UNC SCHOOL OF MEDICINE

September 15, 2016

Food and Drug Administration
Center for Devices and Radiological Health
Document Mail Center - WO66-G609
10903 New Hampshire Avenue
Silver Spring, Maryland 20993

Re.: IDE Supplement to IDE Number G150258

Dear Dr. Mullally,

We are submitting an IDE supplement for the NC NEXUS (North Carolina Newborn Exome Sequencing for Universal Screening) research project/device IDE number G150258 to request FDA approval to perform parental testing to clarify the phase of sequence variants that would otherwise have a result of "uncertain" significance. Enclosed for your review are two identical copies and an ECopy that is an exact duplicate of the paper copies.

If you have any questions about the material included in this IDE supplement, please do not hesitate to contact us.

Please accept our thanks, in advance, for the FDA's review and consideration of this IDE supplement.

Sincerely,

Cynthia M. Powell, MD
Department of Pediatrics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-4202
(919) 966-4202
powellcm@med.unc.edu

Jonathan S. Berg, MD/PhD
Department of Genetics
University of North Carolina at Chapel
Chapel Hill, NC 27599-4202
(919) 966-7043
jonathan_berg@med.unc.edu

Detailed description of change

Parental blood or saliva samples will be requested on a subset of approximately 150 study participants in which the study's Molecular Pathologist (Dr. Karen Weck, MD, PhD) determines that it would be important to conduct family segregation studies in order to most accurately interpret the genetic variant(s) identified in the patient (described in Sections 3.2.3.6.6 and 3.2.6.6.7, Genetic variant interpretation and reporting and CLIA confirmation). Parental samples will have DNA extracted using standard procedures in the UNC McLendon Molecular Diagnostic Laboratory. Custom Sanger sequencing will be performed using the same primers generated to confirm the patient's variant(s). Note that the testing performed in the parents is **site-specific testing** and **not whole exome sequencing.** Indications for performing parental testing include findings identified through the "NGS-NBS" panel of genes, "Indication-based analysis" and the optional categories of additional genomic findings (described in Section 3.1.3, Objectives of the clinical investigation).

Explanation of why the change is being requested

Family segregation studies are a standard of care approach used to determine two key evidence types that are used in the clinical assessment of sequence variants:

1. Determining whether two variants are *in cis* (present on the same chromosome) or *in trans* (present on opposite chromosomes), also called the "phase" of the variants.
2. Determining whether a variant is *de novo* or inherited from a parent.

These pieces of evidence can be used to interpret the pathogenicity of a patient's variant(s) using standard variant interpretation guidelines. For example, in a gene that is associated with a recessive genetic condition: a novel variant, when identified to be *in trans* with a known pathogenic variant, is more likely to be pathogenic. By contrast, inheritance of the same two variants from a single parent (*in cis*) can indicate that they are not, by themselves, causative for a recessive condition. In a

gene that is associated with a dominant genetic condition: a novel variant, when determined to have arisen *de novo*, is more likely to be pathogenic. Likewise, inheritance of a single heterozygous variant from an unaffected parent can argue against pathogenicity, especially when the condition is associated with complete penetrance. It is important to acknowledge that parental studies do not always definitively resolve questions about whether variants are involved in the etiology of an individual's condition, but the ability to use family segregation data to interrogate these questions are considered to be part of the routine standard in molecular diagnostic laboratory practice (described in Section 3.2.3.6.6, Genetic variant interpretation and reporting).

Assessment of the impact of the change on the study

We expect that this change in the study will have a beneficial effect on risk in the study (described in Section 3.3.1, Anticipated risks). Performing parental studies will allow the molecular results in the patients enrolled in the study to be interpreted with greater confidence and accuracy. First, parental studies will allow reclassification of many "variants of uncertain significance" to either pathogenic/likely pathogenic or benign/likely benign. This will have the effect of providing more clinically useful results.

Second, parental studies will help to illuminate specific risks for other family members by determining the segregation pattern of variants in a family. In the case of a recessive condition, this information will allow members on both sides of the family to take advantage of site-specific testing to determine their carrier status, should they desire to do so. In the case of a *de novo* dominant condition, the information will provide specific information about recurrence risk for the parents (low, but non-zero) and other relatives (population risk). In the case of an inherited dominant condition, the information will provide that parent and the family members on their side of the family with highly informative recurrence risk information.

Supporting documentation

Use of family segregation data in sequence variant interpretation guidelines:

Richards S et al.; 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.

Routine use of family segregation data in a clinical laboratory sequence variant interpretation procedures:

Eggington JM et al. A comprehensive laboratory-based program for classification of variants of uncertain significance in hereditary cancer genes. Clin Genet. 2014; 86(3):229-37.

Cheong PL and Caramins M. Approaches for classifying DNA variants found by Sanger sequencing in a medical genetics laboratory. Methods Mol Biol. 2014; 1168:227-50.

Demonstration of enhanced diagnostic specificity when trio sequencing is performed, indicating the utility of including family segregation in variant interpretation procedures:

Lee H et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA. 2014; 312(18):1880-7.

University of North Carolina at Chapel Hill
Consent to Participate in a Research Study
Adult Subjects - Relatives of Study Subject
Biomedical Form

IRB Study # 13-2409
Consent Form Version Date: 8/01/2016

Title of Study: North Carolina Newborn Exome Sequencing as Universal Screening (NC NEXUS)
Principal Investigators: Cynthia Powell, M.D. and Jonathan S. Berg, MD, PhD
UNC-Chapel Hill Department: Genetics
UNC-Chapel Hill Phone number: 919-966-7043
Email Address: powellcm@med.unc.edu; jsberg@med.unc.edu
Co-Investigators: Karen Weck, M.D., Donald Bailey, Ph.D and Christine Rini, Ph.D
Funding Source: National Human Genome Research Institute and National Institutes of Child Health and Development (National Institutes of Health)

Study Contact: Myra I. Roche, M.S., C.G.C.
Study Contact telephone number: (919) 537-3795
Study Contact email: NC_NEXUS@unc.edu

It is important that you understand the information in this consent form so that you can make an informed choice about joining this research study.

What are some general things you should know about research studies?
You are being asked to take part in a research study. To join the study is voluntary.
You may refuse to join, or you may withdraw your consent to be in the study, for any reason.

Research studies are designed to learn new knowledge that may help other people in the future. You may not receive any direct benefits and there may be risks from being in the research study.

Deciding not to be in the study or leaving the study before it is done will not affect your relationship with the researcher, your health care provider, or the University of North Carolina-Chapel Hill.

You will be given a copy of this consent form. You should ask the researchers or staff members, any questions you have about this study at any time.

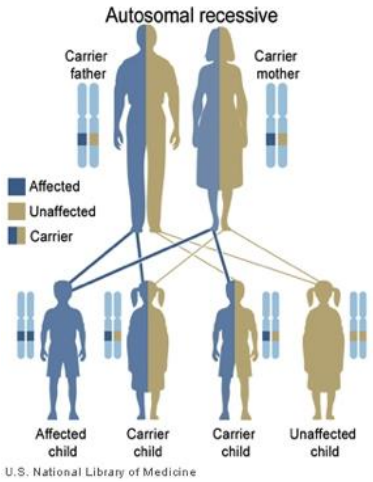
Why are you being invited to participate in this study?

One of your relatives has had genetic testing done as part of the research study called NC NEXUS. This testing looks for differences (variants) in genes. Most genetic variants are harmless, but some are harmful and can cause specific diseases.

When we do genetic testing, sometimes we cannot say for sure if a variant or combination of variants that are found would lead to a genetic condition or not.

For conditions that are inherited in an autosomal recessive pattern, like many of those tested for in NC NEXUS, people with the condition have two genetic variants, one of which they inherited from their

mother and the other they inherited from their father. When testing finds two genetic variants we may not be able to tell whether they were inherited separately (one from each parent) or inherited together from one parent. This information is called "phase." By testing both parents, we may be able to tell whether the variants that were found fit with a recessive inheritance pattern or not.



In other cases, the laboratory cannot immediately determine whether a variant is harmful or not. These variants are called "variants of uncertain clinical significance" (VUS). A VUS result means that a genetic variant was found, but there is not enough information about it to know for sure whether it cause disease or not.

You are being invited to participate in this study, because the results of your relative's genetic testing were uncertain either due to finding a combination of variants for which the phase is not known or due to finding a VUS. By testing your sample to see whether you have the same genetic variant or not, we may be able to learn information that could improve our interpretation of your relative's result.

Your sample will be tested only for the presence or absence of the specific variant(s) identified in your relative.

How long will your participation be expected to take?

Your part in this study is expected to last less than an hour.

What will happen if you take part in the study?

- We will need to obtain saliva samples for the genetic testing.
- We will label your samples with your NC NEXUS ID number and deliver it to the UNC Molecular Genetics laboratory at UNC Hospitals.
- The lab will test your sample to find out whether or not you have the genetic variant(s) identified in your relative.
- When the testing is done, the lab will report the results to a certified genetic counselor or medical geneticist on the research team by a secure email and they will report the results to you by phone.
- Your relative's clinical report will be amended to include the information learned as a result of your testing; however, you will only be identified by your relationship to the participant and not by your name.
- The accuracy of the results depends upon knowing the true biological relationship of the relatives being tested. If the stated father of a participant is not the true biological father, the interpretation may be incorrect.

What are the possible benefits from being in this study?
Research is designed to benefit society by gaining new knowledge. There is little chance you will benefit from being in this research study.

What are the possible risks or discomforts involved with being in this study?
This study requires a saliva sample and will only take a few minutes to obtain.

There may be uncommon or previously unknown risks that might occur. You should report any problems to the researchers.

What if we learn new things or information during the study?
You will be given any new information gained during the course of the study that might affect your medical care or your willingness to continue participating.

What will happen if you are injured by this research?
All research involves a chance that something bad might happen to participants. This may include the risk of personal injury. In spite of all safety measures, you might develop a reaction or injury from having the sample collected. UNC-Chapel Hill has *not* set aside funds to pay for any such reactions or injuries, or for the related medical care. However, by signing this form, you do *not* give up any of your legal rights.

Will there be any cost to you for participating?
You will *not* be charged for the visit or the genetic testing done as part of the study.

Will you receive anything for your participation?
You will receive parking vouchers for the UNC Hospitals parking deck.

What if you have questions about your rights as a research participant?
The IRB reviews all research on human volunteers in order to protect your rights and welfare. If you have questions or concerns about your rights as a research participant you may contact, the IRB at 919-966-3113 or to IRB_subjects@unc.edu. You do not have to use your name.

Participant Agreement:
I have read the information provided above and have asked all the questions I have at this time. I voluntarily agree to my participation.

_____ _____
Signature of Research Participant **Date**

_____
Printed Name of Research Participant

_____ _____
Signature of Research Team Member Obtaining Consent **Date**

_____
Printed Name of Research Team Member Obtaining Consent

Chapter 13

Approaches for Classifying DNA Variants Found by Sanger Sequencing in a Medical Genetics Laboratory

Pak Leng Cheong and Melody Caramins

Abstract

Diagnostic applications of DNA sequencing technologies present a powerful tool for the clinical management of patients. Applications range from better diagnostic classification to identification of therapeutic options, prediction of drug response and toxicity, and carrier testing. Although the advent of massively parallel sequencing technologies has increased the complexity of clinical interpretation of sequence variants by an order of magnitude, the annotation and interpretation of the clinical effects of identified genomic variants remain a challenge regardless of the sequencing technologies used to identify them. Here, we survey methodologies which assist in the diagnostic classification of DNA variants and propose a practical decision analytic protocol to assist in the classification of sequencing variants in a clinical setting. The methods include database queries, software tools for protein consequence, evolutionary conservation and pathogenicity prediction, familial segregation, case–control studies, and literature review. These methods are deliberately pragmatic as diagnostic constraints of clinically useful turnaround times generally preclude obtaining evidence from in vivo or in vitro functional experiments for variant assessment. Clinical considerations require that variant classification is stringent and rigorous, as misinterpretation may lead to inappropriate clinical consequences; thus, multiple parameters and lines of evidence are considered to determine potential biological significance.

Key words Clinical annotation, Databases, Diagnostics, Pathogenicity prediction, Sequencing, Variants

Abbreviations

HGMD	Human Gene Mutation Database
NCBI	National Center for Biotechnology Information
NG	Genomic
NM	mRNA
NP	Protein from RefSeq database
VUS	Variant of unknown significance

Ronald Trent (ed.), *Clinical Bioinformatics*, Methods in Molecular Biology, vol. 1168,
DOI 10.1007/978-1-4939-0847-9_13, © Springer Science+Business Media New York 2014

1 Introduction

The increasingly higher throughput and lower cost of sequencing technologies have facilitated routine mutation identification and characterization in medical laboratories. The ready availability of sequence data has expanded the incorporation of these results into clinical care and decision making. Consequently, the discovery of variants of unknown significance (VUS) is a common occurrence in clinical sequencing, regardless of the sequencing methodology utilized. The focus of a genetic diagnosis in response to a clinical question increasingly revolves around the key challenge of accurate, reliable, and reproducible variant annotation and interpretation.

A review of the concepts of analytical validity, clinical utility, and clinical validity is useful in this context. *Analytic validity* is defined as the process by which the performance of a test system is measured and assessed and often involves addressing inherent issues of quality control, robustness, accuracy, reliability, efficiency, and traceability. In this chapter, we assume that identified DNA variants have been detected in an analytically valid manner and do not directly address this aspect. *Clinical validity* refers to the accuracy with which a test predicts the presence or the absence of the phenotype or, stated as a question, *how accurately does the sequencing result predict the clinical phenotype*. *Clinical utility* of a sequencing test is the capacity of the result to rule a diagnosis in or out and thus make a decision to adopt or to reject a therapeutic course of action possible. Or *does the sequencing result allow the recommendation of a clinical course of action?* Both clinical validity and clinical utility are of great importance when interpreting variants in a diagnostic environment.

The context of Sanger sequencing variant annotation and interpretation usually involves addressing a specific biological hypothesis, which can often be phrased as the following: "Could the patients' signs and symptoms be the result of the detected variant(s) in this particular gene(s)?" Generally, Sanger sequencing will consider only a handful of genes to address this question, and therefore the hypothesis is tested only once or a handful of times. This contrasts with whole-exome or whole-genome sequencing, where multiple testing returns much larger numbers of variants and therefore requires greater interpretive caution due to the increased likelihood of a type I (false positive) error or the risk of increasing a type II error (false negative). Therefore, it is important that patients and physicians who order DNA genetic tests are aware of these limitations.

The diagnostic environment is also necessarily pragmatic; the need for clinically useful turnaround times precludes the ability to develop functional assays to assess directly biological effects of variants. Almost all assessments must be made more or less bioinformatically (in silico) by reference to literature databases, mutation

databases, and variant prediction software. Impressive research efforts such as the Duke University Task Force for Neonatal Genomics, where functional characterization of variants occurs in near real time, thus enabling results to be returned in a time frame which is useful in a neonatal intensive care setting, offer an interesting glimpse to the future.

In this chapter, we refer to many published international best practice guidelines on variant interpretation and classification. The reader is also encouraged to seek further information by consulting local best practice guidelines and interpretation standards, where available.

2 Materials

In order to illustrate clearly the approaches taken, we will use variants in the *CPOX*, *LDLR*, and *BRCA2* genes as examples in a process utilizing resources which can generally be classified into three groups: (1) databases (gene/locus specific and more generic), (2) browsers, and (3) tools.

Most of the resources are available online (*see* Table 1). Each step in variant assessment may use one or more of these. The lack

Table 1
Web resources used in variant classification

Tool	URL address
Align GVGD	http://agvgd.iarc.fr/
IARC breast cancer database	http://brca.iarc.fr/PRIORS/index.php
Breast Cancer Information Core	http://research.nhgri.nih.gov/bic/
COSMIC	www.sanger.ac.uk/genetics/CGP/cosmic/
Gene Ontology	www.geneontology.org
HGMD	www.biobase-international.com/product/hgmd
HGVS recommendations for the description of sequence variants	www.hgvs.org/mutnomen/recs.html
MutPred	http://mutpred.mutdb.org/about.html
NCBI Gene	www.ncbi.nlm.nih.gov/gene
CD-Search on NCBI Conserved Domains Database	www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
NCBI dbSNP	www.ncbi.nlm.nih.gov/projects/SNP/
PFAM	http://pfam.sanger.ac.uk

(continued)

Table 1
(continued)

Tool	URL address
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/
SIFT	http://sift.jcvi.org
SIFT BLink	http://sift.jcvi.org/www/SIFT_BLink_submit.html
SMART	http://smart.embl.de
SNPeffect	http://snpeffect.switchlab.org
SNPs&GO	http://snps.uib.es/snps-and-go/
T-COFFEE regular	www.igs.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi?stage1=1&daction=TCOFFEE::Regular
UCSC Genome Browser	http://genome.ucsc.edu/
UniProt (for Swiss-Prot protein code)	www.uniprot.org
UCL *LDLR* FH database	www.ucl.ac.uk/ldlr/LOVDv.1.1.0/index.php?select_db=LDLR

of standardization and differences in database update and review dates can sometimes be a source of interpretive conflict. In order to overcome this, the Human Variome Project (HVP) is considering the option of accreditation of databases for use in the clinical setting in the future.

3 Methods

There are several steps in establishing the potential clinical and biological significance of a given variant. The initial step in this process hinges on accurate annotation. This facilitates other downstream steps, including in silico predictions, obtaining population frequency data, and literature searches for functional information. A variant in the *CPOX* gene will be used as an example to illustrate these steps, followed by a comparative discussion on how this variant and two other variants in *LDLR* and *BRCA2* are classified based on the evidence of pathogenicity in a clinical setting.

3.1 Annotation of Variant and Visualization of Genomic Context

One of the first annotation steps typically involves contextualizing a particular genomic position within the sequence of known gene/s, transcript/s, or regulatory regions. This facilitates the process of prediction of likely variant effects (if any) on the resulting protein.

The current standard nomenclature system used in the annotation of variants has been developed by the Human Genome Variation Society (HGVS) for the description of genetic variants [1]. Older nomenclature systems may still be in historical use, referred

to in the literature or in databases; this should be noted during review by the user with appropriate caution. Annotation of variants from Sanger sequencing traces can be undertaken using commercially available packages (Mutation Surveyor® from Softgenetics is one such example) or manually by using freely available software. In either instance, a reference sequence is generally a key initial requirement.

It is also important to note at this stage that available software (both commercial and free) will frequently come with a disclaimer that the product should only be used for research purposes and not clinical decision making. This should be recognized as it will mean that a formal evaluation is required to validate the software prior to its use for clinical testing.

Curated reference sequences can be obtained by searching for the gene name, e.g., *CPOX* for coproporphyrinogen-III oxidase, at the National Center for Biotechnology Information (NCBI) Gene website. *G*enomic, *m*RNA, and *p*rotein sequence accessions are listed with the prefixes N*G*, N*M*, and N*P*, respectively, under the *NCBI Reference Sequences* (RefSeq) section (Fig. 1). For mRNAs,

Fig. 1 A view on NCBI Reference Sequences (RefSeq). The link to Sequence Viewer (Graphics) is *arrowed*

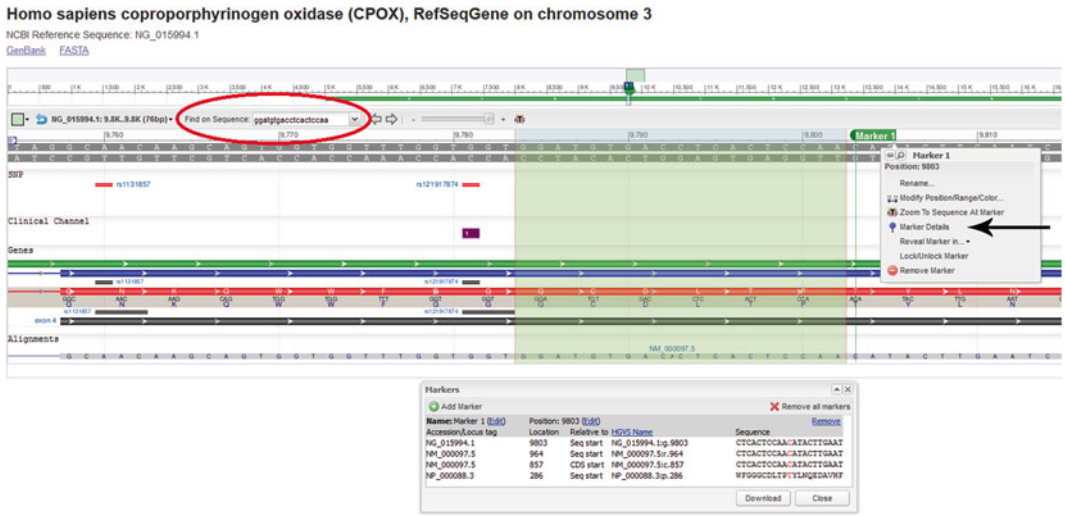


Fig. 2 Searching for sequence and HGVS nomenclature on NCBI Sequence Viewer

the longest transcript is usually used as the reference transcript, although in some cases (for example, the *FECH* gene), the longest transcript may not be the predominant transcript in vivo. Arriving at that site page, the GenBank or FASTA format genomic reference sequences for the gene of interest can be downloaded under the NCBI Reference Sequences tab. A graphical view of the gene is also available by clicking the link "Sequencer Viewer (Graphics)."

One method of locating a variant from the reference sequence involves searching the surrounding sequence on the "Find on Sequence" panel in the Sequencer Viewer (circled at the top of Fig. 2) by entering the sequence into this search box. In our example, for the sequence string containing the C>A change in exon 4 of the *CPOX* gene (…GGATGTGACCTCACTCCAA(C/A)ATACTTGAA…), enter the immediately adjacent sequence "GGATGTGACCTCACTCCAA" in the Sequence Viewer search box. This will pinpoint the region adjacent to the C>A change, and a marker can then be created at the SNP by right clicking on the nucleotide and selecting "Set New Marker At Position." Once the marker is set (Marker 1 in this case), move the cursor to the "Marker 1" label and select "Marker Details" (highlighted by arrow in Fig. 2) to show the HGVS nomenclature of this variant in the fourth column. The amino acid change (ACA>AAA; Thr>Lys) can also be deduced. HGVS nomenclature for this variant is NM_000097.5:c.857C>A or NP_000088.3:p.Thr286Lys depending on whether the coding DNA sequence or amino acid change is emphasized.

Once annotation is completed, the variant needs to be evaluated further according to its sequence context and location. In the *CPOX* example above, there is a non-synonymous variant within the coding region (Table 2).

Table 2
For coding variants the following should also be considered when evaluating biological effects

Type of variant and potential effects	Additional points to consider
Frameshift—insertion or deletion	Is there an alteration in reading frame?
Missense—stop-gain/nonsense (substitution resulting in a stop codon)	In some instances nonsense mutations may not have functional significance, such as the p.Lys3326Ter variant in *BRCA2*, arising from an NM_000059.3:c.9976A>T substitution which results in a stop codon and loss of the final 93 amino acids of the BRCA2 protein. This variant has a reported allele frequency of 0.8 % in some populations and is not considered to be clinically significant [5]
Missense—substitution resulting in loss of stop codon	For example Hb Constant Spring (p.Ter143Gln in *HBA2*) resulting from a stop-loss mutation leading to a lengthened peptide [6]
Insertion/deletion not causing a frameshift	Caution is advised, e.g., in the *LDLR* c.2397_2405delCGTCTTCCT in-frame deletion. This deletion is interesting in that it has no or little effect per se in vitro but becomes functional when found in cis in combination with a non-synonymous variant p.Asn543His [7]
Non-synonymous single-nucleotide variant (SNV) or synonymous SNV	Synonymous SNV can sometimes be pathogenic by affecting splicing. As an example, a critical yet translationally silent C>T variant at position 6 in *SMN2* exon 7 compromises its splicing, causing most of the *SMN2* mRNA (~80 %) to lack exon 7 (SMNΔ7). The resulting unstable molecule is rapidly degraded, leaving patients with SMN deficiency, the degree of which correlates with clinical severity of spinal muscular atrophy [8]

3.2 In Silico Analysis of Annotated Missense Variants

There are several considerations in assessing whether a missense variant is likely to be pathogenic or nonpathogenic.

3.2.1 Location of Variant in Relation to the Transcript

When evaluating biological significance, the following list presents some general considerations. In the absence of other evidence, these are listed from more to less likely predictive of functional effects:

1. Coding region variants—*See* Table 2.
2. Invariant splice sites—Coding nucleotides close to the exon–intron boundaries may not only affect amino acid sequence but also splicing; the first two nucleotides at the start (donor site) or the end (acceptor site) of introns are invariable in 98.71 % of genes—these are called canonical dinucleotides, and mutation to these nucleotides is invariably associated with alternative splicing effects [2].
3. 5′ or 3′ untranslated region (UTR) of a transcript—Variants in these regions *may* have influence on gene expression.
4. Noncoding exon—In some genes not all exons are transcribed, and this alternative transcription may be tissue specific.

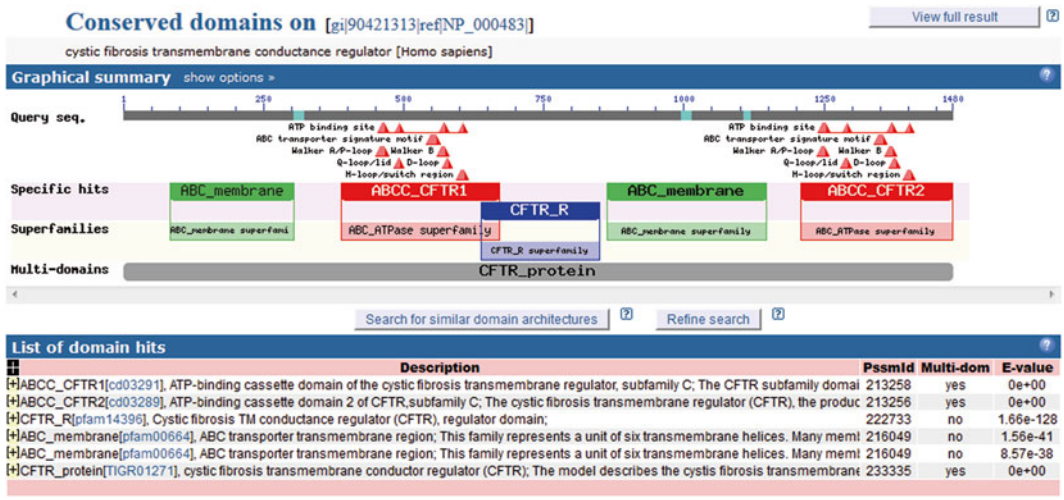


Fig. 3 Conserved protein domain of CFTR on NCBI Conserved Domains Database

For example, exon 1 and part of exon 2 of *HMBS* are transcribed in non-erythroid tissues but not in erythroid tissue [3].

5. Intronic—Intronic variants outside the canonical splice site can also affect splicing. As an example, the variant NM_000140.3:c.315-48T>C, in intron 3 of *FECH*, promotes the use of a cryptic acceptor site, resulting in an aberrant transcript with a premature stop codon [4].
6. Upstream or downstream of transcript start site: An *arbitrary* distance of 500–1,000 bp may be used by some laboratories.

3.2.2 Functional Domains

An important consideration for predicting the functional significance of a variant is its position within known protein domains, such as those defined by PFAM [9], SMART [10, 11], or other domain classification approaches. Conserved residues within functional domains are an indicator of negative evolutionary selection and so considered to provide some indirect evidence that changes will affect protein function. Conserved Domains Database (CDD) on NCBI [12–14] is a curated database that incorporates information from such sources. The CD-Search tool can align a given accession, GenoInfo Identifier (GI) number, or FASTA protein sequence to known domains in the database. An example using cystic fibrosis transmembrane conductance regulator (CFTR) protein sequence is provided in Fig. 3. The two transmembrane domains (ABC_membrane), two ATPase domains (ABCC_CFTR1 and 2), and R domain are shown with conserved amino acids within these domains marked in triangles. It is possible to identify whether variants of interest lie within these domains and if the amino acid of interest is conserved.

3.2.3 In Silico Prediction Strategies

Where no experimental data are available, the effect of a missense variant on protein function may be predicted by using various in silico tools. These predictive tools are generally based on two principles outlined below and should be used very cautiously, especially if predictions are not supported by additional evidence. Examples include the following:

- *Analyses based on evolutionary conservation of the nucleotide or amino acid*. As highlighted, negative evolutionary selection is an important indicator of functional significance of residues. By aligning nucleotide and amino acid sequences of orthologs from different species, the divergence of these residues and implication for putative functional effects may be deduced. Examples of these tools include phyloP, phastCons, SIFT, Align GVGD, Mutation Assessor, PANTHER, and MAPP (detailed discussion in Tavtigian et al. [15]).
- *Structural and biophysical property-based analyses*. This involves an analysis of the difference in biophysical properties between the reference and variant amino acid and predicting the probability that the resulting change will significantly affect protein structure. Examples of these tools include PolyPhen-2, SNPeffect (incorporating FoldX which is a protein stability-based prediction), and LS-SNP/PDB.

Some tools (for example, Align GVGD and SNPs&GO) use a combination of both strategies and may include supervised machine learning to improve their predictions (Table 3).

