



Published in final edited form as:

Genet Med. 2015 October ; 17(10): 782–788. doi:10.1038/gim.2014.196.

Assessment of incidental findings in 232 whole exome sequences from the Baylor-Hopkins Center for Mendelian Genomics

Julie Jurgens, BSc¹, Hua Ling, PhD², Kurt Hetrick, MS², Elizabeth Pugh, PhD², Francois Schiettecatte, PhD³, Kimberly Doheny, PhD², Ada Hamosh, MD, MPH¹, Dimitri Avramopoulos, MD, PhD¹, David Valle, MD¹, and Nara Sobreira, MD, PhD¹

¹McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

²Center for Inherited Disease Research, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21224, USA

³FS Consulting LLC, Salem, MA 01970, USA

Abstract

Purpose—In March 2013, the ACMG published a list of 56 genes with the recommendation that pathogenic and likely pathogenic variants detected incidentally by clinical sequencing should be reported to patients. As an initial step in determining the practical consequences of this recommendation in the research setting, we searched for variants in these genes in 232 whole exome sequences from the Baylor-Hopkins Center for Mendelian Genomics.

Methods—We identified rare, nonsynonymous and splicing SNVs and indels and assessed variant classification using HGMD, Emory and ClinVar databases. We analyzed the burden of mutation in each of the 56 genes and determined which variants should be reported to patients.

Results—Our filtering resulted in 249 distinct variants, with a mean of 1.69 variants per individual. Half of these were novel missense mutations not classified by any of the 3 reference databases. Of 101 variants listed in HGMD, 48 were also in ClinVar and 3 were also in Emory; half of these shared variants were classified discordantly between databases. Some genes consistently had greater variation than others. In total, 0.86% of individuals had a reportable incidental variant.

Conclusion—These observations demonstrate some current challenges of assessing phenotypic consequences of incidental variants for counseling patients.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence, Nara Sobreira, 733 N. Broadway, Broadway Research Building, Room 512, Baltimore, MD 21205, Telephone Number: 410-614-2577, Fax Number: 410-955-7397, nsobrei2@jhmi.edu.

SUPPLEMENTARY MATERIAL

Supplementary information is available at the Genetics in Medicine website.

Keywords

Incidental findings; secondary findings; variant classification; whole exome sequencing; ACMG recommendations

INTRODUCTION

Incidental findings in whole exome and whole genome sequencing are variants of known or possible pathology identified in genes unrelated to the initial reason for which sequencing tests were ordered. As large-scale sequencing is increasingly used in a clinical setting, standardizing procedures for identifying, classifying, and reporting variants is of great importance. In 2013, the American College of Medical Genetics and Genomics (ACMG) released a recommendation suggesting that known pathogenic or expected pathogenic constitutional variants detected in 56 genes representing 24 conditions with well-known guidelines for prevention or treatment should be reported to patients sequenced in a clinical setting.¹ Shortly after the release of these recommendations, the Clinical Sequencing Exploratory Research (CSER) Consortium and Electronic Medical Records and Genomics (eMERGE) Network followed by releasing a set of guidelines for the return of incidental findings within genomics research studies.² These guidelines recommended that medically actionable incidental variants be returned to willing research participants, but exempted researchers from an obligation to search for these variants.

Although incidental variants are unexpected, when variant pathogenicity is certain, the information enables patients to make proactive decisions about their health and inform their physician. However, some variants are of uncertain significance, or initial interpretation of variant pathogenicity may be subject to change.³ Reporting these variants of uncertain significance (VUS) to patients may lead to misinterpretation of the results, produce unnecessary anxiety, and diminish confidence in genetic test results in general. Thus, transmittal of these results requires accurate and thoughtful genetic counseling and careful evaluation of available sources of variant classification to maximize benefits and minimize harm to patients.

At the Baylor-Hopkins Center for Mendelian Genomics (BHCMG, <http://mendeliangenomics.org>),⁴ we utilize whole exome sequencing to detect variants responsible for Mendelian disorders with unknown molecular bases. Since we sequence across the exome, we not only capture variants that are responsible for the patient phenotype, but also incidental variants which may predispose and/or cause other unrelated disorders. Returning meaningful results to research participants through their physicians is a goal of our center, so understanding the clinical significance of incidental findings detected at the research level and knowing which of these should be returned to patients is of special interest.

Although the ACMG guidelines for reporting incidental findings are geared toward a clinical setting, they provide a well-defined, targeted set of genes from among the many interrogated in research sequencing. Thus, our results on this set may inform policies for handling incidental findings. Here we examine the implementation of the ACMG clinical guidelines within a research setting by assessing the spectrum of rare, functional variation in

the 56 genes in whole exome sequences of 232 individuals in 89 families sequenced at the Baylor-Hopkins Center for Mendelian Genomics for a variety of Mendelian disorders with unknown molecular bases, some of which have features that overlap with the disorders covered by the ACMG list. We then assessed variant interpretation by evaluating classification and reportability according to available databases and approaches taken by others in the literature.

