HHS Public Access

Author manuscript

Genet Med. Author manuscript; available in PMC 2015 October 01.

Published in final edited form as:

Genet Med. 2014 October ; 16(10): 741–750. doi:10.1038/gim.2014.29.

The implications of familial incidental findings from exome sequencing: The NIH Undiagnosed Diseases Program experience

Lauren Lawrence[1,2,*], **Murat Sincan**[1,*], **Thomas Markello**[1], **David R Adams**[1], **Fred Gill**[3], **Rena Godfrey**[1], **Gretchen Golas**[1], **Catherine Groden**[1], **Dennis Landis**[1], **Michele Nehrebecky**[1], **Grace Park**[3], **Ariane Soldatos**[1], **Cynthia Tifft**[1], **Camilo Toro**[1], **Colleen Wahl**[1], **Lynne Wolfe**[1], **William A. Gahl**[1], and **Cornelius F. Boerkoel**[1]

[1]NIH Undiagnosed Diseases Program, Common Fund, NIH Office of the Director and NHGRI, Bethesda, MD, USA

[2]University of Southern California Keck School of Medicine, Los Angeles, CA, USA

[3]Internal Medicine Consult Service, NIH Clinical Center, Bethesda, MD, USA

Abstract

Purpose—Using exome sequence data from 159 families participating in the NIH Undiagnosed Diseases Program, we evaluated the number and inheritance of reportable incidental sequence variants.

Methods—Following the ACMG recommendations for reporting of incidental next generation sequencing findings, we extracted variants in 56 genes from the exome sequence data of 543 subjects and determined the reportable incidental findings for each participant. We also defined variant status as inherited or *de novo* for those with available parental sequence data.

Results—We identified 14 independent reportable variants in 159 (8.8%) families. For 9 families with parental sequence data in our cohort, a parent transmitted the variant to one or more children (9 minor children and 4 adult children). The remaining 5 variants occurred in adults for whom parental sequences were unavailable.

Conclusion—Our results are consistent with the expectation that a small percentage of exomes will result in identification of an incidental finding under the ACMG recommendations. Additionally, our analysis of family sequence data highlights that genome and exome sequencing of families has unavoidable implications for immediate family members and therefore requires appropriate counseling of the family.

Correspondence: Murat Sincan, MD, National Institutes of Health, NHGRI, UDP Translational Laboratory, 5625 Fishers Lane, Room 4N-15, Rockville, MD 20852, Tel: 301-594-5182, Fax: 301-480-0804, murat.sincan@nih.gov.

*These authors contributed equally

Keywords

incidental findings; NIH Undiagnosed Diseases Program; exome sequencing; familial; secondary variants

Introduction

'Incidental findings' are defined as genetic variants with medical or social implications that are discovered during genetic testing for an unrelated indication.[1] Based on recent publications,[2] the ACMG Working Group on Incidental Findings in Clinical Exome and Genome Sequencing determined that looking for and reporting some incidental findings would likely have medical benefit for patients and their families. The group therefore recommended, reporting incidental findings from a "minimum list" of 56 genes for individuals having clinical exome or genome sequencing.[3] This recommendation has been widely debated and openly challenged.[4]

Although the return of incidental findings represents an important step forward in the use of sequencing for medical benefit,[5] implementing these recommendations requires the development of infrastructure to support evaluation and reporting.[3] Family members other than the proband are often included in diagnostic exome sequencing, and thus this also has implications for unaffected family members. The typical number of reportable variants that will be generated in practice has not been widely studied. One study of 572 subjects, selected for atherosclerosis phenotypes, found that approximately 1% of exomes may require disclosure of an incidental genetic finding, but the set of genes analyzed in that study did not include all the genes in the ACMG list, and the cohort was non-familial.[2] A more recent study found ~3.4% of European ancestry exomes and 1.2% of African ancestry exomes in the National Heart, Lung, and Blood Institute Exome Sequencing Project bear actionable pathogenic or likely pathogenic incidental findings in 114 genes.[6] More data are needed to assess the possible impact of the ACMG recommendations in a variety of clinical settings. This is an important issue because resources are required to implement the recommendations.

We analyzed research exome sequence data from 543 individuals derived from 159 families. For the recommended 56 genes, this analysis identified 14 independent reportable variants in the exome sequence data of 27 participants. In 9 families with parental sequence data, a parent transmitted the variant to one or more children. These analyses provide data that may be used to refine strategies for the reporting of incidental findings.

Materials and Methods

Subject Cohort

Family members gave informed consent or assent to protocol 76-HG-0238, "Diagnosis and Treatment of Patients with Inborn Errors of Metabolism and Other Genetic Disorders," approved by the NHGRI Institutional Review Board. The exome sequence data were derived from a 159-family cohort consisting of 543 subjects with 188 affected subjects, 137 siblings and 218 parents. The average and median age of the 543 subjects at time of sequencing was

34.0 (standard deviation 20.8) and 37 years, respectively. Some subjects were deceased at the time of sequencing, and for those subjects, projected age at time of sequencing was used, since it is anticipated that incidental findings will only be sought in living subjects. Self-reported ancestry was White/European (89.1%), Black/African American (4.1%), Unknown (3.3%), Asian (2.2%) and Multiracial (1.3%) (Supplementary Table 1). These families included all those admitted to the NIH Undiagnosed Diseases Program and selected for exome analysis as previously described.[7] The sequencing was performed on a research basis, not in a CLIA-certified fashion.

Exome Sequencing

Genomic DNA was extracted from peripheral whole blood using the Gentra Puregene Blood kit (Qiagen) per the manufacturer's protocol. The Illumina TruSeq exome capture kit (Illumina, Inc., San Diego, US), which targets roughly 60 million bases consisting of the Consensus Coding Sequence (CCDS) annotated gene set as well as some structural RNAs, was used. Captured DNA was sequenced on the Illumina HiSeq platform until coverage was sufficient to call high quality genotypes at 85% or more of targeted bases.

Alignment and Genotype Calling

Reads were mapped to NCBI build 37 (hg19) using the Illumina ELAND aligner. When at least one read in a pair mapped to a unique location in the genome, that read and its pair were then aligned with Novoalign (Novocraft, Selangor, Malaysia). These alignments were stored in BAM format, and then fed as input to bam2mpg (http://research.nhgri.nih.gov/software/bam2mpg/index.shtml), which called genotypes using a Bayesian algorithm (Most Probable Genotype, or MPG).[8]

Coverage

Using the UCSC genome browser's hg19 human genome reference exon annotations for the 56 genes, we identified 1257 discrete exon regions including the UTRs. We recorded base-by-base coverage (Supplemental Table 2) and calculated the percent of each exon with coverage of 10, 20 or 30 fold (Supplemental Tables 3–5). We also summarized how many exons had at least 90% of their bases covered to at least each of these coverage thresholds (Table 1).

