ACMG-AMP guidelines was achieved for 70/99 (71%) variants. Twenty-one of the discrepant variants were resolved via email, and the remaining 16 were resolved during phone conferences. The distribution and sources of variant-interpretation consensus can be found in Figure 2; gray outlines show the original distribution of submitted variant interpretations. Figure 1C shows the distribution of types of disagreement among laboratories after the consensus effort. Of the 29 variants that remained discordant, 25 involved only one level of difference (15 differed in confidence differences, three differed between LP and VUS, and seven differed between VUS and LB). Of the four variants with greater than one level of difference, two involved a difference between P and VUS, LB, or B. The final classifications for the 70 variants for which consensus was achieved, and the range of classifications for the remaining 29 discordant variants, are presented in Table S1.

Consensus discussions led to the clarification of the correct use of several ACMG-AMP lines of evidence, some of which included errors in the appropriate use of the rules already described in the guidelines (Table 1). Although the ACMG-AMP guidelines suggest a VUS classification when conflicting pathogenic and benign lines of evidence are identified, some laboratories allowed one line of conflicting benign evidence of only a supporting level (e.g., computational predictions) to override otherwise strong evidence of pathogenicity. In these cases, consensus discussion led to the use of expert judgment, as described in the ACMG-AMP guidelines, for appropriately disregarding the limited conflicting evidence, such as computational predictions. For two variants, achieving concordant interpretations required one laboratory's internal data. It was difficult to resolve the two variants that were excluded from the intra- and inter-laboratory analyses because the ACMG-AMP rules were not designed for low-penetrance variants (risk alleles) or variants in genes not clearly associated with the disorder. Some discrepancies in classification occurred because laboratories were interpreting the same variant for two different associated conditions, which have different disease frequencies. This led to a discordant use of the rules related to allele frequency.

Errors in Using the ACMG-AMP System

Our implementation of a computational tool to assess accuracy of combining the ACMG evidence codes used by the laboratories showed that for 16 out of 353 (5%) variant assessments, the ACMG-AMP codes listed by the laboratories did not support the classification chosen. When the laboratories were queried on these discrepancies, 9 of the 16 were due to tabulation errors, whereas judgment was used to override the ACMG-AMP rules for 7 of the 16 variants. The tabulation errors suggest that using computational tools to calculate the classification will lead to a modest increase in the accuracy of applying the rules.

ACMG-AMP Lines of Evidence Invoked and Modified in Strength

The frequency with which each ACMG-AMP code was invoked is listed in Figure 3. All lines of evidence were used at least once, except for BP7 (a silent, or synonymous, variant) given that no silent variants were submitted. Sixteen lines were used in fewer than 10% of variants, seven were used in 11%–18% of variants, and four were used in over 20% of all variant classifications: PVS1 (20%, predicted to be truncating), BS1 (26%, allele frequency is too high), PP3 (39%, computational evidence), and PM2 (41%, absent in population databases).

We further analyzed sources of discordance in the use of the ACMG-AMP codes to identify those criteria commonly used in an inconsistent manner. For variants where at least one site invoked BA1 (the allele frequency is >5% and too high to cause the disorder), we only evaluated concordance of the use of BA1. This was due to the fact that if a site selected BA1, they did not need to evaluate any other codes. For rules invoked more than ten times overall, PP4 (the phenotype is highly specific to the gene) was used the most inconsistently among the laboratories for a given variant (mean CV = 1.74). This is not surprising given the subjective nature of deciding how specific a phenotype is to a given gene. The most consistently applied rule was

Table 1. ACMG-AMP Rule Clarifications and Suggestions for Modification

Rule	Description	Clarifications and/or Suggestions
PVS1	variant predicted null where LOF is a mechanism of disease	do not apply to variants that are near the 3′ end of the gene and escape nonsense-mediated decay
PS1	variant with the same amino acid change as a previously established pathogenic variant, regardless of nucleotide change	does not include the same variant being assessed because it is not yet pathogenic, and the rule is intended for variants with a different nucleotide change
PS2	de novo variant with confirmed maternity and paternity	apply this rule as moderate or supporting if the variant is mosaic and its frequency in tissue is consistent with the phenotype
PS3	variant shown to have a deleterious effect by a well-established functional study	reduce the strength for assays that are not as well validated or linked to the phenotype
PM1	variant located in a mutational hotspot and/or critical and well-established functional domain	not meant for truncations; more clarification is needed for applying this rule
PM2, BS1	variant absent in population databases or with an allele frequency too high for the disease	cannot assume longer indels would be detected by next-generation sequencing
		use a published control dataset if its size is at least 1,000 individuals
		cannot be applied for low-quality calls or non-covered regions
		must define the condition and inheritance pattern
PM3	for recessive disorders, variant in *trans* with a pathogenic variant	invoke this rule as supporting if the phase is not established
		can upgrade if more than one proband is reported
PM4	protein-length-changing variant	applicable for in-frame deletions, insertions, or stop-loss variants, but not frameshifts, nonsense, and splice variants
PM5	novel missense variant at amino acid with different pathogenic missense change	ensure pathogenicity of previously reported variant
		suggest changing "novel" to "different" because some variants that are not novel might require assessment with this rule
PP3, BP4	variant with multiple lines of computational evidence	all lines must agree
PP4	the patient's phenotype or family history is highly specific to the genotype	not meant to be used for genetically heterogeneous conditions or conditions with unsolved etiology
		not typically applied for an analysis of incidental findings, but it could be applied for prior observations
PP5, BP6	variant called pathogenic or benign by a reputable source	only applicable when evidence is not available (e.g., Sharing Clinical Reports Project)
BS2	variant observed in a healthy adult for a disorder with full penetrance at an early age	populations might not have been screened or excluded for the phenotype
BP1	variant in a gene in which truncations primarily cause disease	clarify the meaning of "primary"; suggest >90%
BP2, BP5	variant in *trans* with a dominant pathogenic variant (BP2) or in an individual with an alternate molecular basis for disease (BP5)	clarify that one should apply BP2 when the pathogenic variant is seen in the same gene as the variant being evaluated and apply BP5 when the pathogenic variant is in a different gene

PVS1 (null variant where loss of function [LOF] is a known mechanism of disease; mean CV = 0.55). The mean and SD of the CV for all lines of evidence used are available in Table S2.

We also evaluated which criteria laboratories had increased or decreased in evidence strength and found that a total of nine lines of evidence were modified at least once. Three criteria were increased in strength (PM3, PP1, and BP2), and seven were decreased in strength (PVS1, PS1, PS2, PS3, PS4, PM3, and BS1). Co-segregation data supporting pathogenicity (PP1) was the most commonly modified line of evidence, whereby laboratories increased the strength from supporting to moderate or strong for nine interpretations on the basis of the strength of the segregation evidence available in the literature or from the laboratory's internal data. The other most common examples of modified strength included the following: PVS1 (a predicted null variant in a gene where LOF is a known mechanism of disease) was downgraded from very strong four times, PS2 (well-established functional studies show a deleterious effect) was downgraded three times, and BS1 (MAF is too high for the disorder) was downgraded three times.