MATERIALS AND METHODS

All research participants were counseled on the possible outcomes of whole exome sequencing and signed a consent form approved by the Johns Hopkins University School of Medicine Institutional Review Board. Participants chose whether they would like to receive information regarding primary or incidental variants found in the course of this study and were given the opportunity to opt out or in at any time.

Whole exome sequences from 232 individuals in 89 families sequenced at Johns Hopkins University (JHU) were analyzed for rare splicing or nonsynonymous SNVs and indels in the 56 reportable genes from the ACMG guidelines using the analysis feature of the web-based system PhenoDB.⁵

Briefly, whole exome sequencing was done on the Illumina HiSeq2000 platform (Illumina, Inc., San Diego, CA) using paired-end 75 or 100 bp reads. 51 Mb of genomic DNA comprised of CCDS exons and adjacent intron sequences were captured with Agilent SureSelect Human All Exon V4 51Mb Kit (Agilent Technologies, Santa Clara, CA). FASTQ files were aligned to the reference genome (GRCh37; Ensembl core database release 50_361)⁶ with the Burrows-Wheeler Alignment (BWA 0.5.10) tool,⁷ resulting in SAM/BAM output.⁸ PCR duplicates were flagged with Picard.⁹ Local realignment around indels, base call quality score recalibration and reduced reads BAMs were performed with GATK 2.3–9.¹⁰ Multi-sample SNV and indel calling was performed with GATK's UnifiedGenotyper on the reduced reads BAM files. Variant sites were filtered with GATK's Variant Quality Score Recalibration best practices,¹¹ and heterozygous genotypes were excluded if they did not have at least 5 alternate allele reads. Sequencing coverage of the 56 genes is listed in Supplementary Table S1.

We selected for nonsynonymous SNVs, exonic indels, and splice junction variants (2 most 5' and 3' intronic bases) in each of the 56 reportable genes. Nonsynonymous SNVs and exonic variants were defined by RefSeq Gene coordinates.^{12,13} Next, we filtered these variants on the basis of their minor allele frequencies (MAF), with exclusion of variants with MAF < 0.01 in the 1000 Genomes Project (April 2012 release),⁶ Exome Variant Server (<http://evs.gs.washington.edu/EVS/>),¹⁴ dbSNP build 137,^{15,16} and our in-house control database (CIDRVar 51Mb), which includes data from 50 individuals sequenced in-house with the Agilent51Mb exome capture kit (Agilent Technologies, Santa Clara, CA).

Next, we analyzed the variants for prior classification in the Human Gene Mutation Database (HGMD),^{17,18} ClinVar,^{19,20} and the Emory Genetics Laboratory Variant Classification Catalog.^{21,22} We then defined the burden of variants in each individual

sequenced and the mutability of each gene (number of rare functional variants/gene and number of rare functional variants/kb of coding region), and the classification and reportability of each variant.

RESULTS

In the 232 exomes, we identified 249 distinct variants in the 56 gene ACMG target set, some of which were seen in more than one individual, for a total of 391 variants. A total of 124 variants were shared between family members, and 18 variants were shared between unrelated individuals. There was an average of 1.69 variants per individual, with a range of 0–7 variants per individual. There were only 45 people (19.4%) with no variants, and in the 232 exomes we found at least one variant in 45 of the 56 genes (80.4%; Supplementary Table S1).

We then stratified these 249 distinct variants by type (Table 1). Most variants were missense (231/249, or 92.8%). There were also 4 nonsynonymous/splice variants (1.6%), 2 splice variants (0.8%), 3 frameshifting indels (1.2%), 5 nonframeshifting indels (2.0%), and 4 nonsense variants (1.6%).

Next, we checked the classification of these variants in three databases: HGMD, ClinVar, and Emory. Of the 249 distinct variants identified, 126 (50.6%) were classified by at least one of the three databases: 54/126 (42.8%) were represented solely in HGMD; 44/126 (34.9%) were in HGMD and ClinVar; 23/126 (18.3%) were in ClinVar alone; 2/126 (1.6%) were in Emory alone; and another 3/126 (2.4%) were in all three databases (Figure 1).