3.2.4 Examples of In Silico Prediction Tools

As mentioned above, the following in silico prediction tools utilize three broad strategies/methodologies: (1) evolutionary conservation, (2) structural and biophysical properties, and (3) machine learning.

- *PhyloP and phastCons*. PhyloP (phylogenetic *P*-value) and phastCons are two phylogenetic scoring systems which quantitatively measure evolutionary conservation [16, 17]. PhyloP scores are based on a measure of conservation at the level of individual nucleotides and are calculated as –log *P*-values, where a positive score indicates conservation. PhastCons relies on identifying elements ("runs") of conserved sites, with scores ranging between 0 and 1, representing the probability of negative selection. These scores are integrated into the University of California Santa Cruz (UCSC) Genome Browser under the conservation track (Fig. 4). PhyloP and phastCons scores can easily be visualized by changing the settings in the Conservation Track to include them (Fig. 5).
- *SIFT*. Another commonly used tool based on sequence homology is *S*ort *I*ntolerant *F*rom *T*olerant (SIFT) [18–21], which comes with a user protocol [22]. One advantage of SIFT is

Table 3
Examples of some classifications used for DNA variants

ACMG classification	IARC classification (with probability of the variant being pathogenic)
Variant previously reported and is a recognized cause of the disorder	Class 1—not pathogenic (<0.1 %)
Variant previously unreported and is expected to cause the disorder	Class 2—likely not pathogenic (likelihood of pathogenicity 0.1–5 %)
Variant previously unreported and may or may not be causative of the disorder	Class 3—uncertain (5–94.9 %)
Variant previously unreported and is probably not causative of disease	Class 4—likely pathogenic (likelihood of pathogenicity 95–99 %)
Variant previously reported and is a recognized natural variant	Class 5—pathogenic (>99 %)
Variant is not known or expected to be causative of disease but is found to be associated with a clinical presentation, e.g., variants associated to particular disease from genome-wide association studies or modifier genes	

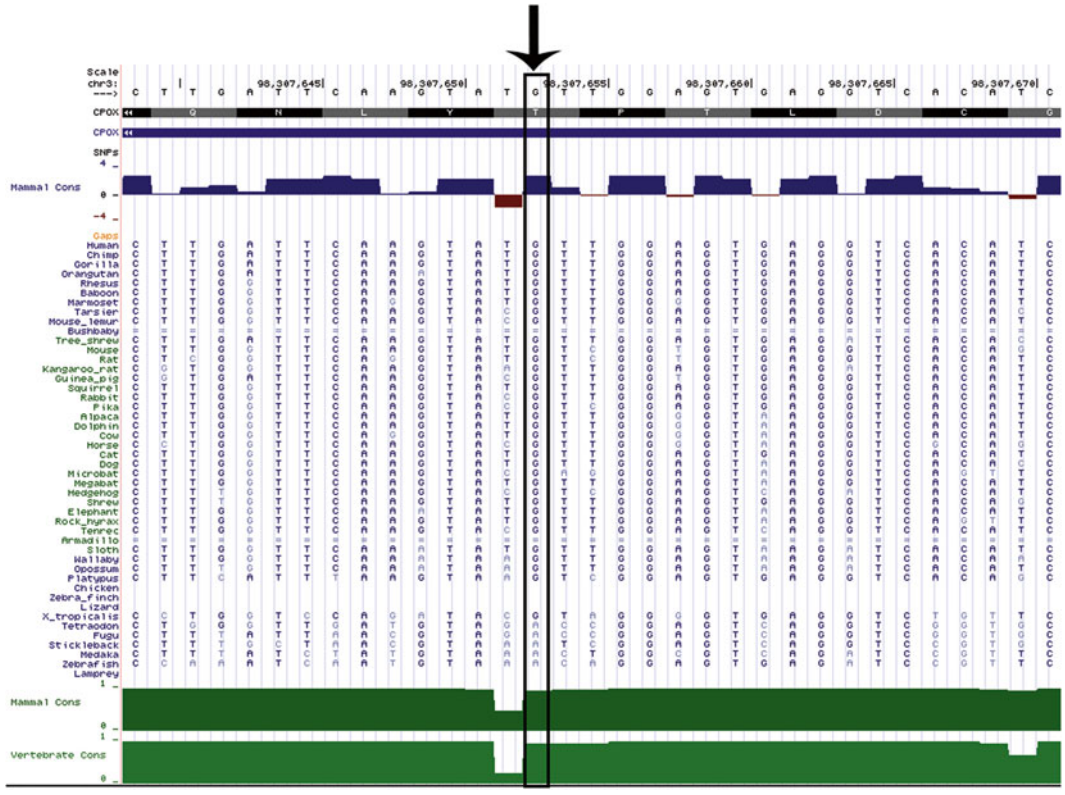


Fig. 4 PhyloP and phastCons scores in UCSC Genome Browser. The nucleotide change leading to CPOX p.Thr286Lys (*arrowed*) is highly conserved as indicated by phyloP (*blue bar* at *top*) and phastCons (*green bars* at the *bottom*). This is in contrast with the adjacent nucleotide where the negative phyloP (in *red*) and low phastCons score indicate accelerated evolution (Color figure online)

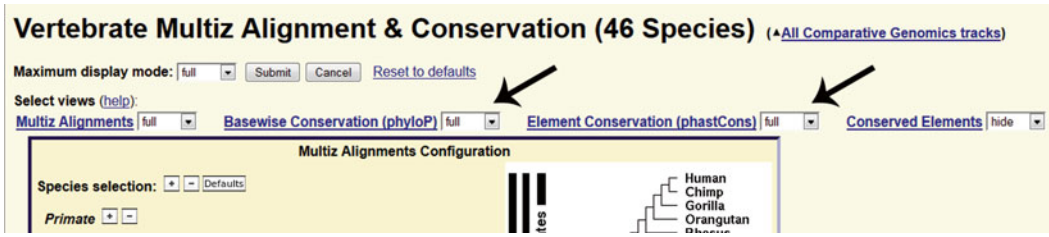


Fig. 5 Conservation track settings in UCSC Genome Browser. Click on "Conservation" under Comparative Genomics in the UCSC Genome Browser to adjust conservation track settings. Select "full" for Basewise Conservation (phyloP) and Element Conservation (phastCons) to display these scores in the Genome Browser (*arrows*)

that it accepts various input formats (e.g., Ensembl protein transcript ID, NCBI GI number, protein FASTA sequence, or RefSeq ID from dbSNP if it is a known SNP) for single-protein or -batch analyses. Users can either allow SIFT to build a multiple sequence alignment or they can submit their own. A SIFT score based on normalized probability of all 20 amino acids appearing in that particular position is calculated, and SIFT will "call" the variation damaging if the score lies below a threshold (predefined at 0.05). It also returns a median sequence conservation score which ranges from 0 (all 20 amino acid substitutions have been observed in multiple sequence alignment at the position) to 4.32 (where only one amino acid is observed at that position). Although a score of >3.25 would ordinarily indicate high conservation, if too few organisms are considered in the alignment, this may simply be reflective of this lack of diversity. Ideally, representation should be as broad as possible, including species from all vertebrate groups, e.g., Mammalia, Primates, Aves, Amphibia, and Reptilia. In our example, we use SIFT BLink, a rapid version of SIFT, as it runs analysis with pre-computed multiple sequence alignment from BLAST search. The GI number for CPOX (41393599) was retrieved from NCBI Protein. Submitting a query for the variant of interest (T286K), SIFT BLink predicted the variant to be tolerated with a score of 0.20 (median sequence conservation score 2.94, with 83 sequences aligned at this position).

- *Align GVGD*. Align GVGD [23] combines biophysical characteristics and multiple protein sequence alignments to predict the pathogenicity of variants. Align GVGD calculates the Grantham variation (GV, variation in the biophysical properties of all amino acids at a particular position in the multiple protein sequence alignment) and Grantham deviation (the deviation in biophysical properties of the altered amino acid from the reference). These scores are based on Grantham scores which measure the volume, polarity, and side chain composition of amino acids [24]. The two scores are combined

to provide a classification of pathogenicity likelihood, ranging from C0 (less likely to be deleterious) to C65 (most likely to be deleterious).

To perform Align GVGD:

- Download FASTA protein sequences for alignment. For CPOX, reference CPOX protein sequences were downloaded from NCBI. There are six NP accessions, i.e., non-predicted protein sequences as prefixed by XP, from *Homo sapiens*, *Sus scrofa*, *Mus musculus*, *Danio rerio*, *Rattus norvegicus*, and *Bos taurus*.
- Submit the downloaded FASTA file to T-COFFEE regular [25], a multiple sequence alignment tool (*see* **Note 1**).
- The multiple sequence alignment is uploaded in FASTA format onto Align GVGD. Enter "T286K" for Substitutions list, and submit the job.

Align GVGD classified the p.Thr286Lys variant as less likely to be deleterious with GV of 144.57 and GD of 8.11. The classification (C0) remained unchanged even when all CPOX protein sequences including predicted sequences were used for alignment.

- *SNPeffect*. SNPeffect [26] analyzes the structural effect of variants on protein using various algorithms (TANGO, WALTZ, LIMBO, and FoldX). Some pre-computed variants are available for search on their database. Using the CPOX p.Thr286Lys example, SNPeffect showed that the variant had no effect on aggregation tendency, amyloid propensity, or chaperone binding. Structural analysis using FoldX however predicted that the Thr-to-Lys change would result in a difference in free energy and hence reduction in protein stability (Fig. 6).

  FoldX prediction is only provided if there is a homologous structural model available. Users can use partial protein sequence, e.g., a particular domain only, to broaden the homology search.
- *MutPred*. MutPred [27] predicts the effect of amino acid changes on (1) protein structure and dynamics, e.g., secondary structure and transmembrane helix; (2) predicted functional properties, e.g., catalytic residues and glycosylation sites; and (3) evolutionary information (based on SIFT). Input requirements include the FASTA sequence of the wild-type protein and the amino acid change. Two scores are returned—a general score to predict whether the variant is deleterious and *P*-values of the top five properties that may be altered as a result. The calling algorithm is based on machine learning with

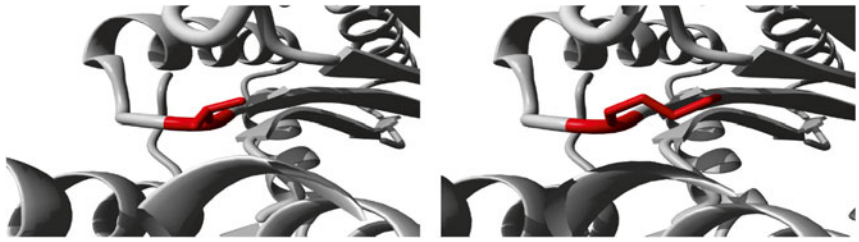


Fig. 6 FoldX prediction on the CPOX p.Thr286Lys variant from SNPeffect. The empirical protein design force-field FoldX is used to calculate the difference in free energy of the mutation: ddG (delta delta G). If the mutation destabilizes the structure, ddG is increased, whereas stabilizing mutations decrease the ddG. Since the FoldX error margin is around 0.5 kcal/mol, changes in this range are considered insignificant. 2aex has 100.00 % homology with the submitted sequence. This pdb is then used to get some more information on the structural effect. The mutation from THR to LYS at position 286 results in a ddG of 3.29 kcal/mol. This implies that the mutation reduces the protein stability. Molecular visualization of the WT (*left*) and variant (*right*) amino acid. The residues colored in *red* represent the wild-type (THR) and variant residue (LYS)

Fig. 7 Output from MutPred. In the p.Thr286Lys example, MutPred returned a general score of 0.555 with three actionable hypotheses, suggesting some evidence that the variant may affect function

a random forest classifier, using data from HGMD and Swiss-Prot as a training set. Depending on the two scores the potential alteration in molecular mechanism is categorized into actionable, confident, or very confident hypotheses (Fig. 7).

- *PolyPhen-2.* PolyPhen-2 uses machine learning to select optimally 11 sequence- and structure-based predictive features for assessment of pathogenicity [28]. Various input formats are allowed. Two datasets (HumDiv and HumVar), both retrieved from UniProt, were used to train the algorithm (*see* supplementary material in ref. 28 for details of the two datasets). PolyPhen-2 returns a probabilistic score on whether the variant is likely to be damaging or not on both datasets. The scoring system for PolyPhen-2 is complex. For example, the HumVar-trained PolyPhen-2 score is more conservative as the HumVar dataset assumes all non-synonymous SNPs with no disease annotation as benign, meaning that variants with mild effect

will be considered benign. On the other hand, the HumDiv-trained PolyPhen-2 score would be more suitable for association discovery albeit a higher false-positive rate. This difference can be highlighted in our example CPOX p.Thr286Lys, where it is *probably damaging* with a score of 0.0995 on HumDiv, but only *possibly damaging* with a score of 0.733 on HumVar.

- *SNPs&GO*. SNPs&GO [29, 30] is an algorithm that makes use of Gene Ontology (GO) functional annotation. It uses support vector machines to incorporate the amino acid change, surrounding protein sequence environment, residue conservation (all of which form the basis of the algorithm PhD-SNP developed by the same laboratory group [31]), multiple sequence alignment (based on another algorithm PANTHER), and GO terms. For an explanation on the actual calculation of GO analysis *see* Kaminker et al. [32]. To use SNPs&GO, find the Swiss-Prot code for the protein of interest from the UniProt website, e.g., for CPOX it is HEM6_HUMAN. If Swiss-Prot code is not available, the FASTA sequence may be utilized but the associated GO terms will need to be entered manually (*see* **Note 2**). GO term associated to a protein can be searched on the Gene Ontology website. Enter the variant, and submit the job. The results will show predictions from SNPs&GO and the two related algorithms (PhD-SNP and PANTHER) (Fig. 8). A variant will be predicted as disease causing (second column) if the probability is above the default setting of 0.5. Measure of the quality of this binary classification (disease or neutral) is provided as a reliability index (RI), which correlates to the accuracy and the Matthews correlation coefficient (MCC; *see* ref. 33 for explanation on using MCC in evaluating the accuracy of predictions). In our example of CPOX p.Thr286Lys, SNPs&GO and PhD-SNP predicted the variant to be disease causing, while PANTHER predicted it to be neutral. The lower probability and reliability index assigned by SNPs&GO as a final score in this instance is reflective of the differing information provided by the PhD-SNP and PANTHER inputs.

3.2.5 Evaluation of In Silico Prediction Tools

There is currently no single consensus method for assessing variant pathogenicity using in silico prediction tools, and no tool alone is universally acknowledged as providing the most accurate prediction in all circumstances. The National Genetics Reference Laboratory (NGRL) at Manchester, UK, has published an evaluation of some prediction tools [33]. This report recommended a consensus approach when the best in silico tool for the particular gene of interest is unknown. The combination of three commonly used in silico prediction tools (PolyPhen-2, SIFT, and Align GVGD) was shown to have inferior prediction, often because the predictions

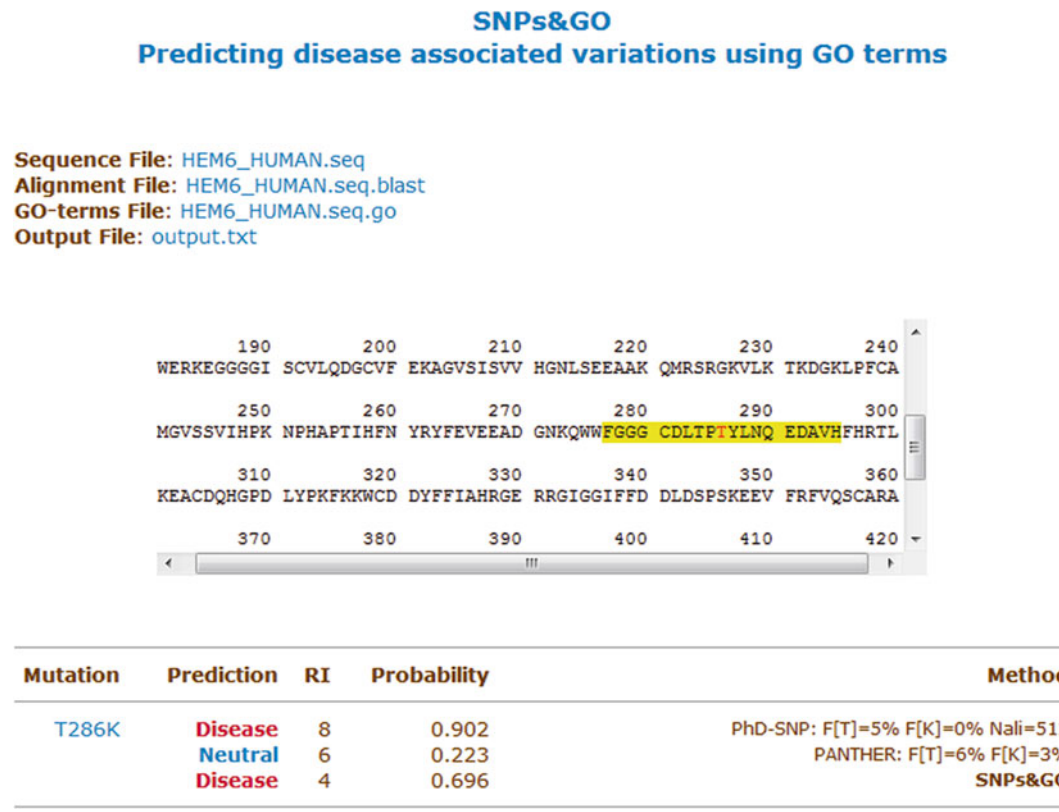
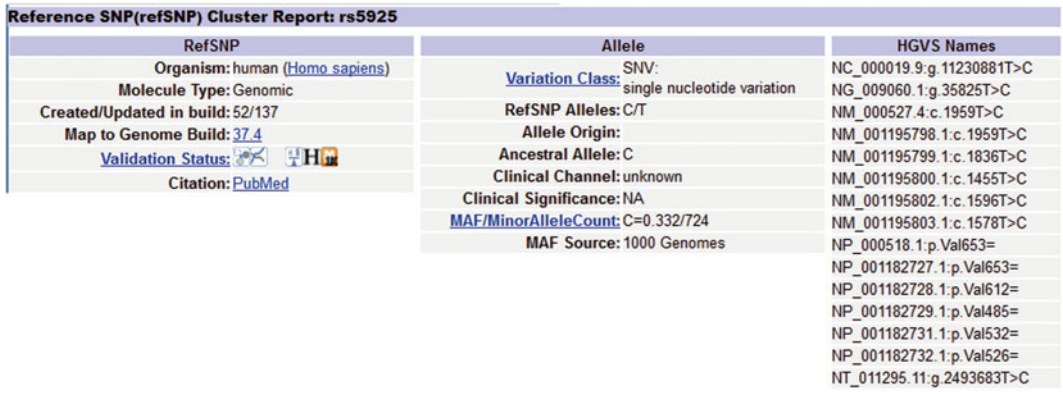


Fig. 8 SNPs&GO output for the CPOX variant p.Thr286Lys

from these tools are contradictory. These three tools are also those accessed directly through Alamut® (Interactive Biosoftware), a commercially available software package commonly used in many diagnostic laboratories. The combination of in silico tools that provided the most accurate predictions for the four genes investigated in the report (*BRCA1*, *BRCA2*, *MLH1*, and *MSH2*) included MutPred, SNPs&GO, and MAPP. MutPred and SNPs&GO were also shown to have the best predictions in over 40,000 pathogenic and neutral variants tested in another study [34].

Evolutionary conservation-based tools are highly sensitive to input multiple sequence alignments. The NGRL report demonstrated that pathogenicity prediction could change substantially depending on input alignment. Align GVGD provides curated alignment for several cancer susceptibility genes. If a laboratory is performing regular assessment of particular genes, building an in-house alignment for these genes should be considered.

The NRGL report did not recommend the use of protein stability-based methods such as FoldX in variant effect prediction due to the variability of tolerance to stability change between proteins.



Reference SNP(refSNP) Cluster Report: rs5925

RefSNP		Allele		HGVS Names
Organism:	human (Homo sapiens)	Variation Class:	SNV: single nucleotide variation	NC_000019.9:g.11230881T>C
Molecule Type:	Genomic			NG_009060.1:g.35825T>C
Created/Updated in build:	52/137	RefSNP Alleles:	C/T	NM_000527.4:c.1959T>C
Map to Genome Build:	37.4	Allele Origin:		NM_001195798.1:c.1959T>C
Validation Status:		Ancestral Allele:	C	NM_001195799.1:c.1836T>C
Citation:	PubMed	Clinical Channel:	unknown	NM_001195800.1:c.1455T>C
		Clinical Significance:	NA	NM_001195802.1:c.1596T>C
		MAF/MinorAlleleCount:	C=0.332/724	NM_001195803.1:c.1578T>C
		MAF Source:	1000 Genomes	NP_000518.1:p.Val653=
				NP_001182727.1:p.Val653=
				NP_001182728.1:p.Val612=
				NP_001182729.1:p.Val485=
				NP_001182731.1:p.Val532=
				NP_001182732.1:p.Val526=
				NT_011295.11:g.2493683T>C

Fig. 9 Summary of SNP information on NCBI dbSNP. Validation status is depicted by various *symbols*. Click on the "Validation Status" link for description on these symbols

3.3 Population Frequencies

Large-scale sequencing projects such as the HapMap project, 1000 Genomes Project, and Exome Sequencing Project from the National Heart, Lung, and Blood Institute (NHLBI-ESP) provide frequency information on polymorphisms that allow inference on pathogenicity. Although common polymorphisms are unlikely to be deleterious, this does not exclude the possibility of milder or modifying effects on protein function.

When accessing these data it is important to (1) understand the source, phenotype, and ethnicity of selected samples in these projects as these will influence variant frequency and (2) be aware of the validation status of reported SNPs. SNPs are considered validated on dbSNP when at least one of the submissions is obtained by experimental methods, the submission contains frequency information, e.g., data from HapMap or 1000 Genomes Project, or there are multiple independent observations [35]. SNPs that have not been validated may be false positives.

The NCBI dbSNP database collates frequency information from various sources. To view SNP summary on dbSNP, users can search for reported SNPs using the HGVS name (under "Search by ID on All Assemblies"). Alternatively, reported SNPs will be highlighted in red under the SNP track in the NCBI Sequence Viewer (see above). The SNP summary shows the minor allelic frequency (MAF) count using data from the 1000 Genomes cohort. Different levels of validation are also shown in symbols (Fig. 9).

Ethnic based population frequencies can be obtained under the "Population Diversity" section (Fig. 10). In the example of rs5925, the allelic frequencies in European (CEU), Asian (HCB and JPT), sub-Saharan African (YRI), and other ethnicities obtained from the HapMap project are shown (red box). Where available, data from NHLBI-ESP are also provided (green box).

	Sample Ascertainment				Genotype Detail				Alleles	
ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	C/C	C/T	T/T	HWP	C	T
ss107937053	ABECASIS_CLINICAL_PANEL		748	AF					0.354	0.646
ss142653712	ENSEMBL_Venter		2	IG	1.000				1.000	
	ENSEMBL_celera		2	IG	1.000				1.000	
ss171497730	PGP		2	IG		1.000			0.500	0.500
ss19402919	CEPH		184	AF					0.370	0.630
	HapMap-CEU	European	226	IG	0.204	0.460	0.336	0.527	0.434	0.566
	HapMap-HCB	Asian	86	IG	0.070	0.326	0.605	0.584	0.233	0.767
	HapMap-JPT	Asian	172	IG	0.023	0.360	0.616	0.317	0.203	0.797
	HapMap-YRI	Sub-Saharan African	226	IG	0.018	0.186	0.796	0.584	0.111	0.889
	HAPMAP-ASW		98	IG	0.020	0.367	0.612	0.371	0.204	0.796
	HAPMAP-CHB	Asian	82	IG	0.073	0.146	0.780	0.010	0.146	0.854
	HAPMAP-CHD		168	IG	0.060	0.333	0.607	0.752	0.226	0.774
	HAPMAP-GIH		176	IG	0.250	0.455	0.295	0.439	0.477	0.523
	HAPMAP-LWK		172	IG	0.035	0.267	0.698	0.752	0.169	0.831
	HAPMAP-MEX		100	IG	0.320	0.480	0.200	1.000	0.560	0.440
	HAPMAP-MKK		282	IG	0.014	0.312	0.674	0.251	0.170	0.830
	HAPMAP-TSI		170	IG	0.141	0.565	0.294	0.150	0.424	0.576
ss342486226	ESP_Cohort_Populations		4550	GF	0.139	0.420	0.441	0.001	0.349	0.651

Fig. 10 Allelic frequency of rs5925 in different ethnicities on dbSNP

The collection of in-house data where no dbSNP entry is available is encouraged. An example is a promoter variant c.-24C>G in *HBA2* or *HBA1* that is commonly found in patients investigated for alpha thalassemia. This variant has been found in association with other known causative *HBA2/HBA1* variants and is also seen on its own in association with a normal phenotype. Therefore this variant would be considered benign in terms of function.

3.4 Family Studies

Family studies provide information about whether a variant of interest is inherited or de novo, about its pattern of inheritance (dominant/recessive/paternal/maternal), and whether the variant co-segregates with the phenotype. This can be especially useful when interpreting variants private to a family which may not be described in the literature, a situation which is frequently present.

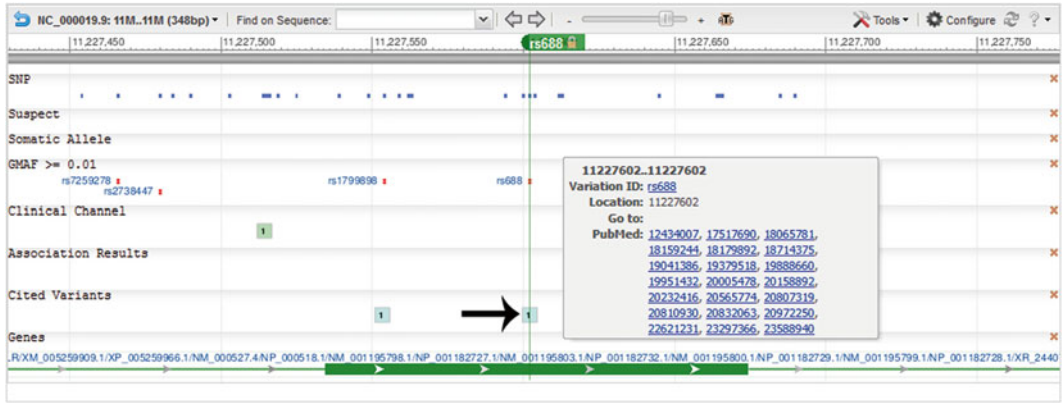


Fig. 11 PubMed IDs for articles related to a specific SNP can be viewed under "Cited Variants" (*arrowed*)

3.5 Literature Search for Published Evidence of Biological Effect (Functional Studies, etc.)

Searching the published literature for functional evidence of a particular variant is essential when attempting to establish potential effects. In dbSNP, PubMed IDs for articles related to a reported SNP are listed under the "Cited Variants" track where available (Fig. 11). Locus-specific databases such as the University College of London *LDLR* familial hypercholesterolemia database for the *LDLR* gene [36], the Breast Cancer Information Core (BIC), and the International Agency for Research on Cancer (IARC) breast cancer database for *BRCA1* and *BRCA2* are also useful. The UK Clinical Molecular Genetics Society (CMGS) guidelines for interpretation and reporting of unclassified variants states under section 4.1 that consulting locus-specific databases when reporting unclassified variants is *essential*, although curatorial rigor of all databases is a consideration [37]. It is envisaged that in the future, larger curated databases such as the Human Gene Mutation Database (HGMD) and Catalogue of Somatic Mutations in Cancer (COSMIC), which also provide links to reference journal articles, may become comprehensive enough to include information available in locus-specific databases and may thus supplant this requirement.

In all instances, the *quality* of source data must be carefully evaluated. It is important to establish whether the evidence presented has been based on in silico, in vitro, or in vivo studies, with robustness of reported results verified by checking for reproducibility by independent groups and/or involving different populations. Too frequently a variant is reported as having been independently found a number of times, but, on closer inspection, the multiple observations are actually based on the one original publication. It may also be useful to search for information other than journal articles such as conference abstracts via search engines, although non-peer-reviewed data are of limited use in a clinical setting.

3.6 Underlying Biological Knowledge

Understanding the underlying molecular mechanisms of disease and their consequences in phenotype causation is essential for variant interpretation. For example, in dominant conditions (such as some inherited cancer syndromes) where homozygous/compound heterozygous allelic loss may be embryonic lethal, co-occurrence of a VUS with another known pathogenic mutation in trans can indicate that the variant is unlikely to be pathogenic.

In other instances, such as in the molecular diagnosis of thalassemia, laboratory phenotypic data may be helpful. In these instances it is important to correlate genotyping results with the phenotype to ascertain whether further studies are required. For example, a hematological profile may be less or more severe than predicted on the basis of a known hemoglobin beta gene mutation. This may be indicative of gene–gene interactions such as a deletion, single-base change, or even a duplication of the hemoglobin alpha globin gene.