Annotations

The variants were annotated using Annovar.[9] Variants and genes listed in Human Gene Mutation Database (HGMD) Professional were added to the annotations. We also used annotations extracted from the supplemental data published by Johnston, et al.,[2] and added annotations for variants listed in ClinVar[10] and locus-specific databases (LSDB) registered in the Leiden Open Variation Database (LOVD).[11] For LSDBs not registered in LOVD, annotations were manually collected from the individual LSDBs and used to annotate the variants on the basis of matching Human Genome Variation Society (HGVS) nomenclature.

Data Extraction

Variants within the 56 genes recommended by the ACMG were considered if they had at least one minor allele call with a minimum coverage of 20 and a minimum mean probable genotype (mpg)/coverage ratio of 0.5.[12]

Data Analysis

The ACMG Recommendations state that "known pathogenic" variants in 56 genes (and "expected pathogenic" variants in a subset of those 56) should be reported to subjects sequenced for unrelated clinical reasons. The LSDBs and catalogs of clinically-relevant variants such as HGMD and ClinVar catalog variants identified in a gene together with annotations of each variant as "pathogenic," "probable pathogenic," "variant of unknown significance," "probable non-pathogenic," or "non-pathogenic" (or similar categories). Such annotations can serve as a foundation for determining whether a variant is "known pathogenic."

An accepted standard for determination of variant pathogenicity (with or without consultation of the databases described above) has not emerged, although several have been proposed.[13] Various methods have been proposed to evaluate the likelihood of pathogenicity for variants of unknown significance in genes associated with disease,[14–16] but we did not use them because they depend on data unavailable to us, i.e., defined penetrance[15,16] or population frequency and phenocopy rate.[14] Additionally, we did not use allele prevalence as supporting criteria because 1) the phenotyping of subjects included in the 1000 Genomes and ESP cohorts is incomplete,[17] 2) many of the disorders are of adult-onset and therefore might not be expressed fully among subjects in the 1000 Genomes and ESP cohorts,[17] 3) some disorders have environmentally-dependent expressivity (e.g., malignant hyperthermia susceptibility) and therefore might not be expressed fully among subjects in the 1000 Genomes and ESP cohorts,[17] and 4) large control cohorts (>10,000) are needed to properly evaluate case-control disparities for rare variants.[13]

Understanding that potential harm is posed both by false positive and false negative incidental findings and that variants discovered in sporadic cases may have a high false-positive rate,[18–20] we chose the following criteria for accepting variants as "known pathogenic": 1) designation in at least one variant database as "pathogenic" or "probable pathogenic" and supporting evidence such as experimental assays or segregation with disease or 2) meeting the criteria for "expected pathogenic" (see below) and a listing in at least one variant database as "pathogenic." This process required review of the literature and required approximately 320 man-hours from individuals knowledgeable of genetics, experimental methodology and medicine. Approximately 200 hours were spent intersecting LSDBs with our variant set and flagging variants for further review. The remaining approximate 120 hours were spent reviewing literature and splice predictions for individual variants under consideration for reporting.

Our minimum acceptable segregation patterns for autosomal dominant disorders were either a confirmed *de novo* variant in an affected child with two unaffected parents or segregation of the variant to three affected family members in two generations. We judged requiring five

informative meioses or positive evidence of linkage as unreasonably stringent criteria [21] and only requiring two affected family members in two generations as too lax a criterion for association of a variant with disease.[18,19] We did not accept clinically identified variants asserted to cause disease as pathogenic without reported functional data or familial segregation.

To define variants as "expected pathogenic" we used the criteria previously described.[22] Briefly, these include mutations leading to premature translation termination, loss of a translation termination codon, loss of a translation initiation codon, and alteration of canonical splice donor or acceptor sites.

Missense variants not previously associated with disease are considered a class of variant that may or may not cause disease and therefore are not automatically disclosed to a patient.[22] Furthermore, the lack of information regarding these variants in an LSDB, HGMD, or ClinVar indicates that they are unlikely to be recognized by the medical genetics community as known pathogenic variants. We therefore designated missense variants not present in these databases as non-reportable.

Both alleles of *MUTYH* must be mutated to meet ACMG reporting recommendations. We therefore selected homozygous non-reference variants and paired compound heterozygous variants. We deemed a variant pair reportable only if each variant of the pair met the criteria of being listed as "pathogenic" in at least one variant database and having supporting evidence such as experimental assays or segregation with disease.

To count the number of reportable incidental findings per independent exome, one subject per family was selected randomly and the number of incidental findings in those subjects was counted. We also counted the number of reportable incidental findings in subjects who are currently minors, and noted whether the disease associated with the variant in question was of adult-onset or childhood-onset.

Phenotype correlation

Family and medical history and pertinent laboratory findings were reviewed where available for individuals with a reportable variant.

Results

For the UDP cohort of 543 exome sequence data, there were 5948 variants in the 56 ACMG recommended genes (Figure 1; see Supplementary Table 2 for a complete list of all variants with annotations) when compared to the human reference sequence (NCBI build 37; hg19) (Table 2). To select variants of sufficient quality, we limited further analyses to those variants with a minimum coverage of 20 reads and a minimum mpg/coverage ratio of 0.5. Of the 5928 variants that remained, 4932 were judged highly unlikely to be reportable under ACMG recommendations because they were not present in LSDBs and localized to introns outside of the canonical spice sites (67%), resided in 3′ untranslated regions (UTR) (13%), encoded synonymous amino acid changes (7.5%), or resided in other non protein-coding regions such as 5′ UTRs or the kilobase flanking the gene (6%) (Figure 1). Two other

classes of variants that we excluded on the basis of absence from LSDBs, predicted functional impact, and per ACMG recommendations[22] were missense variants of unknown significance (6.5%) and variants predicted to affect splicing but outside of the canonical splice sites.