In total, 101/249 variants (40.5%) were listed in HGMD (94 missense SNVs (93.1%), 2 nonsynonymous exonic/splicing variants (2.0%), 2 splicing variants (2.0%), 1 nonsense SNV (1.0%), and 2 nonframeshifting indels (2.0%)). More importantly, HGMD classified 74 (72.8%) as disease-causing mutations (DM), 24 (24.3%) as possible disease-causing mutations (DM?), 1 (1.0%) as a functional polymorphism (FP), and 2 (1.9%) as disease-associated polymorphisms (DP) reported to be in significant association with disease ($p < 0.05$) along with some evidence of functionality.

Of the 148 variants not classified by HGMD, 23 variants in 13 genes were described by ClinVar and 2 variants in 2 genes were listed in Emory. This left a total of 123/249 variants (49.4%) that were not present in any database (Figure 1).

The classification of variants within these three databases was often discordant. Of the 101 variants represented in HGMD, 48 were also in ClinVar. Three of these 48 variants were also in Emory. Almost half of these shared variants (22/48 variants, or 45.8%) were given discordant classifications among databases (Supplementary Table S2).

Next, we examined the effects of gene size and evolutionary constraint on the number of functional variants within each gene, in a manner similar to that described by Petrovski et al., 2013.²³ Some genes consistently had greater variation than others. For instance, *BRCA2*, *APOB*, *CACNA1S*, and *DSP* had the greatest number of variants/gene (Supplementary Table S1). Since increased coding length increases the target size for mutation, larger genes are

expected to have a higher number of variants than smaller genes. To explore this possibility, we analyzed the number of variants per gene against the size (kb) of the consensus coding sequence (CCDS) of the gene. We found that the number of variants per gene generally increased with the size of coding sequence, and the number of variants showed a 62.8% correlation with the size of the consensus coding sequence (Figure 2).

We also found that some genes (e.g. *RBI*, *PMS2*, and *MLH1*) consistently have a lower variant density (variants/kb coding sequence) than others (e.g. *MYBPC3*, *TMEM43*, and *MYL3*) (Supplementary Table S1). The values of variant density range from 0.4 variants per kb in *RBI* to 3.4 variants per kb in *MYL3*. As described by others, a possible explanation for this variation in mutational burden despite correction for coding sequence length is varying degrees of evolutionary constraint.²³ Certain genes are under a higher degree of purifying selection than others, and this is likely reflected by less variation per unit coding length.

Determining Reportable Variants

To determine which variants should be reported to our research participants, we adopted a stringent set of measures outlined in a recent paper by Dorschner et al.²⁴ Their criteria incorporates various lines of evidence in assessing variant pathogenicity, including comparison between variant minor allele frequencies and incidences of the corresponding disorders; familial segregation data; observation in unrelated affected individuals; *de novo* events in a trio; and protein truncation in disorders caused by haploinsufficiency. If a variant is classified as pathogenic or likely pathogenic by these criteria, it is considered reportable back to the patient.

We applied these criteria to all 391 variants in the 56 ACMG genes detected in 232 exomes. In total, we found 2 pathogenic variants (in *MSH2* and *MYLK*) and 3 likely pathogenic variants (in *LMNA*, *MYBPC3*, and *MUTYH*), as shown in Table 2. One of these genes, *MUTYH*, is responsible for autosomal recessive conditions (*MYH*-associated polyposis; adenomas, multiple colorectal, *FAP* type 2; colorectal adenomatous polyposis, autosomal recessive, with pilomatricomas; MIM 608456 and 132600).³⁵ Since current ACMG guidelines recommend reporting only individuals with homozygous variants in *MUTYH*,¹ we would not report this heterozygous variant to the patient. Since the patients in our cohort underwent whole exome sequencing to explain a variety of potential Mendelian disorders, some of which have overlapping features with the disorders covered by the ACMG list, we recognized that our cohort may have an enrichment of reportable variants as compared to the general population. Indeed, two patients, one with a likely pathogenic variant in *LMNA* and one with a likely pathogenic variant in *MYBPC3*, had clinical phenotypes which included dilated cardiomyopathy, so we did not count these two cases toward our final number of incidental variants. In total, this means that 2/232 individuals, or 0.86% of our sample, had a reportable incidental variant in one of the 56 ACMG genes.

DISCUSSION

In June of 2014, Jarvik et al. released recommendations for dealing with incidental variants in the research arena. The authors suggested that researchers be exempted from an obligation to search for variants outside the intended scope of their study, but that actionable

variants be reported to research participants when discovered.² The present study describes the workflow for identification and classification of incidental findings in the Johns Hopkins component of the Baylor-Hopkins Center for Mendelian Genomics based on the ACMG recommendations and on the classification criteria from Dorschner et al.²⁴

In 232 exomes, we identified a total of 391 rare (MAF < 1%) variants in the 56 ACMG genes (1.69 variants per individual), 249 of which were distinct variants. Most of these variants were missense (231/249, or 92.8%), and half were not classified by any of the three databases (HGMD, ClinVar, and Emory) that we used to assess variant classification (123/249 variants, or 49.4%). Since these novel variants were not represented in the variant databases, they could not be utilized for comparative analyses of classifications among the three databases.

