

RESEARCH

Open Access



Prevalence of pathogenic/likely pathogenic variants in the 24 cancer genes of the ACMG Secondary Findings v2.0 list in a large cancer cohort and ethnicity-matched controls

Jung Kim¹, Wen Luo², Mingyi Wang², Talia Wegman-Ostrosky^{1,3}, Megan N. Frone¹, Jennifer J. Johnston⁴, Michael L. Nickerson⁵, Melissa Rotunno⁶, Shengchao A. Li², Maria I. Achatz^{1,22}, Seth A. Brodie², Michael Dean⁵, Kelvin C. de Andrade^{1,7}, Fernanda P. Fortes^{1,7}, Matthew Gianferante¹, Payal Khincha¹, Mary L. McMaster¹, Lisa J. McReynolds¹, Alexander Pemov¹, Maisa Pinheiro¹, Karina M. Santiago^{1,7}, Blanche P. Alter¹, Neil E. Caporaso⁸, Shahinaz M. Gadalla¹, Lynn R. Goldin⁹, Mark H. Greene¹, Jennifer Loud¹, Xiaohong R. Yang⁹, Neal D. Freedman¹⁰, Susan M. Gapstur¹¹, Mia M. Gaudet¹¹, Donato Calista¹², Paola Ghiorzo¹³, Maria Concetta Fagnoli¹⁴, Eduardo Nagore¹⁵, Ketty Peris¹⁶, Susana Puig¹⁷, Maria Teresa Landi⁹, Belynda Hicks², Bin Zhu², Jia Liu², Joshua N. Sampson¹⁸, Stephen J. Chanock¹⁹, Lisa J. Mirabello¹, Lindsay M. Morton²⁰, Leslie G. Biesecker⁴, Margaret A. Tucker²¹, Sharon A. Savage¹, Alisa M. Goldstein^{1*†} and Douglas R. Stewart^{1*†}

Abstract

Background: Prior research has established that the prevalence of pathogenic/likely pathogenic (P/LP) variants across all of the American College of Medical Genetics (ACMG) Secondary Findings (SF) genes is approximately 0.8–5%. We investigated the prevalence of P/LP variants in the 24 ACMG SF v2.0 *cancer* genes in a family-based cancer research cohort ($n = 1173$) and in cancer-free ethnicity-matched controls ($n = 982$).

Methods: We used InterVar to classify variants and subsequently conducted a manual review to further examine variants of unknown significance (VUS).

Results: In the 24 genes on the ACMG SF v2.0 list associated with a *cancer* phenotype, we observed 8 P/LP unique variants (8 individuals; 0.8%) in controls and 11 P/LP unique variants (14 individuals; 1.2%) in cases, a non-significant difference. We reviewed 115 VUS. The median estimated per-variant review time required was 30 min; the first variant within a gene took significantly ($p = 0.0009$) longer to review (median = 60 min) compared with subsequent variants (median = 30 min). The concordance rate was 83.3% for the variants examined by two reviewers.

Conclusion: The 115 VUS required database and literature review, a time- and labor-intensive process hampered by the difficulty in interpreting conflicting P/LP determinations. By rigorously investigating the 24 ACMG SF v2.0 *cancer* genes, our work establishes a benchmark P/LP variant prevalence rate in a familial cancer cohort and controls.

Keywords: ACMG secondary findings, Familial cancer exome, Population study, Variant classification

* Correspondence: goldstea@mail.nih.gov; drstewart@mail.nih.gov

† Alisa M. Goldstein and Douglas R. Stewart contributed equally to this work.

¹Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA

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



Background

In 2013, the American College of Medical Genetics and Genomics (ACMG) recommended that “laboratories performing clinical [exome or genome] sequencing seek and report mutations of the specified classes or types” in a set of 56 genes associated with a severe phenotype, and for which disease risk may be reduced or managed before symptoms arise [1, 2]. These recommendations for reporting of incidental (or secondary) findings (SF) in clinical exome and genome sequencing were later amended to 59 genes (ACMG SF v2.0) [3].