Specific Variant Examples

The *GLA* (MIM: 300644) c.639+919G>A variant (GenBank: NM_000169.2), which has been reported in individuals with late-onset Fabry disease (MIM: 301500) and reduced alpha-galactosidase A enzyme activity,[18,19] was classified by three laboratories. Ranging from VUS to P, the

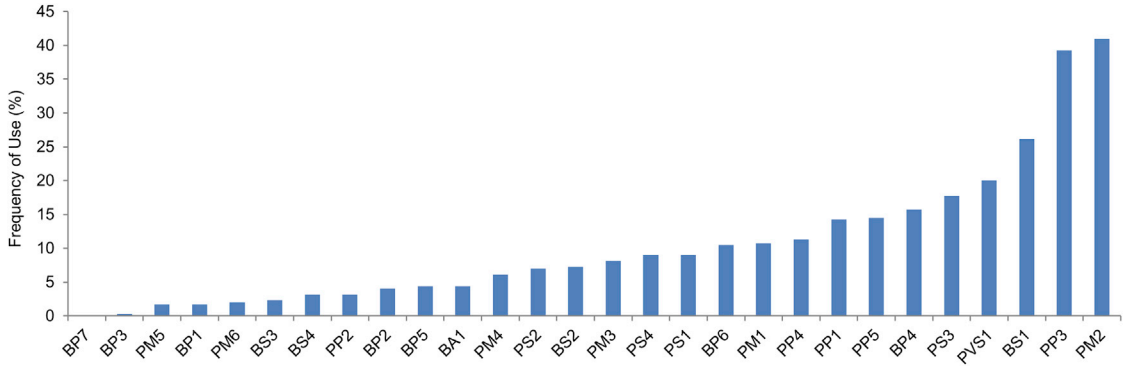

Figure 3. Frequency of Use for Each ACMG Line of Evidence

interpretations based on ACMG-AMP rules were discordant; however, all sites agreed on the classification of P when they used internal rules. This variant was absent from 528 race-matched control individuals across two studies.[18–20] A functional study also supported an effect of this variant on splicing.[19] All three laboratories invoked PS3 (a well-established functional study [clinical alpha-galactosidase enzyme testing] showed a deleterious effect of the variant) and PP1 (evidence of segregation), and one site increased the strength of this line of evidence from supporting to moderate on the basis of three families cited in the literature. PS4 (the prevalence of the variant in affected individuals is statistically greater than that in control individuals) was invoked by the one laboratory that called the variant P by using both their own rules and the ACMG-AMP criteria. Upon further discussion, the group agreed that this rule was applicable on the basis of a single publication citing a significant p value and other studies showing a statistical increase but requiring manual calculation. PVS1 (predicted-null variant in a gene where LOF is a known mechanism of disease) was also applied by all three laboratories after the group decided to downgrade the strength from very strong to strong because of the minor retention of wild-type transcript and the fact that the variant was a deep intronic variant for which a functional study was needed to demonstrate its impact on splicing. Three lines of evidence originally invoked by only one laboratory each were discarded: (1) PM4 (protein-length-changing variant) because this rule is only applicable for in-frame deletions, insertions, and stop-loss variants; (2) PP5 (a reputable source calls the variant pathogenic) because the reputable source's evidence was available for review by the curators and; (3) PP3 and BP4 (multiple lines of computational evidence agree) because this rule applies only when all predictions agree and not simply when some agree and others do not. All three sites came to a consensus that this variant is P on the basis of both their internal laboratory criteria and the ACMG-AMP criteria.

The group reviewed the c.1529C>T (p.Ala510Val) variant (GenBank: NM_003119.2) in *SPG7* (MIM: 602783), which is associated with autosomal-recessive spastic paraplegia (MIM: 607259), a disease that is known to have a variable, but generally adult, age of onset.[21] It was interpreted by all nine laboratories and had a range of pathogenicity classifications from LB to P. This variant was observed in 0.4% (267/66,688) of chromosomes of European ancestry (EU) in the ExAC Browser and has been found to have a 3%–4% heterozygote frequency in the UK and an estimated homozygote frequency of 20–40 per 100,000 individuals.[22] Shared data included that the variant was observed in the heterozygous state in 3 of the first 50 participants sequenced by the submitting laboratory's CSER study. The frequency of *SPG7*-associated spastic paraplegia is conservatively estimated to be 2–6 per 100,000 individuals according to the higher estimate; this yields an estimated frequency of 0.8% for all associated alleles. Multiple publications have cited the identification of homozygous or compound-heterozygous (including this variant) affected individuals.[22–24] It is notable that the laboratories that concluded this was a VUS or LB variant considered the BS1 criteria (the variant is more common than the disease, adjusted for the autosomal-recessive inheritance pattern). Two of the laboratories that concluded the variant was P according to the ACMG-AMP rules used the PM2 criteria (the variant is absent from control individuals or has an extremely low frequency if recessive). The remainder of the laboratories did not use any rules on population-frequency data. An additional line of evidence with conflicting use was PS1 (the variant results in the same amino acid change as a previously established pathogenic variant, regardless of nucleotide change). This rule was invoked by four of nine laboratories; however, after clarification that the intent of this rule, as described in the ACMG-AMP guidelines, is to be applied only for a "different" nucleotide change (i.e., something other than *SPG7* c.1529C>T that still leads to p.Ala510Val), all laboratories agreed that this rule was not applicable. Consensus was not reached for this variant largely because of discordance in applying the population-frequency lines of evidence. Some groups continued to weigh the published literature evidence of pathogenicity, whereas other groups concluded that it could not be a high-penetrance variant given its allele frequency. Given a perceived deficit of affected homozygotes relative to affected compound heterozygotes, some felt it might have low penetrance unless it is paired

Box 1. Recommendations and Additional Resources for Increasing Consistency in the Usage of ACMG-AMP Rules

- Develop disease-specific allele-frequency thresholds to enable lowering of the stand-alone benign criteria from a MAF of $\geq$5% to values specific to each disorder.
- Establish a resource of all genes to define whether LOF is a known mechanism of disease.
- Make recommendations for which computational algorithms are best in practice.
- Better define "well-established" functional data and/or distribute a resource that lists functional assays that meet the well-established threshold. Also define when to use reduced strength of the rule.
- Develop quantitative thresholds of evidence for and against segregation of different strengths.
- Promote the development of software tools that automate computable aspects of the ACMG-AMP guidelines to improve accurate use.

with a more deleterious variant. Resolving the role of this variant in disease might ultimately require a better understanding of the penetrance and possible role of modifiers, and classifying the pathogenicity of lower-penetrance variants was outside the scope of the ACMG-AMP guidelines.

The variant with the largest range of discordance (P to B) after consensus efforts based on the ACMG-AMP guidelines was *BTD* (MIM: 609019) c.1330G>C (p.Asp444His) (GenBank: NM_000060.2), which was interpreted by three laboratories. *BTD* is associated with autosomal-recessive biotinidase deficiency (MIM: 253260), and this variant was detected in one allele of an unaffected individual. The interpretations based on the ACMG-AMP guidelines originally ranged from VUS to P; however, after consensus efforts, the laboratory that classified the variant originally as VUS changed their interpretation to B, whereas the other two laboratories kept their interpretations as LP and P. This variant has been identified in multiple biotinidase-deficiency-affected individuals who have variants associated with profound deficiency on the other allele.[25,26] It was observed in 5.4% of chromosomes of Finnish ancestry and 4.15% of EU chromosomes in the ExAC Browser, and there were 83 reported instances of homozygosity. The laboratory that interpreted this variant as B invoked the MAF > 5% (BA1) as standalone criteria and noted the presence of homozygotes in population databases. Like *SPG7*, this might represent an allele that is more likely to be pathogenic when it is found to be compound heterozygous with a more deleterious allele than when it is found to be homozygous. This is a general problem with recessive disorders, and it might make consideration of the genotype rather than the pathogenicity of each allele more important for disease prediction. The laboratories classifying this variant as P and LP used expert judgment to overrule the use of BA1, which supports a B classification. They cited multiple reputable laboratories that have interpreted this variant as pathogenic and evidence that individuals who are homozygous and compound heterozygous for this variant might have a more mild form of the disease. Consensus efforts brought the group further from agreement on this variant; however, this highlights the importance of employing expert judgment when making interpretations, as well as the challenges that stem from using the ACMG-AMP guidelines to interpret the clinical significance of variants that might be associated with lower penetrance and mild presentations of disease.