Next we used the criteria described by Dorschner et al.²⁴ to classify the 391 total variants identified in 232 exomes and found that 2/232 individuals, or 0.86% of our sample, had an incidental reportable variant in one of the 56 ACMG genes. Overall, this analysis was quite time-consuming, but we were able to automate parts of it such as filtering variants whose frequency was higher than that predicted by the frequency of their associated disorders. This step reduced the number of distinct variants from 249 (obtained in our initial analyses using a 1% MAF cutoff) to 163, so only 65.5% of the original variants required further assessment. We were also able to automate selection of variants in the HGMD, ClinVar, and Emory classification databases using our web-based system, PhenoDB.⁵ Although we did not consider the classifications from these databases in support of variant pathogenicity or reportability, the databases provided useful citations from the literature and clinical observations for some variants. This information reduced the time required for manual review. Protein truncation was automatically predicted, but required manual inspection to confirm that protein truncation is an established cause of the disorder associated with the respective gene. The most time-consuming part of the process was reviewing the literature for each variant to count familial segregations, *de novo* events, and singleton reports. This step required additional expertise regarding diagnoses of the disorders on the ACMG list. Moreover, evaluation of cases in the literature was somewhat subjective.

Several other studies have analyzed reportable incidental findings within whole-exome or whole-genome sequences to-date. The study by Dorschner et al.,²⁴ whose criteria we adopted for our analysis of reportable variants, found that 14/1000 individuals sequenced (1.4%) had a pathogenic or likely pathogenic reportable variant in one of the 52 ACMG genes responsible for adult-onset conditions. Johnston et al. found that 8/572 (1.39%) individuals sequenced through the ClinSeq project had a reportable variant in one of 37 cancer genes; 23 of these are on the ACMG list.³⁶ These numbers are similar to ours and to those in the original ACMG guidelines, which anticipated that 1% of individuals sequenced would have a reportable variant in one of these 56 genes.¹

Another recent report by Lawrence et al. detected 27/543 individuals (~5%) sequenced through the NIH Undiagnosed Diseases Program with an incidental variant in one of the 56 ACMG genes.³⁷ Their method differed from ours in a few ways which may have contributed to the discrepancy in frequency of incidental findings. For instance, they used

reported functional studies as evidence of pathogenicity; required fewer informative meioses to count familial co-segregation events with disease; did not apply minor allele frequency cutoffs; and used designations in variant classification databases as evidence in support of variant pathogenicity. In addition, a few individuals in their cohort had a phenotype related to the reported incidental finding, and some incidental variants were repeated due to familial transmission.³⁴ These differences in methodology may have contributed to the discrepancies in numbers of reportable incidental findings.

During this process we recognized that identifying and classifying the multitude of incidental variants that arise from next-generation sequencing is a time-consuming and subjective process dependent on the particular methodologies of each individual investigator. We confirmed this last observation when we found that almost half (22/48, or 45.8%) of the 101 variants represented in at least 2 databases were given discordant classifications between them. We compared these database classifications to our variant classifications from the Dorschner et al. method²⁴ and found that variant classifications from the ClinVar and Emory databases, but not the HGMD database, were generally consistent with our variant classifications (Supplemental Table S2). This is likely because HGMD aims to provide a comprehensive listing of all reported variants and draws its classifications from reports in the literature^{17,18} that may not be accurate, consistent, up-to-date, or sufficient to provide support of variant pathogenicity. For these reasons, HGMD is generally less conservative than the other databases.

The abundance of rare and novel missense variants in our cohort made accurate assessment of variant pathogenicity more challenging, since all novel and many rare missense variants must initially be classified as variants of unknown significance (VUS). For many rare variants, familial segregation data or *de novo* events in a trio were not available or sufficient to define variants as pathogenic or likely pathogenic.

When we observed *de novo* variants in a trio or co-segregation of a variant with disease, family information could be used as evidence supporting pathogenicity and reportability. Although failure of a variant to segregate might be considered evidence against pathogenicity, we did not use this information in our pathogenicity evaluations, since many disorders on the ACMG list are adult-onset and could be late to manifest in unaffected family members. Family sequence data is also useful for determining the phase of multiple heterozygous variants in a single gene, and thus can be used for finding compound heterozygous variants. When individuals rather than families are sequenced, it may be more difficult to classify variants as pathogenic or likely pathogenic since these sources of information are not available. However, support for pathogenicity can be obtained in other ways, such as literature searches. Thus, family information is helpful but not necessary for determining variant pathogenicity in most cases.