Although both ACMG SF policy statements used the older “known pathogenic” or “expected pathogenic” variant categorization terminology [4], a transition to the newer five-category system of pathogenicity has been urged [5, 6]. To date, multiple studies using the newer pathogenicity scheme to investigate clinical exome sequencing data and publicly available sequence databases in primarily European-American and African-American cohorts have estimated the prevalence of ACMG SF gene list (the original 2013 list and 2017 amendment) pathogenic/likely pathogenic (P/LP) variants to be approximately 0.8–5% [7, 8]. Some, but not all, studies of ethnically diverse cohorts have found higher prevalence of P/LP (5.6–7%) for ACMG SF genes [9–11]. The prevalence of P/LP variants in *cancer* cohorts remains largely uninvestigated, and to our knowledge, prevalence of P/LP variants has not been determined in a large cancer study with ethnicity-matched healthy controls.

The current American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines use conservative methods to classify variants based on numerous criteria including clinical and family history, previous literature, and known population allele frequency [5]. Currently, there are 28 criteria used to determine final variant classification, and the use of these criteria is labor-intensive. One strategy for applying the ACMG/AMP guidelines is a consensus-based tumor board-like review by experts for genes/variants of interest. However, this approach is relatively low-throughput, labor-intensive, and not realistic for large-scale sequencing efforts. Another strategy would employ automated procedures. The software package InterVar was developed as a semi-automated approach to applying the ACMG/AMP guidelines [12]. It incorporates 10 of 28 ACMG/AMP criteria automatically; the remaining 18 criteria can be applied following manual review of a variant in the literature, if published.

In this study, we used the most recent ACMG/AMP criteria and determined the prevalence of P/LP variation in the 24 ACMG SF v2.0 gene list associated with a *cancer* phenotype in a large, family-based heterogeneous cancer research cohort ($n = 1173$ individuals; 738 families) and in ethnicity-matched controls ($n = 982$). (The

remaining 35 *non-cancer* ACMG SF v2.0 genes were not fully investigated.) We used InterVar to classify variants, followed by a manual review to further examine variants of unknown significance (VUS). In addition, we estimated the time to resolve VUS and evaluated the concordance rate between reviewers for 30% of the reviewed variants.

Methods

DCEG familial exome cohort and cancer-free controls, anonymization, and ethics review

Cases were drawn from the NCI Division of Cancer Epidemiology and Genetics (DCEG) Familial Exome cohort, a large, long-term, longitudinal, heterogeneous group of family-based studies with a cancer phenotype and a Mendelian or near-Mendelian pattern of inheritance. The majority of the families lacked a known causative germline genetic variant; the cancer phenotype in the families may or may not overlap with the known cancer phenotype of the 24 ACMG v2.0 cancer genes. Families in which a causative gene was identified were not excluded. Data from 982 controls from 2 cohort studies, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO [13]) and the Cancer Prevention Study (CPSII) of the American Cancer Society [14], and 1 case-control study, the Environment and Genes in Lung Cancer Etiology (EAGLE [15]), were available for inclusion in the current study. Controls were cancer-free at the time of enrollment. Controls in the CPSII and PLCO studies were followed longitudinally, and if cancer developed, this was noted; EAGLE controls were not followed longitudinally. All participants provided written consent and were recruited through IRB-approved protocols. For these analyses, cases and controls underwent irrevocable anonymization. The project was reviewed and approved by the NIH Office of Human Subjects Research Protection, which granted a waiver of the IRB review requirement.

Exome sequencing, quality control, ethnicity determination, and analysis of population stratification

Exome sequencing was performed at the Cancer Genomics Research Laboratory, National Cancer Institute (CGR, NCI), as described [16, 17]. Cases and controls were matched using an ethnicity-informative variation [18]. After the controls and cases were matched, poor quality and contaminated samples were excluded from the dataset. Any variants that were flagged with our pipeline quality control metric (CScorefilter), had a read depth < 10 , ABHet < 0.2 or > 0.8 , or did not pass other quality control filters were excluded from the analysis. All variants were further filtered using $\text{pop_maxfreq} < 0.01$, see Additional file 1: Supplemental Methods for additional details.