The Supplemental Note and Table 1 describe in detail the criteria-specific clarifications that resulted from common usage errors, as well as the challenges and discussion topics related to each of the ACMG-AMP evidence categories that are utilized in implementing the ACMG-AMP guidelines. This material is designed to further clarify numerous rules found in the ACMG-AMP publication. In addition, more general recommendations and additional resources that could increase consistency of the usage of ACMG-AMP rules across laboratories are defined in Box 1.

Discussion

Interpreting the pathogenicity of a genetic variant requires evaluating a large number of heterogeneous types of evidence to arrive at a unitary descriptor of pathogenicity. Given the complexity of the data and uncertainty regarding the validity or utility of some of the data used for these interpretations, it is unsurprising that there would be variation among laboratories regarding these determinations. To that end, the ACMG-AMP system for classifying variant pathogenicity[10] is an important first step in efforts to improve the consistency of variant classification among laboratories. The guidelines include standardized terminology for classifying variants associated with monogenic diseases and a defined series of evidence types that can be used in pathogenicity assessment, enabling a record of the specific evidence type and strength used for determining pathogenicity. This enhances transparency and facilitates resolution of discrepancies in variant interpretation. It also forms a basis for iteratively building on the evidence as new data become available over time.

This study systematically evaluated the implementation of the ACMG-AMP guidelines in the medical practice of variant assessment. Nearly all ACMG-AMP lines of evidence were used, and PVS1 (predicted truncating), BS1 (allele frequency too high), PP3 (computational evidence), and PM2 (absent in population databases) (Figure 3) reflect the spectrum of variants chosen for the exercise and the most

available types of data. We identified differences in the application of the criteria but no difference in classification concordance between the ACMG-AMP system and the laboratory method. In part, the discordance in applying the ACMG-AMP guidelines was due to the subjective process of deciding when certain criteria are met. However, the guidelines provided a valuable framework for subsequent discussion of evidence, often leading to resolution of differences in variant interpretation; achieving this would have been more difficult if each laboratory relied on an independent method for variant assessment. The differences in both the intra- and inter-laboratory analyses identified points of confusion and inaccurate use of the ACMG-AMP criteria, as well as areas where expert judgment is required and additional guidance is needed. It should be noted that this was the first time that most sites had worked with the ACMG-AMP guidelines, and thus familiarity and systems for implementation of the criteria were still evolving. In addition, because the variants were distributed as a pilot evaluation of the ACMG-AMP guidelines and not for clinical reporting, not all sites subjected the variants to their typical CLIA process of review, which includes final review by a board-certified laboratory geneticist or an equivalently trained individual. Thus, the level of discordance reported here might have been inflated by the atypical workflows being deployed. In contrast, the resolution of the discordant variants involved multiple board-certified geneticists and others with long-standing experience in variant assessment, documenting the importance of this level of training in variant interpretation. These study results underscore the need for training in the use of genetic resources, evaluation of variant evidence, and application of the ACMG-AMP guidelines, even among experienced professionals. This study identified areas of confusion regarding the ACMG-AMP criteria, and these will be useful in developing training materials and further guidance for variant assessment. As recently described,[3] the Clinical Genome Resource (ClinGen) consortium is developing tools to aid variant classification based on the ACMG-AMP guidelines and will make this information public through both community availability of the tool and documentation of applied codes with variants submitted to ClinVar.

As described above, our discussion and consensus building led to a decrease in variant discordance from 66% to 29% of the 99 variants analyzed. This underscores the importance of not only having a standardized approach to variant assessment but also sharing variant interpretations for identifying and potentially resolving discordance. Given the rarity of most variants causative for monogenic disease, sharing data and comparing interpretations are imperative for ensuring the greatest opportunity for informed and collaborative variant interpretation. It is important to also reflect on the goals of variant assessment. Although numerous variants have robust evidence that can unequivocally allow classification into discreet categories without debate, many other variants have limited or conflicting evidence, making it difficult to accurately classify these variants. Indeed, for 29 (29%) of the 99 variants assessed in this study, a consensus classification was not achieved, and 5 of the 29 involved a difference between the categories of P or LP and VUS, LB, or B, which could affect medical management. This finding highlights that classifying sequence variants is similar to other fields of medicine in which practitioners can legitimately differ in their assessments of pathogenicity of a laboratory finding. By defining and applying formal criteria that parse these heterogeneous data types, we can better understand and analyze these legitimate differences in expert opinion and at the same time reduce errors and discrepancies.

Supplemental Data

Supplemental Data include a Supplemental Note, Supplemental Acknowledgments, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.03.024.

Conflicts of Interest

Most authors are employed by clinical-service providers. G.P.J. is a scientific advisory board member for ActX, a genetic-testing company. R.C.G. has received compensation from Invitae, Prudential, Illumina, AIA, Helix, and Roche for advisory services or speaking. S.E.P. serves on the scientific advisory board of Baylor Miraca Genetic Laboratory. Baylor College of Medicine and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of the Baylor Miraca Genetics Laboratories, which perform exome sequencing.

Acknowledgments

This work was funded by the National Human Genome Research Institute (NHGRI) and National Cancer Institute (UO1HG0006546, U01HG006485, U01HG006500, U01HG006492, UM1HG007301, UM1HG007292, UM1HG006508, U01HG006487, U01HG006507, U01HG007307, U41HG006834, R21HG006596, R01HG006600, P50HG007257, R01HG006600, R01HG004500, R01CA154517, R01HG006618, R21HG006594-01, R01HG006615-01, R21HG006612, and 5R21HG006613-02) and additional support from the participating laboratories. This research was supported in part by the Intramural Research Program of the NIH NHGRI. We thank the following individuals for contributing to variant assessment at the sites: Hana Zouk, PhD; Steven M. Harrison, PhD; Jillian G. Buchan, PhD; Jessica Booker, PhD; Jim Evans, MD, PhD, Kate Foreman, MS; Gloria Haskell, PhD; Kristy Lee, MS; Julianne O'Daniel, MS; Bradford Powell, MD, PhD; Bryce Seifert, PhD; and Karen Weck, MD.