3.7 Classification Models for Variants Based on Evidence of Pathogenicity

The classification of variant pathogenicity is a complex task that requires professional judgment based on the collective evidence from all the aspects discussed and considered in this chapter. The American College of Medical Genetics (ACMG) has published an approach, detailed in a decision flow chart, on variant classification ([38]; Fig. 12). In both research and medical diagnostics, not all variants will have sufficient information for an unequivocal determination, and not all evidence will have the same strength. As a general principle, variants demonstrated to have biological effects with in vivo and/or in vitro evidence are more convincing than those suggested solely on the basis of in silico predictive effects. Even an in vitro environment, although indicative, may not always be a true reflection of in vivo effects as complex biological interactions cannot be assessed. Cassa et al. [39] have shown that 8.5 % variants classified as disease causing in the manually curated HGMD database are found in asymptomatic individuals. Conversely, a disease phenotype may be due to quantitative or pleiotropic effects of variants beyond the gene of interest. This is especially the case as more and more genotype–phenotype associations are unveiled by genomic scale researches. Interdisciplinary consortia such as evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) are increasingly being formed to harmonize variant interpretation for some clinically important genes [40].

There are various classification systems to categorize variants in the clinical context. The ACMG have proposed a six-category classification [38], and for cancer susceptibility genes the IARC Unclassified Genetic Variants Working Group has suggested a five-class classification defined by the probability of a variant being pathogenic [2, 41] (Table 3).

With limited information, there are likely to be a large proportion of variants classified in the *uncertain* or the *likely pathogenic/likely nonpathogenic categories*, complicating genetic

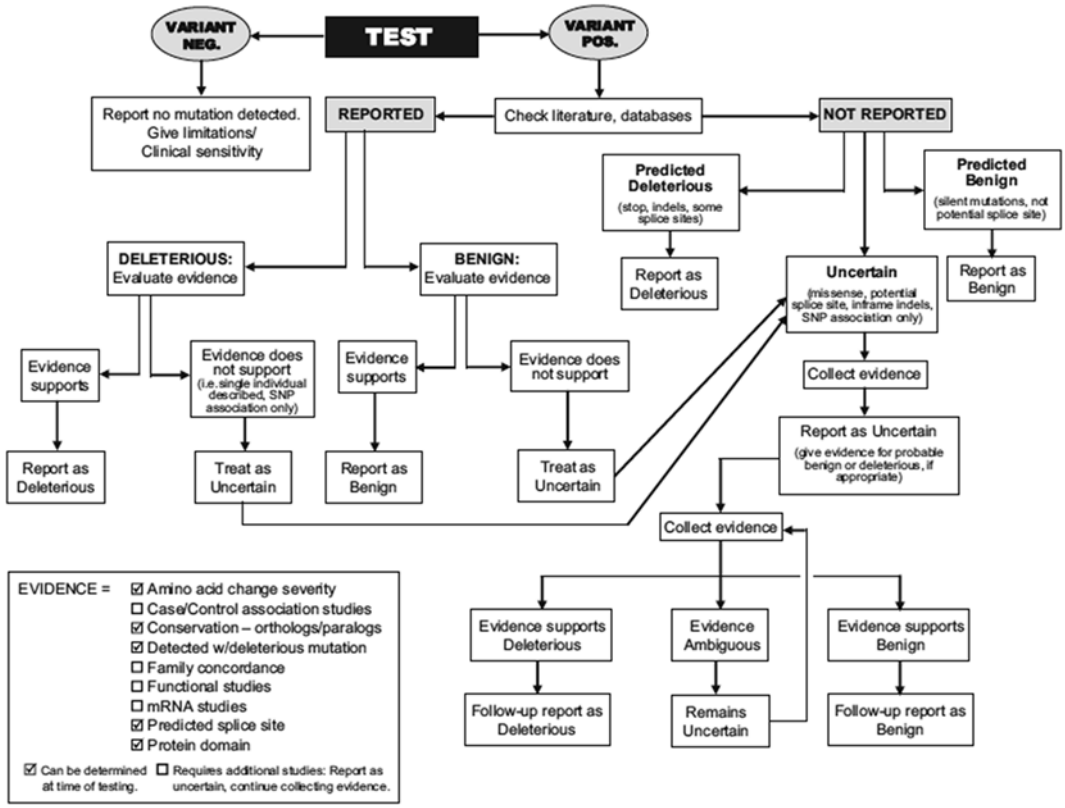


Fig. 12 ACMG flow chart on variant classification and reporting. Adapted from [45] with permission from Nature Publishing Group

counseling, potentially causing ongoing uncertainty for patients, and requiring follow-up studies. Diagnostic laboratories may be justifiably reluctant to allocate the variant into the *likely pathogenic* or the *expected to cause the disorder* categories without overwhelming supportive evidence, particularly if clinical stakes are high. In many instances, calculation of an exact posttest probability of disease for the variant may not be possible. These considerations are highlighted below.

3.8 Case Studies

Following are comparisons of three cases and the findings utilizing the methods described above (summarized in Table 4).

Case 1. CPOX variant (NM_000097.5:c.857C>A, NP_000088.3: p.Thr286Lys).
This variant was detected in two family members who had a biochemical diagnosis of hereditary coproporphyria. The nucleotide and amino acid sequences are highly conserved. In silico evidence using SNPs&GO, MutPred, and Align GVGD was inconsistent. FoldX did suggest reduced stability, although as mentioned above it is not recommended for prediction. The variant is listed on

Table 4
Approaches used to classify variants in the three case studies

	CPOX p.Thr286Lys	LDLR p.Glu101Lys	BRCA2 p.Asn289His
Nucleotide and amino acid conservation (PhyloP and phast Cons)	Conserved	Conserved	Not conserved
Protein domain	No association to functional domain	Calcium-binding site	No association to functional domain
SNPs&GO (probability of being disease causing)	0.696	0.965	0.966
MutPred (probability of being pathogenic)	0.555	0.937	0.163
Align GVGD	C0	C0–C55	C0
SIFT	Tolerated (score = 0.20)	Affect protein function (score = 0)	Tolerated (score = 0.12)
PolyPhen-2 (HumVar)	Possibly damaging (0.733)	Probably damaging (0.985)	Benign (0.075)
dbSNP entry	N/A	rs144172724	rs766173
Population frequency	No frequency information	No frequency information	Found in 5.8 % in population, up to 20 % in Han Chinese
Literature and functional studies	Reported in one patient with hereditary coproporphyria. No functional studies available	Reported in multiple populations with familial hypercholesterolemia. Functional studies showed 15–30 % of normal LDLR activity in homozygous state	Associated with decreased risk to breast cancer. No functional studies available
Variant classification in database	Not available	Disease-causing mutation (HGMD)	Disease-associated polymorphism (HGMD)
Classification	May be pathogenic	Pathogenic	Likely not pathogenic

HGMD database and had been reported once in another patient with hereditary coproporphyria [42]. No functional studies were available. Due to the nature of variable penetrance in porphyria and lack of functional studies, the variant would be classified as *may or may not be causative of the disorder.*

Case 2. LDLR variant (NM_000527.4:c.301G>A (NP_000518.1: p.Glu101Lys)).

This variant was found in a patient with clinical familial hypercholesterolemia. The nucleotide and amino acid sequences are highly conserved. The variant is within a calcium-binding site on NCBI CDD. It is reported in the locus-specific variation database and had been described in various populations [36]. The variant is also known as FH Lancashire or E80K (using a different transcript as reference). Align GVGD prediction ranged from C0 to C55 depending on the input of multiple protein sequence alignments. Manual curation of the sequence alignment (i.e., removing predicted/hypothetical protein or unrelated protein sequences) would see the classification changing from C0 to C55. This highlights the importance of carefully selected alignment. SIFT predicted the variant to be not tolerated. SNPs&GO and MutPred predicted a high probability for the variant to be disease causing. Functional studies showed that LDLR activity was 15–30 % of normal in a homozygous individual [43]. The above evidence was considered to be sufficient to classify this variant as *pathogenic*.

Case 3. BRCA2 variant (NM_000059.3:c.865A>C (NP000050.2: p.Asn289His)).
This variant was detected in a Chinese patient referred for familial breast cancer testing. The variant is listed as a validated SNP (rs766173) on dbSNP with a minor allele frequency of 5.8 % in 1000 Genomes Project. However, the population frequency of this variant is up to 20 % in Han Chinese. It is a polymorphic SNP where a different nucleotide change (c.865A>G, p.Asn289Asp) is also found. The variant is reported in the HGMD database, and there is one article linked to the variant, reporting the variant to be associated with *decreased* risk of breast cancer [44]. In silico studies showed that the variant is not conserved based on PhyloP and phastCons, and it is not associated with any functional domain. Align GVGD using the built-in sequence alignment for BRCA2 indicated that the variant is not likely to affect function (class C0). MutPred predicted the variant to be benign with the probability of it being deleterious at 0.163. However, SNPs&GO called it a disease-causing variant with a high RI of 9 (probability 0.966). Given the high population frequency especially in Han Chinese, it is unlikely that the variant is pathogenic, at least in the Chinese population. Results from in silico studies are inconsistent and therefore inconclusive. There was no positive association of this variant to breast cancer at the moment of reporting. The variant was therefore classified as *likely not pathogenic*.

4 Notes

1. Alternatively, one can perform multiple sequence alignment by SIFT BLink as described previously. SIFT will perform PSI-BLAST, and the FASTA file of the multiple protein sequence alignment can be downloaded. Beware that it may contain

unrelated proteins/predicted proteins which may need to be removed manually.

2. If using FASTA sequence instead of Swiss-Prot code, SNPs&GO will *only* call the CPOX variant disease associated if one includes the two GO terms associated with CPOX (GO:0006779 and GO:0004109) while entering information in SNPs&GO. It will be called neutral if you do not!

References

1. den Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12
2. Burset M, Seledtsov IA, Solovyev VV (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Res 28:4364–4375
3. Chretien S, Dubart A, Beaupain D et al (1988) Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in tissue-specific or in housekeeping expression. Proc Natl Acad Sci U S A 85:6–10
4. Gouya L, Puy H, Robreau AM et al (2002) The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wild-type FECH. Nat Genet 30:27–28
5. Wu K, Hinson SR, Ohashi A et al (2005) Functional evaluation and cancer risk assessment of BRCA2 unclassified variants. Cancer Res 65:417–426
6. Clegg JB, Weatherall DJ (1974) Hemoglobin constant spring, and unusual alpha-chain variant involved in the etiology of hemoglobin H disease. Ann N Y Acad Sci 232:168–178
7. Jensen HK, Jensen TG, Faergeman O et al (1997) Two mutations in the same low-density lipoprotein receptor allele act in synergy to reduce receptor function in heterozygous familial hypercholesterolemia. Hum Mutat 9: 437–444
8. Kashima T, Manley JL (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. Nat Genet 34:460–463
9. Punta M, Coggill PC, Eberhardt RY et al (2012) The Pfam protein families database. Nucleic Acids Res 40:D290–D301
10. Schultz J, Milpetz F, Bork P et al (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95:5857–5864
11. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40:D302–D305
12. Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32:W327–W331
13. Marchler-Bauer A, Anderson JB, Chitsaz F et al (2009) CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res 37:D205–D210
14. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–D229
15. Tavtigian SV, Greenblatt MS, Lesueur F et al (2008) In silico analysis of missense substitutions using sequence-alignment based methods. Hum Mutat 29:1327–1336
16. Pollard KS, Hubisz MJ, Rosenbloom KR et al (2010) Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 20:110–121
17. Siepel A, Bejerano G, Pedersen JS et al (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15:1034–1050
18. Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11:863–874
19. Ng PC, Henikoff S (2002) Accounting for human polymorphisms predicted to affect protein function. Genome Res 12:436–446
20. Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812–3814
21. Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 7:61–80
22. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081
23. Mathe E, Olivier M, Kato S et al (2006) Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. Nucleic Acids Res 34:1317–1325

24. Grantham R (1974) Amino acid difference formula to help explain protein evolution. Science 185:862–864
25. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302: 205–217
26. De Baets G, Van Durme J, Reumers J et al (2012) SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. Nucleic Acids Res 40: D935–D939
27. Li B, Krishnan VG, Mort ME, Xin F et al (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25:2744–2750
28. Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249
29. Calabrese R, Capriotti E, Fariselli P et al (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 30: 1237–1244
30. Capriotti E, Altman RB (2011) Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinformatics 12(Suppl 4):S3
31. Capriotti E, Calabrese R, Casadio R (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 22: 2729–2734
32. Kaminker JS, Zhang Y, Waugh A et al (2007) Distinguishing cancer-associated missense mutations from common polymorphisms. Cancer Res 67:465–473
33. Williams S (2012) Analysis of in silico tools for evaluating missense variants. http://www.ngrl.org.uk/Manchester/sites/default/files/publications/Add-To-Menu/Missense_Prediction_Tool_Report.pdf. Accessed 13 June 2013
34. Thusberg J, Olatubosun A, Vihinen M (2011) Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat 32:358–368
35. Kitts A, Sherry S (2002) The single nucleotide polymorphism database (dbSNP) of nucleotide sequence variation. In: The NCBI handbook [Internet]. http://www.ncbi.nlm.nih.gov/books/NBK21088/. Accessed 10 June 2013
36. Leigh SE, Foster AH, Whittall RA et al (2008) Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. Ann Hum Genet 72:485–498
37. Bell J, Bodmer D, Sistermans E et al (2007) Practice guidelines for the Interpretation and reporting of unclassified variants (UVs) in clinical molecular genetics. http://www.cmgs.org/BPGs/pdfs%20current%20bpgs/UV%20GUIDELINES%20ratified.pdf. Accessed 16 June 2013
38. Richards CS, Bale S, Bellissimo DB et al (2008) ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. Genet Med 10:294–300
39. Cassa CA, Tong MY, Jordan DM (2013) Large numbers of genetic variants considered to be pathogenic are common in asymptomatic individuals. Hum Mutat 34: 1216–1220
40. Spurdle AB, Healey S, Devereau A et al (2012) ENIGMA-evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. Hum Mutat 33:2–7
41. Plon SE, Eccles DM, Easton D et al (2008) Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 29:1282–1291
42. Whatley SD, Mason NG, Woolf JR et al (2009) Diagnostic strategies for autosomal dominant acute porphyrias: retrospective analysis of 467 unrelated patients referred for mutational analysis of the HMBS, CPOX, or PPOX gene. Clin Chem 55:1406–1414
43. Hobbs HH, Brown MS, Goldstein JL (1992) Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Hum Mutat 1:445–466
44. Bhatti P, Struewing JP, Alexander BH et al (2008) Polymorphisms in DNA repair genes, ionizing radiation exposure and risk of breast cancer in U.S. radiologic technologists. Int J Cancer 122:177–182
45. Richards CS, Bale S, Bellissimo DB et al (2008) ACMG recommendations for standards for interpretation and reporting of sequence variations. Genet Med 10:294–300

CLINICAL GENETICS An International Journal of Genetics, Molecular and Personalized Medicine

Clin Genet 2014: 86: 229–237
Printed in Singapore. All rights reserved

CLINICAL GENETICS
doi: 10.1111/cge.12315

Original Article

A comprehensive laboratory-based program for classification of variants of uncertain significance in hereditary cancer genes

Eggington J.M., Bowles K.R., Moyes K., Manley S., Esterling L., Sizemore S., Rosenthal E., Theisen A., Saam J., Arnell C., Pruss D., Bennett J., Burbidge L.A., Roa B., Wenstrup R.J. A comprehensive laboratory-based program for classification of variants of uncertain significance in hereditary cancer genes.
Clin Genet 2014: 86: 229–237. © 2013 The Authors. *Clinical Genetics* published by John Wiley & Sons A/S. Published by John Wiley & Sons Ltd., 2013

Genetic testing has the potential to guide the prevention and treatment of disease in a variety of settings, and recent technical advances have greatly increased our ability to acquire large amounts of genetic data. The interpretation of this data remains challenging, as the clinical significance of genetic variation detected in the laboratory is not always clear. Although regulatory agencies and professional societies provide some guidance regarding the classification, reporting, and long-term follow-up of variants, few protocols for the implementation of these guidelines have been described. Because the primary aim of clinical testing is to provide results to inform medical management, a variant classification program that offers timely, accurate, confident and cost-effective interpretation of variants should be an integral component of the laboratory process. Here we describe the components of our laboratory's current variant classification program (VCP), based on 20 years of experience and over one million samples tested, using the *BRCA1/2* genes as a model. Our VCP has lowered the percentage of tests in which one or more *BRCA1/2* variants of uncertain significance (VUSs) are detected to 2.1% in the absence of a pathogenic mutation, demonstrating how the coordinated application of resources toward classification and reclassification significantly impacts the clinical utility of testing.

Conflict of interest

All authors are employees of Myriad Genetics, Inc. and Myriad Genetic Laboratories, Inc. and receive salaries and stock options as compensation.

J.M. Eggington[a], K.R. Bowles[a], K. Moyes[a], S. Manley[a], L. Esterling[a], S. Sizemore[a], E. Rosenthal[a], A. Theisen[a], J. Saam[a], C. Arnell[a], D. Pruss[b], J. Bennett[a], L.A. Burbidge[a], B. Roa[a] and R.J. Wenstrup[a]

[a]Myriad Genetic Laboratories, Inc., Salt Lake City, UT, USA and [b]Myriad Genetics, Inc., Salt Lake City, UT, USA

Key words: *BRCA1* – *BRCA2* – co-segregation – HBOC – hereditary breast and ovarian cancer – variant classification – variants of uncertain significance – VUS

Corresponding author: Karla Bowles, PhD, Myriad Genetic Laboratories, Inc., 320 Wakara Way, Salt Lake City, UT 84108, USA.
Tel.: +1 801 584 3062;
fax: +1 801 584 3615;
e-mail: kbowles@myriad.com

Received 13 August 2013, revised and accepted for publication 5 November 2013

Sequencing and large rearrangement analyses detect DNA changes within hereditary cancer genes and are offered to individuals with a personal and/or family history of cancer to identify pathogenic mutation carriers. Early identification of mutation carriers allows for increased clinical surveillance and early detection, and may prompt more aggressive prevention strategies, such as prophylactic surgery or chemoprevention, to reduce risk. For example, patients with pathogenic mutations in the genes *BRCA1* and *BRCA2* have a diagnosis of hereditary breast and ovarian cancer syndrome (HBOC), a condition for which there are extensive medical management guidelines aimed at the prevention and early detection of breast and ovarian cancer (1).

Once a genetic variant is detected in the laboratory, its clinical significance must be determined. Guidelines for the classification of variants have been proposed by the American College of Medical Genetics (2), the

The copyright line for this article was changed on January 23, 2015 after original online publication.

International Agency for Research on Cancer (IARC) Unclassified Genetic Variants Working Group (3) and other researchers (4). These guidelines recommend a multi-tier classification system, grouping variants based upon perceived risk of disease association. Our laboratory has developed and currently utilizes a similar five-tier variant classification system composed of the following variant classification categories: 'deleterious' (pathogenic), 'suspected deleterious' (likely pathogenic), 'variant of uncertain clinical significance' (VUS), 'genetic variant, favor polymorphism' (likely not pathogenic), and 'polymorphism' (not pathogenic).

In a small proportion of patients, genetic testing will identify a VUS, which confounds the clinical interpretation of the result. VUSs consist primarily of missense substitutions that result in single amino acid changes, but also include variants that have the potential to alter RNA splicing (5) and other changes that have the potential to alter the production of fully functional protein (2).

VUSs present a diagnostic challenge to the clinician. Similar to other non-informative results – for example, a 'no mutation detected' result in an individual with no family history of a specific mutation – clinical management of individuals carrying a VUS should be based upon personal and family history and not the presence or absence of the variant itself (6). However, non-informative results including VUSs often increase anxiety among patients, family members, and providers who cannot take advantage of the risk assessment, prevention, and therapeutic measures that are available to carriers of known deleterious mutations to modify behavior or lifestyle, or to make important clinical decisions that may, in many cases, involve prophylactic surgery (7). In addition, all first-degree relatives including non-carriers are considered at risk as long as the contribution of the variant to disease cannot be assessed, resulting in frequent unnecessary anxiety and prophylactic screening. However, unlike other non-informative results, the presence of a VUS may provoke anxiety that testing is not complete until the pathogenicity of the variant is determined (8).

The overall interpretation of VUS is currently reported in 2.1% of patients undergoing genetic analysis for HBOC at Myriad Genetic Laboratories (9, 10). This represents a decline from around 13% over the past decade. The dramatic decline in the percentage of patients receiving a VUS result reflects both the impact of targeted efforts directed at determining the pathogenicity of variants, as well as the availability of data from an increased number of individuals undergoing testing for HBOC (10).

The primary aim of clinical genetic testing is to provide results that inform medical management, so it is vital that diagnostic laboratories have in place a robust variant classification program that offers timely, accurate, confident and cost-effective interpretation of variants as an integral part of their testing services. Here we describe our current laboratory-based variant classification program which integrates multiple sources of both passively and proactively ascertained data in a coordinated fashion for use in a clinical setting. We use *BRCA1/2* here as a model, but similar techniques for variant classification can be applied to other genes.

Materials and methods

Novel variant interpretation

Myriad's New Mutations Committee (NMC) consists of American Board of Medical Genetics (ABMG)-certified laboratory directors, the chief medical officer, clinical variant specialists, genetic counselors, and other representatives with expertise in clinical care, statistical genetics, biochemistry and structural biology. The NMC is responsible for the initial classification of new variants on a daily basis. The initial classification of new variants follows the guidelines set forth by the American College of Medical Genetics (2) in addition to a set of internal guidelines developed based on NMC

Table 1. Myriad Genetic Laboratories' current categorization of variant reclassification methods[a]

	Initial variant classification	Variant reclassification	Variant reclassification			
			Primary upgrade	Secondary upgrade	Primary downgrade	Secondary downgrade
ACMG guidelines	X	–	–	–	–	–
in trans[b]	–	X	–	–	X	X
Mutation co-occurrence	–	X	–	–	X	X
Segregation	X[c]	X	X	X	X	X
History-weighting algorithm	–	X	X	X	X	X
Evolutionary conservation	–	X	–	–	–	X
Functional or mRNA splice-site assays	X[c]	X	X	X	X	X
Population frequency	X[c]	X	–	–	X	X

ACMG, American College of Medical Genetics and Genomics.

[a]*Note*: this table represents a framework and does not replace the expert review process that is needed before implementing any particular methodology. See the main text for clarification.

[b]*in trans*: identification of homozygous and compound heterozygous individuals.

[c]Infrequently these data are available in the literature or through publicly accessible population frequency databases at the time of a variant's initial classification.

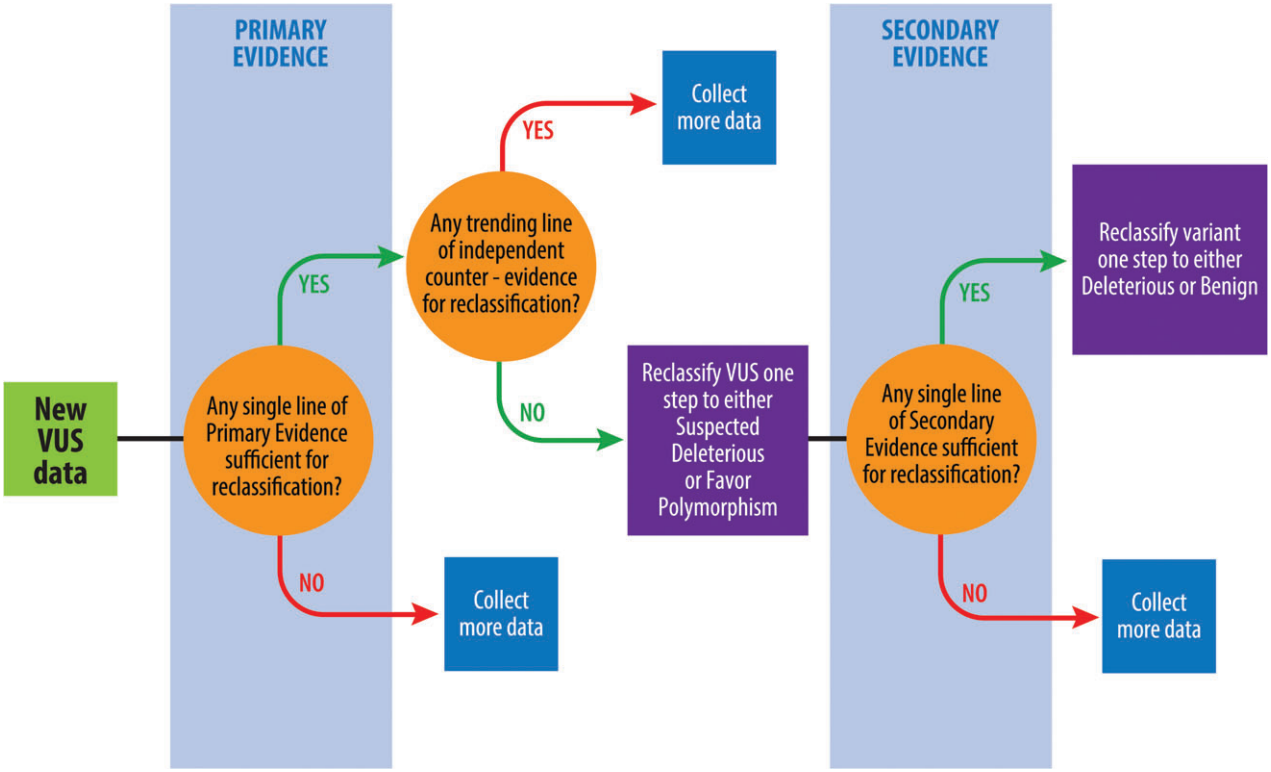


Fig. 1. Conceptual flowchart for *BRCA1/2* VUS reclassification. See Table 1 for primary and secondary reclassification method types. Methods used are determined by the chronology of which data becomes significant.

expertise and experience. The NMC also reclassifies previously identified variants as new data become available.

Variant reclassification

The monitoring of variant data is embedded in our laboratory operations. Automation protocols alert the NMC when there is sufficient statistical evidence to consider reclassification of a variant. The literature and public databases are also continuously monitored to determine whether additional data have been made available that would inform the reclassification of a variant. The NMC meets to discuss reclassification data, and variants may be reclassified following a thorough evaluation of all relevant data. Table 1 outlines the different reclassification methods that are currently employed in our laboratory. Amended reports are sent weekly to healthcare providers who have patients for whom a variant reclassification affects their report.

Our laboratory VCP utilizes multiple lines of evidence, described below and in Table 1, for reclassification of a variant's disease status. Statistical methods must reach an acceptable level of certainty before they can be used. This level of certainty currently exceeds 99% positive and 99% negative predictive values in our laboratory. Semi-quantitative and qualitative methods are thoroughly evaluated by experts in the relevant field prior to use. Independent methods are used only once to obtain a single step in reclassification (e.g. from 'VUS' to 'suspected deleterious'). Figure 1 represents a simplified flow chart for the use of primary and secondary evidence in VUS reclassification (see also Table 1). Primary evidence is data that can be used by itself to upgrade or downgrade a variant by one classification step (e.g. from 'VUS' to 'variant favor polymorphism') as long as no significant contradictory evidence exists. Secondary evidence is data that can be used in conjunction with primary evidence for a full two-step upgrade or downgrade of a variant. Additional supporting evidence should be considered when variants are reclassified, but it is not considered strong enough to alter a classification. Unless otherwise specified, the methods described below can be used for primary lines of evidence for a variant reclassification. The reclassification process for any particular variant is initiated when new evidence is either generated in the production laboratory or made available in the public literature. All lines of evidence available at that time are then considered, whether they are independently sufficient for a reclassification (primary evidence), are supportive (secondary evidence) or are contradictory.

Literature review

Our laboratory employs scientists in a variety of fields who continually evaluate the literature to determine if there is sufficient evidence to reclassify a variant. In some instances, information from the literature is sufficient to be used on its own as primary evidence

for a one- or two-step upgrade or downgrade of a variant. If the literature were to be used for a full two-step upgrade or downgrade of a variant, two separate methodologies with independent and significant findings would be required. Some of the reclassification methods described below can be found in the published literature.

Population frequency

Databases containing whole-exome sequencing data of control populations have recently become publicly available (11, 12). Our laboratory classifies variants as benign polymorphisms if they are present in $>2\%$ of a control population with a sample size >200 individuals without significant evidence to the contrary. These populations primarily consist of families in which hereditary cancer-predisposing syndromes are not indicated. Our laboratory also uses a comparative approach to statistically evaluate the affected population tested at our laboratory against control populations. A variant present in statistically equal frequencies in the two populations is considered benign. A variant enriched in the affected population is evaluated further for potential causality but is not reclassified on these data alone.

mRNA splice-site assays

The general mechanisms of RNA splicing are well understood (13), so novel genetic variants occurring at canonical splice acceptor and donor sites can be classified at their first observation based upon this knowledge. However, other variants not immediately at the splice-site junction may also impair RNA splicing. Biochemical analysis of potential mRNA splicing variants can provide evidence for variant reclassification. Analysis of patient mRNA or a minigene assay demonstrating a particular variant results in abnormal mRNA splicing may provide evidence that the variant is deleterious (14, 15). In our laboratory, evidence from splice-site assays in the literature is primarily used, after expert review, to upgrade a variant provided the assay clearly shows complete loss of the functional mRNA isoform(s) transcribed from the variant allele. Because of this requirement, the application of this method is limited.