Automated and manual review of variation in the 24 ACMG SF v2.0 cancer genes

Variation in the 24 ACMG SF v2.0 genes primarily associated with a *cancer* phenotype (“ACMG SF v2.0 cancer”: *APC*, *BMPRIA*, *BRCA1*, *BRCA2*, *MEN1*, *MLH1*, *MSH2*, *MSH6*, *MUTYH*, *NF2*, *PMS2*, *PTEN*, *RBI*, *RET*, *SDHB*, *SDHC*, *SDHD*, *SMAD4*, *STK11*, *TP53*, *TSC1*, *TSC2*, *VHL*, *WT1*) was annotated using ANNOVAR [19], which included InterVar, a semi-automated software tool which applies the ACMG-AMP guidelines [12]. To more fully classify potentially pathogenic variants, all ACMG SF v2.0 *cancer* gene variants listed in the Human Gene Mutation Database (HGMD; version 2015.2; Qiagen, Cardiff, Wales, UK) as “disease mutation” (DM) underwent manual review, regardless of the InterVar assertion and without knowledge of case or control status. In addition, we used Google Scholar to search the published literature for information on variants designated VUS by InterVar which were not listed in HGMD. The primary literature was then reviewed by 17 reviewers (including oncologists, hematologists, clinical geneticists, genetic counselors, geneticists, or genetic epidemiologists). The reviewers were assigned specific gene(s) after variant review training and classified the variants according to the ACMG/AMP guidelines using a pre-populated Excel file that contained needed variant annotation information. Reviewers were also asked to provide comments for each score provided and to estimate the time needed to review each variant. We noted which variant within each gene was the first one evaluated by each reviewer. After initial review, variants were subject to a quality control (QC) process in which the criteria for scoring and reviewer comments were compared for agreement. As a second QC check, 31% ($n=36$) of the 115 variants initially reviewed were re-evaluated by a second independent reviewer. If there was discordance between the primary and secondary reviewers on variant classification, discussion was initiated to reach consensus. The ACMG/AMP combining criteria were implemented using the Genetic Variant Interpretation Tool available online (http://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/) [20]. Graph and p values (t test) were calculated using GraphPad Prism 7 (GraphPad Software Inc., La Jolla, CA), and 95% confidence intervals were calculated using STATA 14 (StataCorp LLC, College Station, TX).

Results

Sequence quality, demographics, and matching cases and controls

For the entire DCEG Familial Exome cohort (plus controls), exome sequencing was performed such that 88% of coding sequence from the University of California Santa Cruz (UCSC) human genome (hg) 19 transcripts

database had ≥ 15 reads with an average coverage of 61 \times . After the sample quality control, there were 982 control individuals from the PLCO, EAGLE, and CPSII cohorts and 1173 cases (738 families) from 15 cancer-based studies (Tables 1 and 2, Additional file 2: Table S1). Population stratification for Northern and Western European ancestry (CEU) > 0.80 (Additional file 2: Figure S1) resulted in well-matched cases and controls by principal component analysis (Additional file 2: Figure S2).

InterVar classification of ACMG SF v2.0 cancer and non-cancer genes prior to expert review

We used InterVar to classify all filtered variants into 6 categories (pathogenic (P), likely pathogenic (LP), variant of unknown significance (VUS), likely benign (LB), benign (B), and no classification) for cases and controls. Since our cohort includes family members, we performed 2 separate analyses: first, we used all cases, and second, we randomly selected 1 affected individual per family. Table 3 shows the InterVar classification of the variants for *all* ACMG SF v2.0 genes, divided into “cancer genes” and “non-cancer genes” columns. In cancer genes, there were 760 variants deemed VUS or “no classification”; “no classification” variants were primarily intronic, located in the 5’ or 3’ untranslated regions, or indels. There were 8 unique P variants (controls and cases); 2 were in *MUTYH*. *MUTYH* is the only ACMG SF v2.0 *cancer* gene in which the phenotype is associated with an autosomal recessive pattern of inheritance and is therefore reportable only for compound heterozygotes or homozygotes [21]. Since all subjects in this study harbored only 1 P/LP *MUTYH* variant, we excluded this gene from our prevalence calculation.

InterVar classification of ACMG SF v2.0 cancer genes after expert review

Of the InterVar-determined cancer gene VUS ($n=297$) or “no classification” variants ($n=463$) in cases and controls, 115 (15%) had been reported previously, as per queries of HGMD and Google Scholar (VUS variants only). A total of 77 variants were classified as “DM” in HGMD, and an additional 38 VUS variants were identified in the searchable published literature queried

Table 1 Demographic characteristics of cases

Variable	Cases
Ancestry	European (CEU > 0.80)
Sex	M: 682 (58%) F: 491 (42%)
Number of families	738 families
Number of individuals	1173
Involved studies	See Additional file 2: Table S1
CEU Northern and Western European ancestry	

Table 2 Demographic characteristics of controls

Variable	CPSII	PLCO	EAGLE
Ancestry		European (CEU > 0.80)	
Sex	M: 106 (49%) F: 111 (51%)	M: 217 (59%) F: 153 (41%)	M: 314 (79%) F: 81 (21%)
Number of families		No families	
Number of individuals	217	370	395
Average age (years)	71	67	66
Average follow-up (years)	11.1	9.3	N/A
Number of individuals with a cancer found during follow-up	7	32	Were not followed