Received: December 21, 2015
Accepted: March 22, 2016
Published: May 12, 2016

Web Resources

1000 Genomes, http://www.1000genomes.org/
ClinGen Pathogenicity Calculator, http://calculator.clinicalgenome.org/site/cg-calculator
CSER Consortium, https://cser-consortium.org/
ExAC Browser, http://exac.broadinstitute.org

Gene Tests, http://www.ncbi.nlm.nih.gov/books/NBK1107/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
OMIM, http://www.omim.org/
RefSeq, http://www.ncbi.nlm.nih.gov/refseq/
Sharing Clinical Reports Project, https://www.clinicalgenome.org/data-sharing/sharing-clinical-reports-project-scrp/

References

1. Kohane, I.S., Hsing, M., and Kong, S.W. (2012). Taxonomizing, sizing, and overcoming the incidentalome. Genet. Med. *14*, 399–404.
2. Evans, B.J., Burke, W., and Jarvik, G.P. (2015). The FDA and genomic tests–getting regulation right. N. Engl. J. Med. *372*, 2258–2264.
3. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al.; ClinGen (2015). ClinGen–the Clinical Genome Resource. N. Engl. J. Med. *372*, 2235–2242.
4. Amendola, L.M., Dorschner, M.O., Robertson, P.D., Salama, J.S., Hart, R., Shirts, B.H., Murray, M.L., Tokita, M.J., Gallego, C.J., Kim, D.S., et al. (2015). Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res. *25*, 305–315.
5. Yorczyk, A., Robinson, L.S., and Ross, T.S. (2015). Use of panel tests in place of single gene tests in the cancer genetics clinic. Clin. Genet. *88*, 278–282.
6. Richards, C.S., Bale, S., Bellissimo, D.B., Das, S., Grody, W.W., Hegde, M.R., Lyon, E., and Ward, B.E.; Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee (2008). ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet. Med. *10*, 294–300.
7. Karbassi, I., Maston, G.A., Love, A., DiVincenzo, C., Braastad, C.D., Elzinga, C.D., Bright, A.R., Previte, D., Zhang, K., Rowland, C.M., et al. (2016). A Standardized DNA Variant Scoring System for Pathogenicity Assessments in Mendelian Disorders. Hum. Mutat. *37*, 127–134.
8. Thompson, B.A., Spurdle, A.B., Plazzer, J.P., Greenblatt, M.S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capellá, G., den Dunnen, J.T., et al.; InSiGHT (2014). Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat. Genet. *46*, 107–115.
9. Goldgar, D.E., Easton, D.F., Byrnes, G.B., Spurdle, A.B., Iversen, E.S., and Greenblatt, M.S.; IARC Unclassified Genetic Variants Working Group (2008). Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. Hum. Mutat. *29*, 1265–1272.
10. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al.; ACMG Laboratory Quality Assurance Committee (2015). 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. *17*, 405–424.
11. Biesecker, L.G., Mullikin, J.C., Facio, F.M., Turner, C., Cherukuri, P.F., Blakesley, R.W., Bouffard, G.G., Chines, P.S., Cruz, P., Hansen, N.F., et al.; NISC Comparative Sequencing Program (2009). The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. Genome Res. *19*, 1665–1674.
12. Hayes, A.F., and Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. Commun. Methods Meas. *1*, 77–89.
13. Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. Hum. Commun. Res. *30*, 411–433.
14. Krippendorff, K. (2004). Content analysis: An introduction to its methodology, Second Edition (Sage).
15. Brennan, R.L., and Prediger, D.J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. Educ. Psychol. Meas. *41*, 687–699.
16. Zwick, R. (1988). Another look at interrater agreement. Psychol. Bull. *103*, 374–378.
17. Krippendorff, K. (2013). Bootstrapping Distributions for Krippendorff's Alpha for Coding of Predefined Units: Single-Valued cα and multi-valued $_{mv}\alpha$. http://web.asc.upenn.edu/usr/krippendorff/boot.c-Alpha.pdf.
18. Lin, H.Y., Huang, C.H., Yu, H.C., Chong, K.W., Hsu, J.H., Lee, P.C., Cheng, K.H., Chiang, C.C., Ho, H.J., Lin, S.P., et al. (2010). Enzyme assay and clinical assessment in subjects with a Chinese hotspot late-onset Fabry mutation (IVS4c+c919G→A). J. Inherit. Metab. Dis. *33*, 619–624.
19. Ishii, S., Nakao, S., Minamikawa-Tachino, R., Desnick, R.J., and Fan, J.Q. (2002). Alternative splicing in the alpha-galactosidase A gene: increased exon inclusion results in the Fabry cardiac phenotype. Am. J. Hum. Genet. *70*, 994–1002.
20. Hwu, W.L., Chien, Y.H., Lee, N.C., Chiang, S.C., Dobrovolny, R., Huang, A.C., Yeh, H.Y., Chao, M.C., Lin, S.J., Kitagawa, T., et al. (2009). Newborn screening for Fabry disease in Taiwan reveals a high incidence of the later-onset GLA mutation c.936+919G>A (IVS4+919G>A). Hum. Mutat. *30*, 1397–1405.
21. Casari, G., De Fusco, M., Ciarmatori, S., Zeviani, M., Mora, M., Fernandez, P., De Michele, G., Filla, A., Cocozza, S., Marconi, R., et al. (1998). Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. Cell *93*, 973–983.
22. Roxburgh, R.H., Marquis-Nicholson, R., Ashton, F., George, A.M., Lea, R.A., Eccles, D., Mossman, S., Bird, T., van Gassen, K.L., Kamsteeg, E.J., and Love, D.R. (2013). The p.Ala510Val mutation in the SPG7 (paraplegin) gene is the most common mutation causing adult onset neurogenetic disease in patients of British ancestry. J. Neurol. *260*, 1286–1294.
23. Elleuch, N., Depienne, C., Benomar, A., Hernandez, A.M., Ferrer, X., Fontaine, B., Grid, D., Tallaksen, C.M., Zemmouri, R., Stevanin, G., et al. (2006). Mutation analysis of the paraplegin gene (SPG7) in patients with hereditary spastic paraplegia. Neurology *66*, 654–659.
24. van Gassen, K.L., van der Heijden, C.D., de Bot, S.T., den Dunnen, W.F., van den Berg, L.H., Verschuuren-Bemelmans, C.C., Kremer, H.P., Veldink, J.H., Kamsteeg, E.J., Scheffer, H., and van de Warrenburg, B.P. (2012). Genotype-phenotype correlations in spastic paraplegia type 7: a study in a large Dutch cohort. Brain *135*, 2994–3004.
25. Norrgard, K.J., Pomponio, R.J., Swango, K.L., Hymes, J., Reynolds, T.R., Buck, G.A., and Wolf, B. (1997). Mutation (Q456H) is the most common cause of profound biotinidase deficiency in children ascertained by newborn screening in the United States. Biochem. Mol. Med. *61*, 22–27.
26. Hymes, J., Stanley, C.M., and Wolf, B. (2001). Mutations in BTD causing biotinidase deficiency. Hum. Mutat. *18*, 375–381.

The American Journal of Human Genetics, Volume *98*

Supplemental Data

Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium

Laura M. Amendola, Gail P. Jarvik, Michael C. Leo, Heather M. McLaughlin, Yassmine Akkari, Michelle D. Amaral, Jonathan S. Berg, Sawona Biswas, Kevin M. Bowling, Laura K. Conlin, Greg M. Cooper, Michael O. Dorschner, Matthew C. Dulik, Arezou A. Ghazani, Rajarshi Ghosh, Robert C. Green, Ragan Hart, Carrie Horton, Jennifer J. Johnston, Matthew S. Lebo, Aleksandar Milosavljevic, Jeffrey Ou, Christine M. Pak, Ronak Y. Patel, Sumit Punj, Carolyn Sue Richards, Joseph Salama, Natasha T. Strande, Yaping Yang, Sharon E. Plon, Leslie G. Biesecker, and Heidi L. Rehm

Supplemental Note

The following section describes in more detail specific points of discussion that arose in evaluating different types of evidence and their associated codes.