Functional assays

Because cancer predisposition in HBOC results from the inheritance of alterations that result in loss of function of tumor suppressor genes, *in vitro* detection of a decrease in activity of a tumor suppressor may correspond to increased cancer predisposition. Functional assays assess the effects of *BRCA1/2* missense variants on known protein function. Existing assays include, but are not limited to, those designed to measure variant effects on centrosome number control (16), homologous recombination (17), transcription, protein-folding, and phosphopeptide binding (18). These techniques include homologous recombination and centrosome amplification assays (19). In addition, analysis of the solved portion of the BRCA2 crystal structure can sometimes be utilized to determine the effect that a variant may have on BRCA2 function (20). Because functional assays have only been tested on a limited number of variants, our laboratory may use published data from these assays as supporting evidence in conjunction with primary data for the upgrade or downgrade of a variant.

Evolutionary conservation

Evaluation of species conservation may provide supportive evidence for variant reclassification. Phylogenetic conservation of protein sequence throughout evolution often reflects the requirement for certain amino acids for protein activity. Multiple computational algorithms, including Sorting Intolerant from Tolerant (SIFT), Polymorphism Phenotyping (Poly Phen) and Align-GVGD (21–24), have been designed to evaluate the evolutionary/functional significance of an amino acid change through analysis of multiple species protein alignments. However, because of high false positive and false negative rates (4), our laboratory does not currently use these algorithms to reclassify a variant to a 'deleterious' or 'suspected deleterious' category. Identification of a particular amino acid change in multiple species does provide supportive evidence that a variant is benign; therefore, in our laboratory, we currently use conservation analysis as secondary evidence to support a downgrade if the exact missense change seen in a patient sample exists in multiple reference sequences of higher species (primarily vertebrates). Protein sequences for different species are added to the alignments as they become publicly available. No fewer than 15 species are used. It is important to note that these methods are subject to the quality of the species alignment used and the context of the particular missense mutation in question. Regions of poorly aligned sequences are considered insufficient for this analysis. Once a missense variant has achieved a classification of favor polymorphism (FP, that is, a variant with one significant line of evidence in favor of benign) based upon an independent line of evidence, if that particular missense variant is seen repeatedly in other species in well-aligned and conserved regions of the protein, the variant may be downgraded to 'polymorphism'.

Segregation analysis

Segregation analysis measures whether or not a variant segregates with cancer in one or more families. It has traditionally relied upon obtaining one or more large pedigrees with multiple affected family members available for analysis. Myriad uses a modified approach to segregation analysis that allows for analysis of small families, similar to that described by Thompson, Easton, and Goldgar (25). A 500:1 likelihood ratio for deleterious or benign is considered sufficient for a reclassification of a VUS to suspected deleterious or favor

polymorphism. Statistical data obtained from a series of small families sharing the same variant are combined to assess the clinical significance of the specific variant. Almost without exception, this approach requires active participation of multiple families before a variant can be reclassified.

Variant test results from our laboratory are frequently accompanied by an offer of no-cost testing after evaluation of each proband's pedigree. Our protocol directs that testing offers be typically made to the most informative individuals in the pedigree, such as older unaffected females and younger affected women who are 1° or 2° relatives. Results for multiple families with the same variant are combined. Owing to the relative high phenocopy rates for HBOC, this method's primary limitation is the large amount of data required to statistically overcome the phenocopy observations.

Identification of homozygous and compound heterozygous individuals (*in trans*)

Given the severe phenotypes associated with homozygosity for a *BRCA1* or *BRCA2* deleterious mutation (26–29), observation of a homozygous variant or a variant *in trans* with a deleterious mutation (i.e. compound heterozygosity) in a healthy individual or an individual with later-onset cancer provides significant evidence that the variant itself does not represent a deleterious mutation. For example, with few exceptions (30), biallelic *BRCA1* mutations are embryonic lethal; therefore, if a known pathogenic mutation is present, a VUS is highly unlikely to be deleterious. Conversely, biallelic *BRCA2* mutations result in Fanconi anemia, an autosomal recessive syndrome characterized by congenital anomalies, bone marrow failure, cellular sensitivity to DNA cross-linking agents, and predisposition to cancer. Therefore, the non-pathogenicity of a novel *BRCA2* variant co-occurring *in trans* with a known pathogenic *BRCA2* mutation can be assessed by the absence of features of the Fanconi anemia phenotype. A limitation of this method is that attenuated biallelic disease states can exist (30). For example, in rare cases of *BRCA2* biallelic states, Fanconi anemia is not obvious in the patient. To address this limitation, where attenuated forms of Fanconi anemia may exist, Myriad facilitates the Chromosome Breakage Analysis test, at no-cost to the patient, in order to obtain a definitive diagnosis.

Mutation co-occurrence

Mutation co-occurrence (MCO), similar to 'ascertainment ratio', is a statistical technique based on the empirical observation that if a pathogenic mutation is identified in a family, that mutation is usually found to be the primary cause of disease in the family (31). Therefore, the presence of a known pathogenic mutation in a biochemical pathway reduces the likelihood that a VUS in the same pathway is clinically relevant (24, 31). The known deleterious mutation can either be in the same gene (*in cis* or *in trans*) or in other genes in the same pathway (e.g. a *BRCA1* mutation and a *BRCA2* VUS). One limitation of this method is that the development of MCO for a particular gene and pathway requires large sets of empirical data to account for the clinical consequences of carrying two pathogenic mutations of the gene(s) in question and to account for ascertainment bias. To ensure that each reclassification methodology used for a particular variant is independent of any other reclassification method already used for a variant, care must be taken in the use of MCO such that compound heterozygous observations which have been used previously to reclassify a variant by one step are not used again for MCO to obtain a second reclassification downgrade.

History-weighting algorithm

Our laboratory's history-weighting algorithm is based on the premise that individuals with deleterious mutations are expected to have more severe personal and family histories than individuals with benign polymorphisms (32). The technique is an advance, facilitated by the quantity of data analyzed since its publication, on the method described by Easton et al. (33) in which the probability of deleterious mutation is calculated for each proband based upon their personal and family history. The combined history-weighting scores for unrelated probands carrying a particular VUS are compared against the observed clinical population tested by the laboratory which accounts for patient ascertainment bias.

The history-weighting score is validated for both single steps in upgrades and downgrades of a classification. One limitation of this method is that vanishingly rare variants cannot be reclassified using this method. However, owing to the large volume of testing at our laboratory, this method accounts for the majority of variant reclassifications (Fig. 4, Table 2).

Table 2. Relative effectiveness of reclassification methods by proband count for HBOC (October 2011–November 2012)

Reclassification method	Total reclassification events	Average number of probands per downgrade (minimum[a])	Average number of probands per upgrade (minimum[a])
in trans	84 (16%)	17.5 (1)	NA
Mutation co-occurrence	24 (4.7%)	31.2 (2)	NA
Segregation	18 (3.5%)	41.8 (9)	32 (13)
History-weighting algorithm	290 (56%)	19.4 (6)	33.7 (28)
Evolutionary conservation	90 (18%)	15.4 (NA)	NA
Functional or mRNA splice-site assays	6 (1.0%)	7 (NA)	12.6 (NA)
Population frequency	2 (0.4%)	8.5 (NA)	NA
Total	514	–	–

HBOC, hereditary breast and ovarian cancer syndrome; NA, not applicable.

[a]The minimum numbers are representative of the empirical data during the specified time period and are not necessarily theoretical minimums.

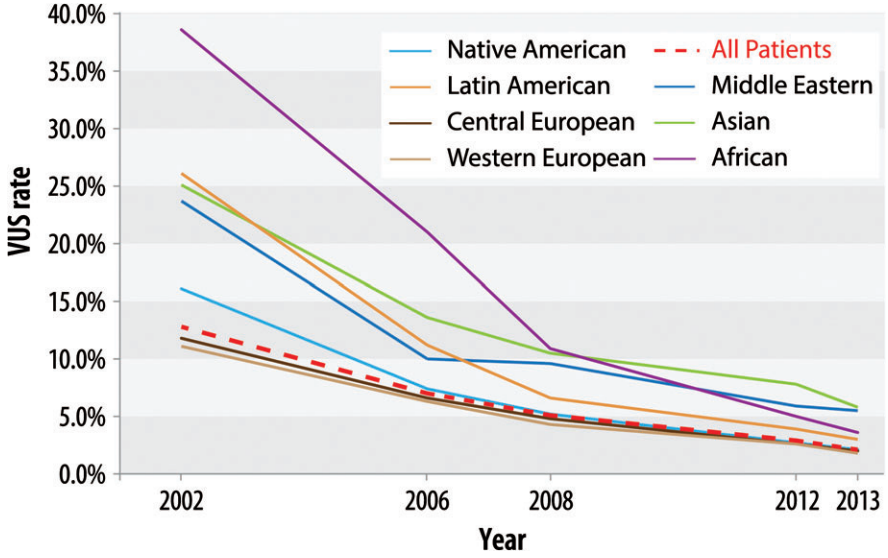


Fig. 2. Decline in rate of HBOC variants of uncertain significance. Myriad's data was analyzed periodically from 2002 to 2013 to establish the percentage of tests reported with an overall interpretation of VUS and subdivided by ancestry. Inclusion criteria consisted of individuals who were referred for clinical genetic testing of *BRCA1/2*. Patients for which no ancestry was selected or for which multiple ancestries were selected were excluded from the ancestry breakdown but were included in the total count.

Results

Overall outcomes of variant reclassification process

Myriad's mutation data set was analyzed periodically from 2002 to 2013 to establish the percentage of overall tests reported with an interpretation of VUS and subdivided by ancestry (Fig. 2). Inclusion criteria consisted of individuals who were referred for clinical genetic testing of *BRCA1/2*. Patients for whom no ancestry was selected or for whom multiple ancestries were selected were excluded from the ancestry breakdown but were included in the total count.

The VUS rate is defined as the percentage of *BRCA1/2* patients comprehensively tested in the entire Myriad test history that have an overall test report of VUS at the time point specified. From 2002 to 2013, the VUS rate declined from 12.8% of all *BRCA1/2* test results to 2.1% of all results (84% decline). During this time period there was also a decline in the VUS rate across all ancestries. It is important to note that a patient who has a suspected deleterious or deleterious mutation accompanying a VUS will have an overall test result of suspected deleterious or deleterious, respectively. A patient who has a favor polymorphism or polymorphism (benign) in addition to a VUS will have an overall test result of VUS. More than one million *BRCA1/2* test reports were evaluated to attain the 2013 VUS rate.

VCP segregation analysis/family testing uptake

Between October 2011 and August 2012, Myriad recorded family history submissions from 16.8% of the VUS/FP results (Fig. 3). Family testing was offered to an average of 2.3 relatives per family history submission, with a 24.1% response rate.

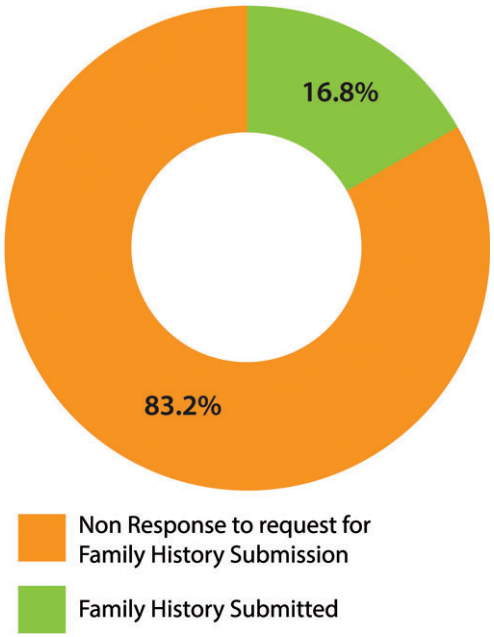


Fig. 3. Overall participation rate in submission of family histories for Myriad's variant classification program segregation analysis for HBOC (October 2011–August 2012).

Relative contribution of variant reclassification techniques

During October 2011 to November 2012, segregation analysis resulted in only 3.5% of reclassification events (Table 2). The history-weighting algorithm and MCO achieved 61% of all reclassification events. Both techniques are powered by our laboratory's data set and do not require any additional follow-up family testing beyond the proband's results. Of reclassification events, 16% were a result of *in trans* observations,

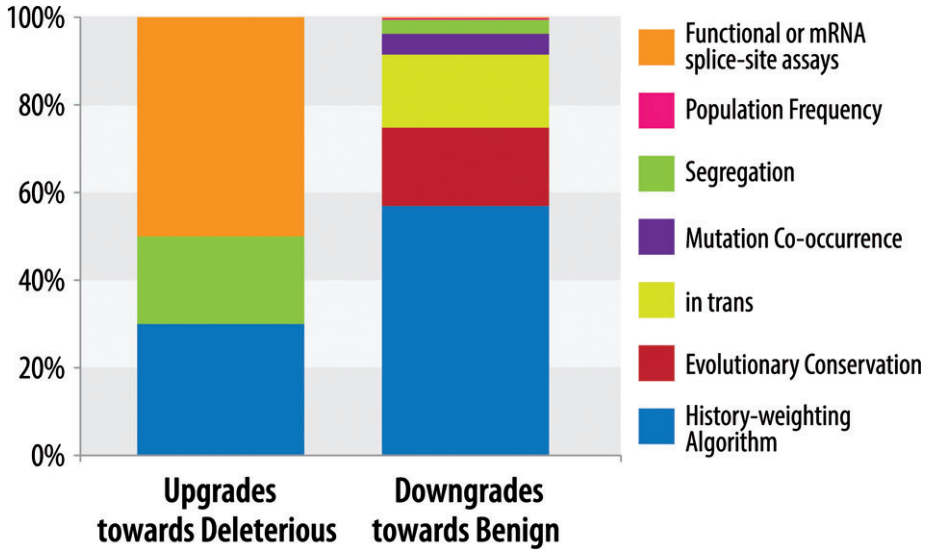


Fig. 4. Percentage of reclassification upgrades and downgrades achieved by Myriad's variant reclassification methods for HBOC (October 2011–November 2012).

which occasionally require follow-up testing of family members to determine phase. Of reclassifications, 19% were achieved through methods using publicly available data: evolutionary conservation evaluation (18%), functional and mRNA splice-site assays published in literature (1%) and population frequency data (<1%).

Historically, most VUSs are determined to be benign (33), representing not only a true bias, but also the greater number of methods to discover benign variants compared to deleterious mutations (Table 2 and Fig. 4). Of the reclassification events that were upgrades to deleterious or suspected deleterious in the 13-month time period reported here, 20% were based on segregation data, 30% were based on the history-weighting algorithm, and 50% were based upon functional or mRNA splice-site assays reported in the literature (Fig. 4). The history-weighting algorithm is dependent on the total number of probands carrying a particular variant. For segregation analysis, the likelihood of obtaining sufficient participating families to achieve a reclassification increases as the number of probands for a variant increases. Of the upgrades to 'suspected deleterious' or 'deleterious', the history-weighting algorithm and segregation analysis had similar averages of the number of proband carriers at time of reclassification (Table 2).

Discussion

On the basis of analysis of our dataset of more than one million patients, our laboratory has developed an extensive variant classification program which utilizes multiple variant classification techniques; new classification methods are continually being assessed by our scientists. Although variations of some of these techniques have been utilized by geneticists for many years, others have only become feasible within recent years following the accumulation of the large data set required for their development and use, and they are unique to our variant classification process. The program is particularly effective for autosomal dominant genes such as *BRCA1/2*. The techniques described here would need to be altered to account for sex-linked or autosomal recessive genes.

From 2002 to 2013, the VUS rate declined from 12.8% of all *BRCA1/2* test results to 2.1% of all results (84% decline, Fig. 2). The substantial decline in the VUS rate in these ancestries is a result of improved methods for establishing the clinical significance of variants and increased utilization of testing in these populations, which provides more data for analysis.

Uncertainty remains about the clinical relevance of VUS. Although our laboratory recommends that clinical management of VUS carriers should be based upon personal and family history and not the presence or absence of the variant itself, some healthcare providers increase surveillance or pursue treatment options beyond that recommended for such variants (34). However, our results show the majority of *BRCA1/2* VUSs are discovered to be benign through a variety of methods, with history-weighting analysis the most robust method. Segregation analysis shows particular power in identifying deleterious variants rather than benign variants. As shown by the minimums in Table 2, segregation analysis is able to achieve a reclassification with fewer probands per variant compared with the history-weighting algorithm. Although segregation analysis is labor intensive, its success rate is limited; only 3.5% of variant reclassifications are achieved using this technique. Such limitations emphasize the utility of a variant classification program that can weigh other forms of evidence in addition to segregation data. Considering laboratories and community research centers have finite resources, these data therefore suggest that the tailoring of family analysis to specific families with higher likelihoods of having a deleterious mutation may be the most productive use of resources. These data show

that while segregation analysis is a comparatively poor tool for discovering benign variants, it is a powerful tool for discovering deleterious mutations (Fig. 4).

The decrease in our laboratory's VUS rate over time is a function of increased data sets and increased expertise in evaluating variants. Our laboratory has classified thousands of VUS and continues to identify dozens of novel VUS every week (9). Thus, the importance of experience and technical acumen of the laboratory classifying the variants cannot be overstated. Because of both the need for timely test interpretation and the pace at which new variants are identified and existing variants reclassified, we believe that variant classification is an integral part of the testing process to be performed by a CLIA-approved (for USA laboratories), quality-assured laboratory as standard operating procedure.

New technologies, with their attendant increases in tests ordered, and, thus, variants identified, will challenge variant classification programs – for example, in reporting VUSs and sending amended reports when variants are reclassified. The amount of *BRCA1/2* testing is likely to increase if therapies specifically targeting *BRCA1/2* mutated tumors, such as poly-(ADP-ribose) polymerase (PARP) inhibitors, move out of clinical trials and into practice. In addition, the advent of next-generation massively parallel sequencing will necessitate reliable, high-throughput variant classification programs. As costs decrease for whole-genome sequencing experiments, it is likely that a larger number of individuals show genetic variation beyond the exons and intron/exon boundaries routinely covered by current clinical tests (35). One value of our model-based approach is that it provides a quantitative output that can be used to categorize variants into defined classification categories and so minimizes subjectivity and decreases turnaround time (3).

On the basis of 20 years of experience, our laboratory has developed a robust program for classification of variants. Our variant classification program has lowered the percentage of tests in which one or more *BRCA1/2* variants of uncertain significance (VUSs) are detected to 2.1%, showing how the coordinated application of resources toward classification and reclassification significantly impacts the clinical utility of testing.

Acknowledgements

We acknowledge the efforts of the clinicians and patients who have participated in Myriad Genetic Laboratories' Variant Classification Program. We also thank the past and present directors, genetic counselors, scientists, statisticians, clinical data and variant specialists of Myriad Genetic Laboratories, Inc. who have participated in the reclassification of genetic variants. Research was funded by Myriad Genetic Laboratories, Inc.

References

1. Daly M. Genetic/familial high-risk assessment: breast and ovarian. NCCN Clinical Practice Guidelines in Oncology, National Comprehensive Cancer Network, 2013.
2. Richards CS, Bale S, Bellissimo DB et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. Genet Med 2008: 10: 294–300.
3. Plon SE, Eccles DM, Easton D et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 2008: 29: 1282–1291.
4. Lindor NM, Guidugli L, Wang X et al. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). Hum Mutat 2012: 33: 8–21.
5. Szabo CI, Worley T, Monteiro AN. Understanding germ-line mutations in BRCA1. Cancer Biol Ther 2004: 3: 515–520.
6. National Comprehensive Cancer Network. Hereditary breast and/or ovarian cancer syndrome. NCCN Guidelines, 2013.
7. Burke W, Daly M, Garber J et al. Recommendations for follow-up care of individuals with an inherited predisposition to cancer. II. BRCA1 and BRCA2. Cancer Genetics Studies Consortium. JAMA 1997: 277: 997–1003.
8. Culver J, Brinkerhoff C, Clague J et al. Variants of uncertain significance in BRCA testing: evaluation of surgical decisions, risk perception, and cancer distress. Clin Genet 2013: 84: 464–472.
9. Myriad Genetic Laboratories. Internal Data. 2011.
10. Eggington JM. Perplexing variants of uncertain significance. National Society of Genetic Counselors 2011 Annual Education Conference, San Diego, CA, 2011: 13.
11. Goh G, Choi M. Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. Genom Inform 2012: 10: 214–219.
12. Krumm N, Sudmant PH, Ko A et al. Copy number variation detection and genotyping from exome sequence data. Genome Res 2012: 22: 1525–1532.
13. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 1987: 15: 7155–7174.
14. Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. J Med Genet 2005: 42: 737–748.
15. Piton A, Jouan L, Rochefort D et al. Analysis of the effects of rare variants on splicing identifies alterations in GABA(A) receptor genes in autism spectrum disorder individuals. Eur J Hum Genet 2012: 21: 749–756.
16. Kais Z, Chiba N, Ishioka C et al. Functional differences among BRCA1 missense mutations in the control of centrosome duplication. Oncogene 2012: 31: 799–804.
17. Parvin J, Chiba N, Ransburgh D. Identifying the effects of BRCA1 mutations on homologous recombination using cells that express endogenous wild-type BRCA1. J Vis Exp 2011: 48: 2468.
18. Lee MS, Green R, Marsillac SM et al. Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays. Cancer Res 2010: 70: 4880–4890.
19. Farrugia DJ, Agarwal MK, Pankratz VS et al. Functional assays for classification of BRCA2 variants of uncertain significance. Cancer Res 2008: 68: 3523–3531.
20. Karchin R, Monteiro AN, Tavtigian SV et al. Functional impact of missense variants in BRCA1 predicted by supervised learning. PLoS Comput Biol 2007: 3: e26.
21. Adzhubei IA, Schmidt S, Peshkin L et al. A method and server for predicting damaging missense mutations. Nat Methods 2010: 7: 248–249.
22. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001: 11: 863–874.
23. Sunyaev S, Ramensky V, Koch I et al. Prediction of deleterious human alleles. Hum Mol Genet 2001: 10: 591–597.
24. Tavtigian SV, Byrnes GB, Goldgar DE et al. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. Hum Mutat 2008: 29: 1342–1354.
25. Thompson D, Easton DF, Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. Am J Hum Genet 2003: 73: 652–655.
26. Abkevich V, Zharkikh A, Deffenbaugh AM et al. Analysis of missense variation in human BRCA1 in the context of interspecific sequence variation. J Med Genet 2004: 41: 492–507.
27. Frank TS, Deffenbaugh AM, Reid JE et al. Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals. J Clin Oncol 2002: 20: 1480–1490.
28. Hirsch B, Shimamura A, Moreau L et al. Association of biallelic BRCA2/FANCD1 mutations with spontaneous chromosomal instability and solid tumors of childhood. Blood 2004: 103: 2554–2559.

29. Wagner JE, Tolar J, Levran O et al. Germline mutations in BRCA2: shared genetic susceptibility to breast cancer, early onset leukemia, and Fanconi anemia. Blood 2004: 103: 3226–3229.
30. Domchek SM, Tang J, Stopfer J et al. Biallelic deleterious BRCA1 mutations in a woman with early-onset ovarian cancer. Cancer Discov 2013: 3: 399–405.
31. Tavtigian SV, Samollow PB, de Silva D et al. An analysis of unclassified missense substitutions in human BRCA1. Fam Cancer 2006: 5: 77–88.
32. Bowles KR, Morris B, Hughes E et al. A clinical history weighting algorithm accurately classifies *BRCA1* and *BRCA2* variants. ACMG Annual Clinical Genetics Meeting, Phoenix, AZ, 2013.
33. Easton DF, Deffenbaugh AM, Pruss D et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet 2007: 81: 873–883.
34. Plon SE, Cooper HP, Parks B et al. Genetic testing and cancer risk management recommendations by physicians for at-risk relatives. Genet Med 2011: 13: 148–154.
35. Walsh T, Lee MK, Casadei S et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proc Natl Acad Sci U S A 2010: 107: 12629–12633.

Original Investigation

Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders

Hane Lee, PhD; Joshua L. Deignan, PhD; Naghmeh Dorrani, MS, CGC; Samuel P. Strom, PhD; Sibel Kantarci, PhD; Fabiola Quintero-Rivera, MD; Kingshuk Das, MD; Traci Toy, BS; Bret Harry, BS; Michael Yourshaw, PhD; Michelle Fox, MS, CGC; Brent L. Fogel, MD, PhD; Julian A. Martinez-Agosto, MD, PhD; Derek A. Wong, MD; Vivian Y. Chang, MD, MS; Perry B. Shieh, MD, PhD; Christina G. S. Palmer, PhD, CGC; Katrina M. Dipple, MD, PhD; Wayne W. Grody, MD, PhD; Eric Vilain, MD, PhD; Stanley F. Nelson, MD

IMPORTANCE Clinical exome sequencing (CES) is rapidly becoming a common molecular diagnostic test for individuals with rare genetic disorders.

OBJECTIVE To report on initial clinical indications for CES referrals and molecular diagnostic rates for different indications and for different test types.

DESIGN, SETTING, AND PARTICIPANTS Clinical exome sequencing was performed on 814 consecutive patients with undiagnosed, suspected genetic conditions at the University of California, Los Angeles, Clinical Genomics Center between January 2012 and August 2014. Clinical exome sequencing was conducted as trio-CES (both parents and their affected child sequenced simultaneously) to effectively detect de novo and compound heterozygous variants or as proband-CES (only the affected individual sequenced) when parental samples were not available.

MAIN OUTCOMES AND MEASURES Clinical indications for CES requests, molecular diagnostic rates of CES overall and for phenotypic subgroups, and differences in molecular diagnostic rates between trio-CES and proband-CES.

RESULTS Of the 814 cases, the overall molecular diagnosis rate was 26% (213 of 814; 95% CI, 23%-29%). The molecular diagnosis rate for trio-CES was 31% (127 of 410 cases; 95% CI, 27%-36%) and 22% (74 of 338 cases; 95% CI, 18%-27%) for proband-CES. In cases of developmental delay in children (<5 years, n = 138), the molecular diagnosis rate was 41% (45 of 109; 95% CI, 32%-51%) for trio-CES cases and 9% (2 of 23, 95% CI, 1%-28%) for proband-CES cases. The significantly higher diagnostic yield (*P* value = .002; odds ratio, 7.4 [95% CI, 1.6-33.1]) of trio-CES was due to the identification of de novo and compound heterozygous variants.

CONCLUSIONS AND RELEVANCE In this sample of patients with undiagnosed, suspected genetic conditions, trio-CES was associated with higher molecular diagnostic yield than proband-CES or traditional molecular diagnostic methods. Additional studies designed to validate these findings and to explore the effect of this approach on clinical and economic outcomes are warranted.

Editorial page 1865

Related article page 1870

Supplemental content at jama.com

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Stanley F. Nelson, MD, Clinical Genomics Center, David Geffen School of Medicine, University of California, Los Angeles, 695 Charles E. Young Dr S, GONDA 5506A, Los Angeles, CA 90095 (snelson@ucla.edu).

JAMA. 2014;312(18):1880-1887. doi:10.1001/jama.2014.14604
Published online October 18, 2014.

Over the last few years, advances in next-generation sequencing technologies have decreased the cost of sequencing per base pair about 10-fold, improved accuracy, and greatly increased the speed of generating sequence data. Exome sequencing, which sequences the protein-coding region of the genome, has been rapidly applied to variant discovery in research settings and recent increases in accuracy have enabled development of clinical exome sequencing (CES) for mutation identification in patients with suspected genetic diseases.

Early in 2012, our center launched a CES program with the goal of delivering a more comprehensive method for determining a molecular diagnosis for patients with presumed rare Mendelian disorders that have remained undiagnosed despite exhaustive genetic, biochemical, and radiological testing. We introduced a new test, called trio-CES, in which the whole exome of the affected proband and both parents are sequenced. The trio-CES test has the potential benefit of permitting more sensitive identification of de novo variants and compound heterozygotes and removing from consideration the many heterozygous rare variants observed in each exome from being considered causal in the affected individual because transmission is observed from an unaffected parent. This has not been routinely implemented by other centers due to costs and potential concerns for incidental findings in the unaffected parents. This study reports the first sequential 814 cases tested by our laboratory and investigates diagnostic yields from different implementations of exome sequencing.

Methods

Our CES test was validated according to the Clinical Laboratory Improvement Amendments regulations and College of American Pathologists guidelines as a single test from DNA extraction to result reporting. The cases were ascertained between January 17, 2012, and August 31, 2014. All work was performed within the University of California, Los Angeles (UCLA), Clinical Genomics Center. This study was approved by the UCLA institutional review boards.

Clinical Exome Sequencing Test

The cases are from a consecutive set of clinical cases referred for exome sequencing from clinics at a single university health system as well as from outside referring physicians. All samples required completion of a requisition form, pretest genetic counseling with clinical consent, and provision of a recent clinical note related to the reason that the clinician ordered exome sequencing. Patients were not required to undergo standardized clinical examinations or diagnostic testing prior to referral. Clinical characteristics and prior laboratory investigations of patients reported by referring clinicians were not systematically confirmed by study investigators. Most cases were submitted from geneticists for which substantial prior genetic investigation had been performed and no clear resolution was determined. Patients determined to be appropriate for exome sequencing often presented with clinical symptoms that either involved more than 1 body system or were deemed to be highly genetically heterogeneous. For instance, mutations in more than 1000 different disease genes can manifest as developmental delay. Thus, no specific genetic test is clearly defined for this group of patients. In addition, some cases were referred because other molecular diagnostic testing (such as microarray analysis) or specific gene sequencing (either isolated or panels) were either inconclusive or not available clinically.