CEU Northern and Western European ancestry, CPSII Cancer Prevention Study II (American Cancer Society), EAGLE Environment and Genes in Lung Cancer Etiology, F female, M male, N/A not applicable, PLCO Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial

through Google Scholar. Of the remaining 645 variants, there was little or no additional published or online information available, and therefore, these variants were not further evaluated. After review by 1 cancer expert, 36 (31%) randomly selected variants underwent review by a second cancer expert. The concordance rate between the primary and secondary reviewers for the pathogenicity category of these 36 variants was 83.3%. Discussion between reviewers led to the resolution of the 6 discrepant variants from the 36 re-reviewed variants (16.7%) in this study. Among the 115 variants reviewed, 2 unique variants were promoted to P from VUS and 5 unique variants were promoted to P from “no classification.” Two unique variants were promoted to LP from VUS, and 1 unique variant was promoted to LP from “no classification” (Additional file 3: Table S2).

Prevalence of P/LP variation in cases and controls and estimated time to review

The allele and total counts of P/LP variants for the 24 ACMG SF v2.0 *cancer* genes after expert review for cases and controls are shown in Table 4. The prevalence of P/LP variants among controls was 0.8% (95% confidence interval (CI) 0.3–1.4%), among cases, 1.2% (95% CI 0.6–1.8%), and for one case per family, 1.1% (95% CI 0.3–1.8%). In controls, the P/LP alleles were in *BRCA2* (five unique), *MSH2* (one), *PMS2* (one), and *TP53* (one)

(Additional file 3: Table S2). In cases, the P/LP alleles were in *BRCA1* (one) *BRCA2* (one), *PMS2* (one), and *TP53* (eight unique) (Additional file 3: Table S2). There were no significant differences in the prevalence of P/LP variants between controls and either case set (Table 4). Reviewers needed an estimated median of 30 min (range = 5–240 min) per variant to review the pertinent literature, to consult the ACMG/AMP guidelines, and to make a judgment on the classification criteria (Fig. 1). The first variant examined within a gene took significantly longer ($p = 0.0009$) to review (median = 60 min; range = 10–240 min) compared with subsequent variants in the same gene (median = 30 min; range = 5–117 min). However, these estimated times did not account for the time required to run InterVar, perform a QC check, conduct secondary reviewer validation, and resolve discordances. Incorporating these additional tasks into the review process would result in a much higher time requirement to classify variants.

Discussion

In 1173 individuals from a heterogeneous, family-based study of inherited cancer predisposition, the prevalence of P/LP variants in the 24 ACMG SF v2.0 *cancer* genes was 1.2%, not significantly different from P/LP variant prevalence in 982 ethnicity-matched controls (0.8%). Our study is notable for the large cohort size, the use of

Table 3 ACMG SF v2.0 genes classified per ACMG/AMP guidelines using InterVar software

Classification	Controls (982 exomes)		All cases (1173 exomes)		1 case/family (738 exomes)	
	Cancer genes	Non-cancer genes	Cancer genes	Non-cancer genes	Cancer genes	Non-cancer genes
Pathogenic	4 (17)	0	4 (13)	0	4 (7)	0
Likely pathogenic	0 (0)	8 (8)	0 (0)	11 (17)	0 (0)	10 (11)
Variant of unknown significance	141 (174)	403 (503)	179 (260)	440 (713)	140 (167)	336 (454)
Likely benign	248 (463)	398 (671)	242 (570)	417 (865)	202 (355)	332 (541)
Benign	12 (42)	30 (208)	12 (106)	28 (266)	11 (69)	26 (167)
No classification*	252 (3477)	832 (8181)	290 (4315)	814 (10393)	225 (2714)	668 (6535)

Numbers represent unique variant count and number in parenthesis represents allele counts

*Majority of the “no classification” variants were intronic, 5'/3' untranslated regions, and 22 indels

Table 4 Pathogenic and likely pathogenic variants in 24 ACMG SF v2.0 *cancer* genes after expert review