Each of the remaining 996 variants was then annotated with information available from HGMD, ClinVar and LSDBs and for the predicted consequence (e.g., frameshift, splicing and termination). Of these, 250 were listed as known pathogenic or probable pathogenic in at least one database or were a premature translation termination, loss of a translation termination codon, loss of a translation initiation codon, or alteration of canonical splice donor or acceptor site. After reviewing the literature for supporting evidence to justify designating these 250 variants as pathogenic, 3 variants met criteria as “expected pathogenic” and 11 as “known pathogenic” (Table 3 and Figure 1c). These 14 variants were present in 27 subjects from 14 families. No reportable variant was observed in more than one family. Thus 5.0% (27/543) of the exomes in our cohort had a finding that would result in disclosure under the ACMG recommendations.

To determine how many of the variants arose *de novo* as opposed to being inherited, we analyzed the parental sequences in 9 of the 14 families where parental sequences were available. For all 9 families (9 minor children and 4 adult children), one parent transmitted the variant to one or more children. The remaining 5 variants were identified in an adult for whom parental sequence was not available.

We identified a reportable incidental finding in 9 minor subjects in our cohort. For these 9 subjects, 5 had incidental findings associated with adult-onset conditions, and 4 had incidental findings associated with childhood-onset conditions.

A review of family and personal medical history revealed pertinent medical findings in only two cases. An adult subject with an *SCN5A* mutation had a history of exercise-induced fatigue and a first degree relative with an unspecified early onset cardiac condition; this relative was not enrolled in our study and, therefore, we could not evaluate segregation of the variant or verify phenotypic relevance. Another adult subject had an *APOB* mutation with a normal lipid profile: serum cholesterol 161 mg/dL (normal <200), LDL 93 mg/dL (normal <100) and HDL 56 mg/dL (high risk <40, low risk ≥60).

Discussion

By analysis of exome sequence data from 543 individuals distributed among 159 families, we clarify the reporting burden for the recommendations of the ACMG Working Group on Incidental Findings in Clinical Exome and Genome Sequencing.[3] We discovered 14 reportable variants for 27 individuals in 14 families. Therefore 8.8% of families enrolled for exome sequencing under the NIH UDP protocol had incidental findings requiring disclosure if the sequencing had been performed by a CLIA-certified laboratory.

Compared to the 1% rate of reportable incidental findings observed for the 23 of the 56 genes analyzed by Johnston *et al.*[2] and the 1.2–3.4% rate for 114 genes analyzed by Dorschner *et al.*,[6] we find a higher rate of reportable incidental findings. This increased rate

of reportable incidental findings could arise for several reasons including 1) increased coverage and quality of sequencing of the exome, 2) differences in variant selection, 3) differences in the subject cohort or 4) higher frequency of reportable variants in the ACMG recommended genes compared to the previously studied genes.

Regarding the sequence coverage and quality, the study of Johnston *et al.*, analyzed a smaller portion of the exome and aligned the sequences against an earlier version of the human reference genome. These two factors suggest that inclusion of more of the human exome and refinement of the reference genome might increase the number of detectable reportable variants. Testing of this by a detailed analysis of exons sequenced and not sequenced in the two data sets was, however, beyond the scope of this work since we did not have access to the exome sequences of Johnston *et al.*.[2] To enable future comparative investigations, we have provided details of coverage for our exome sequence data (Supplementary Tables 3–6)

Regarding differences in variant selection, the ACMG's estimation of a 1% rate of reportable incidental findings was based on an allele frequency within the cohort of > 0.5% and an allele frequency of >0.015% in dbSNP as exclusionary criteria for a pathogenic designation.[2] We did not use allele frequency as an exclusionary criterion for pathogenicity for two reasons. First, deleterious alleles occasionally exhibit higher prevalence in some populations.[23,24] Second, as discussed above, phenotyping is incomplete in cohorts from which most frequency data are derived.

To classify as variant as reportable, Dorschner *et al.* required an allelic frequency of less than a pre-determined disease-specific maximum prevalence plus various permutations of independently observed segregation with disease. Compared to our study, their criterion was 4 versus 3 segregations of the variant with disease; however, on the other hand, they did not consider functional assays as evidence for pathogenicity and only considered protein truncation as pathogenic if it occurred in the first 90% of the amino acid sequence. These differences likely contributed to the differences in our rates (5% vs 1.2–3.4%) of incidental findings. For example, their more stringent segregation requirements and lack of consideration of functional experimental (e.g. patch-clamp) evidence likely led to their classification of three variants that we considered as "known pathogenic" as "variants of unknown significance", i.e., *CACNA1S* p.T1354S, *SCN5A* p.T220I, and *SCN5A* p.E428K.

In this context, we expect that judicious comparison of variant classification may demonstrate that even reasonable parties disagree as to the benefits and risks of reporting such variants as incidental findings. The ACMG recommendations try to balance the need and ability to return highly beneficial risk information to the patients (true positives) while at the same time limiting the potential harm by not returning false positive results. The recommendations are written quite conservatively to strike a good balance between these two competing goals. Consequently, the recommendations clearly state that "variants that are previously unreported but are of the type which is expected to cause the disorder, as defined by prior ACMG guidelines, should be reported." The aforementioned guidelines are "ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007" and can be found at https://www.acmg.net/StaticContent/SGs/

ACMG_recommendations_for_standards_for.9. pdf. These guidelines state that if a variant is not previously reported to cause the disease only two paths lead to classification of a variant as reportable. One predicted deleteriousness (stop, indels, some splice sites) or in case of uncertainty (missense, potential splice site, inframe indels, SNP association only) the researchers need to collect supporting evidence to favor the deleteriousness of the variant.

Although one might advocate for an even stricter criteria, the criteria we have selected for our study is more stringent than the criteria provided by both the "ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing" and "ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007." We also acknowledge that the supporting evidence for these uncertain variants will vary in its quality and quantity and that the evidence will never be unequivocal for the simple fact that in light of unequivocal evidence, the variant in question would otherwise have been previously reported as disease causing. These variants and supporting evidence need to be returned to the clinician who ordered the sequencing and it is the clinician's duty to put these test results in the context of the patient's clinical background. Clinicians do this for other tests, and the clinician's understanding of the test characteristics is more important in the correct interpretation of the test than the test characteristics themselves. A test with high false positive rate but also with high sensitivity can be quite useful and desirable if used in the correct context with the right information to interpret the results. Our approach is therefore in agreement with "ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing" although until all possible changes in the human genome are annotated with unequivocal evidence to either support or refute the pathogenicity of each variant, there will always be a risk to make a false positive call. A priori the sensitivity or specificity of our methods cannot be determined, although higher specificity might be achieved with the use of very demanding requirements with respect to segregation or case-control disparities. The higher rate of incidental findings in our cohort as compared to Johnston *et al.*[2] and Dorschner *et al.*[6] highlights a possible limitation of our study in that our criteria may have a high false positive rate. More research is needed to compare the sensitivity and specificity of different filtering strategies, ideally with long-term follow-up. In any case, incidental findings should be worked up in accordance with the degree of confidence in their deleteriousness, with a conservative approach taken to those variants with a minimum of evidence supporting pathogenicity.