Patients were tested after pretest counseling to describe the goals of the test, turnaround time, cost, limitations, the current rate of diagnostic success, the potential to identify variants of uncertain clinical significance and their meaning, and the possibility to opt in or out of receiving the results of incidental findings. This discussion was documented by signature of the referring physician and the patient or patient representative.

On the CES test requisition form, physicians were asked to choose primary clinical indications, report ethnicity, provide any family history, and offer any differential diagnoses, suspected causative genes, or both. Ethnicity information, provided by the physician, was used for quality control and validating genetic findings. For instance, more rare variants will be observed in African American populations. Two test options were offered: a trio-CES and a proband-CES. Proband-CES is exome sequencing of the patient only and trio-CES is exome sequencing of up to 3 family members, including the patient. Typically trio-CES tests include both unaffected parents and the patient, but other combinations of family members were accepted if appropriate, depending on the family history and individual availability.

Sequencing, Data Analysis, and Quality Control

Whole blood collected in tubes with EDTA or purified genomic DNA from tissue sources was accepted for testing. Genomic DNA extraction, library preparation, sequencing, and data analysis were performed using validated protocols (eMethods 1-3 and eFigure 1 in the Supplement). For the patients receiving trio-CES, library preparation and sequencing were only performed after all 3 blood samples had arrived to minimize potential differences due to batch effects. Exome capture was performed using SureSelect Human All Exon V2 Kit (Agilent Technologies) and sequencing was performed using the HiSeq 2000 for a 50-bp paired-end run or HiSeq 2500 for a 100-bp paired-end run (both from Illumina). An average of 60 million independent paired reads or 9.7 Gb of sequence data were generated per sample to provide a mean 100-fold coverage across the RefSeq protein-coding exons and flanking intronic sequence (±2 bp) with more than 93% of these bases and 94% of all reported Human Gene Mutation Database (HGMD) variant positions with a depth of coverage 10 × or more. Thus, we estimated that CES has a more than 93% chance of observing clinically relevant single nucleotide or small indel (insertions and deletions) variant(s). The mitochondrial genome is not specifically captured, but as a byproduct of being present at a high copy number, 99% of the base positions of the protein-

coding sequences in the mitochondrial genome are sequenced simultaneously to detect homoplasmic and highly heteroplasmic variants[1] (although we have not reported any clinically significant mitochondrial genome variant, we observed mitochondrial polymorphisms in almost all cases).

Various quality metrics were recorded for each sample for quality control and these metrics were highly consistent from sample to sample (eTable 1 in the Supplement) providing a sensitive indicator of possible sample contamination, errors in library preparation, or inadequate sequencing. Of 1734 samples sequenced, no samples failed these initial quality metrics. Normalized coverage was used to search for evidence of deletions or duplications of an entire chromosomal arm and to confirm the sex of the individuals from the X and the Y chromosome coverage. Because the relative depth of coverage of each exon was generally consistent, variance from the normal coverage was a sensitive means to assess for poor capture quality. Of 1734 samples, 3 samples were rejected due to atypical exon coverage and repeated successfully with a new blood specimen. Consanguinity was estimated and possible uniparental disomy was identified on the basis of long (>5 Mb) stretches of homozygosity.[2,3] Finally, when 1 or more additional family members were sequenced at the same time as the proband, the inheritance pattern and variant sharing across the exome was queried to confirm genetic relationships.

For the first approximately 300 cases, all reported variants were confirmed by Sanger sequencing and more than 99% confirmed. After empirically determining that variants with a QUAL score (a scaled probability of a variant existing at a given site based on the sequencing data calculated by the Genome Analysis Toolkit [GATK] variant caller) 500 or higher to be highly accurate,[4] only clinically significant variants with a QUAL score lower than 500 and all small indels were confirmed by Sanger sequencing before reporting. In current practice, approximately 20% of all reported variants, which are deemed to be of lower certainty, undergo Sanger sequencing confirmation and of these more than 99% are confirmed.

Variant Analysis and Interpretation

Variants were annotated using Variant Annotator eXtras (VAX, custom annotator)[5] to provide information regarding their effect on protein function, allele frequency in the general population, and prior evidence of disease causality and filtered to select likely pathogenic DNA variants from an average of 21 259 DNA variants per exome (eMethods 4 in the Supplement). Variants with minor allele frequency greater than 1% were removed and deemed to be likely benign (eFigure 2 in the Supplement).[6-8] Of the remaining variants, variants that resulted in amino acid substitutions, microdeletions, microduplications, splice-site changes, or premature protein terminations in the canonical transcript or the most commonly referenced transcript in the literature were selected. For trio-CES cases, variants were further filtered into 4 categories: de novo (new variants not observed in either parent, usually heterozygous in patient, and potentially causing an autosomal dominant condition), homozygous (both parents are heterozygous for the same variant and the child inherited a rare allele from both parents, potentially causing a recessive condition), compound heterozygous (the affected individual has 1 rare variant from the mother and a different rare variant inherited from the father, potentially causing a recessive condition), and inherited variants (this is the largest group of variants and are inherited from a parent and are usually not disease causing; eFigure 2 in the Supplement). Variants were called high quality if each individual in the trio had minimum of 10 × depth of coverage, the parents and unrelated samples in the batch lacked evidence of the alternate allele being present, and the QUAL score was 500 or higher.

Given the heterogeneity of genetic conditions (more than 4000 disease-gene relationships have been described in the Online Mendelian Inheritance in Man [OMIM]), we developed a permissive approach to use the clinician notes to generate a list of phenotypic key words (examples shown in eTables 2-3 in the Supplement). Each phenotypic key word was then searched in the HGMD (professional version) or OMIM database to generate a list of potential disease genes for each patient. From this process we generated a primary gene list (PGL), and the filtered variants were then further annotated as to whether or not they occurred in the PGL. For proband-CES cases, all homozygous and potential compound heterozygous variants with a minor allele frequency less than 1% and all heterozygous variants with a minor allele frequency less than 0.1% within the PGL were examined in depth and prioritized by how well the patient's phenotypes matched to the gene description (eMethods 4-5 and eFigures 2-3 in the Supplement). In addition, heterozygous variants with a minor allele frequency less than 1% that were identified in genes known to cause recessive disorders similar to the patient's phenotype were searched for and highlighted. For trio-CES cases, all de novo, homozygous, and compound heterozygous variants were examined. Inherited heterozygous variants in genes in the PGL that were known to be imprinted or pathogenic (ie, reported in HGMD) were inspected as well. All variants in HGMD are not necessarily disease causing.[9] For this reason, even if a variant was annotated to be present in HGMD, the literature supporting the evidence of each variant was manually curated by our bioinformaticians.

A variant list of all CES cases was presented by the bioinformatician and reviewed at the weekly genomic data board meeting consisting of pathologists, geneticists, genetic counselors, bioinformaticians, and the ordering physician, whenever possible. Non-UCLA physicians had the option of joining the genomic data board meeting in person or through a prescheduled encrypted online conference. Our interpretation process consisted of a brief synopsis of the patient course, followed by a presentation of the annotated filtered variant list; this permitted an informed discussion among the members of the board to evaluate each variant and decide which, if any, to report as pathogenic, likely pathogenic, or variant of uncertain clinical significance (eMethods 5 and eFigure 3 in the Supplement) and determine a molecular diagnosis. Typically, 12 to 16 cases were reviewed within each 2-hour meeting.

Table 1. Overall Molecular Diagnosis Rate

	Total (N = 814)		CES Test					
			Proband (n = 338)		Trio (n = 410)		Other (n = 66)[a]	
	No. of Patients	% (95% CI)	No. of Patients	% (95% CI)	No. of Patients	% (95% CI)	No. of Patients	% (95% CI)
Diagnosis	213	26 (23-29)	74	22 (18-27)	127	31 (27-36)	12	18 (11-29)
Potential diagnosis	228	28 (25-31)	121	36 (31-41)	84	20 (17-25)	23	35 (24-47)
No significant variant	342	42 (39-46)	139	41 (36-46)	173	42 (38-47)	30	45 (34-57)
Other[b]	31	4 (3-5)	4	1 (0-3)	26	6 (4-9)	1	2 (0-9)

Abbreviation: CES, clinical exome sequencing.

[a] The other CES group includes cases in which only 1 or no parent was sequenced but other family members were sequenced.

[b] Other includes cases in which we confirmed previously reported microarray finding that the genomic data board determined to be likely pathogenic, cases in which we reported variants of uncertain significance in novel genes, and cases in which CES was ordered to rule out a clinical diagnosis.

Statistical Analysis

To test the significance of a higher diagnostic rate of trio-CES than proband-CES, *P* values were calculated by 1-tailed Fisher exact test.[10] All other comparisons were done by a 2-tailed Fisher exact test. A *P* value of .05 was used as a significance threshold. The 95% CI for proportion was calculated using an online calculator.[11] The odds ratio and 95% CI for the significance of difference in diagnostic rate were also calculated using an online calculator.[12]

Because there were more proband-CES cases in the adult group (due to the lack of availability of older parents), and there were differences in clinical indications based on age, we also assessed the relative diagnostic yield only in the highly genetically heterogeneous group of developmental delay cases and only when CES was performed in children. Age groups of younger than 5 years, 5 to 18 years, and older than 18 years at testing were used for a more direct comparison with published results.[13]

Results

Study Population Characteristics

Patients were referred from a total of 143 different physicians from 42 different institutions. A UCLA clinic ordered 459 of the total 814 cases. Of the 814 patients, 520 patients (64%) were children (18 years or younger) and 254 (49%) of those children were younger than 5 years at testing (eTable 4 in the Supplement). More cases were ordered for trio-CES (353 of 520 patients, 68% [95% CI, 64%-72%]) in the childhood group than in the adult group (57 of 294 patients, 19% [95% CI, 15%-24%], $P < .001$) reflecting parental availability. In total, 453 were males (56%) and 361 were females (44%). The CES was requested for males more than females (*P* value = .003) in the childhood group (male: 310 of 520 patients, 60% [95% CI, 55%-64%]; female: 210 of 520 patients, 40% [95% CI, 36%-45%]) than in the adult group (men: 143 of 294 patients, 49% [95% CI, 43%-54%]; women: 151 of 294 patients, 51% [95% CI, 46%-57%]). The most common clinical indication was developmental delay overall (298 of 814 patients, 37% [95% CI, 33%-40%]) and in the childhood group (274 of 520 patients, 53% [95% CI, 48%-57%]), whereas ataxia was the most common clinical indication in the adult group (77 of 294 patients, 26% [95% CI, 21%-32%]).

Overall Diagnostic Rate

Each individual case from CES generated a PGL and required substantial filtering to determine causality. Of 814 patients, approximately 6% had genetic evidence of consanguinity. For nonconsanguineous families, there were typically fewer than 10 genes with homozygous variants in the proband that were not also homozygous in the parents and fewer than 10 genes with compound heterozygous variants. For proband-CES samples, it was not possible to reliably infer which genes contained de novo or compound heterozygous variants; in a typical case approximately 10 genes with homozygous variants, 40 to 50 genes with potential compound heterozygous variants (in which 2 heterozygous variants in the same gene were observed), and 250 to 300 genes with a single heterozygous variant with minor allele frequency less than 0.1% were considered for initial interpretation (eFigure 2 in the Supplement). In contrast, there were an average of 1.1 amino acid-altering, novel (never observed in the general population), high-quality de novo variants in each trio, which was consistent with prior observations.[14]

Table 1 summarizes the diagnostic rate of our sequential series of cases. Overall, a molecular diagnosis (with the causative variant(s) identified in a well-established clinical gene) was provided for 213 of the 814 total cases (26% [95% CI, 23%-29%]). The breadth of molecular diagnoses was large, and the list of the 213 cases is provided in eTable 2 with detailed variant information and its pathogenicity. Of 264 reported variants in 213 cases, 188 were reported as likely pathogenic and 73 were reported as pathogenic variants (eTable 2 in the Supplement). In addition, 228 of the 814 total cases (28% [95% CI, 25%-31%]) had potential molecular diagnoses based on the results from the CES (Table 1). These are variants identified in clinically relevant genes that were classified as variants of uncertain significance because (1) additional phenotyping (~ 25%) or segregation analysis (~ 50%) was needed, or (2) only 1 heterozygous protein-damaging variant was identified in a gene known to cause a recessive disorder consistent with the patient's phenotype but not covered 100% by CES (~ 25%). Thus, other methods to detect additional variant alleles (exonic deletion or duplication or variants in noncoding regions) were recommended. Some of the cases described in this report have been previously reported.[15]

Diagnostic Rate for Trio-CES vs Proband-CES and in Phenotypic Subgroups

Overall, 410 of the 814 patients (50%) were trio-CES, with both parents sequenced, and 338 patients (42%) were proband-CES. Sixty-six patients (8%) were submitted with only 1 parent or other family members without the parents. There was a significantly higher molecular diagnostic yield from cases performed as trio-CES (127 of 410 cases; 31% [95% CI, 27%-36%]) relative to proband-CES (74 of 338 cases; 22% [95 CI, 18%-27%], P = .003) in the overall cohort of cases. Among the 127 trio-CES tests with a conclusive molecular diagnosis, 50% (63 of 127 cases; 95% CI, 41%-58%) had a de novo variant, 16% (20 of 127 cases; 95% CI, 10%-23%) had a homozygous variant, 20% (26 of 127 cases; 95% CI, 14%-28%) had compound heterozygous variants, and 8% (10 of 127; 95% CI, 4%-14%) had an X-linked hemizygous variant (**Table 2**).

The most common phenotype of patients described by the referring physician was developmental delay (298 of 814 cases, 37%). The 5 most common comorbid phenotypes with developmental delay included hypotonia (111 of 298 cases, 37%), epilepsy or seizures (108 of 298 cases, 36%), dysmorphic features (144 of 298 cases, 48%), autism (69 of 298 cases, 23%), or congenital heart disorder or defect (36 of 298 cases, 12%). The molecular diagnosis rate for each of these phenotypes is shown in **Table 3**. Although overall 28% (83 of 298 cases; 95% CI, 23%-33%) of developmental delay resulted in a molecular diagnosis, only 16% (11 of 69 cases; 95% CI, 9%-27%) of developmental delay with autism resulted in a molecular diagnosis. The diagnostic rate of developmental delay with autism relative to developmental delay and dysmorphic features of 31% (44 of 144; 95% CI, 24%-39%) indicates differences in diagnostic success with these different presentations (P = .03). In order to explore the diagnostic success rate with trio-CES relative to proband-CES accounting for age, we analyzed diagnostic yield in children younger than 5 years at testing (n = 138). For this category, trio-CES was requested for 79% (109 of 138 cases) of all developmental delay cases and 17% (23 of 138 cases) were requested as proband-CES. There was an improvement in the diagnostic yield of trio-CES, with 45 of 109 cases (41% [95% CI: 32%-51%]) receiving a molecular diagnosis relative to the proband-CES cases, which had a molecular diagnostic yield of 9% (2 of 23 cases; 95% CI, 1%-28%, P value = .002; odds ratio, 7.4 [95% CI, 1.6-33.1]; **Table 4**).

Of all diseases for which CES was applied, the group of patients with retinal disorders had the highest molecular diagnostic rate (15 of 31 cases, 48% [95% CI, 32%-65%]), suggesting that a larger fraction of all possible Mendelian disease genes for this family of disorders have been described at this point in the literature. In contrast, ataxia (11 of 86 cases, 13% [95% CI, 7%-22%], P value < .001) and disorder of sexual development (6 of 35 cases, 17% [95% CI, 8%-33%], P value = .009) cases had much lower diagnostic rates, suggesting a larger fraction of unknown genes or nongenetic underlying mechanisms (Table 3).

Two of the 410 trio-CES cases (0.5%) had mislabeled blood tubes (the blood sample labeled as 1 of the parent's was from

Table 2. Distribution of Mutation Types for Trio–Clinical Exome Sequencing Cases With Conclusive Molecular Diagnosis

Mutation Type	No. of Cases (n = 127)	% (95% CI)
De novo	63	50 (41-58)
Homozygous	20	16 (10-23)
Compound heterozygous	26	20 (14-28)
Inherited heterozygous[a]	5	4 (1-9)
Copy number variant or uniparental disomy	3	2 (1-7)
X-linked hemizygous	10	8 (4-14)

[a] Inherited from 1 of the affected or unaffected parents in an autosomal dominant disorder gene with unknown or lower penetrance.

Table 3. Overall Molecular Diagnosis Rate of Phenotypic Subgroups by Clinical Exome Sequencing Test Type

	All		CES Test					
			Proband		Trio[a]		Other[b]	
Phenotypic Subgroup	Rate	% (95% CI)	Rate	% (95% CI)	Rate	% (95% CI)	Rate	% (95% CI)
DD	83/298	28 (23-33)	12/63	19 (11-31)	68/211	32 (26-39)	3/24	13 (4-32)
DD + hypotonia	28/111	25 (18-34)	2/27	7 (1-24)	26/80	33 (23-43)	0/4	0 (0-55)
DD + epilepsy or seizures	32/108	30 (22-39)	4/26	15 (6-34)	27/73	37 (27-48)	1/9	11 (0-46)
DD + dysmorphic features	44/144	31 (24-39)	4/30	13 (5-30)	37/99	37 (28-47)	3/15	20 (6-46)
DD + autism	11/69	16 (9-27)	0/10	0 (0-32)	10/47	21 (12-35)	1/12	8 (0-38)
DD + heart disorder	11/36	31 (18-47)	1/4	25 (3-71)	9/30	30 (17-48)	1/2	50 (9-91)
Ataxia and related neurological disorders	11/86	13 (7-22)	10/77	13 (7-22)	1/6	17 (1-58)	0/3	0 (0-62)
Muscular dystrophy and related disorders	22/74	30 (20-41)	14/57	25 (15-37)	8/15	53 (30-75)	0/2	0 (0-71)
Cardiomyopathy and arrhythmia	10/39	26 (14-41)	7/23	30 (15-51)	3/14	21 (7-48)	0/2	0 (0-71)
Cancer predisposition	7/36	19 (9-35)	5/15	33 (15-59)	1/16	6 (0-30)	1/5	20 (2-64)
Disorder of sexual development	6/35	17 (8-33)	5/18	28 (12-51)	0/14	0 (0-25)	1/3	33 (6-80)
Retinal disorders	15/31	48 (32-65)	5/12	42 (19-68)	7/11	64 (35-85)	3/8	38 (13-70)

Abbreviations: CES, clinical exome sequencing; DD, developmental delay.

[a] The trio-CES group includes 12 quartets (trio + sibling) and 1 quintet (trio + 2 siblings).

[b] The other CES group includes cases where only 1 or no parent was sequenced but other family members were sequenced.

Table 4. Molecular Diagnosis Rate of Phenotypic Subgroups by Age Group

Phenotypic Subgroup	Age Groups											
	<5 y				5-18 y				>18 y			
	Proband-CES		Trio-CES		Proband-CES		Trio-CES		Proband-CES		Trio-CES	
	Rate	% (95% CI)	Rate	% (95% CI)	Rate	% (95% CI)	Rate	% (95% CI)	Rate	% (95% CI)	Rate	% (95% CI)
DD	2/23	9 (1-28)	45/109	41 (32-51)[a]	8/30	27 (14-45)	22/93	24 (16-33)	2/10	20 (5-52)	1/9	11 (0-46)
DD + hypotonia	0/17	0 (0-22)	20/44	45 (32-60)[a]	2/7	29 (8-65)	6/34	18 (8-34)	0/3	0 (0-62)	0/2	0 (0-71)
DD + epilepsy or seizures	0/8	0 (0-37)	15/29	52 (34-69)[a]	3/14	21 (7-48)	11/39	28 (16-44)	1/4	25 (3-71)	1/5	20 (2-64)
DD + dysmorphic features	2/13	15 (3-43)	25/57	44 (32-57)	1/14	7 (0-34)	11/38	29 (17-45)	1/3	33 (6-80)	1/4	25 (3-71)
DD + autism	0/3	0 (0-62)	5/20	25 (11-47)	0/6	0 (0-44)	4/23	17 (6-38)	0/1	0 (0-83)	1/4	25 (3-71)
DD + heart disorder	0/1	0 (0-83)	7/18	39 (20-61)	0/1	0 (0-83)	2/11	18 (4-49)	1/2	50 (9-91)	0/1	0 (0-83)
Ataxia and related neurological disorders					0/3	0 (0-62)	0/4	0 (0-55)	10/74	14 (7-23)	1/2	50 (9-91)
Muscular dystrophy and related disorders	0/1	0 (0-83)	4/5	80 (36-98)	3/11	27 (9-57)	1/4	25 (3-71)	11/45	24 (14-39)	3/6	50 (19-81)
Cardiomyopathy and arrhythmia	1/4	25 (3-71)	0/4	0 (0-55)	2/4	50 (15-85)	2/8	25 (6-60)	4/15	27 (10-52)	1/2	50 (9-91)
Cancer predisposition			0/5	0 (0-49)	0/2	0 (0-71)	0/6	0 (0-44)	5/13	38 (18-65)	1/5	20 (2-64)
Disorder of sexual development	3/8	38 (13-70)	0/8	0 (0-37)	0/7	0 (0-40)	0/5	0 (0-49)	2/3	67 (20-94)	0/1	0 (0-83)
Retinal disorders							3/5	60 (23-88)	5/12	42 (19-68)	4/6	67 (30-91)

Abbreviations: CES, clinical exome sequencing; DD, developmental delay.

[a] *P* values for testing if the diagnostic rate of trio-CES was higher than proband-CES were .01 or less. The other CES group was not included in the calculation. *P* value was only calculated for these 3 groups that had significant (80%) power based on the observed proportions.[16]

the proband from the submitting blood draw facility), which were detected by assessing relatedness of the samples, and both instances were remedied by obtaining new specimens.

Illustrative Cases

A benefit of performing trio-CES is also underscored for identification of de novo variants in genes that are not yet associated with any human disorder, and the significance of identifying these de novo variants has already proven to be critical in a few cases. For instance, an infant (case 75 in eTable 2 in the Supplement) with multifocal complex partial epilepsy and regression of developmental milestones had a novel de novo missense variant in *KCNT1* detected, which was not known at the time to be associated with any human disease. However, soon after completion of CES for this infant, *KCNT1* de novo variants were reported in the literature as a cause of infantile epileptic encephalopathy (OMIM 614959), a condition that matches the patient phenotype well permitting update of the PGL and provision of a conclusive molecular diagnosis.[17] Another infant (case 113 in eTable 2 in the Supplement) with developmental delay, seizures, perisylvian polymicrogyria, and microcephaly had a novel de novo missense variant in *TUBB2A*, recently identified as a causal gene.[18] Furthermore, we reviewed case 107 (eTable 2 in the Supplement), which was reported 3 months prior with a de novo variant in *TUBB2A* as a variant of uncertain significance, and generated an addendum classifying the variant as likely pathogenic.

Trio-CES is also broadening the clinical phenotype of various rare Mendelian genetic diseases. An example is a 9-year-old girl with developmental delay, mild intellectual disability, hypotonia, dysmorphic features, early tooth eruption, and premature adult teeth in whom trio-CES identified a novel missense de novo variant in the *KMT2A (MLL)* gene that is known to cause autosomal dominant Wiedemann-Steiner syndrome (OMIM 605130; case 130 in eTable 2 in the Supplement). Wiedemann-Steiner syndrome had not been considered by the referring clinicians because the patient had not manifested the hallmark phenotype of the syndrome, namely hairy elbows. However, upon discovery of the de novo variant and another review of the clinical presentation, the ordering physician was able to note the manifestation of the excess growth of terminal hair analogous to hairy elbows.[19]

Although CES is not intended to identify copy number variant, uniparental disomy, somatic mosaic heterozygous variants or variants in the mitochondrial genome, in selected cases we observed evidence of large homozygous or hemizygous exonic deletions and duplications (del and dup). A total of 7 such observations were made, and 5 of 7 were confirmed by an outside laboratory (eTable 3 in the Supplement). Examples include the identification of pater-

nal uniparental disomy in a trio-CES case due to 2 homozygous regions of 5Mb and 19Mb on chromosome 6 with no decrease in coverage and no maternally inherited variants on the entire chromosome 6 (data not shown). We have also observed potential somatic mosaic heterozygous variants with significantly more reads from the reference allele than the alternate allele in 2 cases (case 1, 176 reference reads and 40 alternate reads; case 2, 193 reference reads and 35 alternate reads; *P* value <.001 for both cases). Both were confirmed by Sanger sequencing.

Incidental Findings

Return of "incidental" or "off-target" findings remains an area of debate in medical genetics.[20-24] Although we used published guidelines[20] to determine which genes and diseases may be considered for reporting, we did not restrict ourselves to this gene list nor did we actively search for incidental variants for every case. Instead, we have chosen to create a set of criteria to define an incidental variant in any gene. These criteria include considering only variants that are (1) present in a gene that is unrelated to the primary clinical concern(s) of the patient (typically not in PGL) and are also (2) predicted to be pathogenic or likely pathogenic according to current American College of Medical Genetics sequence interpretation guidelines. The final decision to report was made by the genomic data board. In concert with recent modifications,[25] we allowed for the patient to opt out of receiving such incidental findings via our consent form, though 97% of patients or parents (252 of 260) have chosen to receive them. Incidental variants have only been reported in 5% of cases and include likely pathogenic variants found in *BRCA1* and *BRCA2*, Lynch syndrome genes, and cardiomyopathies or hereditary arrhythmias.

Discussion

Clinical exome sequencing has rapidly become a component of the clinical approach to individuals with rare diseases and is being applied to a wide range of clinical presentations that require a broad search for causal variants across the spectrum of genetically heterogeneous Mendelian disorders. Similarly to the initial description of CES performed at Baylor College of Medicine,[13,26] the current study describes a molecular diagnostic rate of about 26%. Referrals to our center are for a much broader range of potential Mendelian genetic diseases than in the prior study with more than half of the sequencing requested for nonneurological diseases. However, developmental delay remains the most common reason for testing. In the instance of a new presentation of an affected child with multiorgan syndromic features and without any prior family history of a possible genetic syndrome, our data support that trio-CES is more sensitive than proband-CES, especially for genetically heterogeneous conditions such as developmental delay. This is primarily because of increased sensitivity to observe de novo variants and compound heterozygous variants. Because on average only a single de novo variant is observed per trio-CES, de novo variants have the potential to highlight novel disease-causing genes. The de novo mutations in *KCNT1* and *TUBB2A*, prior to their description as clinical genes, serve as examples of this potential. However, trio-CES did not increase the diagnostic rate for cardiomyopathy, cancer predisposition, and disorder of sex development patient groups. This may be due to ordering physicians preferentially selecting proband-CES if they have a suspected gene in mind for these nonsyndromic disorders and reserving trio-CES for more complicated cases, in which there is greater uncertainty.

This study has a number of important limitations. For example, physicians may not order trio-CES due to concerns about increased cost of trio-CES or concerns about incidental findings in unaffected parents. Because the trio-CES and proband-CES are not randomized, other unobserved confounding factors may also be affecting the diagnostic yield. Additionally, diagnostic usage of CES is not able to detect all causal mutation types, and thus specific mutations will be unobserved by this test. For example, CES does not detect pathogenic repeat expansions (for disorders such as spinocerebellar ataxia) or most copy number variants. Also, CES does not sample all protein-coding bases: the average sequence coverage information for each gene is available online.[27] The relative coverage of any given considered gene list by referring physicians is an important consideration for the appropriateness of CES. Our study is reporting a consecutive case series referred for CES from a wide variety of clinical practices. Thus, we do not have a full accounting of prior genetic or other phenotypic testing to allow assessment of cost-effectiveness.

Another challenging part of the CES test is the interpretation of the variants in the context of the phenotypic data provided. Although every effort is made to collect as much clinical information as possible from the referring clinicians, there are instances in which the clinician's input at the genomic data board discussion is useful to assess the plausibility of a given variant in the context of a more complete clinical description. Reanalysis of a negative exome data with updated PGL is only performed upon patient's or physician's request. There was only 1 such request and the additional information did not result in identification of a significant variant. Challenges remain in the interpretation of many cases. Improved knowledge of rare allele frequencies of healthy individuals, improved coverage of the genome by sequencing, enhanced methods for detecting all types of genetic variation, and more routine use of trio-CES will improve molecular diagnostic success rates as this field matures.

Conclusions

In this sample of patients with undiagnosed, suspected genetic conditions, trio-CES was associated with higher molecular diagnostic yield than proband-CES or traditional molecular diagnostic methods. Additional studies designed to validate these findings and to explore the effect of this approach on clinical and economic outcomes are warranted.

ARTICLE INFORMATION

Published Online: October 18, 2014. doi:10.1001/jama.2014.14604.