We found that prediction is even more difficult in genes that are subject to more variation because of their increased size or tolerance for variation. Since functional information was not available for most variants, 152/249 distinct variants (61.0%) were classified as VUS.

Some variants had been previously described in the dbSNP, EVS, and/or 1000 Genomes databases, and of 249 distinct variants, 126 (50.6%) were classified by at least one of the three databases used to assess variant classification (HGMD, ClinVar, and/or Emory). The available classifications in these cases were helpful but not always clear, mainly due to discrepancies among bioinformatic prediction databases or functional studies; low numbers of unrelated affected individuals with the same variant; misreporting of variant pathogenicity in the literature or classification databases; or occurrence of variants in large or highly mutable genes. These inconsistencies all point to the need for a single, centralized database and methodology for variant classification.

Our study has several limitations. First of all, many individuals in our cohort were ascertained for various Mendelian disorders. Because of this, two of the reportable findings could not be considered incidental, in that they may actually have contributed to the phenotype of the patients. In addition, our cohort was relatively small, and 73% of individuals self-identified with European descent. Moreover, we may have missed some variants due to low coverage in some regions of the 56 ACMG genes. Another limitation was the minor allele frequency cut-offs that we employed. For our early comparisons of variant classification between databases, we used a 1% MAF cut-off, which may be too high for some disorders and necessitate more time for manual review of variants. On the other hand, our later assessments of variant reportability used more stringent minor allele frequency cut-offs based on expected disease prevalences. This increased stringency may have resulted in exclusion of variants which were in fact pathogenic. In addition, the method we used to find reportable variants fails to consider functional studies. Because of this, we may have excluded some variants which have been functionally characterized as pathogenic or included variants which show milder or less conclusive functional consequences.

An alternative sequencing method like whole-genome sequencing would increase detection of incidental findings as well as VUS, owing to the increased coverage of this method and the challenge of interpreting noncoding variants. Whole-genome sequencing will uncover structural variants and copy-number variants as well. However, the ACMG recommendations currently apply only to SNVs and indels, so this additional variation—some of which may be significant—would not be considered under the current guidelines.¹

Based on this experience, we suggest that clinicians and researchers interested in returning incidental variants found by next-generation sequencing adopt a uniform, well-defined set of criteria for variant classification. Here we applied the ACMG recommendations and the classification criteria defined by Dorschner et al. (2013),²⁴ but we suggest that the addition of bioinformatic prediction tools, genic intolerance scores, and functional studies to the criteria would make variant classification more complete. We also expect that research and clinical laboratories performing next-generation sequencing will soon be more engaged in submitting variants to manually curated classification databases, improving interpretation of variants and counseling of patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We would like to thank all of our research participants who made this study possible. J.J. is a trainee of the Predoctoral Training Program in Human Genetics at Johns Hopkins University and is supported by the NIH grant T32GM07814. This work was funded by a grant from the National Human Genome Research Institute (U54HG006542).

REFERENCES

- 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. [PubMed: 23788249]
- Jarvik GP, Amendola LM, Berg JS, et al. Return of Genomic Results to Research Participants: The Floor, the Ceiling, and the Choices In Between. *Am J Hum Genet*. 2014; 94:818–826. [PubMed: 24814192]
- Bell CJ, Dinwiddie DL, Miller NA, et al. Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing. *Sci Transl Med*. 2011; 3(65):65ra4.
- Baylor-Hopkins Center for Mendelian Genomics (BHCMG). Available at: <http://mendeliangenomics.org>.
- Hamosh A, Sobreira N, Hoover-Fong J, Sutton VR, Boehm C, Schiettecatte F, Valle D. PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum. Mutat*. 2013; 34(4):566–571. [PubMed: 23378291]
- The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. [PubMed: 23128226]
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009a; 25(14):1754–1760. [PubMed: 19451168]
- Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAM tools. *Bioinformatics*. 2009b; 25(16):2078–2079. [PubMed: 19505943]
- Picard software. Available at <http://picard.sourceforge.net>.
- McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a Map Reduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20(9):1297–1303. [PubMed: 20644199]
- DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43(5):491–498. [PubMed: 21478889]
- Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2005 Jan 1; 33(Database issue):D501–D504. [PubMed: 15608248]
- RefSeq Genes Track. [Accessed April 25, 2012] The Cancer Genome Workbench. Available at: <https://cgwbncinoh.gov/cgi-bin/hgTrackUi?hgsid=79902&c=chr21&g=refGene>.
- Exome Variant Server (EVS). Seattle: NHLBI GO Exome Sequencing Project (ESP); <http://evs.gs.washington.edu/EVS/>. [Accessed June–August 2013]
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001; 29(1):308–311. [PubMed: 11125122]
- Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine; (dbSNP Build 137). Available from: <http://www.ncbi.nlm.nih.gov/SNP/>
- Stenson PD, Ball EV, Mort M, et al. The human gene mutation database (HGMD): 2003 update. *Hum Mutat*. 2003; 21:577–581. [PubMed: 12754702]
- The Human Gene Mutation Database. Available at <http://www.hgmd.cf.ac.uk/ac/>.
- Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014; 42(1):D980–D985. [PubMed: 24234437]
- [Accessed June–August 2013] ClinVar variant database. <http://www.ncbi.nlm.nih.gov/clinvar/>.