Classification	Controls (982 exomes)	All cases (1173 exomes)	1 case/family (738 exomes)
Pathogenic	6 (6)	5 (6)	3 (3)
Likely pathogenic	2 (2)	6 (8)	4 (5)
Prevalence of pathogenic and likely pathogenic variants	8/982 0.8% 95% CI 0.3–1.4%	14/1173 1.2% 95% CI 0.6–1.8%	8/738 1.1% 95% CI 0.3–1.8%
<i>p</i> value (controls vs. all cases)		0.5196	
<i>p</i> value (controls vs. 1 case/families)			0.6171

Variants in *MUTYH* were excluded from counts since it underlies a recessive disorder and no *MUTYH* homozygotes or compound heterozygotes were observed in cases or controls. The first number represents unique variant count and number in parenthesis represents allele counts
CI confidence interval

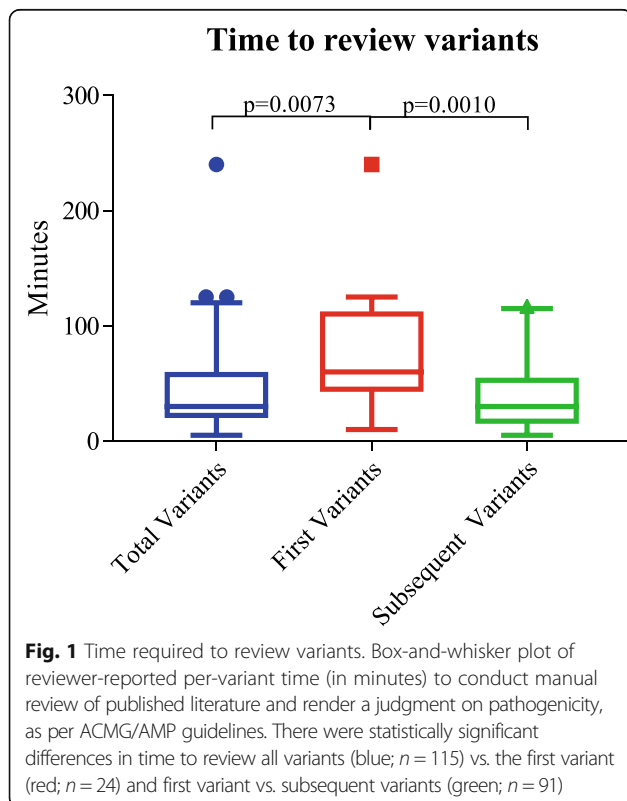
variation-based ethnicity-matching of cases and controls, thorough expert-driven review of variants by ACMG/AMP criteria, and an exclusive focus on the 24 ACMG SF v2.0 *cancer* genes.

Direct comparison of our results with previous studies is challenging because of the differences in methodology and study populations. We acknowledge that our familial cancer cohort is heterogeneous since it is comprised of individuals drawn from a wide variety of familial tumor-predisposition studies, making comparison difficult. Analyses of the 1000 Genomes and the NHLBI GO Exome Sequencing Project cohorts for P/LP variants in a list of “medically actionable” genes (larger than the ACMG SF v2.0 list) found a prevalence of 2.2–3.4% [11, 22]. The P/LP prevalence rate (for the original (v1.0)

ACMG SF 56-gene list) in smaller, single-institution research cohorts spanned an order of magnitude from 0.86% (Baylor-Hopkins Center for Mendelian Genomics) [23] to 8.8% (Undiagnosed Disease Project) [24].

Data on the prevalence of P/LP variants in cancer cohorts are sparse. To be comprehensive in this research study, we considered both P and LP variants in our analysis, although the threshold to report LP variation from ACMG SF v2.0 genes in a clinical setting is under debate [6]. One study of 439 individuals undergoing tumor-germline dyad sequencing found that 4.3% harbored a germline variant (in a panel of 247 genes) indicative of hereditary cancer predisposition [25]. A study of 392 patients with pancreatic cancer undergoing tumor/normal sequencing found a prevalence rate of pathogenic variation of 5.1% from a panel of 130 genes [26]. We were not able to find publications that reported prevalence of P/LP for all ACMG SF genes in cancer cohorts. The lower prevalence rate we observed in our study compared with prior publications may be attributable to our evaluating only a subset of known cancer susceptibility genes. In addition, the 24 ACMG SF v2.0 *cancer* genes largely underlie risk in common cancers (e.g., breast, ovarian, and colon cancer) and well-known genetic disorders (e.g., Li-Fraumeni syndrome, retinoblastoma) and are not necessarily associated with the disorders constituting our study cohort (Additional file 2: Table S1). Although one of our studies recruited individuals with a history of familial breast and ovarian cancer, eligibility required documentation of negative germline *BRCA1/2* genetic testing (Additional file 2: Table S1).