Author Affiliations: Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles (Lee, Deignan, Strom, Kantarci, Quintero-Rivera, Das, Toy, Grody, Nelson); Clinical Genomics Center, David Geffen School of Medicine, University of California, Los Angeles (Lee, Deignan, Dorrani, Strom, Kantarci, Quintero-Rivera, Das, Toy, Grody, Vilain, Nelson); Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles (Dorrani, Yourshaw, Fox, Martinez-Agosto, Wong, Chang, Dipple, Grody, Vilain); Institute for Digital Research and Education, University of California, Los Angeles (Harry); Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles (Fogel, Shieh); Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles (Martinez-Agosto, Palmer, Dipple, Grody, Vilain, Nelson); Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles (Palmer, Nelson).

Author Contributions: Drs Lee and Nelson had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Lee, Kantarci, Dipple, Grody, Vilain, Nelson.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Lee, Quintero-Rivera, Das, Martinez-Agosto, Dipple, Nelson.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Lee, Yourshaw, Nelson.
Obtained funding: Chang, Nelson.
Administrative, technical, or material support: Deignan, Quintero-Rivera, Das, Harry, Fox, Fogel, Shieh, Palmer, Dipple, Grody, Vilain, Nelson.
Study supervision: Das, Martinez-Agosto, Grody, Vilain, Nelson.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Drs Lee, Deignan, Dorrani, Strom, Kantarci, Quintero-Rivera, Das, Toy, Grody, and Nelson are part of the Department of Pathology and Laboratory Medicine, which derives revenue from the clinical use of the clinical exome sequencing test. Dr Fogel reports receiving grant funding from the National Institute of Mental Health, the National Institute for Neurological Disorders and Stroke, and the National Ataxia Foundation. Dr Shieh reports that he is a member of the speakers bureau for Grifols. Dr Vilain reports receiving grant funding from the University of California, Los Angeles.

Funding/Support: Part of the study was supported by the National Institutes of Health's National Center for Advancing Translational Science (UCLA CTSI UL1TR000124) and National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIH-P30-5P30AR057230), the California Institute of Regenerative Medicine (RT2-01985), the K12 Child Health Research Career Development Award (2K12HD034610-16), and the Hyundai Hope on Wheels Scholar Award.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and decision to submit the manuscript for publication.

Additional Contributions: We thank the staffs at the University of California, Los Angeles, Clinical Genomics Center for their compensated assistance in performing the clinical exome sequencing test, routing samples, and administrative tasks.

REFERENCES

1. Griffin HR, Pyle A, Blakely EL, et al. Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations [published online June 5, 2014]. *Genet Med*. doi:10.1038/gim.2014.66.

2. Rehder CW, David KL, Hirsch B, Toriello HV, Wilson CM, Kearney HM. American College of Medical Genetics and Genomics: standards and guidelines for documenting suspected consanguinity as an incidental finding of genomic testing. *Genet Med*. 2013;15(2):150-152.

3. Papenhausen P, Schwartz S, Risheg H, et al. UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am J Med Genet A*. 2011;155A(4):757-768.

4. Strom SP, Lee H, Das K, et al. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genet Med*. 2014;16(7):510-515.

5. Yourshaw M, Taylor SP, Rao AR, Martín MG, Nelson SF. Rich annotation of DNA sequencing variants by leveraging the Ensembl Variant Effect Predictor with plugins [published online March 12, 2014]. *Brief Bioinform*. doi:10.1093/bib/bbu008.

6. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;508(7497):469-476.

7. Duzkale H, Shen J, McLaughlin H, et al. A systematic approach to assessing the clinical significance of genetic variants. *Clin Genet*. 2013;84(5):453-463.

8. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations. *Genet Med*. 2008;10(4):294-300.

9. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database. *Hum Genet*. 2014;133(1):1-9.

10. GraphPad Software. Data Analysis Resource Center. http://graphpad.com/quickcalcs/contingency1.cfm. Accessed October 8, 2014.

11. GraphPad Software. QuickCalcs. http://graphpad.com/quickcalcs/ConfInterval1.cfm. Accessed October 8, 2014.

12. Hutchon DJR. Calculator for confidence intervals of odds ratio in an unmatched case control study. http://www.hutchon.net/ConfidOR.htm. Accessed October 8, 2014.

13. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013;369(16):1502-1511.

14. O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 2012;485(7397):246-250.

15. Fogel BL, Lee H, Deignan JL, et al. Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia [published online August 18, 2014]. *JAMA Neurol*. doi:10.1001/jamaneurol.2014.1944.

16. DSS Research. Calculators. https://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/samplesizecalculators.aspx. Accessed October 8, 2014.

17. Heron SE, Smith KR, Bahlo M, et al. Missense mutations in the sodium-gated potassium channel gene *KCNT1* cause severe autosomal dominant nocturnal frontal lobe epilepsy. *Nat Genet*. 2012;44(11):1188-1190.

18. Cushion TD, Paciorkowski AR, Pilz DT, et al. De novo mutations in the β-tubulin gene *TUBB2A* cause simplified gyral patterning and infantile-onset epilepsy. *Am J Hum Genet*. 2014;94(4):634-641.

19. Strom SP, Lozano R, Lee H, et al. De Novo variants in the *KMT2A* (*MLL*) gene causing atypical Wiedemann-Steiner syndrome in 2 unrelated individuals identified by clinical exome sequencing. *BMC Med Genet*. 2014;15(1):49.

20. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15(7):565-574.

21. Holtzman NA. ACMG recommendations on incidental findings are flawed scientifically and ethically. *Genet Med*. 2013;15(9):750-751.

22. Allyse M, Michie M. Not-so-incidental findings. *Trends Biotechnol*. 2013;31(8):439-441.

23. Yu JH, Harrell TM, Jamal SM, Tabor HK, Bamshad MJ. Attitudes of genetics professionals toward the return of incidental results from exome and whole-genome sequencing. *Am J Hum Genet*. 2014;95(1):77-84.

24. Tabor HK, Auer PL, Jamal SM, et al; NHLBI Exome Sequencing Project. Pathogenic variants for Mendelian and complex traits in exomes of 6517 European and African Americans . *Am J Hum Genet*. 2014;95(2):183-193.

25. American College of Medical Genetics and Genomics. ACMG updates recommendation on "opt out" for genome sequencing return of results. https://www.acmg.net/docs/Release_ACMGUpdatesRecommendations_final.pdf. Accessed May 30, 2014.

26. Jacob HJ. Next-generation sequencing for clinical diagnostics. *N Engl J Med*. 2013;369(16):1557-1558.

27. UCLA Health. UCLA Clinical Genomics Center. http://pathology.ucla.edu/genomics. Accessed October 8, 2014.

Genetics inMedicine

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

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

Disclaimer: These ACMG Standards and Guidelines were developed primarily as an educational resource for clinical laboratory geneticists to help them provide quality clinical laboratory services. Adherence to these standards and guidelines is voluntary and does not necessarily assure a successful medical outcome. These Standards and Guidelines should not be considered inclusive of all proper procedures and tests or exclusive of other procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific procedure or test, the clinical laboratory geneticist should apply his or her own professional judgment to the specific circumstances presented by the individual patient or specimen. Clinical laboratory geneticists are encouraged to document in the patient's record the rationale for the use of a particular procedure or test, whether or not it is in conformance with these Standards and Guidelines. They also are advised to take notice of the date any particular guideline was adopted and to consider other relevant medical and scientific information that becomes available after that date. It also would be prudent to consider whether intellectual property interests may restrict the performance of certain tests and other procedures.

The American College of Medical Genetics and Genomics (ACMG) previously developed guidance for the interpretation of sequence variants.[1] In the past decade, sequencing technology has evolved rapidly with the advent of high-throughput next-generation sequencing. By adopting and leveraging next-generation sequencing, clinical laboratories are now performing an ever-increasing catalogue of genetic testing spanning genotyping, single genes, gene panels, exomes, genomes, transcriptomes, and epigenetic assays for genetic disorders. By virtue of increased complexity, this shift in genetic testing has been accompanied by new challenges in sequence interpretation. In this context the ACMG convened a workgroup in 2013 comprising representatives from the ACMG, the Association for Molecular Pathology (AMP), and the College of American Pathologists to revisit and revise the standards and guidelines for the interpretation of sequence variants. The group consisted of clinical laboratory directors and clinicians. This report represents expert opinion of the workgroup with input from ACMG, AMP, and College of American Pathologists stakeholders. These recommendations primarily apply to the breadth of genetic tests used in clinical laboratories, including genotyping, single genes, panels, exomes, and genomes. This report recommends the use of specific standard terminology—"pathogenic," "likely pathogenic," "uncertain significance," "likely benign," and "benign"—to describe variants identified in genes that cause Mendelian disorders. Moreover, this recommendation describes a process for classifying variants into these five categories based on criteria using typical types of variant evidence (e.g., population data, computational data, functional data, segregation data). Because of the increased complexity of analysis and interpretation of clinical genetic testing described in this report, the ACMG strongly recommends that clinical molecular genetic testing should be performed in a Clinical Laboratory Improvement Amendments–approved laboratory, with results interpreted by a board-certified clinical molecular geneticist or molecular genetic pathologist or the equivalent.

Genet Med advance online publication 5 March 2015

Key Words: ACMG laboratory guideline; clinical genetic testing; interpretation; reporting; sequence variant terminology; variant reporting

[1]Department of Molecular and Medical Genetics, Knight Diagnostic Laboratories, Oregon Health & Science University, Portland, Oregon, USA; [2]College of American Pathologists, Chicago, Illinois, USA; [3]GeneDx, Gaithersburg, Maryland, USA; [4]Department of Pediatrics, Section of Genetics, Medical College of Wisconsin, Milwaukee, Wisconsin, USA; [5]Department of Human Genetics, Clinical Molecular Genetics Laboratory, The University of Chicago, Chicago, Illinois, USA; [6]Cytogenetics/Molecular Genetics Laboratory, Nationwide Children's Hospital, Columbus, Ohio, USA; [7]Department of Pathology, Ohio State University College of Medicine, Columbus, Ohio, USA; [8]Department of Pediatrics, Ohio State University College of Medicine, Columbus, Ohio, USA; [9]Department of Pathology and Laboratory Medicine, University of California Los Angeles School of Medicine, Los Angeles, California, USA; [10]Department of Pediatrics, University of California Los Angeles School of Medicine, Los Angeles, California, USA; [11]Department of Human Genetics, University of California Los Angeles School of Medicine, Los Angeles, California, USA; [12]Department of Human Genetics, Emory Genetics Laboratory, Emory University, Atlanta, Georgia, USA; [13]Department of Pathology, ARUP Institute for Clinical and Experimental Pathology, University of Utah, Salt Lake City, Utah, USA; [14]Department of Pediatrics, Molecular Genetics Laboratory, Children's Hospital Colorado, University of Colorado Anschutz Medical School, Denver, Colorado, USA; [15]Partners Laboratory for Molecular Medicine and Department of Pathology, Brigham & Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [16]Current affiliation: Phoenix Children's Hospital, Phoenix, Arizona, USA. Correspondence: Sue Richards (richarsu@ohsu.edu)

Approved by the ACMG Board of Directors on 15 December 2014 and the AMP Board of Directors on 9 January 2015.

Vol. 1 000049

Submitted 28 January 2015; accepted 28 January 2015; advance online publication 5 March 2015. doi:10.1038/gim.2015.30

INTRODUCTION

Clinical molecular laboratories are increasingly detecting novel sequence variants in the course of testing patient specimens for a rapidly increasing number of genes associated with genetic disorders. While some phenotypes are associated with a single gene, many are associated with multiple genes. Our understanding of the clinical significance of any given sequence variant falls along a gradient, ranging from those in which the variant is almost certainly pathogenic for a disorder to those that are almost certainly benign. While the previous American College of Medical Genetics and Genomics (ACMG) recommendations provided interpretative categories of sequence variants and an algorithm for interpretation, the recommendations did not provide defined terms or detailed variant classification guidance.[1] This report describes updated standards and guidelines for the classification of sequence variants using criteria informed by expert opinion and empirical data.

METHODS

In 2013 a workgroup consisting of ACMG, Association for Molecular Pathology (AMP), and College of American Pathologists members, representing clinical laboratory directors and clinicians, was formed with the goal of developing a recommendation for the use of standard terminology for classifying sequence variants using available evidence weighted according to a system developed through expert opinion, workgroup consensus, and community input. To assess the views of the clinical laboratory community, surveys were sent to over 100 sequencing laboratories in the United States and Canada that were listed in GeneTests.org, requesting input on terminology preferences and evaluation of evidence for classifying variants. Laboratory testing experience included rare disease as well as pharmacogenomics and somatic cancer testing. The first survey, aimed at assessing terminology preferences, was sent in February 2013, and the results were presented in an open forum at the 2013 ACMG annual meeting including over 75 attendees. Survey respondents represented more than 45 laboratories in North America. The outcome of the survey and open forum indicated that (i) a five-tier terminology system using the terms "pathogenic," "likely pathogenic," "uncertain significance," "likely benign," and "benign" was preferred and already in use by a majority of laboratories, and (ii) the first effort of the workgroup should focus on Mendelian and mitochondrial variants.

In the first survey, laboratories also were asked to provide their protocols for variant assessment, and 11 shared their methods. By analyzing all the protocols submitted, the workgroup developed a set of criteria to weight variant evidence and a set of rules for combining criteria to arrive at one of the five classification tiers. Workgroup members tested the scheme within their laboratories for several weeks using variants already classified in their laboratories and/or by the broader community. In addition, typical examples of variants harboring the most common types of evidence were tested for classification assignment to ensure the system would classify those variants according to current approaches consistently applied by workgroup members. A second survey was sent in August 2013 to the same laboratories identified through GeneTests.org as well as through AMP's listserv of ~2,000 members, along with the proposed classification scheme and a detailed supplement describing how to use each of the criteria. Laboratories were asked to use the scheme and to provide feedback as to the suitability and relative weighting of each criteria, the ease of use of the classification system, and whether they would adopt such a system in their own laboratory. Responses from over 33 laboratories indicated majority support for the proposed approach, and feedback further guided the development of the proposed standards and guidelines.

In November 2013 the workgroup held a workshop at the AMP meeting with more than 50 attendees, presenting the revised classification criteria and two potential scoring systems. One system is consistent with the approach presented here and the other is a point system whereby each criterion is given a number of points, assigning positive points for pathogenic criteria and negative points for benign criteria, with the total defining the variant class. With an audience-response system, the participants were asked how they would weight each criterion (as strong, moderate or supporting, or not used) during evaluation of variant evidence. Again, the responses were incorporated into the classification system presented here. It should be noted that while the majority of respondents did favor a point system, the workgroup felt that the assignment of specific points for each criterion implied a quantitative level of understanding of each criterion that is currently not supported scientifically and does not take into account the complexity of interpreting genetic evidence.

The workgroup also evaluated the literature for recommendations from other professional societies and working groups that have developed variant classification guidelines for well-studied genes in breast cancer, colon cancer, and cystic fibrosis and statistical analysis programs for quantitative evaluation of variants in select diseases.[2–5] While those variant analysis guidelines are useful in a specific setting, it was difficult to apply their proposed criteria to all genes and in different laboratory settings. The variant classification approach described in this article is meant to be applicable to variants in all Mendelian genes, whether identified by single gene tests, multigene panels, exome sequencing, or genome sequencing. We expect that this variant classification approach will evolve as technology and knowledge improve. We should also note that those working in specific disease groups should continue to develop more focused guidance regarding the classification of variants in specific genes given that the applicability and weight assigned to certain criteria may vary by gene and disease.

GENERAL CONSIDERATIONS

Terminology

A mutation is defined as a permanent change in the nucleotide sequence, whereas a polymorphism is defined as a variant with a frequency above 1%. The terms "mutation" and "polymorphism," however, which have been used widely,

often lead to confusion because of incorrect assumptions of pathogenic and benign effects, respectively. Thus, it is recommended that both terms be replaced by the term "variant" with the following modifiers: (i) pathogenic, (ii) likely pathogenic, (iii) uncertain significance, (iv) likely benign, or (v) benign. Although these modifiers may not address all human phenotypes, they comprise a five-tier system of classification for variants relevant to Mendelian disease as addressed in this guidance. It is recommended that all assertions of pathogenicity (including "likely pathogenic") be reported with respect to a condition and inheritance pattern (e.g., c.1521_1523delCTT (p.Phe508del), pathogenic, cystic fibrosis, autosomal recessive).

It should be noted that some laboratories may choose to have additional tiers (e.g., subclassification of variants of uncertain significance, particularly for internal use), and this practice is not considered inconsistent with these recommendations. It should also be noted that the terms recommended here differ somewhat from the current recommendations for classifying copy-number variants detected by cytogenetic microarray.[6] The schema recommended for copy-number variants, while also including five tiers, uses "uncertain clinical significance—likely pathogenic" and "uncertain clinical significance—likely benign." The majority of the workgroup was not supportive of using "uncertain significance" to modify the terms "likely pathogenic" or "likely benign" given that it was felt that the criteria presented here to classify variants into the "likely" categories included stronger evidence than outlined in the copy-number variant guideline and that combining these two categories would create confusion for the health-care providers and individuals receiving clinical reports. However, it was felt that the use of the term "likely" should be restricted to variants where the data support a high likelihood that it is pathogenic or a high likelihood that it is benign. Although there is no quantitative definition of the term "likely," guidance has been proposed in certain variant classification settings. A survey of the community during an ACMG open forum, however, suggested a much wider range of uses of the term "likely." Recognizing this, we propose that the terms "likely pathogenic" and "likely benign" be used to mean greater than 90% certainty of a variant either being disease-causing or benign to provide laboratories with a common, albeit arbitrary, definition. Similarly, the International Agency for Research on Cancer guideline[2] supports a 95% level of certainty of pathogenicity, but the workgroup (confirmed by feedback during the ACMG open forum) felt that clinicians and patients were willing to tolerate a slightly higher chance of error, leading to the 90% decision. It should also be noted that at present most variants do not have data to support a quantitative assignment of variant certainty to any of the five categories given the heterogeneous nature of most diseases. It is hoped that over time experimental and statistical approaches to objectively assign pathogenicity confidence to variants will be developed and that more rigorous approaches to defining what the clinical community desires in terms of confidence will more fully inform terminologies and likelihoods.

The use of new terminologies may require education of the community. Professional societies are encouraged to engage in educating all laboratories as well as health-care providers on the use of these terms, and laboratories also are encouraged to directly educate their ordering physicians.

Nomenclature

A uniform nomenclature, informed by a set of standardized criteria, is recommended to ensure the unambiguous designation of a variant and enable effective sharing and downstream use of genomic information. A standard gene variant nomenclature (http://www.hgvs.org/mutnomen) is maintained and versioned by the Human Genome Variation Society (HGVS),[7] and its use is recommended as the primary guideline for determining variant nomenclature except as noted.[6] Laboratories should note the version being used in their test methods. Tools are available to provide correct HGVS nomenclature for describing variants (https://mutalyzer.nl).[8] Clinical reports should include sequence reference(s) to ensure unambiguous naming of the variant at the DNA level, as well as to provide coding and protein nomenclature to assist in functional interpretations (e.g., "g." for genomic sequence, "c." for coding DNA sequence, "p." for protein, "m." for mitochondria). The coding nomenclature should be described using the "A" of the ATG translation initiation codon as position number 1. Where historical alternate nomenclature has been used, current nomenclature should be used with an additional notation of the historical naming. The reference sequence should be complete and derived from either the National Center for Biotechnology Information RefSeq database (http://www.ncbi.nlm.nih.gov/RefSeq/)[9] with the version number or the Locus Reference Genomic database (http://www.lrg-sequence.org).[10] Genomic coordinates should be used and defined according to a standard genome build (e.g., hg19) or a genomic reference sequence that covers the entire gene (including the 5′ and 3′ untranslated regions and promoter). A reference transcript for each gene should be used and provided in the report when describing coding variants. The transcript should represent either the longest known transcript and/or the most clinically relevant transcript. Community-supported reference transcripts can often be identified through Locus Reference Genomic,[10] the Consensus CDS Database,[11] the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar), or a locus-specific database. However, laboratories should evaluate the impact of the variant on all clinically relevant transcripts, including alternate transcripts that contain additional exons or extended untranslated regions, when there are known variants in these regions that are clinically interpretable.

Not all types of variants (e.g., complex variants) are covered by the HGVS recommendations, but possible descriptions for complex variants have been reported.[7,12] In addition, this ACMG recommendation supports three specific exceptions to the HGVS nomenclature rules: (i) "X" is still considered acceptable for use in reporting nonsense variants in addition to the current HGVS recommendation of "*" and "Ter"; (ii) it is

recommended that exons be numbered according to the chosen reference transcript used to designate the variant; and (iii) the term "pathogenic" is recommended instead of "affects function" because clinical interpretation is typically directly evaluating pathogenicity.

Literature and database use

A large number of databases contain a growing number of variants that are continuously being discovered in the human genome. When classifying and reporting a variant, clinical laboratories may find valuable information in databases, as well as in the published literature. As noted above, sequence databases can also be used to identify appropriate reference sequences. Databases can be useful for gathering information but should be used with caution.

Population databases (**Table 1**) are useful in obtaining the frequencies of variants in large populations. Population databases cannot be assumed to include only healthy individuals and are known to contain pathogenic variants. These population databases do not contain extensive information regarding the functional effect of these variants or any possible associated phenotypes. When using population databases, one must determine whether healthy or disease cohorts were used and, if possible, whether more than one individual in a family was included, as well as the age range of the subjects.

Disease databases (**Table 1**) primarily contain variants found in patients with disease and assessment of the variants' pathogenicity. Disease and gene-specific databases often contain variants that are incorrectly classified, including incorrect

Table 1 Population, disease-specific, and sequence databases

Population databases	
Exome Aggregation Consortium http://exac.broadinstitute.org/	Database of variants found during exome sequencing of 61,486 unrelated individuals sequenced as part of various disease-specific and population genetic studies. Pediatric disease subjects as well as related individuals were excluded.
Exome Variant Server http://evs.gs.washington.edu/EVS	Database of variants found during exome sequencing of several large cohorts of individuals of European and African American ancestry. Includes coverage data to inform the absence of variation.
1000 Genomes Project http://browser.1000genomes.org	Database of variants found during low-coverage and high-coverage genomic and targeted sequencing from 26 populations. Provides more diversity compared to the Exome Variant Server but also contains lower-quality data, and some cohorts contain related individuals.
dbSNP http://www.ncbi.nlm.nih.gov/snp	Database of short genetic variations (typically ≤50 bp) submitted from many sources. May lack details of the originating study and may contain pathogenic variants.
dbVar http://www.ncbi.nlm.nih.gov/dbvar	Database of structural variation (typically >50 bp) submitted from many sources.
Disease databases	
ClinVar http://www.ncbi.nlm.nih.gov/clinvar	Database of assertions about the clinical significance and phenotype relationship of human variations.
OMIM http://www.omim.org	Database of human genes and genetic conditions that also contains a representative sampling of disease-associated genetic variants.
Human Gene Mutation Database http://www.hgmd.org	Database of variant annotations published in the literature. Requires fee-based subscription to access much of the content.
Locus/disease/ethnic/other-specific databases	
Human Genome Variation Society http://www.hgvs.org/dblist/dblist.html Leiden Open Variation Database http://www.lovd.nl	The Human Genome Variation Society site developed a list of thousands of databases that provide variant annotations on specific subsets of human variation. A large percentage of databases are built in the Leiden Open Variation Database system.
DECIPHER http://decipher.sanger.ac.uk	A molecular cytogenetic database for clinicians and researchers linking genomic microarray data with phenotype using the Ensembl genome browser.
Sequence databases	
NCBI Genome http://www.ncbi.nlm.nih.gov/genome	Source of full human genome reference sequences.
RefSeqGene http://www.ncbi.nlm.nih.gov/refseq/rsg Locus Reference Genomic (LRG) http://www.lrg-sequence.org	Medically relevant gene reference sequence resource.
MitoMap http://www.mitomap.org/MITOMAP/HumanMitoSeq	Revised Cambridge reference sequence for human mitochondrial DNA.

Vol. 1 000052

claims published in the peer-reviewed literature, because many databases do not perform a primary review of evidence. When using disease databases, it is important to consider how patients were ascertained, as described below.

When using databases, clinical laboratories should (i) determine how frequently the database is updated, whether data curation is supported, and what methods were used for curation; (ii) confirm the use of HGVS nomenclature and determine the genome build and transcript references used for naming variants; (iii) determine the degree to which data are validated for analytical accuracy (e.g., low-pass next-generation sequencing versus Sanger-validated variants) and evaluate any quality metrics that are provided to assess data accuracy, which may require reading associated publications; and (iv) determine the source and independence of the observations listed.

Variant assessment also includes searching the scientific and medical literature. Literature using older nomenclature and classification or based on a single observation should be used with caution. When identifying individuals and families with a variant, along with associated phenotypes, it is important to consider how patients were ascertained. This caveat is important when assessing data from publications because affected individuals and related individuals are often reported multiple times, depending on the context and size of the study. This may be due to authorship overlap, interlaboratory collaborations, or a proband and family members being followed across different clinical systems. This may mistakenly lead to duplicate counting of affected patients and a false increase in variant frequency. Overlapping authorship or institutions is the first clue to the potential for overlapping data sets.

Clinical laboratories should implement an internal system to track all sequence variants identified in each gene and clinical assertions when reported. This is important for tracking genotype–phenotype correlations and the frequency of variants in affected and normal populations. Clinical laboratories are encouraged to contribute to variant databases, such as ClinVar, including clinical assertions and evidence used for the variant classification, to aid in the continued understanding of the impact of human variation. Whenever possible, clinical information should be provided following Health Insurance Portability and Accountability Act regulations for privacy. Clinical laboratories are encouraged to form collaborations with clinicians to provide clinical information to better understand how genotype influences clinical phenotype and to resolve differences in variant interpretation between laboratories. Because of the great potential to aid clinical laboratory practice, efforts are underway for clinical variant databases to be expanded and standardized. Standardization will provide easier access to updated information as well as facilitate submission from the clinical laboratory. For example, the ClinVar database allows for the deposition of variants with clinical observations and assertions, with review status tracked to enable a more transparent view of the levels of quality of the curation.

Computational (in silico) predictive programs

A variety of in silico tools, both publicly and commercially available, can aid in the interpretation of sequence variants. The algorithms used by each tool may differ but can include determination of the effect of the sequence variant at the nucleotide and amino acid level, including determination of the effect of the variant on the primary and alternative gene transcripts, other genomic elements, as well as the potential impact of the variant on the protein. The two main categories of such tools include those that predict whether a missense change is damaging to the resultant protein function or structure and those that predict whether there is an effect on splicing (**Table 2**). Newer tools are beginning to address additional noncoding sequences.[13]

The impact of a missense change depends on criteria such as the evolutionary conservation of an amino acid or nucleotide, the location and context within the protein sequence, and the biochemical consequence of the amino acid substitution. The measurement of one or a combination of these criteria is used in various in silico algorithms that assess the predicted impact of a missense change. Several efforts have evaluated the performance of available prediction software to compare them with each other and to assess their ability to predict "known" disease-causing variants.[14–17] In general, most algorithms for missense variant prediction are 65–80% accurate when examining known disease variants.[16] Most tools also tend to have low specificity, resulting in overprediction of missense changes as deleterious, and are not as reliable at predicting missense variants with a milder effect.[18] The in silico tools more commonly used for missense variant interpretation in clinical laboratories include PolyPhen2,[19] SIFT,[20] and MutationTaster.[21] A list of in silico tools used to predict missense variants can be found in **Table 2**.

Multiple software programs have been developed to predict splicing as it relates to the creation or loss of splice sites at the exonic or intronic level.[22] In general, splice site prediction tools have higher sensitivity (~90–100%) relative to specificity (~60–80%) in predicting splice site abnormalities.[23,24] Some of the in silico tools commonly used for splice site variant interpretation are listed in **Table 2**.

While many of the different software programs use different algorithms for their predictions, they have similarities in their underlying basis; therefore, predictions combined from different in silico tools are considered as a single piece of evidence in sequence interpretation as opposed to independent pieces of evidence. The use of multiple software programs for sequence variant interpretation is also recommended because the different programs each have their own strengths and weaknesses, depending on the algorithm; in many cases performance can vary by the gene and protein sequence. These are only predictions, however, and their use in sequence variant interpretation should be implemented carefully. It is not recommended that these predictions be used as the sole source of evidence to make a clinical assertion.