21. Bean LJH, 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. [PubMed: 23757202]
22. Emory Genetics Laboratory. <http://genetics.emory.edu/egl/emvclass/emvclass.php>.
23. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* 2013; 9(8):e1003709. [PubMed: 23990802]
24. Dorschner MO, Amendola LM, Turner EH, et al. Actionable, Pathogenic Incidental Findings in 1,000 Participants' Exomes. *Am J Hum Genet.* 2013; 93:1–10.
25. Astejada MN, Goto K, Nagano A, Ura S, Noguchi S, Nonaka I, Nishino I, Hayashi YK. Emerinopathy and laminopathy clinical, pathological and molecular features of muscular dystrophy with nuclear envelopathy in Japan. *Acta Myol.* 2007; 26:159–164. [PubMed: 18646565]
26. Parks SB, Kushner JD, Nauman D, et al. Lamin A/C mutation analysis in a cohort of 324 unrelated patients with idiopathic or familial dilated cardiomyopathy. *Am Heart J.* 2008; 156:161–169. [PubMed: 18585512]
27. Cowan J, Li D, Gonzalez Quintana J, Morales A, Hershberger RE. Morphological analysis of 13 LMNA variants identified in a cohort of 324 unrelated patients with idiopathic or familial dilated cardiomyopathy. *Circ Cardiovasc Genet.* 2010; 3(1):6–14. [PubMed: 20160190]
28. Payne SR, Kemp CJ. Tumor suppressor genetics. *Carcinogenesis.* 2005; 26(12):2031–2045. [PubMed: 16150895]
29. Morak M, Massdorf T, Sykora H, Kerscher M, Holinski-Feder E. First evidence for digenic inheritance in hereditary colorectal cancer by mutations in the base excision repair genes. *Eur J Cancer.* 2011; 47:1046–1055. [PubMed: 21195604]
30. Sieber OM, Lipton L, Crabtree M, et al. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N Engl J Med.* 2003; 348:791–799. [PubMed: 12606733]
31. Ehlermann P, Weichenhan D, Zehelein J, Steen H, Pribe R, Zeller R, Lehrke S, Zugck C, Ivandic BT, Katus HA. Adverse events in families with hypertrophic or dilated cardiomyopathy and mutations in the MYBPC3 gene. *BMC Med Genet.* 2008; 9:95. [PubMed: 18957093]
32. Rodriguez-Garcia MI, Monserrat L, Ortiz M, et al. Screening mutations in myosin binding protein C3 gene in a cohort of patients with Hypertrophic Cardiomyopathy. *BMC Med Genet.* 2010; 11:67. [PubMed: 20433692]
33. Olivetto I, Girolami F, Sciagra R, et al. Myofilament protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. *Mayo Clin Proc.* 2008; 83:630–638. [PubMed: 18533079]
34. Wang L, Guo DC, Cao J, et al. Mutations in myosin light chain kinase cause familial aortic dissections. *Am J Hum Genet.* 2010; 87(5):701–707. [PubMed: 21055718]
35. Online Mendelian Inheritance in Man, OMIM[®]. Baltimore, MD: McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University; 2014. World Wide Web URL: <http://omim.org/>
36. Johnston JJ, Rubinstein WS, Facio FM, et al. Secondary variants in individuals undergoing exome sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. *Am J Hum Genet.* 2012; 91(1):97–108. [PubMed: 22703879]
37. Lawrence L, Sincan M, Markello T, et al. The implications of familial incidental findings from exome sequencing: the NIH Undiagnosed Diseases Program experience. *Genet Med.* 2014; 16(10):741–750. [PubMed: 24784157]