Relevant to differences in the study populations, the cohort reported by Johnston *et al.* was selected for atherosclerotic phenotypes (including unrelated controls) and was not a familial cohort. The cohort reported by Dorschner *et al.* was selected from among the NHLBI ESP on the basis of European and African ancestry. Our cohort is largely of European ancestry. Transmission within our cohort increased the number of individuals at risk from 14 to 27. With undiagnosed disorders, there is also the possibility of an antecedent hypermutable disorder; however, no one individual in our cohort had an increased number of reportable variants and our prior analyses of numbers of exome sequence variants within the UDP families did not identify marked differences from those reported for other cohorts.[25]

As for differences in the gene lists employed, Johnston *et al.* analyzed only a subset of the genes recommended by the ACMG Working Group on Incidental Findings in Clinical Exome and Genome Sequencing, i.e., the 23 associated with cancer syndromes.[2] In contrast, the ACMG list also encompasses genes associated with cardiac arrhythmias and myopathies, connective tissue disorders, familial hypercholesterolemia, and malignant hyperthermia susceptibility. Dorschner *et al.* analyzed 114 genes including 52 of the 56 genes on the ACMG list.[6]

Another variable in estimating the rate of reportable incidental findings is the thoroughness with which a disease and gene have been studied. In other words, the more individuals who have been identified with a disorder and checked for mutations in a gene, the more disease-causing mutations are likely to have been characterized. Reviewing our data, *SCN5A* (n=4) and *BRCA2* (n=2) had the most reportable variants. For *SCN5A*, this may reflect the fact that more variants are entered in databases because 1) both gain and loss of function variants in *SCN5A* can cause disease and 2) functional testing for pathogenicity is relatively accessible using patch-clamping experiments.

Four additional issues arising during our analysis were 1) defining the level of disease penetrance warranting reporting of a potential disease-causing variant, 2) determining how to weight variants deposited by clinical laboratories without corroborating evidence of pathogenicity, 3) the need for clinical correlation, and 4) obligations to extended family members. Relevant to the first issue, the ACMG recommendations state that variants with "higher" penetrance should be reported, but they leave the determination of "higher" to the clinical laboratory. For example, we identified a *TP53* variant (p.R337H/chr17:g. 7574017C>T, see Table 3) with 2.5–9.9% penetrance for pediatric adrenocortical carcinoma (ACC),[26,27] and newborn screening programs in Brazil have shown that screening for carriers of this mutation reduces morbidity and mortality.[26] This reporting conundrum was not resolved by the relationship of *TP53* to Li-Fraumeni Syndrome because this variant has not been associated with Li-Fraumeni Syndrome. Consequently, the reporting of a variant is difficult to code bioinformatically and will require human interpretation and possibly clinical consultation.

Regarding delineation of the pathogenicity of variants deposited by clinical laboratories, *BRCA1* and *BRCA2* variants provide an excellent illustration. Although our criteria for pathogenicity are scientifically sound, many *BRCA1* and *BRCA2* variants in public databases lack information on segregation with disease or experimental functional assays. Because variants lacking this information would not be considered pathogenic in our paradigm, our approach may well under-report the *BRCA1* and *BRCA2* associated cancer risks.

Another issue arising from this analysis speaks to the fact that a molecular finding is not a clinical diagnosis. Clinical records are often not available to testing labs, though in some cases they may substantiate or cast doubt on a variant's pathogenicity. The subject, in whom we identified a pathogenic *APOB* mutation (p.R3527W/chr2:g. 21229161G>A), a conclusion supported by functional assays demonstrating reduced LDLR binding,[28] had a favorable serum cholesterol and lipoprotein profile. A similar finding was also reported by Andreasen *et al.*[20] on "causative variants" for cardiomyopathies. This highlights that even

conservative standards to determine pathogenicity do not obviate the need for clinical interpretation and correlation.

The last issue is that of obligation to provide potentially helpful medical information to extended family members. For example, the person with an *SCN5A* variant and exercise-induced fatigue had a brother with an unspecified early-onset cardiac condition. If this brother carried the *SCN5A* variant, then this information might be diagnostically and therapeutically useful to him. Possible ethical approaches to notification include encouraging the subject in our cohort to discuss this finding with his brother, with or without provision of counseling to the brother, or direct notification of the brother. The American Medical Association's Code of Medical Ethics endorses encouraging the subject to notify at-risk relatives, with provision of assistance to the subject regarding communication of opportunities for testing and counseling.[29] This serves as a reminder that genetic testing may generate professional ethical obligations extending beyond the subject being tested.

Discussion on whether to inform individuals enrolled under the NIH UDP protocol about the identified variants focused on the delineated and perceived obligations defined by the language of the consent document and the process by which the consent was explained. In conclusion, whether to return or not return the incidental findings was deferred to the choices the individual or guardian had made when completing the written informed consent.

An issue raised by our study was the amount of work needed to determine which variants are reportable. We found that variants were listed occasionally as mutations or known pathogenic alleles in LSDBs without published evidence of segregation with disease or functional assays to support pathogenicity. Consequently, it is incumbent on the reporting laboratory to assemble and determine the credibility of the evidence used to determine the pathogenicity of a variant. Confounding this is the failure of many LSDBs to provide access to variants in a format that is easily applied to datasets derived from exome and genome sequencing. In contrast, ClinVar provides the required annotations as readily usable VCFs. Deposition of variants and their clinical significance in ClinVar would improve the efficiency of the recommended analysis.

Our analysis had some limitations. First, the exome sequencing that produced the variants for analysis was research-grade rather than clinical-grade and therefore not all exons in the 56 recommend genes had sufficient sequence coverage to call variants in all individuals. In addition, we did not validate the variants by Sanger sequence but rather inspected the alignments of short reads using IGV, a method that we have found more sensitive than Sanger sequencing. Second, our curation of variants was limited by the availability of annotations in public databases; we expect that the number and quality of these annotations will improve with time, as will the number of reportable variants. This raises the question of whether exome and genome sequence data should be reanalyzed at regular intervals to take into account the increasing information.