Population data (PS4, PM2, BA1, BS1 and BS2)
Whether or not to apply PM2 (variant is absent in population databases or at an extremely low frequency), was a point of discussion for several variants with discordant interpretations. The group struggled with quantifying an 'extremely low frequency' with some sites only invoking this rule when the variant was not present in population databases while others created an arbitrary cut off (for example <0.001%). The development of disease and/or gene-specific allele frequency thresholds would help to standardize the use of this rule. Some sites incorrectly invoked PM2 for a 20 bp indel variant not seen in ExAC (http://exac.broadinstitute.org), ESP (http://evs.gs.washington.edu/EVS/) or 1000 Genomes (http://www.1000genomes.org/). As stated in the ACMG/AMP rules, indel detection by next generation sequencing technologies can be challenging and therefore one should not assume accurate detection, particularly of longer indels. The guidelines also highlight consideration of whether a variant call in a population database passed quality metrics. In this exercise PM2 was mis-applied when a variant appeared to be missing from ExAC, but instead had been filtered out based on a low quality call. Similarly, some regions are not sequenced by exome-capture methods, so variants maybe missed and therefore not be found in exome-based databases. One must check coverage data and whether mapping was reliable before assuming an allele is absent. It should be noted that many of these points of caution relate to the other codes using population databases. For example, BS1 and BS2 also cannot be applied based on frequency data from low quality calls.

A point of confusion arose related to which disease frequency to use when analyzing whether to invoke rule BS1 (the allele frequency of the variant is too common to cause disease) for genes associated with more than one disease. In this case, it was concluded by discussion that if calling a variant benign for all diseases, one must use the most common disease prevalence. On the other hand, one can call a variant benign with respect to a single condition (relevant to the diagnostic test being performed for example), but make no claim on another condition. For example, in the *PTPN11* [MIM:176876] gene, primarily associated with Noonan syndrome [MIM: 163950], a variant may be under consideration for either Noonan syndrome, or, if observed in a patient with cardiomyopathy, could be evaluated with respect to this alternate condition. Because cardiomyopathy is more common, one can either evaluate the variant for all conditions by using the more common condition (cardiomyopathy), or consider its role only in one condition and use that prevalence.

It was common that laboratories used their own criteria to call a variant B or LB when the ACMG/AMP rules were less able to achieve these interpretations and labeled more VUS. For example, the ACMG/AMP rules contain, BA1, a stand-alone rule that if the allele frequency is >5% the variant can be called B. These discrepancies were not unexpected given that the ACMG/AMP rules had to be designed in a generic manner to address all genes and diseases. The 5% criterion is very conservative for highly penetrant alleles, particularly for rare autosomal dominant disorders. However, for very rare diseases, many laboratories have implemented lower cutoffs as stand-alone classification criteria for B and LB classifications, particularly if no other evidence is present to implicate the variant in disease. It was concluded that a widespread effort to define appropriate allele frequency cutoffs for each disease would be a useful resource for the community and aid in improved consistency of variant interpretation. These disease-specific efforts would require expertise to address possible underestimation of disease prevalence and reduced or age-dependent penetrance and take into account the genetic

heterogeneity of each disease to refine a more realistic highest possible allele frequency for a novel pathogenic variant. Special consideration would also be necessary for autosomal recessive conditions where the incidence of disease is not clearly known.

For certain variants laboratories struggled with defining 'fully penetrant' regarding the BS2 line of evidence (the variant is observed in a healthy adult for a disorder with full penetrance at an early age), leading to discordance. For instance some sites applied this rule for gene-condition pairs such as *DSP* [MIM:125647] and cardiomyopathy and *BRCA1* [MIM:113705] and breast cancer; however, pathogenic variants in these genes are not associated with full penetrance or presence at an early age. Consensus concluded that the BS2 rule should not be applied based on individuals in population databases, as these individuals are not well-characterized and could have cardiomyopathy or breast cancer.

Computational and predictive data (PVS1, PS1, PM5, PM4, PP3, BP7 BP4, BP3 BP1)
PVS1, a null variant where LOF is a known mechanism of disease, was incorrectly invoked by laboratories for a variant near the 3' end of the gene that would likely escape nonsense mediated decay[20; 21]. The ACMG/AMP guidelines warn against invoking PVS1 in this context, reinforcing the importance of considering the caveats already published in the guideline for each line of evidence. For some variants, laboratories disagreed on whether or not to apply PVS1 because not all agreed LOF was a known mechanism for the condition. Development of a resource to define which genes LOF is an established mechanism for disease would be useful and improve consistent application of this rule.

One source of discrepancy occurred in one site's usage of PS1, the variant results in the same amino acid as a previously described pathogenic variant regardless of the nucleotide change. One site erroneously used this to apply prior publication of the same exact variant while this rule, as described in more depth on the ACMG/AMP guideline, only applies when the established pathogenic variant has a different nucleotide change than the variant being interpreted. Discussion clarified that analysis of prior cases was more appropriately applied to rule PS4, where multiple occurrences of the same variant in cases can be considered evidence if there is a statistically significant increased occurrence in cases compared to controls. In addition to clarifying the intended usage of PS1, another point of discordance in applying both PS1 and PM5 was in circumstances where the other published variants were not universally agreed to be "well-established" as pathogenic, a requirement for usage of both rules.

For criteria PP3 and BP4 (multiple lines of computational evidence do, or do not, support a deleterious effect on the protein), it was clarified that these lines of evidence can only be invoked when ALL lines of evidence evaluated are consistent (for either missense or splicing evaluation, both is not necessary). This is clear in the more detailed text of the ACMG/AMP guidelines, but is not listed in the brief rule description. Clarifying this point resolved some discordant use of these rules; however, discordance remained when laboratories used different computational programs. Consistent use of these computational lines of evidence would be aided by the development of recommendations for which computational evidence programs are best to use and what thresholds are appropriate. Use of such algorithms and standards may also be gene specific or only applicable to a type of variation. There were several variants that were discordant by ACMG/AMP rule usage simply because laboratories allowed a computational line of evidence that conflicted with strong evidence supporting pathogenicity, to be used to call the variant a VUS. While the ACMG/AMP guidelines do state that conflicting lines of evidence (e.g. segregation supports pathogenicity but functional evidence does not) generally result in a VUS classification, it was noted that professional judgement must be used. All sites agreed that computational predictions of missense or splice variants are well known to

have reduced accuracy and therefore should not be used to override other strong evidence. For example, if a variant has ample criteria to be called pathogenic, it should not be reduced to VUS just because computational algorithms did not predict an impact.

For several variants the definition of “primarily” was a source of discordance with respect to whether or not to invoke BP1 (a missense variant in a gene where primarily truncations cause disease); “primarily” may range from a simple majority (>50%) to nearly all (e.g., >90%). If the former, some laboratories felt that this was not the level of truncations that would lead them to determine a SNV was less likely to cause disease. A member of the ACMG/AMP guideline committee noted that the rule was intended to focus on disorders where all, or nearly all, (e.g. > 90%) were due to truncations. All sites agreed that more quantitative guidance to establish a threshold to invoke this rule would increase pathogenicity classification concordance.

Functional data (PS3, PM1, PP2, BS3)
Whether or not laboratories invoked PS3, well-established functional studies support a damaging effect, was dependent on whether the lab trusted the clinical relevance of the results of such studies. Defining ‘well-established’ was the critical component and familiarity with the assay seemed to be a factor when deciding to invoke this rule. There was considerable variability in the functional study thresholds among groups. After discussion, sites agreed that, at a minimum, the assay must be validated with known pathogenic and benign variants and the output of the assay must have an established mechanistic relevance to the associated phenotype. Developing a resource curated by disease experts which lists functional assays that meet the ‘well-established’ threshold would increase consistency in applying the PS3 rule. Of note, the strength of PS3 can be reduced to moderate or supporting if a lab does not feel confident enough in the functional study to invoke it as a strong line of evidence.