Table 2 In silico predictive algorithms

Category	Name	Website	Basis
Missense prediction	ConSurf	http://consurftest.tau.ac.il	Evolutionary conservation
	FATHMM	http://fathmm.biocompute.org.uk	Evolutionary conservation
	MutationAssessor	http://mutationassessor.org	Evolutionary conservation
	PANTHER	http://www.pantherdb.org/tools/csnpScoreForm.jsp	Evolutionary conservation
	PhD-SNP	http://snps.biofold.org/phd-snp/phd-snp.html	Evolutionary conservation
	SIFT	http://sift.jcvi.org	Evolutionary conservation
	SNPs&GO	http://snps-and-go.biocomp.unibo.it/snps-and-go	Protein structure/function
	Align GVGD	http://agvgd.iarc.fr/agvgd_input.php	Protein structure/function and evolutionary conservation
	MAPP	http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html	Protein structure/function and evolutionary conservation
	MutationTaster	http://www.mutationtaster.org	Protein structure/function and evolutionary conservation
	MutPred	http://mutpred.mutdb.org	Protein structure/function and evolutionary conservation
	PolyPhen-2	http://genetics.bwh.harvard.edu/pph2	Protein structure/function and evolutionary conservation
	PROVEAN	http://provean.jcvi.org/index.php	Alignment and measurement of similarity between variant sequence and protein sequence homolog
	nsSNPAnalyzer	http://snpanalyzer.uthsc.edu	Multiple sequence alignment and protein structure analysis
	Condel	http://bg.upf.edu/fannsdb/	Combines SIFT, PolyPhen-2, and MutationAssessor
	CADD	http://cadd.gs.washington.edu	Contrasts annotations of fixed/nearly fixed derived alleles in humans with simulated variants
Splice site prediction	GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml	Markov models
	Human Splicing Finder	http://www.umd.be/HSF/	Position-dependent logic
	MaxEntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html	Maximum entropy principle
	NetGene2	http://www.cbs.dtu.dk/services/NetGene2	Neural networks
	NNSplice	http://www.fruitfly.org/seq_tools/splice.html	Neural networks
	FSPLICE	http://www.softberry.com/berry.phtml?topic=fsplice&group=programs&subgroup=gfind	Species-specific predictor for splice sites based on weight matrices model
Nucleotide conservation prediction	GERP	http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html	Genomic evolutionary rate profiling
	PhastCons	http://compgen.bscb.cornell.edu/phast/	Conservation scoring and identification of conserved elements
	PhyloP	http://compgen.bscb.cornell.edu/phast/	
		http://compgen.bscb.cornell.edu/phast/help-pages/phyloP.txt	Alignment and phylogenetic trees: Computation of *P* values for conservation or acceleration, either lineage-specific or across all branches

In silico tools/software prediction programs used for sequence variant interpretation.

PROPOSED CRITERIA FOR INTERPRETATION OF SEQUENCE VARIANTS

The following approach to evaluating evidence for a variant is intended for interpretation of variants observed in patients with suspected inherited (primarily Mendelian) disorders in a clinical diagnostic laboratory setting. It is not intended for the interpretation of somatic variation, pharmacogenomic (PGx) variants, or variants in genes associated with multigenic non-Mendelian complex disorders. Care must be taken when applying these rules to candidate genes ("genes of uncertain significance" (GUS)) in the context of exome or genome studies (see Special Considerations below) because this guidance is not intended to fulfill the needs of the research community in its effort to identify new genes in disease.

Although these approaches can be used for evaluating variants found in healthy individuals or secondary to the indication for testing, further caution must be used, as noted in several parts of the guideline, given the low prior likelihood that most

variants unrelated to the indication are pathogenic. Although we expect that, in general, these guidelines will apply for variant classification regardless of whether the variant was identified through analysis of a single gene, gene panel, exome, genome, or transcriptome, it is important to consider the differences between implicating a variant as pathogenic (i.e., causative) for a disease and a variant that may be predicted to be disruptive/damaging to the protein for which it codes, but is not necessarily implicated in a disease. These rules are intended to determine whether a variant in a gene with a definitive role in a Mendelian disorder may be pathogenic for that disorder. Pathogenicity determination should be independent of interpreting the cause of disease in a given patient. For example, a variant should not be reported as pathogenic in one case and not pathogenic in another simply because the variant is not thought to explain disease in a given case. Pathogenicity should be determined by the entire body of evidence in aggregate, including all cases studied, arriving at a single conclusion.

This classification approach may be somewhat more stringent than laboratories have applied to date. They may result in a larger proportion of variants being categorized as uncertain significance. It is hoped that this approach will reduce the substantial number of variants being reported as "causative" of disease without having sufficient supporting evidence for that classification. It is important to keep in mind that when a clinical laboratory reports a variant as pathogenic, health-care providers are highly likely to take that as "actionable" and to alter the treatment or surveillance of a patient[25] or remove such management in a genotype-negative family member, based on that determination (see How Should Health-Care Providers Use These Guidelines and Recommendations, below).

We have provided two sets of criteria: one for classification of pathogenic or likely pathogenic variants (**Table 3**) and one for classification of benign or likely benign variants (**Table 4**). Each pathogenic criterion is weighted as very strong (PVS1), strong (PS1–4); moderate (PM1–6), or supporting (PP1–5), and each benign criterion is weighted as stand-alone (BA1), strong (BS1–4), or supporting (BP1–6). The numbering within each category does not convey any differences of weight and is merely labeled to help refer to the different criteria. For a given variant, the user selects the criteria based on the evidence observed for the variant. The criteria then are combined according to the scoring rules in **Table 5** to choose a classification from the five-tier system. The rules apply to all available data on a variant, whether gathered from the current case under investigation or from well-vetted previously published data. Unpublished case data may also be obtained through public resources (e.g., ClinVar or locus specific databases) and from a laboratory's own database. To provide critical flexibility to variant classification, some criteria listed as one weight can be moved to another weight using professional judgment, depending on the evidence collected. For example, rule PM3 could be upgraded to strong if there were multiple observations of detection of the variant in *trans* (on opposite chromosomes) with other pathogenic variants (see PM3 BP2 cis/trans Testing for further guidance). By contrast, in situations when the data are not as strong as described, judgment can be used to consider the evidence as fulfilling a lower level (e.g., see PS4, Note 2 in **Table 3**). If a variant does not fulfill criteria using either of these sets (pathogenic or benign), or the evidence for benign and pathogenic is conflicting, the variant defaults to uncertain significance. The criteria, organized by type and strength, is shown in **Figure 1**. Please note that expert judgment must be applied when evaluating the full body of evidence to account for differences in the strength of variant evidence.

The following is provided to more thoroughly explain certain concepts noted in the criteria for variant classification (**Tables 3** and **4**) and to provide examples and/or caveats or pitfalls in their use. This section should be read in concert with **Tables 3** and **4**.

PVS1 null variants

Certain types of variants (e.g., nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single exon or multiexon deletion) can often be assumed to disrupt gene function by leading to a complete absence of the gene product by lack of transcription or nonsense-mediated decay of an altered transcript. One must, however, exercise caution when classifying these variants as pathogenic by considering the following principles:

(i) When classifying such variants as pathogenic, one must ensure that null variants are a known mechanism of pathogenicity consistent with the established inheritance pattern for the disease. For example, there are genes for which only heterozygous missense variants cause disease and null variants are benign in a heterozygous state (e.g., many hypertrophic cardiomyopathy genes). A novel heterozygous nonsense variant in the *MYH7* gene would not be considered pathogenic for dominant hypertrophic cardiomyopathy based solely on this evidence, whereas a novel heterozygous nonsense variant in the *CFTR* gene would likely be considered a recessive pathogenic variant.

(ii) One must also be cautious when interpreting truncating variants downstream of the most 3′ truncating variant established as pathogenic in the literature. This is especially true if the predicted stop codon occurs in the last exon or in the last 50 base pairs of the penultimate exon, such that nonsense-mediated decay[26] would not be predicted, and there is a higher likelihood of an expressed protein. The length of the predicted truncated protein would also factor into the pathogenicity assignment, however, and such variants cannot be interpreted without a functional assay.

(iii) For splice-site variants, the variant may lead to exon skipping, shortening, or inclusion of intronic material as a result of alternative donor/acceptor site usage or creation of new sites. Although splice-site variants are predicted to lead to a null effect, confirmation of impact requires functional analysis by either RNA or protein analysis. One must also consider the possibility of an in-frame deletion/insertion, which could retain the critical domains of the protein and hence lead to

Table 3 Criteria for classifying pathogenic variants

Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.
Moderate	PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation PM2 Absent from controls (or at extremely low frequency if recessive) (**Table 6**) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing. PM3 For recessive disorders, detected in *trans* with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase. PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level. PM6 Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant. PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

LOF, loss of function; OR, odds ratio.

Table 4 Criteria for classifying benign variants

Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	BS1 Allele frequency is greater than expected for disorder (see **Table 6**)
	BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age
	BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing
	BS4 Lack of segregation in affected members of a family
	Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting	BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease
	BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in *cis* with a pathogenic variant in any inheritance pattern
	BP3 In-frame deletions/insertions in a repetitive region without a known function
	BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)
	Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
	BP5 Variant found in a case with an alternate molecular basis for disease
	BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation
	BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

either a mild or neutral effect with a minor length change (PM4) or a gain-of-function effect.

(iv) Considering the presence of alternate gene transcripts and understanding which are biologically relevant, and in which tissues the products are expressed, are important. If a truncating variant is confined to only one or not all transcripts, one must be cautious about over-interpreting variant impact given the presence of the other protein isoforms.

(v) One must also be cautious in assuming that a null variant will lead to disease if found in an exon where no other pathogenic variants have been described, given the possibility that the exon may be alternatively spliced. This is particularly true if the predicted truncating variant is identified as an incidental finding (unrelated to the indication for testing), given the low prior likelihood of finding a pathogenic variant in that setting.

PS1 same amino acid change

In most cases, when one missense variant is known to be pathogenic, a different nucleotide change that results in the same amino acid (e.g., c.34G>C (p.Val12Leu) and c.34G>T (p.Val12Leu)) can also be assumed to be pathogenic, particularly if the mechanism of pathogenicity occurs through altered protein function. However, it is important to assess the possibility that the variant may act directly through the specific DNA change (e.g., through splicing disruption as assessed by at least computational analysis) instead of through the amino acid change, in which case the assumption of pathogenicity may no longer be valid.

PS2 PM6 de novo variants

A variant observed to have arisen de novo (parental samples testing negative) is considered strong support for pathogenicity if the following conditions are met:

(i) Both parental samples were shown through identity testing to be from the biological parents of the patient. Note that PM6 applies if identity is assumed but not confirmed.

(ii) The patient has a family history of disease that is consistent with de novo inheritance (e.g., unaffected parents for a dominant disorder). It is possible, however, that more than one sibling may be affected because of germ-line mosaicism.

(iii) The phenotype in the patient matches the gene's disease association with reasonable specificity. For example, this argument is strong for a patient with a de novo variant in the *NIPBL* gene who has distinctive facial features, hirsutism, and upper-limb defects (i.e., Cornelia de Lange syndrome), whereas it would be weaker for a de novo variant found by exome sequencing in a child with nonspecific features such as developmental delay.

PS3 BS3 functional studies

Functional studies can be a powerful tool in support of pathogenicity; however, not all functional studies are effective in

predicting an impact on a gene or protein function. For example, certain enzymatic assays offer well-established approaches to assess the impact of a missense variant on enzymatic function in a metabolic pathway (e.g., α-galactosidase enzyme function). On the other hand, some functional assays may be less consistent predictors of the effect of variants on protein function. To assess the validity of a functional assay, one must consider how closely the functional assay reflects the biological environment. For example, assaying enzymatic function directly from biopsied tissue from the patient or an animal model provides stronger evidence than expressing the protein in vitro. Likewise, evidence is stronger if the assay reflects the full biological function of the protein (e.g., substrate breakdown by an enzyme) compared with only one component of function (e.g., adenosine triphosphate hydrolysis for a protein with additional binding properties). Validation, reproducibility, and robustness data that assess the analytical performance of the assay and account for specimen integrity, which can be affected by the method and time of acquisition, as well as storage and transport, are important factors to consider. These factors are mitigated in the case of an assay in a Clinical Laboratory Improvement Amendments laboratory–developed test or commercially available kit. Assays that assess the impact of variants at the messenger RNA level can be highly informative when evaluating the effects of variants at splice junctions and within coding sequences and untranslated regions, as well as deeper intronic regions (e.g., messenger RNA stability, processing, or translation). Technical approaches include direct analysis of RNA and/or complementary DNA derivatives and in vitro minigene splicing assays.

Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

PS4 PM2 BA1 BS1 BS2 variant frequency and use of control populations

Assessing the frequency of a variant in a control or general population is useful in assessing its potential pathogenicity. This can be accomplished by searching publicly available population databases (e.g., 1000 Genomes Project, National Heart, Lung, and Blood Institute Exome Sequencing Project Exome Variant Server, Exome Aggregation Consortium; **Table 1**), as well as using race-matched control data that often are published in the literature. The Exome Sequencing Project data set is useful for Caucasian and African American populations and has coverage data to determine whether a variant is absent. Although the 1000 Genomes Project data cannot be used to assess the absence of a variant, it has a broader representation of different racial populations. The Exome Aggregation Consortium more recently released allele frequency data from >60,000 exomes from a diverse set of populations that includes approximately two-thirds of the Exome Sequencing Project data. In general, an allele frequency in a control population that is greater than expected for the disorder (**Table 6**) is considered strong support for a benign interpretation for a rare Mendelian disorder (BS1) or, if over 5%, it is considered as stand-alone support (BA1). Furthermore, if the disease under investigation is fully penetrant at an early age and the variant is observed in a well-documented healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) condition, then this is considered strong evidence for a benign interpretation (BS2). If the variant is absent, one should confirm that the read depth in the database is sufficient for an accurate call at the variant site. If a variant is absent from (or below the expected carrier frequency if recessive) a large general population or a control cohort (>1,000 individuals) and the population is race-matched to the patient harboring the identified variant, then this observation can be considered a moderate piece of evidence for pathogenicity (PM2). Many benign variants are "private" (unique to individuals or families), however, and therefore absence in a race-matched population is not considered sufficient or even strong evidence for pathogenicity.

The use of population data for case–control comparisons is most useful when the populations are well phenotyped, have

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Figure 1 Evidence framework. This chart organizes each of the criteria by the type of evidence as well as the strength of the criteria for a benign (left side) or pathogenic (right side) assertion. Evidence code descriptions can be found in **Tables 3** and **4**. BS, benign strong; BP, benign supporting; FH, family history; LOF, loss of function; MAF, minor allele frequency; path., pathogenic; PM, pathogenic moderate; PP, pathogenic supporting; PS, pathogenic strong; PVS, pathogenic very strong.

large frequency differences, and the Mendelian disease under study is early onset. Patients referred to a clinical laboratory for testing are likely to include individuals sent to "rule out" a disorder, and thus they may not qualify as well-phenotyped cases. When using a general population as a control cohort, the presence of individuals with subclinical disease is always a possibility. In both of these scenarios, however, a case–control comparison will be underpowered with respect to detecting a difference; as such, showing a statistically significant difference can still be assumed to provide supportive evidence for pathogenicity, as noted above. By contrast, the absence of a statistical difference, particularly with extremely rare variants and less penetrant phenotypes, should be interpreted cautiously.

Odds ratios (ORs) or relative risk is a measure of association between a genotype (i.e., the variant is present in the genome) and a phenotype (i.e., affected with the disease/outcome) and can be used for either Mendelian diseases or complex traits. In this guideline we are addressing only its use in Mendelian disease. While relative risk is different from the OR, relative risk asymptotically approaches ORs for small probabilities. An OR of 1.0 means that the variant does not affect the odds of having the disease, values above 1.0 mean there is an association between the variant and the risk of disease, and those below 1.0 mean there is a negative association between the variant and the risk of disease. In general, variants with a modest Mendelian effect size will have an OR of 3 or greater, whereas highly penetrant variants will have very high ORs; for example, *APOE* E4/E4 homozygotes compared with E3/E3 homozygotes have an OR of 13 (https://www.tgen.org/home/education-outreach/past-summer-interns/2012-summer-interns/erika-kollitz.aspx#.VOSi3C7G_vY). However, the confidence interval (CI) around the OR is as important as the measure of association itself. If the CI includes 1.0 (e.g., OR = 2.5, CI = 0.9–7.4), there is little confidence in

Table 6 Assessment of variant frequency in the general population for curated variant classification

Disease	Gene	Inheritance	Population	Incidence	Carrier frequency	Common mutation	Variant classification	ESP6500 AA MAF	ESP6500 EA MAF	ESP6500 All MAF	Concordance	Criteria to support classification
Cystic fibrosis	CFTR	AR	Caucasian	0.031%	3.6%	p.F508del	Ex24:p.F508del (Pathogenic)	N/A	N/A	N/A	N/A	ref. 57 (note variant not available in EVS)
							Ex11:c.1523T>G / p.F508C (Benign)	0.070%	0.150%	0.120%	No	ref. 58
							Ex23:c.3870A>G / p.(=) (Benign)	15.090%	2.970%	7.070%	Partial	AA MAF
							5′ UTR:c.-8G>C (Benign)	1.160%	5.550%	4.060%	Partial	EA MAF
							IVS6:c.743+40A>G (Benign)	0.700%	5.190%	3.670%	Partial	EA MAF
Phenylketoneuria	PAH	AR	North European	0.010%	2.0%		Ex12:c.1242C>T / p.(=) (Benign)	0.360%	1.310%	0.990%	No	PAH database (http://www.pahdb.mcgill.ca/)
							Ex12:c.1278T>C / p.(=) (Benign)	13.550%	0.090%	4.650%	Partial	AA MAF
							IVS12:c.1316-35C>T (Benign)	0.320%	2.630%	1.850%	Partial	EA MAF
							Ex9:c.963C>T / p.(=) (Benign)	5.170%	0.000%	1.750%	Partial	AA MAF
MCADD	ACADM	AR	Not specific	0.006%	1.5%	p.K329E aka p.K304E	Ex7:c.489T>G / p.(=) (Benign)	7.010%	0.050%	2.410%	Partial	AA MAF
ARPKD	PKHD1	AR	Not specific	0.005%	1.4%		IVS20:c.1964+17G>T (Benign)	0.200%	0.810%	0.610%	No	AA MAF
							Ex61:c.10515C>A / p.S3505R (Benign)	0.230%	1.130%	0.820%	No	EA MAF
							Ex66:c.11738G>A / p.R3913H (Benign)	1.270%	0.000%	0.430%	No	AA MAF
							Ex17:c.1587T>C / p.(=) (Benign)	1.380%	6.860%	5.010%	Partial	EA MAF
							Ex65:c.11525G>T / p.R3842L (Benign)	0.360%	2.430%	1.730%	Partial	EA MAF
							Ex61:c.10585G>C / p.E3529Q (Benign)	3.950%	0.010%	1.350%	Partial	AA MAF
Rett syndrome	MECP2	X Linked	Not specific	0.012%	De novo		Ex4:c.1161C>T / p.(=) (Benign)	0.030%	0.000%	0.010%	Partial	AA MAF
							Ex4:c.608C>T / p.T203M (Benign)	0.000%	0.060%	0.040%	Partial	http://www.ncbi.nlm.nih.gov/clinvar/variation/95198/
							Ex4:c.683C>G / p.T228S (Pathogenic)	0.830%	0.000%	0.300%	Partial	RETT database
Kabuki syndrome	KMT2D (MLL2)	AD	Not specific	0.003%	De novo		IVS31:c.8047-15C>T (Benign)	0.000%	0.020%	0.020%	Partial	EA MAF
							Ex31:c.6836G>A / p.Gly2279E (Benign)	0.000%	0.120%	0.080%	Partial	EA MAF
CHARGE syndrome	CHD7	AD	Not specific	0.010%	De novo		Ex2:c.309G>A / p.(=) (Benign)	1.460%	0.000%	0.490%	Partial	AA MAF
							Ex31:c.6478G>A / p.A2160T (Benign)	1.250%	0.000%	0.390%	Partial	AA MAF
							Ex2:c.856A>G / p.R286Gly (Benign)	0.780%	0.000%	0.250%	Partial	AA MAF
GJB2 associated hearing loss	GJB2	AR	Not specific	0.067%	2.5%	c.35delG	Ex2:c.35delG (Pathogenic)	0.090%	1.080%	0.740%	No	ref. 59
Hemochromatosis	HFE	AR	All	0.040%	8.3%	p.C282Y	Ex4:c.845G>A / p.C282Y (Other Reportable)	1.520%	6.410%	4.750%	No	http://www.ncbi.nlm.nih.gov/sites/GeneTests/review/gene/HFE

Variants in some common genetic disorders with their known incidences for dominant (Kabuki syndrome and CHARGE syndrome), X-linked diseases (Rett syndrome), and carrier frequencies for recessive disorders (cystic fibrosis, phenylketonuria (PKU), medium co-acyl dehydrogenase deficiency (MCADD), autosomal recessive polycystic kidney disease (ARPKD), GJB2-associated hearing loss, and hemochromatosis) are shown. Some variants in these genes found in the Exome Sequencing Project-GO database are listed with the variant classifications based on evidence and concordance with allele frequencies in African American (AA) and European American (EA) subpopulations. Variants that have an allele frequency greater than expected for the disorder in at least one subpopulation (AA or EA) of individuals not known to have the disorder can be considered to have strong evidence of benign impact (BS1). Variants known to be pathogenic for dominant disorders should have allele frequencies in the general population below the population disease incidence, and pathogenic variants for recessive disorders should have heterozygous frequencies consistent with their disease incidence. The allele frequencies shown for certain variants are lower for certain benign and higher for certain pathogenic variants than the disease incidence/carrier frequency; therefore, other data are required to classify these variants. Variants represented in the concordance column are designated as partial when a subpopulation frequency does not conform as above or when concordance is not achieved. Variants are designated as having partial concordance when a single subpopulation frequency is inconsistent with the classification.

the assertion of association. In the above *APOE* example the CI was ~10–16. Very simple OR calculators are available on the Internet (e.g., http://www.hutchon.net/ConfidOR.htm/ and http://easycalculation.com/statistics/odds-ratio.php/).[27,28]

PM1 mutational hot spot and/or critical and well-established functional domain

Certain protein domains are known to be critical to protein function, and all missense variants in these domains identified to date have been shown to be pathogenic. These domains must also lack benign variants. In addition, mutational hotspots in less well-characterized regions of genes are reported, in which pathogenic variants in one or several nearby residues have been observed with greater frequency. Either evidence can be considered moderate evidence of pathogenicity.

PM3 BP2 *cis*/*trans* testing

Testing parental samples to determine whether the variant occurs in *cis* (the same copy of the gene) or in *trans* (different copies of the gene) can be important for assessing pathogenicity. For example, when two heterozygous variants are identified in a gene for a recessive disorder, if one variant is known to be pathogenic, then determining that the other variant is in *trans* can be considered moderate evidence for pathogenicity of the latter variant (PM3). In addition, this evidence could be upgraded to strong if there are multiple observations of the variant in *trans* with other pathogenic variants. If the variant is present among the general population, however, a statistical approach would be needed to control for random co-occurrence. By contrast, finding the second variant in *cis* would be supporting, though not definitive, evidence for a benign role (BP2). In the case of uncertain pathogenicity of two heterozygous variants identified in a recessive gene, then the determination of the *cis* versus *trans* nature of the variants does not necessarily provide additional information with regard to the pathogenicity of either variant. However, the likelihood that both copies of the gene are impacted is reduced if the variants are found in *cis*.

In the context of dominant disorders the detection of a variant in *trans* with a pathogenic variant can be considered supporting evidence for a benign impact (BP2) or, in certain well-developed disease models, may even be considered standalone evidence, as has been validated for use in assessing *CFTR* variants.[3]

PM4 BP3 protein length changes due to in-frame deletions/insertions and stop losses

The deletion or insertion of one or more amino acids as well as the extension of a protein by changing the stop codon to an amino acid codon (e.g., a stop loss variant) is more likely to disrupt protein function compared with a missense change alone as a result of length changes in the protein. Therefore, in-frame deletions/insertions and stop losses are considered moderate evidence of pathogenicity. The larger the deletion, insertion, or extension, and the more conserved the amino acids are in a deleted region, the more substantial is the evidence to support pathogenicity. By contrast, small in-frame deletions/insertions in repetitive regions, or regions that are not well conserved in evolution, are less likely to be pathogenic.

PM5 novel missense at the same position

A novel missense amino acid change occurring at the same position as another pathogenic missense change (e.g., Trp38Ser and Trp38Leu) is considered moderate evidence but cannot be assumed to be pathogenic. This is especially true if the novel change is more conservative compared with the established pathogenic missense variant. Also, the different amino acid change could lead to a different phenotype. For example, different substitutions of the Lys650 residue of the *FGFR3* gene are associated with a wide range of clinical phenotypes: p.Lys650Gln or p.Lys650Asn causes mild hypochondroplasia; p.Lys650Met causes severe achondroplasia with developmental delay and acanthosis nigricans; and thanatophoric dysplasia type 2, a lethal skeletal dysplasia, arises from p.Lys650Glu.

PP1 BS4 segregation analysis

Care must be taken when using segregation of a variant in a family as evidence for pathogenicity. In fact, segregation of a particular variant with a phenotype in a family is evidence for linkage of the locus to the disorder but not evidence of the pathogenicity of the variant itself. A statistical approach has been published[29,30] with the caveat that the identified variant may be in linkage disequilibrium with the true pathogenic variant in that family. Statistical modeling takes into account age-related penetrance and phenocopy rates, with advanced methods also incorporating in silico predictions and co-occurrence with a known pathogenic variant into a single quantitative measure of pathogenicity.[31] Distant relatives are important to include because they are less likely to have both the disease and the variant by chance than members within a nuclear family. Full gene sequencing (including entire introns and 5′ and 3′ untranslated regions) may provide greater evidence that another variant is not involved or identify additional variants to consider as possibly causative. Unless the genetic locus is evaluated carefully, one risks misclassifying a nonpathogenic variant as pathogenic.

When a specific variant in the target gene segregates with a phenotype or disease in multiple affected family members and multiple families from diverse ethnic backgrounds, linkage disequilibrium and ascertainment bias are less likely to confound the evidence for pathogenicity. In this case, this criterion may be taken as moderate or strong evidence, depending on the extent of segregation, rather than supporting evidence.

On the other hand, lack of segregation of a variant with a phenotype provides strong evidence against pathogenicity. Careful clinical evaluation is needed to rule out mild symptoms of reportedly unaffected individuals, as well as possible phenocopies (affected individuals with disease due to a nongenetic or different genetic cause). Also, biological family relationships need to be confirmed to rule out adoption, nonpaternity,

sperm and egg donation, and other nonbiological relationships. Decreased and age-dependent penetrance also must be considered to ensure that asymptomatic family members are truly unaffected.

Statistical evaluation of cosegregation may be difficult in the clinical laboratory setting. If appropriate families are identified, clinical laboratories are encouraged to work with experts in statistical or population genetics to ensure proper modeling and to avoid incorrect conclusions of the relevance of the variant to the disease.

PP2 BP1 variant spectrum

Many genes have a defined spectrum of pathogenic and benign variation. For genes in which missense variation is a common cause of disease and there is very little benign variation in the gene, a novel missense variant can be considered supporting evidence for pathogenicity (PP2). By contrast, for genes in which truncating variants are the only known mechanism of variant pathogenicity, missense variants can be considered supporting evidence for a benign impact (BP1). For example, truncating variants in *ASPM* are the primary type of pathogenic variant in this gene, which causes autosomal recessive primary microcephaly, and the gene has a high rate of missense polymorphic variants. Therefore missense variants in *ASPM* can be considered to have this line of supporting evidence for a benign impact.

PP3 BP4 computational (in silico) data

Not overestimating computational evidence is important, particularly given that different algorithms may rely on the same (or similar) data to support predictions and most algorithms have not been validated against well-established pathogenic variants. In addition, algorithms can have vastly different predictive capabilities for different genes. If all of the in silico programs tested agree on the prediction, then this evidence can be counted as supporting. If in silico predictions disagree, however, then this evidence should not be used in classifying a variant. The variant amino acid change being present in multiple nonhuman mammalian species in an otherwise well-conserved region, suggesting the amino acid change would not compromise function, can be considered strong evidence for a benign interpretation. One must, however, be cautious about assuming a benign impact in a nonconserved region if the gene has recently evolved in humans (e.g., genes involved in immune function).

PP4 using phenotype to support variant claims

In general, the fact that a patient has a phenotype that matches the known spectrum of clinical features for a gene is not considered evidence for pathogenicity given that nearly all patients undergoing disease-targeted tests have the phenotype in question. If the following criteria are met, however, the patient's phenotype can be considered supporting evidence: (i) the clinical sensitivity of testing is high, with most patients testing positive for a pathogenic variant in that gene; (ii) the patient has a well-defined syndrome with little overlap with other clinical presentations (e.g., Gorlin syndrome including basal cell carcinoma, palmoplantar pits, odontogenic keratocysts); (iii) the gene is not subject to substantial benign variation, which can be determined through large general population cohorts (e.g., Exome Sequencing Project); and (iv) family history is consistent with the mode of inheritance of the disorder.

PP5 BP6 reputable source

There are increasing examples where pathogenicity classifications from a reputable source (e.g., a clinical laboratory with long-standing expertise in the disease area) have been shared in databases, yet the evidence that formed the basis for classification was not provided and may not be easily obtainable. In this case, the classification, if recently submitted, can be used as a single piece of supporting evidence. However, laboratories are encouraged to share the basis for classification as well as communicate with submitters to enable the underlying evidence to be evaluated and built upon. If the evidence is available, this criterion should not be used; instead, the criteria relevant to the evidence should be used.