In summary, clinical exome and genome sequencing are cost effective methods for identifying the molecular bases of genetic conditions. These untargeted approaches, however, also uncover genetic variants with medical or social implications unrelated to the

indication for testing. In this context, the ACMG Working Group on Incidental Findings in Clinical Exome and Genome Sequencing recently recommended reporting “known pathogenic” and “expected pathogenic” mutations for 56 genes. Approximately 5% of all exomes in the NIH Undiagnosed Diseases Program familial cohort, and 8.8% of families in our cohort, had a reportable finding. The most time consuming aspect of fulfilling these recommendations was assembling the evidence for “pathogenicity” or “probable pathogenicity” because no well curated comprehensive public database is currently available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Patricia Birch and Shelin Adam for critical review of the manuscript. We thank the NHGRI Intramural Sequencing Center for their sequencing, alignment, genotyping and annotation services. This work was supported in part by the Common Fund, Office of the Director, and the Intramural Research Program of the National Human Genome Research Institute (NIH, Bethesda, Maryland)

Abbreviations

1KG	Thousand Genomes Project
ACC	adrenocortical carcinoma
ACMG	American College of Medical Genetics
ACT	adrenocortical tumor
ARVD/ARVC	arrhythmogenic right ventricular cardiodysplasia/cardiomyopathy
BIC	NHGRI Breast Cancer Information Database
EP	expected pathogenic
ERT	enzyme replacement therapy
ESP	Exome Sequencing Project
HGMD	Human Gene Mutation Database
KP	known pathogenic
LF-like	Li-Fraumeni-like
LOVD	Leiden Open Variation Database
LSDB	locus-specific database
NISC	NIH Intramural Sequencing Center

References

1. Wolf SM, Crock BN, Van Ness B, et al. Managing incidental findings and research results in genomic research involving biobanks and archived data sets. Genetics in medicine : official journal of the American College of Medical Genetics. 2012; 14(4):361–384. [PubMed: 22436882]

2. Johnston J, Rubinstein W, Facio F, et al. Secondary variants in individuals undergoing exome sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. American journal of human genetics. 2012; 91(1):97–108. [PubMed: 22703879]

3. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genetics in medicine : official journal of the American College of Medical Genetics. 2013; 15(7):565–574. [PubMed: 23788249]

4. Burke W, Matheny Antommaria AH, Bennett R, et al. Recommendations for returning genomic incidental findings? We need to talk! Genet Med. 2013

5. Christenhusz G, Devriendt K, Dierickx K. To tell or not to tell?; A systematic review of ethical reflections on incidental findings arising in genetics contexts. European Journal of Human Genetics. 2012; 21(3):248–255. [PubMed: 22739341]

6. Dorschner MO, Amendola LM, Turner EH, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am J Hum Genet. 2013; 93(4):631–640. [PubMed: 24055113]

7. Gahl WA, Markello TC, Toro C, et al. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. Genet Med. 2012; 14(1):51–59. [PubMed: 22237431]

8. Teer JK, Bonnycastle LL, Chines PS, et al. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. Genome Res. 2010; 20(10):1420–1431. [PubMed: 20810667]

9. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38(16):e164. [PubMed: 20601685]

10. ClinVar. [Accessed 17 June 2013] ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf/clinvar_00-latest.vcf.gz

11. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Human Mutation. 2011; 32(5):557–563. [PubMed: 21520333]

12. Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. Accurate and comprehensive sequencing of personal genomes. Genome Res. 2011; 21(9):1498–1505. [PubMed: 21771779]

13. Sunyaev SR. Inferring causality and functional significance of human coding DNA variants. Hum Mol Genet. 2012; 21(R1):R10–17. [PubMed: 22990389]

14. Petersen GM, Parmigiani G, Thomas D. Missense mutations in disease genes: a Bayesian approach to evaluate causality. Am J Hum Genet. 1998; 62(6):1516–1524. [PubMed: 9585599]

15. 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(3):652–655. [PubMed: 12900794]

16. Mohammadi L, Vreeswijk MP, Oldenburg R, et al. A simple method for co-segregation analysis to evaluate the pathogenicity of unclassified variants; BRCA1 and BRCA2 as an example. BMC Cancer. 2009; 9:211. [PubMed: 19563646]

17. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467(7319):1061–1073. [PubMed: 20981092]

18. Norton N, Robertson PD, Rieder MJ, et al. Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. Circulation. Cardiovascular genetics. 2012; 5(2):167–174. [PubMed: 22337857]

19. Cassa CA, Tong MY, Jordan DM. Large Numbers of Genetic Variants Considered to be Pathogenic are Common in Asymptomatic Individuals. Hum Mutat. 2013

20. Andreasen C, Nielsen JB, Refsgaard L, et al. New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. Eur J Hum Genet. 2013; 21(9):918–928. [PubMed: 23299917]

21. Jordan DM, Kiezun A, Baxter SM, et al. Development and validation of a computational method for assessment of missense variants in hypertrophic cardiomyopathy. Am J Hum Genet. 2011; 88(2):183–192. [PubMed: 21310275]

22. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genetics in medicine : official journal of the American College of Medical Genetics. 2008; 10(4):294–300. [PubMed: 18414213]