One source of discrepancy identified that the PM1 rule (variant is located in a mutational hot spot and/or a critical and well-established functional domain), should only be invoked for missense variants, not truncations, and it should only be applied if the variant occurs in domains that are devoid of benign variation as described in more detail in the ACMG/AMP guideline. Defining ‘well-established’ also led to discordance in PM1 rule usage. The group defined a mutational hot spot as a location where there are multiple changes in the same domain that are known to be pathogenic; however, there was still disagreement regarding how many pathogenic variants constitute ‘multiple’, how well-defined the domain must be and how close other benign variants can be to the domain and variant in question.

Laboratories used various techniques to determine whether or not to apply PP2, a missense variant in a gene with a low rate of benign missense variation and in which missense variants are a common mechanism of disease. Establishing a quantitative metric for “a low rate” would clarify when to correctly invoke PP2.

Segregation data (PP1, BS4)
PP1, cosegregation in affected family members, and BS4, lack of cosegregation, were inconsistently invoked. This is likely due to the absence of a quantitative metric to establish whether or not these rules apply. For example, one lab invoked PP1 based on a single family with two affected individuals shown to carry a variant, but other sites did not deem this evidence sufficient. PP1 was the most commonly modified line of evidence illustrating that laboratories did consider how many affected individuals in a family tested positive for a variant and/or how many families with the variant showed segregation; however, whether PP1 was made a moderate or strong line of evidence, or modified at all, was based solely on the opinion of the laboratories. As stated above, the ACMG/AMP guidelines support the use of expert judgment when

classifying variants; however, quantitatively analyzing segregation data would increase concordance in using these rules. For one variant (NM_017636.3 (TRPM4) [MIM:606936]: c.2531G>A (p.Gly844Asp)) the segregation data in the literature was from a family with a different phenotype (right bundle branch block, RBBB) than the phenotype of the individual in whom the variant was found (long QT syndrome, LQTS). The group concluded that in this context PP1 could not be invoked since the variant was being interpreted for the LQTS phenotype; however, this data was used to invoke BS2, observed in a healthy adult, since the affected individuals in the literature did not have LQTS, but downgraded it from strong to supporting based on incomplete penetrance of the LQTS phenotype.

De novo data (PS2, PM6)
De novo data was not commonly found for variants reviewed in this project; however, the group did discuss how and when to downgrade PS2, the variant is *de novo* in a patient with a disease and no family history with both maternity and paternity confirmed, for one analyzed variant. The guidelines support invoking PM6 when maternity and/or paternity is not confirmed, and the discussions supported use of PS2 downgraded to moderate if the individual is mosaic for a variant that is thus presumed *de novo* and high enough frequency to be associated with the phenotype. PS2 could also be invoked even if maternity and paternity were not confirmed if there were multiple *de novo* occurrences published or observed (i.e. parental testing was performed but maternity and paternity assessment was either not performed or not documented in the literature.)

Allelic data (PM3, BP2)
Laboratories discussed when to modify the strength of PM3, the variant is seen in *trans* with a pathogenic variant for recessive disorders. Published literature may not always explicitly state the phase of variants found in affected individuals which raises a challenge for invoking PM3. When phase has not been established, some felt that PM3 could be invoked as supporting evidence. Also, if the variant is seen in *trans* with a pathogenic variant in more than one individual it was felt that PM3 can be upgraded to strong. However, sites did not agree on how many additional observations were necessary to call the evidence strong (2 vs. 3) but concluded that such guidance would be useful.

Other databases (PP5, BP6)
Both PP5 and BP6 (a reputable source reports the variant as pathogenic or benign respectively), was commonly invoked incorrectly when any P or B interpretation was present in a database, for example ClinVar. As stated in the ACMG/AMP guideline, these rules should only be invoked when the supporting evidence for the assertions in the database is not available for review, for example interpretations from the Sharing Clinical Reports Project (https://www.clinicalgenome.org/data-sharing/sharing-clinical-reports-project-scrp/) could correctly invoke these rules. Invoking PP5 when the evidence to support the P classification made by the reputable source is the same as the evidence the lab is using to evaluate the variant's pathogenicity would be counting the same evidence twice. Furthermore, it was clarified that ClinVar is not a reputable source itself and this judgement must be placed on the individual ClinVar submitters.

Other data (PP4, BP5)
As noted above, for rules invoked more than 10 times overall, PP4 (the patient's phenotype or family history is highly specific for a disease with a single genetic etiology), was used the most inconsistently. Laboratories discussed several contexts in which PP4 should not be applied. As stated in the ACMG/AMP guidelines, PP4 should not be invoked for cases where the phenotype has locus heterogeneity such as intellectual disability, breast cancer, or hearing loss. This rule

was misapplied by laboratories that interpreted it to mean the gene is known to be associated with the disorder, rather than the gene was the sole gene known to cause the disorder. PP4 will also typically not be invoked when interpreting a variant that has been identified as an incidental finding since it is unlikely the individual has the phenotype specific to the gene of interest.

For two variants, sites discussed and clarified the use of BP5 vs BP2. Invoking BP5 (the variant was found in a case with an alternate molecular basis for disease), requires that a pathogenic variant in a *different* gene has been found in an individual who also has the variant being evaluated. This differs from BP2 (the variant was observed in *trans* with a dominant variant or in *cis* with a recessive variant), which should be invoked if that pathogenic variant is seen in the *same* gene as the variant being evaluated.

Supplemental Acknowledgments

We also thank the members of the CSER Actionability and Return of Results working group who are not listed as authors: Michelle D. Amaral, PhD; Katrina Armstrong, MD; Benjamin Berkman, JD, MPH; Wendy Chung, MD, PhD; Jessica Everett, MS, CGC; Judy Garber, MD, MPH; Michele C. Gornick, PhD; Stacy W. Gray, MD, AM; Lucia Hindorff, MPH, PhD; Ingrid A. Holm, MD, MPH; Dave Kaufman, PhD; Barbara A. Koenig, PhD; Sek Won Kong, MD; Katie Lewis, MSc; Edward J. Lose, MD; Jeffrey Pennington; Katie Porter, JD,MPH; Dan Robinson, PhD; Brian Shirts, MD, PhD; Elian Silverman; Nancy B. Spinner, PhD; Holly K. Tabor, PhD; Ellen Tsai, PhD; Pankaj Vats, MS; David L. Veenstra, PharmD, PhD; Benjamin S. Wilfond, MD; and Susan M. Wolf, JD.