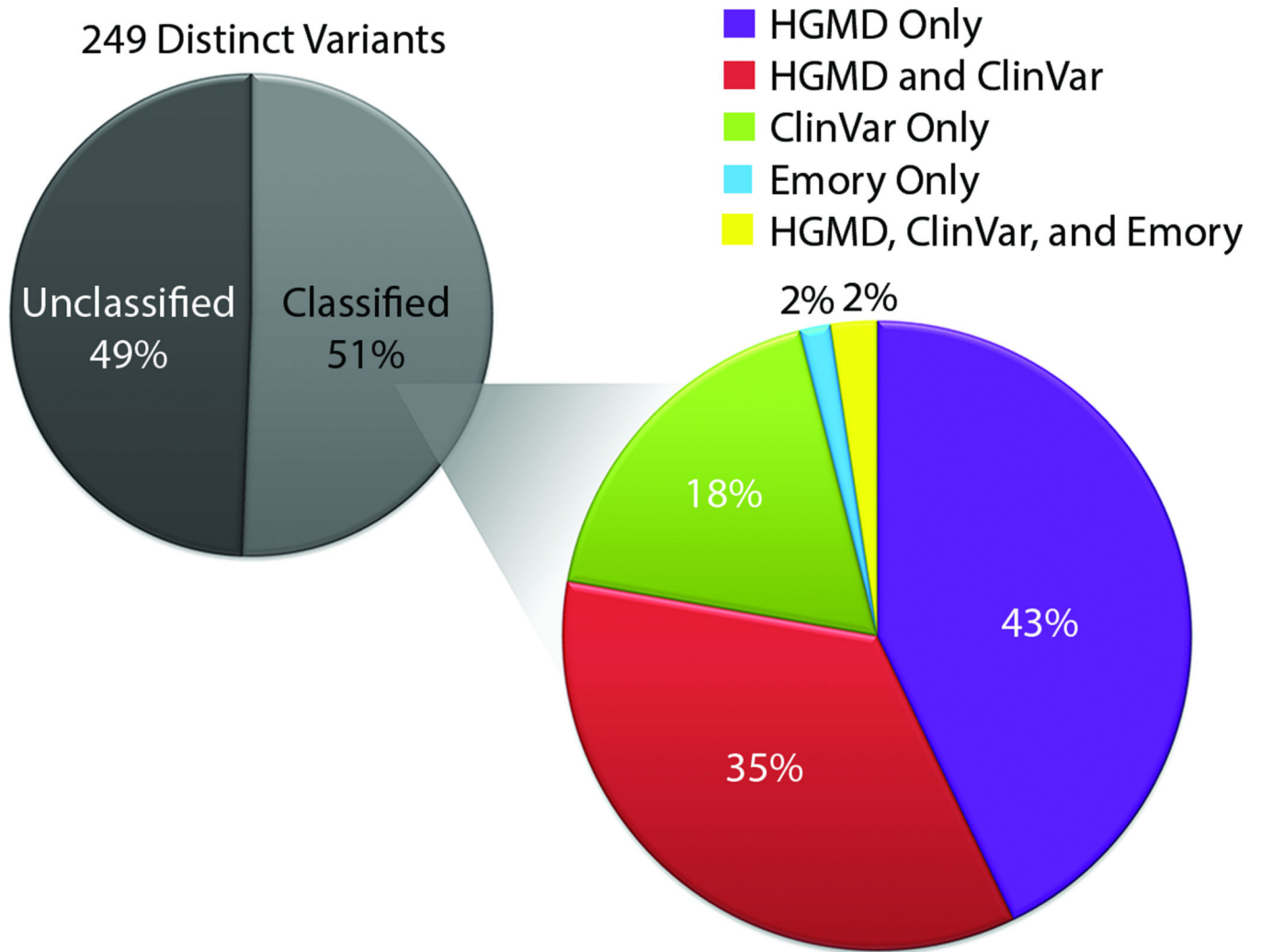


Figure 1. Classification of the 249 variants by the HGMD, ClinVar, and Emory databases.