BP5 alternate locus observations

When a variant is observed in a case with a clear alternate genetic cause of disease, this is generally considered supporting evidence to classify the variant as benign. However, there are exceptions. An individual can be a carrier of an unrelated pathogenic variant for a recessive disorder; therefore, this evidence is much stronger support for a likely benign variant classification in a gene for a dominant disorder compared with a gene for a recessive disorder. In addition, there are disorders in which having multiple variants can contribute to more severe disease. For example, two variants, one pathogenic and one novel, are identified in a patient with a severe presentation of a dominant disease. A parent also has mild disease. In this case, one must consider the possibility that the novel variant could also be pathogenic and contributing to the increased severity of disease in the proband. In this clinical scenario, observing the novel variant as the second variant would not support a benign classification of the novel variant (though it is also not considered support for a pathogenic classification without further evidence). Finally, there are certain diseases in which multigenic inheritance is known to occur, such as Bardet-Beidel syndrome, in which case the additional variant in the second locus may also be pathogenic but should be reported with caution.

BP7 synonymous variants

There is increasing recognition that splicing defects, beyond disruption of the splice consensus sequence, can be an important mechanism of pathogenicity, particularly for genes in which loss of function is a common mechanism of disease. Therefore, one should be cautious in assuming that a synonymous nucleotide change will have no effect. However, if the nucleotide position is not conserved over evolution and splicing assessment algorithms predict neither an impact to a splice consensus sequence nor the creation of a new alternate splice consensus sequence, then a splicing impact is less likely. Therefore, if

supported by computational evidence (BP4), one can classify novel synonymous variants as likely benign. However, if computational evidence suggests a possible impact on splicing or there is raised suspicion for an impact (e.g., the variant occurs in *trans* with a known pathogenic variant in a gene for a recessive disorder), then the variant should be classified as uncertain significance until a functional evaluation can provide a more definitive assessment of impact or other evidence is provided to rule out a pathogenic role.

REPORTING SEQUENCE VARIANTS

Writing succinct yet informative clinical reports can be a challenge as the complexity of the content grows from reporting variants in single genes to multigene panels to exomes and genomes. Several guidance documents have been developed for reporting, including full sample reports of the ACMG clinical laboratory standards for next-generation sequencing guidance.[32–35] Clinical reports are the final product of laboratory testing and often are integrated into a patient's electronic health record. Therefore, effective reports are concise, yet easy to understand. Reports should be written in clear language that avoids medical genetics jargon or defines such terms when used. The report should contain all of the essential elements of the test performed, including structured results, an interpretation, references, methodology, and appropriate disclaimers. These essential elements of the report also are emphasized by Clinical Laboratory Improvement Amendments regulations and the College of American Pathologists laboratory standards for next-generation sequencing clinical tests.[36]

Results

The results section should list variants using HGVS nomenclature (see Nomenclature). Given the increasing number of variants found in genetic tests, presenting the variants in tabular form with essential components may best convey the information. These components include nomenclature at both the nucleotide (genomic and complementary DNA) and protein level, gene name, disease, inheritance, exon, zygosity, and variant classification. An example of a table to report structured elements of a variant is found in the **Supplementary Appendix S1** online. Parental origin may also be included if known. In addition, if specific variants are analyzed in a genotyping test, the laboratory should specifically note the variants interrogated, with their full description and historical nomenclature if it exists. Furthermore, when reporting results from exome or genome sequencing, or occasionally very large disease-targeted panels, grouping variants into categories such as "Variants in Disease Genes with an Established Association with the Reported Phenotype," "Variants in Disease Genes with a Likely Association with the Reported Phenotype," and (where appropriate) "Incidental (Secondary) Findings" may be beneficial.

Interpretation

The interpretation should contain the evidence supporting the variant classification, including its predicted effect on the resultant protein and whether any variants identified are likely to fully or partially explain the patient's indication for testing. The report also should include any recommendations to the clinician for supplemental clinical testing, such as enzymatic/functional testing of the patient's cells and variant testing of family members, to further inform variant interpretation. The interpretation section should address all variants described in the results section but may contain additional information. It should be noted whether the variant has been reported previously in the literature or in disease or control databases. The references, if any, that contributed to the classification should be cited where discussed and listed at the end of the report. The additional information described in the interpretation section may include a summarized conclusion of the results of in silico analyses and evolutionary conservation analyses. However, individual computational predictions (e.g., scores, terms such as "damaging") should be avoided given the high likelihood of misinterpretation by health-care providers who may be unfamiliar with the limitations of predictive algorithms (see In Silico Predictive Programs, above). A discussion of decreased penetrance and variable expressivity of the disorder, if relevant, should be included in the final report. Examples of how to describe evidence for variant classification on clinical reports are found in the **Supplementary Appendix S1** online.

Methodology

The methods and types of variants detected by the assay and those refractory to detection should be provided in the report. Limitations of the assay used to detect the variants also should be reported. Methods should include those used to obtain nucleic acids (e.g., polymerase chain reaction, capture, whole-genome amplification), as well as those to analyze the nucleic acids (e.g., bidirectional Sanger sequencing, next-generation sequencing, chromosomal microarray, genotyping technologies), because this may provide the health-care provider with the necessary information to decide whether additional testing is required to follow up on the results. The methodology section should also give the official gene names approved by the Human Genome Organization Gene Nomenclature Committee, RefSeq accession numbers for transcripts, and genome build, including versions. For large panels, gene-level information may be posted and referenced by URL. The laboratory may choose to add a disclaimer that addresses general pitfalls in laboratory testing, such as sample quality and sample mix-up.

Access to patient advocacy groups, clinical trials, and research

Although specific clinical guidance for a patient is not recommended for laboratory reports, provision of general information for categories of results (e.g., all positives) is appropriate and helpful. A large number of patient advocacy groups and clinical trials are now available for support and treatment of many diseases. Laboratories may choose to add this information to the body of the report or attach the information so it is sent to the health-care provider along with the report.

Laboratories may make an effort to connect the health-care provider to research groups working on specific diseases when a variant's effect is classified as "uncertain," as long as Health Insurance Portability and Accountability Act patient privacy requirements are followed.

Variant reanalysis

As evidence on variants evolves, previous classifications may later require modification. For example, the availability of variant frequency data among large populations has led many uncertain significance variants to be reclassified as benign, and testing additional family members may result in the reclassification of variants.

As the content of sequencing tests expands and the number of variants identified grows, expanding to thousands and millions of variants from exome and genome sequencing, the ability for laboratories to update reports as variant knowledge changes will be untenable without appropriate mechanisms and resources to sustain those updates. To set appropriate expectations with health-care providers and patients, laboratories should provide clear policies on the reanalysis of data from genetic testing and whether additional charges for reanalysis may apply. Laboratories are encouraged to explore innovative approaches to give patients and providers more efficient access to updated information.[37,38]

For reports containing variants of uncertain significance in genes related to the primary indication, and in the absence of updates that may be proactively provided by the laboratory, it is recommended that laboratories suggest periodic inquiry by health-care providers to determine whether knowledge of any variants of uncertain significance, including variants reported as likely pathogenic, has changed. By contrast, laboratories are encouraged to consider proactive amendment of cases when a variant reported with a near-definitive classification (pathogenic or benign) must be reclassified. Regarding physician responsibility, see the ACMG guidelines on the duty to recontact.[39]

Confirmation of findings

Recommendations for the confirmation of reported variants is addressed elsewhere.[35,36] Except as noted, confirmation studies using an orthogonal method are recommended for all sequence variants that are considered to be pathogenic or likely pathogenic for a Mendelian disorder. These methods may include, but are not limited to, re-extraction of the sample and testing, testing of parents, restriction enzyme digestion, sequencing the area of interest a second time, or using an alternate genotyping technology.

SPECIAL CONSIDERATIONS

Evaluating and reporting variants in GUS based on the indication for testing

Genome and exome sequencing are identifying new genotype–phenotype connections. When the laboratory finds a variant in a gene without a validated association to the patient's phenotype, it is a GUS. This can occur when a gene has never been associated with any patient phenotype or when the gene has been associated with a different phenotype from that under consideration. Special care must be taken when applying the recommended guidelines to a GUS. In such situations, utilizing variant classification rules developed for recognized genotype–phenotype associations is not appropriate. For example, when looking across the exome or genome, a de novo observation is no longer strong evidence for pathogenicity given that all individuals are expected to have approximately one de novo variant in their exome or 100 in their genome. Likewise, thousands of variants across a genome could segregate with a significant logarithm of the odds (LOD) score. Furthermore, many deleterious variants that are clearly disruptive to a gene or its resultant protein (nonsense, frameshift, canonical ±1,2 splice site, exon-level deletion) may be detected; however, this is insufficient evidence for a causative role in any given disease presentation.

Variants found in a GUS may be considered as candidates and reported as "variants in a gene of uncertain significance." These variants, if reported, should always be classified as uncertain significance. Additional evidence would be required to support the gene's association to disease before any variant in the gene itself can be considered pathogenic for that disease.[5] For example, additional cases with matching rare phenotypes and deleterious variants in the same gene would enable the individual variants to be classified according to the recommendations presented here.

Evaluating variants in healthy individuals or as incidental findings

Caution must be exercised when using these guidelines to evaluate variants in healthy or asymptomatic individuals or to interpret incidental findings unrelated to the primary indication for testing. In these cases the likelihood of any identified variant being pathogenic may be far less than when performing disease-targeted testing. As such, the required evidence to call a variant pathogenic should be higher, and extra caution should be exercised. In addition, the predicted penetrance of pathogenic variants found in the absence of a phenotype or family history may be far less than predicted based on historical data from patients ascertained as having disease.

Mitochondrial variants

The interpretation of mitochondrial variants other than well-established pathogenic variants is complex and remains challenging; several special considerations are addressed here.

The nomenclature differs from standard nomenclature for nuclear genes, using gene name and m. numbering (e.g., m.8993T>C) and p. numbering, but not the standard c. numbering (see also Nomenclature). The current accepted reference sequence is the Revised Cambridge Reference Sequence of the Human Mitochondrial DNA: GenBank sequence NC_012920 gi:251831106.[40,41]

Heteroplasmy or homoplasmy should be reported, along with an estimate of heteroplasmy of the variant if the test has

been validated to determine heteroplasmy levels. Heteroplasmy percentages in different tissue types may vary from the sample tested; therefore, low heteroplasmic levels also must be interpreted in the context of the tissue tested, and they may be meaningful only in the affected tissue such as muscle. Over 275 mitochondrial DNA variants relating to disease have been recorded (http://mitomap.org/bin/view.pl/MITOMAP/WebHome).[42] MitoMap is considered the main source of information related to mitochondrial variants as well as haplotypes. Other resources, such as frequency information (http://www.mtdb.igp.uu.se/),[43] secondary structures, sequences, and alignment of mitochondrial transfer RNAs (http://mamit-trna.u-strasbg.fr/),[44] mitochondrial haplogroups (http://www.phylotree.org/)[45]and other information (http://www.mtdna-community.org/default.aspx),[46] may prove useful in interpreting mitochondrial variants.

Given the difficulty in assessing mitochondrial variants, a separate evidence checklist has not been included. However, any evidence needs to be applied with additional caution (for a review, see ref. [47]). The genes in the mitochondrial genome encode for transfer RNA as well as for protein; therefore, evaluating amino acid changes is relevant only for genes encoding proteins. Similarly, because many mitochondrial variants are missense variants, evidence criteria for truncating variants likely will not be helpful. Because truncating variants do not fit the known variant spectrum in most mitochondrial genes, their significance may be uncertain. Although mitochondrial variants are typically maternally inherited, they can be sporadic, yet de novo variants are difficult to assess because of heteroplasmy that may be below an assay's detection level or different between tissues. The level of heteroplasmy may contribute to the variable expression and reduced penetrance that occurs within families. Nevertheless, there remains a lack of correlation between the percentage of heteroplasmy and disease severity.[47] Muscle, liver, or urine may be additional specimen types useful for clinical evaluation. Undetected heteroplasmy may also affect outcomes of case, case–control, and familial concordance studies. In addition, functional studies are not readily available, although evaluating muscle morphology may be helpful (i.e., the presence of ragged red fibers). Frequency data and published studies demonstrating causality may often be the only assessable criteria on the checklist. An additional tool for mitochondrial diseases may be haplogroup analysis, but this may not represent a routine method that clinical laboratories have used, and the clinical correlation is not easy to interpret.

Consideration should be given to testing nuclear genes associated with mitochondrial disorders because variants in nuclear genes could be causative of oxidative disorders or modulating the mitochondrial variants.

Pharmacogenomics

Establishing the effects of variants in genes involved with drug metabolism is challenging, in part because a phenotype is only apparent upon exposure to a drug. Still, variants in genes related to drug efficacy and risk for adverse events have been described and are increasingly used in clinical care. Gene summaries and clinically relevant variants can be found in the Pharmacogenomics Knowledge Base (http://www.pharmgkb.org/).[48] Alleles and nomenclature for the cytochrome P450 gene family is available at http://www.cypalleles.ki.se/.[49] Although the interpretation of PGx variants is beyond the scope of this document, we include a discussion of the challenges and distinctions associated with the interpretation and reporting of PGx results.

The traditional nomenclature of PGx alleles uses star (*) alleles, which often represent haplotypes, or a combination of variants on the same allele. Traditional nucleotide numbering using outdated reference sequences is still being applied. Converting traditional nomenclature to standardized nomenclature using current reference sequences is an arduous task, but it is necessary for informatics applications with next-generation sequencing.

Many types of variants have been identified in PGx genes, such as truncating, missense, deletions, duplications (of functional as well as nonfunctional alleles), and gene conversions, resulting in functional, partially functional (decreased or reduced function), and nonfunctional (null) alleles. Interpreting sequence variants often requires determining haplotype from a combination of variants detected. Haplotypes are typically presumed based on population frequencies and known variant associations rather than testing directly for chromosomal phase (molecular haplotyping).

In addition, for many PGx genes (particularly variants in genes coding for enzymes), the overall phenotype is derived from a diplotype, which is the combination of variants or haplotypes on both alleles. Because PGx variants do not directly cause disease, using terms related to metabolism (rapid, intermediate, poor); efficacy (resistant, responsive, sensitive); or "risk," rather than pathogenic, may be more appropriate. Further nomenclature and interpretation guidelines are needed to establish consistency in this field.

Common complex disorders

Unlike Mendelian diseases, the identification of common, complex disease genes, such as those contributing to type 2 diabetes, coronary artery disease, and hypertension, has largely relied on population-based approaches (e.g., genome-wide association studies) rather than family-based studies.[50,51] Currently, numerous genome-wide association study reports have resulted in the cataloguing of over 1,200 risk alleles for common, complex diseases and traits. Most of these variants are in nongenic regions, however, and additional studies are required to determine whether any of the variants are directly causal through effects on regulatory elements, for example, or are in linkage disequilibrium with causal variants.[52]

Common, complex risk alleles typically confer low relative risk and are meager in their predictive power.[53] To date, the utility of common, complex risk allele testing for patient care[54] has been unclear, and models to combine multiple markers into a cumulative risk score often are flawed and are usually no better than traditional risk factors such as family

history, demographics, and nongenetic clinical phenotypes.[55,56] Moreover, in almost all of the common diseases the risk alleles can explain only up to 10% of the variance in the population, even when the disease has high heritability. Given the complexity of issues, this recommendation does not address the interpretation and reporting of complex trait alleles. We recognize, however, that some of these alleles are identified during the course of sequencing Mendelian genes, and therefore guidance on how to report such alleles when found incidentally is needed. The terms "pathogenic" and "likely pathogenic" are not appropriate in this context, even when the association is statistically valid. Until better guidance is developed, an interim solution is to report these variants as "risk alleles" or under a separate "other reportable" category in the diagnostic report. The evidence for the risk, as identified in the case–control/genome-wide association studies, can be expressed by modifying the terms, such as "established risk allele," "likely risk allele," or "uncertain risk allele," if desired.

Somatic variants

The description of somatic variants, primarily those observed in cancer cells, includes complexities not encountered with constitutional variants, because the allele ratios are highly variable and tumor heterogeneity can cause sampling differences. Interpretation helps select therapy and predicts treatment response or the prognosis of overall survival or tumor progression–free survival, further complicating variant classification. For the interpretation of negative results, understanding the limit of detection of the sequencing assay (at what allele frequency the variant can be detected by the assay) is important and requires specific knowledge of the tumor content of the sample. Variant classification categories are also different, with somatic variants compared with germ-line variants, with terms such as "responsive," "resistant," "driver," and "passenger" often used. Whether a variant is truly somatic is confirmed by sequence analysis of the patient's germ-line DNA. A different set of interpretation guidelines is needed for somatic variants, with tumor-specific databases used for reference, in addition to databases used for constitutional findings. To address this, a workgroup has recently been formed by the AMP.

HOW SHOULD HEALTH-CARE PROVIDERS USE THESE GUIDELINES AND RECOMMENDATIONS?

The primary purpose of clinical laboratory testing is to support medical decision making. In the clinic, genetic testing is generally used to identify or confirm the cause of disease and to help the health-care provider make individualized treatment decisions including the choice of medication. Given the complexity of genetic testing, results are best realized when the referring health-care provider and the clinical laboratory work collaboratively in the testing process.

When a health-care provider orders genetic testing, the patient's clinical information is integral to the laboratory's analysis. As health-care providers increasingly utilize genomic (exome or genome) sequencing, the need for detailed clinical information to aid in interpretation assumes increasing importance. For example, when a laboratory finds a rare or novel variant in a genomic sequencing sample, the director cannot assume it is relevant to a patient just because it is rare, novel, or de novo. The laboratory must evaluate the variant and the gene in the context of the patient's and family's history, physical examinations, and previous laboratory tests to distinguish between variants that cause the patient's disorder and those that are incidental (secondary) findings or benign. Indeed, accurate and complete clinical information is so essential for the interpretation of genome-level DNA sequence findings that the laboratory can reasonably refuse to proceed with the testing if such information is not provided with the test sample.

For tests that cover a broad range of phenotypes (large panels, exome and genome sequencing) the laboratory may find candidate causative variants. Further follow-up with the health-care provider and patient may uncover additional evidence to support a variant. These additional phenotypes may be subclinical, requiring additional clinical evaluation to detect (e.g., temporal bone abnormalities detected by computed tomography in a hearing-impaired patient with an uncertain variant in *SLC26A4*, the gene associated with Pendred syndrome). In addition, testing other family members to establish when a variant is de novo, when a variant cosegregates with disease in the family, and when a variant is in *trans* with a pathogenic variant in the same recessive disease-causing gene is valuable. Filtering out or discounting the vast majority of variants for dominant diseases when they can be observed in healthy relatives is possible, making the interpretation much more efficient and conclusive. To this end, it is strongly recommended that every effort be made to include parental samples along with that of the proband, so-called "trio" testing (mother, father, affected child), in the setting of exome and genome sequencing, particularly for suspected recessive or de novo causes. Obviously this will be easier to achieve for pediatric patients than for affected adults. In the absence of one or both parents, the inclusion of affected and unaffected siblings can be of value.

Many genetic variants can result in a range of phenotypic expression (variable expressivity), and the chance of disease developing may not be 100% (reduced penetrance), further underscoring the importance of providing comprehensive clinical data to the clinical laboratory to aid in variant interpretation. Ideally, it is recommended that clinical data be deposited into, and shared via, centralized repositories as allowable by Health Insurance Portability and Accountability Act and institutional review board regulations. Importantly, referring health-care providers can further assist clinical laboratories by recruiting DNA from family members in scenarios where their participation will be required to interpret results, (e.g., when evaluating cosegregation with disease using affected family members, genotyping parents to assess for de novo occurrence and determining the phase of variants in recessive disorders using first-degree relatives).

A key issue for health-care providers is how to use the evidence provided by genetic testing in medical management decisions. Variant analysis is, at present, imperfect, and the variant category reported does not imply 100% certainty. In general, a variant classified as pathogenic using the proposed classification scheme has met criteria informed by empirical data such that a health-care provider can use the molecular testing information in clinical decision making. Efforts should be made to avoid using this as the sole evidence of Mendelian disease; it should be used in conjunction with other clinical information when possible. Typically, a variant classified as likely pathogenic has sufficient evidence that a health-care provider can use the molecular testing information in clinical decision making when combined with other evidence of the disease in question. For example, in the prenatal setting an ultrasound may show a key confirmatory finding; in postnatal cases, other data such as enzyme assays, physical findings, or imaging studies may conclusively support decision making. However, it is recommended that all possible follow-up testing, as described above, be pursued to generate additional evidence related to a likely pathogenic variant because this may permit the variant to be reclassified as pathogenic. A variant of uncertain significance should not be used in clinical decision making. Efforts to resolve the classification of the variant as pathogenic or benign should be undertaken. While this effort to reclassify the variant is underway, additional monitoring of the patient for the disorder in question may be prudent. A variant considered likely benign has sufficient evidence that a health-care provider can conclude that it is not the cause of the patient's disorder when combined with other information, for example, if the variant does not segregate in an affected family member and complex inheritance patterns are unlikely. A variant considered benign has sufficient evidence that a health-care provider can conclude that it is not the cause of the patient's disorder.

How the genetic testing evidence is used is also dependent on the clinical context and indication for testing. In a prenatal diagnostic case where a family is considering irrevocable decisions such as fetal treatment or pregnancy termination, the weight of evidence from the report and other sources such as fetal ultrasound needs to be considered before action is taken. When a genetic test result is the *only* evidence in a prenatal setting, variants considered likely pathogenic must be explained carefully to families. It is therefore critical for referring health-care providers to communicate with the clinical laboratory to gain an understanding of how variants are classified to assist in patient counseling and management.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

ACKNOWLEDGMENTS

The workgroup thanks our many colleagues in the genetics community who helped to develop this "consensus" guideline through their responses to surveys, participation at workshops, insightful discussions, and constructive comments. H.L.R., S.B., S.D., M.H., and E.L. were supported in part by National Institutes of Health HG006834. H.L.R. was supported in part by National Institutes of Health grants HG006500 and HD077671. These Standards and Guidelines were approved by the ACMG Board of Directors on 15 December 2014 and the AMP Board of Directors on 9 January 2015.

DISCLOSURE

All workgroup members are clinical service providers. The following workgroup members have a commercial conflict of interest: S.B. (GeneDx, BioReference (stock), advisory boards for RainDance, Ingenuity); M.H. (advisor for Oxford Genetic Technologies, Tessarae, Ingenuity/Qiagen); E.L. (advisory board for Complete Genomics); and H.L.R. (scientific advisory boards for Ingenuity/Qiagen, Complete Genomics, Knome, Focused Genomics). The other authors declare no conflict of interest.

REFERENCES

1. Richards CS, Bale S, Bellissimo DB, et al.; Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 2008;10:294–300.
2. Plon SE, Eccles DM, Easton D, et al.; IARC Unclassified Genetic Variants Working Group. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 2008;29:1282–1291.
3. Sosnay PR, Siklosi KR, Van Goor F, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet* 2013;45:1160–1167.
4. Thompson BA, Spurdle AB, Plazzer JP, et al.; InSiGHT. 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* 2014;46:107–115.
5. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014;508:469–476.
6. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST; Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med* 2011;13:680–685.
7. Nomenclature for the description of sequence variants: Human Genome Variation Society. https://www.hgvs.org/mutnomen. Accessed 4 November 2014.
8. Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat* 2008;29:6–13.
9. RefSeq: NCBI Reference Sequence Database. https://www.ncbi.nlm.nih.gov/refseq. Accessed 4 November 2014.
10. Locus Reference Genomic. https://www.lrg-sequence.org. Accessed 4 November 2014.
11. Concensus CDS Database NCBI. https://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi. Accessed 4 November 2014.
12. Taschner PE, den Dunnen JT. Describing structural changes by extending HGVS sequence variation nomenclature. *Hum Mutat* 2011;32:507–511.
13. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–315.
14. Hicks S, Wheeler DA, Plon SE, Kimmel M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat* 2011;32:661–668.
15. Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB; IARC Unclassified Genetic Variants Working Group. In silico analysis of missense substitutions using sequence-alignment based methods. *Hum Mutat* 2008;29:1327–1336.
16. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* 2011;32:358–368.

17. Thompson BA, Greenblatt MS, Vallee MP, et al. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum Mutat* 2013;34:255–265.
18. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012;7:e46688.
19. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–249.
20. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073–1081.
21. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010;7:575–576.
22. Jian X, Boerwinkle E, Liu X. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med* 2014;16:497–503.
23. Houdayer C, Caux-Moncoutier V, Krieger S, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 2012;33:1228–1238.
24. Vreeswijk MP, Kraan JN, van der Klift HM, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum Mutat* 2009;30:107–114.
25. Henderson LB, Applegate CD, Wohler E, Sheridan MB, Hoover-Fong J, Batista DA. The impact of chromosomal microarray on clinical management: a retrospective analysis. *Genet Med* 2014;16:657–664.
26. Popp MW, Maquat LE. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet* 2013;47:139–165.
27. Calculator for confidence intervals of odds ratio in an unmatched case control study. For example groups of cases and controls studied to assess a treatment or exposure to a suspected causal factor. http://www.hutchon.net/confidor.htm. Accessed 4 November 2014.
28. Odds Ratio Confidence Interval Calculator. http://www.easycalculation.com/statistics/odds-ratio. Accessed 4 November 2014.
29. Bayrak-Toydemir P, McDonald J, Mao R, et al. Likelihood ratios to assess genetic evidence for clinical significance of uncertain variants: hereditary hemorrhagic telangiectasia as a model. *Exp Mol Pathol* 2008;85:45–49.
30. Thompson D, Easton DF, Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet* 2003;73:652–655.
31. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ; Breast Cancer Information Core (BIC) Steering Committee. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet* 2004;75:535–544.
32. Scheuner MT, Hilborne L, Brown J, Lubin IM; members of the RAND Molecular Genetic Test Report Advisory Board. A report template for molecular genetic tests designed to improve communication between the clinician and laboratory. *Genet Test Mol Biomarkers* 2012;16:761–769.
33. Lubin IM, Caggana M, Constantin C, et al. Ordering molecular genetic tests and reporting results: practices in laboratory and clinical settings. *J Mol Diagn* 2008;10:459–468.
34. Lubin IM, McGovern MM, Gibson Z, et al. Clinician perspectives about molecular genetic testing for heritable conditions and development of a clinician-friendly laboratory report. *J Mol Diagn* 2009;11:162–171.
35. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al.; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013;15:733–747.
36. Aziz N, Zhao Q, Bry L, et al. College of American Pathologists' Laboratory Standards for Next-Generation Sequencing Clinical Tests. *Arch Pathol Lab Med* 2014. doi:10.5858/arpa.2014-0250-CP.
37. Aronson SJ, Clark EH, Varugheese M, Baxter S, Babb LJ, Rehm HL. Communicating new knowledge on previously reported genetic variants. *Genet Med* 2012;14:713–719.
38. Bean LJ, Tinker SW, da Silva C, Hegde MR. Free the data: one laboratory's approach to knowledge-based genomic variant classification and preparation for EMR integration of genomic data. *Hum Mutat* 2013;34:1183–1188.
39. Hirschhorn K, Fleisher LD, Godmilow L, et al. Duty to re-contact. *Genet Med* 1999;1:171–172.
40. Revised Cambridge Reference Sequence (rCRS) of the Human Mitochondrial DNA: GenBank sequence NC_012920 gi:251831106. http://www.mitomap.org/MITOMAP/HumanMitoSeq.
41. Behar DM, van Oven M, Rosset S, et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* 2012;90:675–684.
42. MITOMAP: A human mitochondrial genome database. http://mitomap.org/bin/view.pl/MITOMAP/WebHome. Accessed 4 November 2014.
43. mtDB-Human Mitochondrial Genome Database. http://www.mtdb.igp.uu.se. Accessed 4 November 2014.
44. Compilation of mammalian mitochondrial tRNA genes. http://mamit-trna.u-strasbg.fr/Summary.asp. Accessed 4 November 2014.
45. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009;30:E386–E394.
46. mtDNA Community. http://www.mtdnacommunity.org/default.aspx/. Accessed 4 November 2014.
47. Wong LJ. Diagnostic challenges of mitochondrial DNA disorders. *Mitochondrion* 2007;7:45–52.
48. PharmGKB. http://www.pharmgkb.org. Accessed 4 November 2014.
49. The Human Cytochrome P450 (CYP) Allele Nomenclature Database. http://www.cypalleles.ki.se/. Accessed 4 November 2014.
50. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;363:166–176.
51. Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009;360:1759–1768.
52. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999;22:139–144.
53. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010;141:210–217.
54. Topol EJ, Damani SB. The KIF6 collapse. *J Am Coll Cardiol* 2010;56:1564–1566.
55. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008;359:2208–2219.
56. Paynter NP, Chasman DI, Paré G, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 2010;303:631–637.
57. Kerem E, Corey M, Kerem BS, et al. The relation between genotype and phenotype in cystic fibrosis–analysis of the most common mutation (delta F508). *N Engl J Med*. 1990 Nov 29;323(22):1517–22.
58. Kobayashi K, Knowles MR, Boucher RC, et al. Benign missense variations in the cystic fibrosis gene. *Am J Hum Genet*. 1990 Oct;47(4):611–615.
59. Zelante L, Gasparini P, Estivill X, et al. Connexin26 mutations associated with the most common form of non-syndromic neurosensory autosomal recessive deafness (DFNB1) in Mediterraneans. *Hum Mol Genet*. 1997 Sep;6(9):1605–9.