23. Roa BB, Boyd AA, Volcik K, Richards CS. Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. Nat Genet. 1996; 14(2):185–187. [PubMed: 8841191]
24. Miserez AR, Laager R, Chiodetti N, Keller U. High prevalence of familial defective apolipoprotein B-100 in Switzerland. J Lipid Res. 1994; 35(4):574–583. [PubMed: 8006512]
25. Adams DR, Sincan M, Fajardo KF, et al. Analysis of DNA sequence variants detected by high-throughput sequencing. Human Mutation. 2012; 33(4):599–608. [PubMed: 22290882]
26. Custodio G, Parise GA, Kiesel Filho N, et al. Impact of Neonatal Screening and Surveillance for the TP53 R337H Mutation on Early Detection of Childhood Adrenocortical Tumors. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013; 31(20): 2619–2626. [PubMed: 23733769]
27. Figueiredo BC, Sandrini R, Zambetti GP, et al. Penetrance of adrenocortical tumours associated with the germline TP53 R337H mutation. Journal of Medical Genetics. 2006; 43(1):91–96. [PubMed: 16033918]
28. Fisher E, Scharnagl H, Hoffmann MM, et al. Mutations in the apolipoprotein (apo) B-100 receptor-binding region: Detection of apo B-100 (Arg(3500)--> Trp) associated with two new haplotypes and evidence that apo B-100 (Glu(3405)-> Gln) diminishes receptor-mediated uptake of LDL. Clinical Chemistry. 1999; 45(7):1026–1038. [PubMed: 10388479]
29. Code of Medical Ethics. [Accessed 8/22/13] Opinion 2.131 - Disclosure of Familial Risk in Genetic Testing. http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion2131.page
30. DiGiammarino EL, Lee AS, Cadwell C, et al. A novel mechanism of tumorigenesis involving pH-dependent destabilization of a mutant p53 tetramer. Nat Struct Biol. 2002; 9(1):12–16. [PubMed: 11753428]
31. Makita N, Behr E, Shimizu W, et al. The E1784K mutation in SCN5A is associated with mixed clinical phenotype of type 3 long QT syndrome. The Journal of clinical investigation. 2008; 118(6):2219–2229. [PubMed: 18451998]
32. Wei J, Wang DW, Alings M, et al. Congenital long-QT syndrome caused by a novel mutation in a conserved acidic domain of the cardiac Na+ channel. Circulation. 1999; 99(24):3165–3171. [PubMed: 10377081]
33. Wang Q, Chen S, Chen Q, et al. The common SCN5A mutation R1193Q causes LQTS-type electrophysiological alterations of the cardiac sodium channel. Journal of Medical Genetics. 2004; 41(5):e66. [PubMed: 15121794]
34. Hwang HW, Chen JJ, Lin YJ, et al. R1193Q of SCN5A, a Brugada and long QT mutation, is a common polymorphism in Han Chinese. Journal of Medical Genetics. 2005; 42(2):e7. author reply e8. [PubMed: 15689442]
35. Darbar D, Kannankeril PJ, Donahue BS, et al. Cardiac sodium channel (SCN5A) variants associated with atrial fibrillation. Circulation. 2008; 117(15):1927–1935. [PubMed: 18378609]
36. Olesen MS, Yuan L, Liang B, et al. High prevalence of long QT syndrome-associated SCN5A variants in patients with early-onset lone atrial fibrillation. Circ Cardiovasc Genet. 2012; 5(4):450–459. [PubMed: 22685113]
37. Benson DW, Wang DW, Dyment M, et al. Congenital sick sinus syndrome caused by recessive mutations in the cardiac sodium channel gene (SCN5A). J Clin Invest. 2003; 112(7):1019–1028. [PubMed: 14523039]
38. Gui J, Wang T, Jones RP, Trump D, Zimmer T, Lei M. Multiple loss-of-function mechanisms contribute to SCN5A-related familial sick sinus syndrome. PloS one. 2010; 5(6):e10985. [PubMed: 20539757]
39. Kapplinger JD, Tester DJ, Alders M, et al. An international compendium of mutations in the SCN5A-encoded cardiac sodium channel in patients referred for Brugada syndrome genetic testing. Heart rhythm : the official journal of the Heart Rhythm Society. 2010; 7(1):33–46. [PubMed: 20129283]
40. Olson TM, Michels VV, Ballew JD, et al. Sodium channel mutations and susceptibility to heart failure and atrial fibrillation. JAMA : the journal of the American Medical Association. 2005; 293(4):447–454. [PubMed: 15671429]
41. http://www.ncbi.nlm.nih.gov/clinvar/RCV000030362/#evidence

42. Dalal D, James C, Devanagondi R, et al. Penetrance of mutations in plakophilin-2 among families with arrhythmogenic right ventricular dysplasia/cardiomyopathy. Journal of the American College of Cardiology. 2006; 48(7):1416–1424. [PubMed: 17010805]
43. Andersen PS, Hedley PL, Page SP, et al. A novel Myosin essential light chain mutation causes hypertrophic cardiomyopathy with late onset and low expressivity. Biochemistry research international. 2012; 2012:685108. [PubMed: 22957257]
44. Maron BJ. Hypertrophic cardiomyopathy: a systematic review. JAMA. 2002; 287(10):1308–1320. [PubMed: 11886323]
45. Gersh BJ, Maron BJ, Bonow RO, et al. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2011; 124(24):2761–2796. [PubMed: 22068435]
46. Spada M, Pagliardini S, Yasuda M, et al. High incidence of later-onset fabry disease revealed by newborn screening. American journal of human genetics. 2006; 79(1):31–40. [PubMed: 16773563]
47. De Brabander I, Yperzeele L, Ceuterick-De Groote C, et al. Phenotypical characterization of alpha-galactosidase A gene mutations identified in a large Fabry disease screening program in stroke in the young. Clinical neurology and neurosurgery. 2013; 115(7):1088–1093. [PubMed: 23219219]
48. Terryn W, Vanholder R, Hemelsoet D, et al. Questioning the Pathogenic Role of the GLA p.Ala143Thr “Mutation” in Fabry Disease: Implications for Screening Studies and ERT. JIMD reports. 2013; 8:101–108. [PubMed: 23430526]
49. Pirone A, Schredelseker J, Tuluc P, et al. Identification and functional characterization of malignant hyperthermia mutation T1354S in the outer pore of the Cavalpha1S-subunit. American journal of physiology. Cell physiology. 2010; 299(6):C1345–1354. [PubMed: 20861472]
50. Sharing Clinical Reports. http://sharingclinicalreports.org/
51. Meindl A. Comprehensive analysis of 989 patients with breast or ovarian cancer provides BRCA1 and BRCA2 mutation profiles and frequencies for the German population. International journal of cancer. Journal international du cancer. 2002; 97(4):472–480. [PubMed: 11802209]
52. Gaffney D, Reid JM, Cameron IM, et al. Independent Mutations at Codon-3500 of the Apolipoprotein-B Gene Are Associated with Hyperlipidemia. Arteriosclerosis Thrombosis and Vascular Biology. 1995; 15(8):1025–1029.
53. Choong ML, Koay ES, Khoo KL, Khaw MC, Sethi SK. Denaturing gradient-gel electrophoresis screening of familial defective apolipoprotein B-100 in a mixed Asian cohort: two cases of arginine3500-->tryptophan mutation associated with a unique haplotype. Clinical Chemistry. 1997; 43(6 Pt 1):916–923. [PubMed: 9191540]
54. Tai DY, Pan JP, Lee-Chen GJ. Identification and haplotype analysis of apolipoprotein B-100 Arg(3500)-> Trp mutation in hyperlipidemic Chinese. Clinical Chemistry. 1998; 44(8):1659–1665. [PubMed: 9702952]