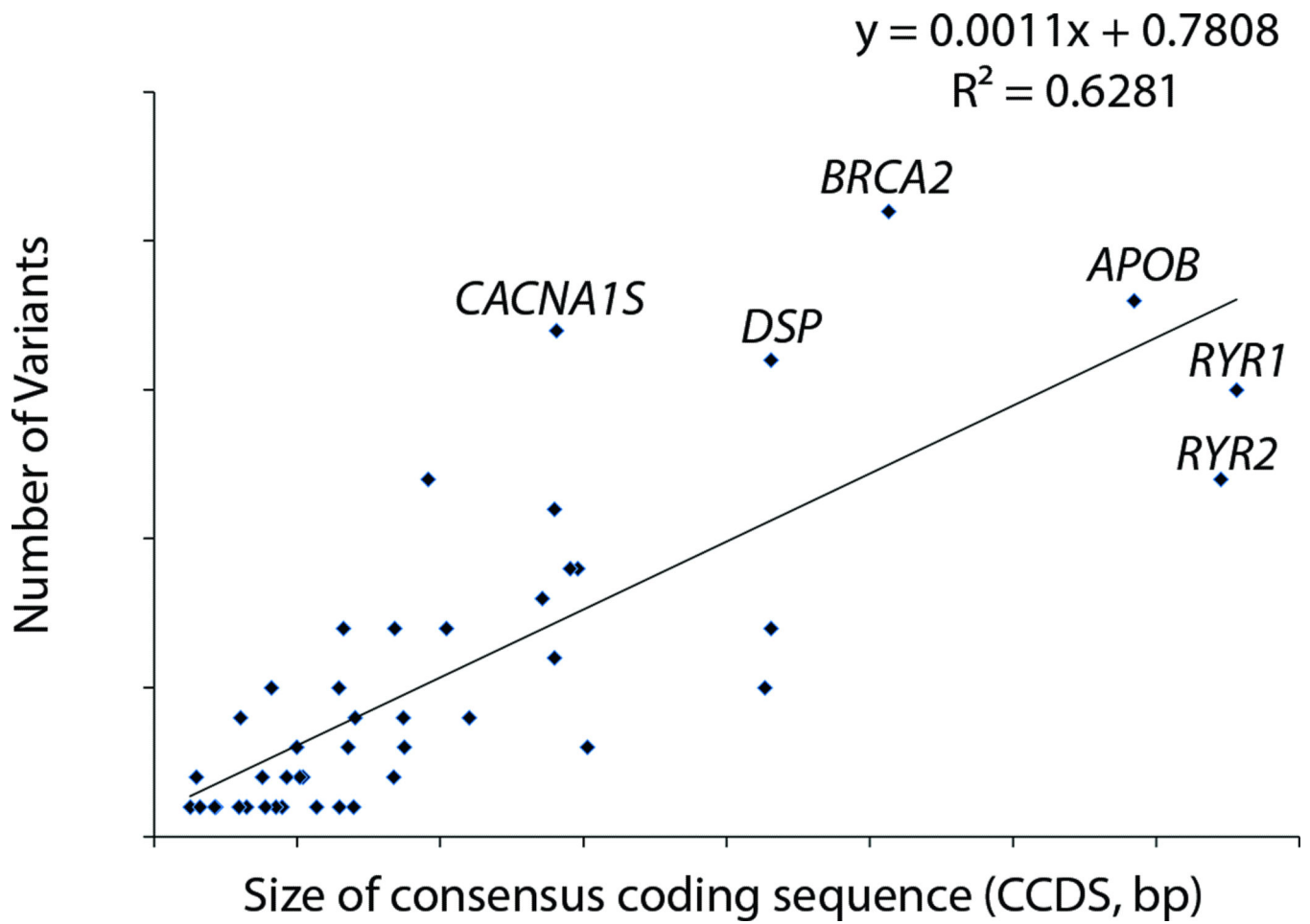


Figure 2.
The number of variants per gene plotted against the size (kb) of the consensus coding sequence (CCDS) of the gene.

Table 1

Classification of the 249 variants by mutation type.

Variant Type	Number of Variants
Missense SNV	231
Nonframeshifting indel	5
Nonsense	4
Exonic/Splicing	4
Frameshifting indel	3
Splicing	2
Total	249

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Reportable incidental variants identified in 232 individuals.

Gene	Chromosome	Genomic Change	cDNA & Protein Change	Criteria met	Classification	Reportable?	Incidental?	Patient ancestry
<i>LMNA</i>	1	156106743G>A	NM_005572.3: c.1412G>A; p.Arg471His	MAF < cutoff In 3 unrelated affected individuals ^{25, 26, 27}	Likely pathogenic VUS	Yes	No	European
<i>MSH2</i>	2	47656969C>T	NM_000251.2: c.1165C>T; p.Arg389*	MAF < cutoff Protein truncation in disease caused by haploinsufficiency ²⁸	Pathogenic	Yes	Yes	Admixed African/ European
<i>MUTYH</i>	1	45798269T>C	NM_001048173.1: c.583A>G; p.Ile195Val	MAF < cutoff Segregation in one affected family ²⁹ In one unrelated singleton ³⁰	Likely pathogenic VUS	No	Yes	European
<i>MYBPC3</i>	11	47356628G>C	NM_000256.3: c.2870C>G; p.Thr957Ser	MAF < cutoff In 3 unrelated affected individuals ^{31,32,33}	Likely pathogenic VUS	Yes	No	European
<i>MYLK</i>	3	123512573delG	NM_053025.3: c.116delC; p.Pro39Leufs*26	MAF < cutoff Protein truncation in disease caused by haploinsufficiency ³⁴	Pathogenic	Yes	Yes	Latino