Table S1. Final ACMG classifications for consensus variants and range of classifications for discordant variants

	Variant	Number of sites	Classification consensus or range	How consensus achieved (if applicable)
1	NM_005228.3(*EGFR*)[MIM:131550]:c.2369C>T(p.Thr790Met)	9	Pathogenic	Conference Call
2	NM_000465.3(*BARD1*)[MIM:601593]:c.1075_1095delTTGCCTGAATGTTCTTCACCA(p.Leu359_Pro365del)	3	Benign	Originally agreed
3	NM_000122.1(*ERCC3*)[MIM:133510]:c.325C>T(p.Arg109*)	3	Likely Pathogenic	Originally agreed
4	NM_000546.5(*TP53)* [MIM:191170]:c.743G>A(p.Arg248Gln)	3	Pathogenic	Originally agreed
5	NM_001127510.2(*APC*)[MIM:611731]:c.3920T>A(p.Ile1307Lys)	3	Benign/VUS	NA
6	NM_000038.5(*APC*)[MIM:611731]:c.3386T>C(p.Leu1129Ser)	3	Likely Benign/Benign	NA
7	NM_004360.3(*CDH1*)[MIM:192090]:c.1568A>G(p.Tyr523Cys)	3	VUS	Originally agreed
8	NM_033084.3(*FANCD2*)[MIM:613984]:c.1278+1G>T	3	Likely Pathogenic/VUS	NA
9	NM_000257.3(*MYH7*)[MIM:160760]:c.2717A>G(p.Asp906Gly)	3	Pathogenic/Likely Pathogenic	NA
10	NM_000535.5(*PMS2*)[MIM:600259]:c.1532C>T(p.Thr511Met)	3	Benign	Conference Call
11	NM_003242.5(*TGFBR2*)[MIM:190182]:c.383delA(p.Lys128Serfs*35)	3	Benign	Conference Call
12	NM_001163817.1(*DHCR7*)[MIM:602858]: c.964-1G>C	3	Pathogenic	Originally agreed
13	NM_001042351.2(*G6PD*)[MIM:305900]:c.202G>A (p.Val68Met)	3	Pathogenic	Email
14	NM_031844.2(*HNRNPU*)[MIM:602869]:c.2304_2305del (p.Gly769Glufs*83)	3	Likely Pathogenic	Originally agreed
15	NM_000363.4(*TNNI3*)[MIM:191044]:c.485G>A (p.Arg162Gln)	3	Pathogenic/Likely Pathogenic	NA
16	NM_078485.3(*COL9A1*)[MIM:120210]: c.70C>A (p.Gln24Lys)	3	VUS	Originally agreed
17	NM_024022.2(*TMPRSS3*)[MIM:605511]:c.1152G>T (p.Met384Ile)	3	VUS	Email
18	NM_001089.2(*ABCA3*)[MIM:601615]:c.2614A>G(p.Ser872Gly)	3	VUS	Email
19	NM_000238.3(*KCNH2*)[MIM:152427]:c.442C>T (p.Arg148Trp)	3	VUS	Originally agreed
20	NM_017636.3(*TRPM4*)[MIM:606936]:c.2531G>A (p.Gly844Asp)	9	Likely Benign	Conference Call
21	NM_007294.3(*BRCA1*)[MIM:113705]:c.3119G>A (p.Ser1040Asn)	3	Pathogenic/Likely Pathogenic	NA
22	NM_018848.3(*MKKS*)[MIM:604896]: c.724G>T (p.Ala242Ser)	3	VUS/Likely Benign	NA
23	NM_000546.5(*TP53*)[MIM:191170]:c.455C>T (p.Pro152Leu)	3	Pathogenic	Originally agreed
24	NM_007294.3(*BRCA2*)[MIM:600185]:c.7762_7764delinsTT (p.I2588Ffs*60)	3	Pathogenic	Email
25	NM_000535.5(*PMS2*)[MIM:600259]:c.1096G>C (p.Asp366His)	3	VUS	Originally agreed
26	NM_024642.3(*GALNT12*)[MIM:610290]: c.1278_1293delGTGGTTCTTGGAGACT (p.Trp427Cysfs*23)	3	VUS	Email
27	NM_006231.2(*POLE*)[MIM:174762]:c.2214G>C (p.Lys738Asn)	3	VUS	Originally agreed
28	NM_001943.3(*DSG2*)[MIM:125671]:c.2568A>C (p.Lys856Asn)	3	Likely Benign	Conference Call
29	NM_000059.3(*BRCA2*)[MIM:600185]:c.4779A>C (p.Glu1593Asp)	9	VUS/Likely Benign	NA
30	NM_000138.4(*FBN1*)[MIM:134797]:c.2956G>A (p.Ala986Thr)	3	Likely Benign	Email
31	NM_000540.2(*RYR1*)[MIM:180901]:c.4178A>G (p.Lys1393Arg)	3	Likely Benign	Email
32	NM_000256.3(*MYBPC3*)[MIM:600958]:c.977G>A (p.Arg326Gln)	3	Benign	Originally agreed
33	NM_000540.2(*RYR1*)[MIM:180901]:c.13513G>C (p.Asp4505His)	3	VUS/Likely Benign	NA
34	NM_1048171.1 (*MUTYH*)[MIM:604933]:c.536A>G (p.Tyr179Cys)	3	Pathogenic	Originally agreed
35	NM_003119.23 (*SPG7*)[MIM:602783]:c.1529C>T (p.Ala510Val)	9	Pathogenic/VUS	NA
36	NM_000262.2 (*NAGA*)[MIM:104170]:c.606C>A (p.Tyr202*)	3	Likely Pathogenic	Email
37	NM_003060.3 (*SLC22A5*):c.1463G>A (p.Arg488His)	3	VUS	Email
38	NM_133259.3 (LRPPRC):c.3286delC (p.His1096Thrfs7*)	3	Likely Pathogenic	Email
39	NM_000124.3 (ERCC6):c.3289A>G (p.Met1097Val)	3	Benign	Originally agreed
40	NM_000059.3 (*BRCA2*)[MIM:600185]:c.4061C>T (p.Thr1354Met)	3	Likely Benign/Benign	NA
41	NM_000506.3(*F2*)[MIM:176930]:c.598G>A (p.Glu200Lys)	3	VUS	Originally agreed
42	NM_000060.2 (*BTD*)[MIM:609019]:c.1330G>C (p.Asp444His)	3	Pathogenic/Benign	NA
43	NM_000445.3 (*PLEC*)[MIM:601282]:c.4732C>T (p.Arg1578Cys)	3	VUS	Originally agreed
44	NM_015506.2 (*MMACHC*)[MIM:609831]:c.271dupA, (p.Arg91Lysfs14*)	3	Pathogenic	Originally agreed
45	NM_153717(*BBS2*)[MIM:606151]:c.1864C>T(p.Arg622*)	3	Pathogenic	Email
46	NM_005859(*PURA*)[MIM:600473]:c.698T>C(p.Phe233Ser)	3	VUS	Email
47	NM_078480(*PUF60*)[MIM:604819]:c.436C>T(p.Arg146Cys)	3	VUS	Originally agreed
48	NM_015560(*OPA1*)[MIM:605290]:c.113_130del18(p.R38_S43del)	3	VUS/Likely Benign	NA
49	NM_001197104(*KMT2A*)[MIM:159555]:c.6572G>A(p.Arg2191Gln)	4	Likely Benign	Conference Call
50	NM_004541(*NDUFA1*)[MIM:300078]:c.G94C(p.G32R)	3	Likely Benign/Benign	NA
51	NM_000531(*OTC*)[MIM:300461]:c.118C>T(p.Arg40Cys)	3	Pathogenic/Likely Pathogenic	NA
52	NM_000138(*FBN1*)[MIM:134797]:c.1328-2A>G	3	Pathogenic	Originally agreed
53	NM_004006(*DMD*)[MIM:300377]:c.4233+2C>T	9	Likely Benign	Conference Call
54	NM_000179(*MSH6*)[MIM:600678]:c.4068_4071dupGATT(p.Lys1358fs)	3	Benign	Conference Call
55	NM_000492(*CFTR*)[MIM:602421]:c.3705T>G(p.Ser1235Arg)	3	Benign	Email
56	NM_004992.3(*MECP2*)[MIM:300005]:c.27-6C>G	3	Pathogenic/Likely Pathogenic	NA
57	NM_013334 (*GMPPB*)[MIM:615320]: c.860G>A(Arg287Gln)	3	Pathogenic/Likely Pathogenic	NA
58	NM_000059 (*BRCA2*)[MIM:600185]:delTG (p.V220IfsX3)	3	Pathogenic	Originally agreed
59	NM_000833 (*GRIN2A*)[MIM:138253]:c.4375T>C (p.Ser1459Gly)	9	Pathogenic/Likely Pathogenic	NA
60	NM_005445 (*SMC3*)[MIM:606062]:c.283G>A (p.Glu95Lys)	3	Likely Pathogenic	Originally agreed
61	NM_007325 (*GRIA3*)[MIM:305915]:c.466T>C (Tyr156His)	3	VUS	Originally agreed
62	NM_000138 (*FBN1*)[MIM:134797]:c.8176C>T (Arg2726Trp)	3	VUS/Likely Benign	NA
63	NM_001005463.2 (*EBF3*)[MIM:607407] c.1101+1G>T	3	VUS	Email
64	NM_014801 (*PCNXL2*):c.3526C>T (His1176Tyr)	3	VUS	Email
65	NM_000249 (*MLH1*)[MIM:120436]:c.394G>C (p.Asp132His)	3	Likely Benign	Email
66	NM_000492 (*CFTR*) [MIM:602421]:c.2991G>C (p.Lys997Phe)	3	Benign	Originally agreed
67	NM_144612.6(*LOXHD1*)[MIM:613072]: c.1028G>A (p.Arg343His)	3	Likely Benign	Email
68	NM_000257.2(*MYH7*)[MIM:160760]:c.327C>T (p.Tyr109Tyr)	3	Likely Benign/Benign	NA
69	NM_001114753.1(ENG)[MIM:131195]:c.818C>T (p.Thr273Ile)	3	VUS	Originally agreed
70	NM_024422.3(*DSC2*)[MIM:125645]: c.631-2A>G	3	Pathogenic	Email
71	NM_000169.2(*GLA*)[MIM:300644]:c.639+919G>A	3	Pathogenic	Conference Call
72	NM_001369.2(*DNAH5*)[MIM:603335]:c.7468_7488del (p.Trp2490_Leu2496del)	9	Likely Pathogenic/VUS	NA
73	NM_000484.3(*APP*)[MIM:104760]:c.2137G>A (p.Ala713Thr)	3	Likely Pathogenic/VUS	NA
74	NM_174916.2(*UBR1*)[MIM:605981]:c.4107T>A (p.Cys1369*)	3	Likely Pathogenic	Email
75	NM_000391.3(*TPP1*)[MIM:607998]:c.1678_1679delCT (p.Leu560ThrfsX47)	3	Likely Pathogenic	Email
76	NM_000053.3(*ATP7B*)[MIM:606882]:c.2972C>T (p.Thr991Met)	3	VUS	Originally agreed
77	NM_005633.3(*SOS1*)[MIM:182530]:c.1010A>G (p.Tyr337Cys)	3	VUS	Originally agreed
78	NM_003476.4(*CSRP3*)[MIM:600824]:c.286_287delCC (p.Pro96fs)	3	VUS	Conference Call