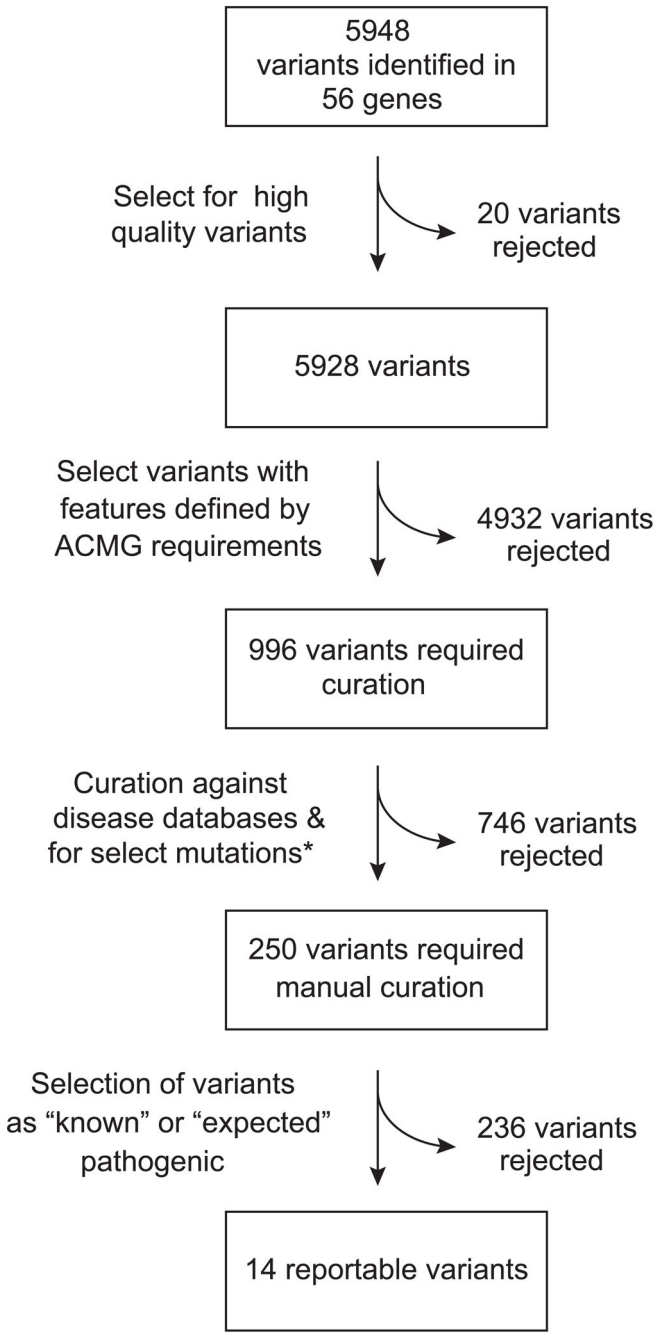

Figure 1.

Flow chart summarizing the NIH Undiagnosed Diseases Program analysis of and observations for the 56 genes recommended for interrogation by the ACMG Working Group on Incidental Findings in Clinical Exome and Genome Sequencing. The observations were derived from analysis of exome sequence data derived from a 159-family cohort consisting of 543 subjects with 188 affected subjects, 137 siblings and 218 parents. * Mutations recommended for reporting as “expected pathogenic” include premature translation

termination, loss of a translation termination codon, loss of a translation initiation codon, or alteration of canonical splice donor or acceptor site.

Table 1

Summary coverage statistics for exome sequence included in the study

	Threshold		
	10x	**20x**	**30x**
Percent of exons for which >90% of the subjects had ≥95% coverage of the exon at ≥threshold	65.5 %	45.4 %	23.4 %
Percent of exons for which >90% of the subjects had 100% coverage of the exon at ≥threshold	63 %	41.6 %	20 %

Table 2

Variants analyzed

Type of variant	Number of variants
Total Variants in ACMG Genes	5948*
Variants meeting minimum quality standards	5928
Variants rejected for absence from databases and for mutation properties	4932
Intronic	3300
Exonic synonymous	700
3′ UTR	655
5′ UTR	100
5′ Flanking	40
3′ Flanking	49
Non-canonical splice	4
3′ UTR ncRNA	78
5′ UTR ncRNA	6
Variants requiring curation	996
Variants requiring manual curation	250
Variants designated reportable	14

*Multi-allelic variants were counted as a single variant in the numbers listed in this paper, but in Table 3 and in Supplementary Table 2, they are provided as individual allelic variants

Abbreviations: ncRNA, noncoding RNA; UTR, untranslated region.

Table 3

Reportable variants detected in the NIH UDP exome cohort

Gene	Disease	Chr	Variant			ClinVar Access. No.	dbSNP		No of var Chr*	Rationale
			genome	cDNA	Protein		rsID	Minor allele freq.		
TP53	Pediatric adrenocortical carcinoma	17	7574017C>T	NM_000546.5 : c.1010G>A	p.R337H	SCV000115376	rs121912664	NA	2	Meets criteria for known pathogenic variant as a functional assay has shown reduced function at physiologic pH.[30] Although the variant is associated with pediatric ACC rather than Li-Fraumeni syndrome, the diseases are related and similarly amenable to medical intervention. Indeed, recent use of neonatal screening for this allele in Southern Brazil has demonstrated utility, with authors stating “Without screening and surveillance, only 50% of children with ACTs survive, and many require intensive, toxic chemotherapy.”[26]
SCN5A	Long QT Syn type 3 Brugada Syn type 1	3	38592513C>T	NM_000335.4 : c.5347G>A	p.E1783K	SCV000115377	rs137854601	NA	1	Meets known pathogenic criteria with electrophysiologic and patch-clamp experiments demonstrating negative inactivation shift and enhanced flecainide block in one study,[31] and segregation with disease and small but prolonged inward current during long depolarizations in another study.[32]
SCN5A	Long QT Syn type 3 Brugada Syn type 1	3	38616876C>T	NM_000335.4 : c.3575G>A	p.R1192Q	SCV000115378	rs41261344	0.012	3	Meets known pathogenic criteria as the variant (identified in subjects with Long QT and Brugada Syndrome) has been shown to produce late inactivating current relative to wild-type channels.[33] Subsequent reports have identified the variant in 6% of a small sample of Han Chinese people; the authors of this most recent paper suggest it may still be causal but with reduced penetrance since 1 of 9 carriers did have prolonged QT_c and another 1 of 9 had an intermediate-range Recent panels of persons of QT_c.[34] East Asian ancestry have demonstrated prevalence of this variant varying from 0.2–12.5% (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=rs41261344). While the upper range of this prevalence certainly casts doubt as to the pathogenicity of the variant, the carriers in these panels were not phenotyped, and this evidence therefore cannot be used to definitively disprove the above-cited functional study.
SCN5A	Lone atrial fibrillation	3	38647498C>T	NM_000335.4 : c.1282G>A	p.E428K	SCV000115379	rs199473111	NA	2	Meets known pathogenic criteria by segregation with lone atrial fibrillation in 3/3 family members.[35] The third family member had atrial fibrillation by history alone, but we give the benefit of the doubt to the authors. Although lone atrial fibrillation is not recognized as a reportable condition for mutations in SCN5A by the ACMG, recent studies have found mutations known to cause long QT or Brugada syndrome in families with early onset lone atrial fibrillation.[35,36] The diseases are related and similarly amenable to medical intervention, so we consider the variant reportable under the spirit of the guidelines.
SCN5A	Sick Sinus Syn	3	38655278G> A	NM_000335.4 : c.659C>T	p.T220I	SCV000115380	rs45620037	0.000	1	Meets known pathogenic criteria as patch-clamp experiments have found reduced peak current, delayed recovery from inactivation, and delayed inactivation.[37,38] The variant has been

Gene	Disease	Chr	Variant			ClinVar Access. No.	dbSNP		No of var Chr*	Rationale
			genome	cDNA	Protein		rsID	Minor allele freq.		
										identified in subjects with sick sinus syndrome, though evidence of segregation with dis
										identified in subjects with sick sinus[39,40] syndrome, though evidence of segregation with dis
PKP2	Arrhythmogenic right ventricular cardio-myopathy	12	32955491C>G	NM_004572.3 : c.2146-1G>C		SCV000115381	rs193922674	NA	2	Meets known pathogenic criteria as a splicing mutation that has been identified in over 20 subjects with ARVD.[41] Does not segregate perfectly with disease in published reports of two families, but segregates with disease in 2/3 subjects in two families.[42]
MYL3	Hypertrophic cardio-myopathy	3	46902238C>T	NM_000258.2 : c.235G>A	p.V79I	SCV000115382			1	Meets known pathogenic criteria as the variant segregates with disease in 4/6 post-adolescent carriers in one family.[43] HCM often displays onset during adolescence, thus carriers under 18 would not necessarily be expected to display the phenotype.[44] In this study, 3 family members demonstrate a borderline phenotype, but their findings are consistent with an HCM spectrum of disease (t-wave inversions, left axis deviation, angulated septum and diastolic dysfunction). These finding are compatible with a scenario in which only a portion of the left ventricular septum is hypertrophied, or early emergence of clinical disease, both possibilities recognized by the 2011 ACCF/AHA Guideline for the Diagnosis and Treatment of Hypertrophic Cardiomyopathy.[45]
GLA	Fabry Disease	X	100656740C>T	NM_000169.2 : c.427G>A	p.A143T	SCV000115383	rs104894845	NA	4	Meets known pathogenic criteria as the variant has been shown to produce low but residual (36% wild-type) α-Gal A activity in a transfection assay.[46] Earlier interpretations of these findings were that this represented a late-onset variant with a non-classical phenotype,[46,47] but a recent paper has called into question whether this variant is pathogenic at all.[48] Although the recent arguments are compelling, some patients with this allele are on ERT[48]; we therefore feel that clinical navigation of this complex medical research is best conducted between the carrier subjects in our cohort and their physicians, and that reporting the variant as an incidental finding is not precluded by recent publications arguing against the variant's pathogenicity.
DSP	Arrhythmogenic right ventricular cardio-myopathy	6	7583973C>T	NM_004415.2 : c.6478C>T	p.R2160X	SCV000115384			2	Meets expected pathogenic criteria as a stop-gain mutation. Not present in LSDBs.
CACNA 1S	Malignant Hyperthermia Susceptibility	1	201020165T> A	NM_000069.2 : c.4060A>T	p.T1354S	SCV000115385			2	Meets known pathogenic criteria as the variant segregated with in vitro contracture test in 7/9, with remaining 2/9 equivocal on the contracture test.[49] A tenth carrier in the family was not biopsied.[49] The same study also used patch-clamp to demonstrate accelerated inward Ca^{2+} current and increased sensitization of RYR1 under caffeine exposure in a transfection model.[49]
BRCA2	Breast and ovarian cancer susceptibility	13	32914529A>T	NM_000059.3 : c.6037A>T	p.K2013X	SCV000115386	rs80358840	NA	1	Meets known pathogenic criteria as a stop-gain observed in affected subjects. Submitted by 2 subjects in Sharing Clinical Reports[50]; also identified in a German study in 1 individual.[51]

Gene	Disease	Chr	Variant			ClinVar Access. No.	dbSNP		No of var Chr*	Rationale
			genome	cDNA	Protein		rsID	Minor allele freq.		
BRCA2	Breast and ovarian cancer susceptibility	13	32929240delAC	NM_000059.3 : c.7251_7252del	p.His2417Glnfs*3	SCV000115387			3	Meets expected pathogenic criteria as a frameshift mutation. Not present in LOVD, BIC, or UMD, but frameshift mutations in this but frameshift mutations in this region in BIC are listed as clinically relevant.
BRCA1	Breast and ovarian cancer susceptibility	17	41197713insG	NM_007294.3 : c.5578dup	p.His1860Profs*20	SCV000115388			1	Meets expected pathogenic criteria as a frameshift mutation. Although it is very near the end of the coding sequence, many frameshift mutations in these exons are cited in BIC as pathogenic.
APOB	Familial hypercholesterolemia	2	21229161G>A	NM_000384.2 : c.10579C>T	p.R3527W	SCV000115389			2	Meets known pathogenic criteria as functional evidence supports reduced LDL binding.[28,52–54] The effects of this variant are thought to be milder than a Gln substitution at the same codon.[28,52–54]

*Number of variant chromosomes in the UDP dataset. All individuals were heterozygous or hemizygous for the variant.

Abbreviations: Chr, chromosome; No., number; Syn, syndrome; Var, variant