79	NM_000142.4(*FGFR3*)[MIM:134934]:c.2310C>G (p.Tyr770*)	3	VUS	Originally agreed
80	NM_001001431(*TNNT2*)[MIM:191045]:c.391C>T (p.Arg131Trp)	3	Pathogenic/Likely Pathogenic	NA
81	NM_001142605.1(*EFTUD2*)[MIM:603892]:c.1258G>A (p.Gly420Ser)	3	Benign/VUS	NA
82	NM_ 032335.3(*PHF6*)[MIM:300414]:c.865A>G (p.Thr289Ala)	3	VUS	Originally agreed
83	NM_022068.2(*PIEZO2*)[MIM:613629]: c.8057G>A (p.Arg2686His)	3	Pathogenic	Originally agreed
84	NM_000552.3(*VWF*):[MIM:613160]c.6937C>T (p.Arg2313Cys)	3	VUS	Conference Call
85	NM_001008844.1(*DSP*)[MIM:125647]:c.3701A>T (p.Glu1234Val)	9	Likely Benign/Benign	NA
86	NM_000321.2(*RB1*)[MIM:614041]:c.920C>T (p.Thr307Ile)	3	Likely Benign/Benign	NA
87	NM_198056.2(*SCN5A*)[MIM:600163]:c.3956G>T (p.Gly1319Val)	3	Likely Pathogenic	Conference Call
88	NM_003620.3(*PPM1D*)[MIM:605100]:c.1437dupT(p.Lys480fs)	3	Pathogenic	Conference Call
89	NM_000069.2 (*CACNA1S*)[MIM:114208]:c.4060A>T (p.Thr1354Ser)	3	VUS	Originally agreed
90	NM_000540.2 (*RYR1*)[MIM:180901]:c.1840C>T (p.Arg614Cys)	3	Pathogenic	Originally agreed
91	NM_144997.5 (*FLCN*)[MIM:607273]:c.1285dupC (p.His429Profs*27)	3	Pathogenic	Email
92	NM_000257.2 (*MYH7*)[MIM:160760]:c.2359C>T (p.Arg787Cys)	3	VUS	Conference Call
93	NM_001103.2 (*ACTN2*)[MIM:102573]:c.26A>G (p.Gln9Arg)	3	VUS	Originally agreed
94	NM_133378.4 (*TTN*)[MIM:188840]:c.94398G>A (p.Asp31467Asn)	3	VUS/Likely Benign	NA
95	NM_015311.1 (*OBSL1*)[MIM:610991]:c.4951G>T (p.Glu1651*)	3	VUS	Originally agreed
96	NM_015166.3 (*MLC1*)[MIM:605908]:c.353C>T (p.Thr118Met)	3	Pathogenic/Likely Pathogenic	NA
97	NM_000018.2 (*ACADVL*)[MIM:609575]:c.1844G>A (p.Arg615Gln)	3	VUS/Likely Benign	NA
98	NM_007103.3 (*NDUFV1*)[MIM:161015]:c.753_756del (p.Pro252Glnfs*44)	9	Likely Pathogenic	Conference Call
99	NM_000057.2 (*BLM*)[MIM:210900]:c.2603C>T (p.Pro868Leu)	3	Benign	Conference Call

Table S2. Mean and standard deviation of the coefficient of variation for all lines of evidence

Line of Evidence	Used	Mean CV	SD CV
BP3	1	3.00	.
PP2	9	2.07	0.54
BP1	5	1.81	0.76
PM6	6	1.80	0.68
PP4	30	1.74	0.58
BS3	8	1.73	0.00
BS4	9	1.71	0.08
BP5	11	1.71	0.19
PS1	23	1.68	0.54
BS2	19	1.62	0.52
PM1	25	1.58	0.65
PM5	5	1.56	0.39
PM4	11	1.53	0.75
PS4	21	1.52	0.36
BP2	10	1.51	0.79
PM3	15	1.44	0.40
PP5	31	1.43	0.61
BP6	22	1.36	0.45
BA1	10	1.30	0.74
BP4	28	1.27	0.68
PM2	71	1.19	0.78
PP3	65	1.08	0.80
PS1	11	1.08	0.80
BS2	41	1.08	0.67
PP1	21	1.06	0.56
PS3	29	1.06	0.67
PVS1	25	0.55	0.65