The ethnicity-matched controls (PLCO/EAGLE/CPSII) were on average 70 years of age, healthy adults without a history of cancer (other than non-melanoma skin cancer) at the time of study enrollment and sample collection. Interestingly, 0.8% of this control sample harbored a P/LP variant in one of the 24 ACMG SF v2.0 *cancer* genes, not significantly different from the cancer cohort ($p = 0.5$) (Table 4). Furthermore, participants in the CPSII and PLCO study have been followed for an average of 10 years after sample collection. During this follow-up, out of 586



participants from CPSII and PLCO, 39 participants developed cancer. Considering only controls who did not develop cancer after research follow-up, we found a similar prevalence of P/LP variants compared with all controls (1.2% vs. 1.5%, respectively).

By rigorously investigating P/LP variation in the 24 ACMG SF v2.0 *cancer* genes, our work establishes a clinically useful benchmark prevalence rate, especially in controls. Recent studies have shown that pathogenic variation in single genes like *DICER1* [27] and *TP53* [28] (in public datasets like non-TCGA ExAC, 1000G, and ESP) have a higher prevalence than the known or expected population frequency of their associated syndromes. In the case of *DICER1* and *TP53*, the recognition that pathogenic variation in recognized cancer genes is more common than expected is an important, emerging, and unanticipated finding from population-based exome sequencing, one that has significant clinical implications. In this study, we observed P/LP variation in the 24 ACMG SF v2.0 *cancer* genes (specifically, *BRCA2*, *MLH1*, *MSH2*, *PMS2*, and *TP53*) in 0.8% of our 982 controls, who, by a mean age of ~70 years, had not developed any malignancy. Thus, in our controls, the prevalence of P/LP germline variation in *BRCA2* was 0.5% (all subjects 5/982, females only 1/345; none were common Ashkenazi variants). In Lynch syndrome genes, the prevalence was 0.4% (*MLH1*, *MSH2*, *PMS2*; excluding *MUTYH* 4/982), and for Li-Fraumeni syndrome, it was 0.1% (*TP53*; 1/982). These frequencies are comparable to other published estimates (*BRCA2* 0.45% in cancer-free Australian women [29]; 0.31% in women of European non-Finnish descent in the Exome Aggregation Consortium, excluding The Cancer Genome Atlas data [30]; Lynch 0.2% [31]). We acknowledge that our controls may not be representative of the entire general population since, as volunteers, the controls may have an interest in cancer studies perhaps due to a family history of cancer.

Reviewers were required to track the estimated amount of time needed to classify variants. Our study is the first to distinguish between the amount of time to review first and subsequent variants within a gene. We found that the first variant took significantly longer to review when compared with subsequent variants, a reflection of the learning curve inherent in applying these new, complex classification algorithms. Although our team was composed of cancer experts, they were not necessarily experts on the specific genes they were reviewing. This could potentially have led to the additional time for familiarization with the gene(s) to be reviewed. Our overall finding that variant review was time-consuming is consistent with previous studies [32]. However, in some clinical labs, a more sophisticated automated pipeline and highly trained variant specialists

would likely result in shorter review times. We note that our measurements reflect only the estimated time to review the primary literature and do not include the time required to conduct InterVar classification, secondary review, and consensus-seeking or summation of the ACMG/AMP scores. Since the 24 ACMG SF v2.0 *cancer* genes are recognized and generally well-studied, the amount of available literature (and time spent reviewing it) may be greater than for lesser-known (non-ACMG SF v2.0) genes. In addition, our study population was restricted to people of non-Finnish European ancestry. Published work has highlighted the additional challenges in interpreting genetic variation in non-European populations [33]. Thus, our variant interpretation times may have been shorter compared with those of non-European populations.

Our experience with InterVar and the ACMG/AMP guidelines deserves a brief comment. We found that InterVar was a useful tool to start the initial variant classification using the ACMG/AMP guidelines. Despite the use of InterVar and manual review, most variants remain unresolved due to the lack of published literature and for our study and limited clinical information. Proper classification of variants, especially those used in clinical decision-making, is a time-consuming and laborious process that, for now, requires human expertise and judgment. Currently, this process is more subjective, and yields less reproducible results, than is optimal. In the future, this may be streamlined with more extensive, comprehensive electronic databases of definitively classified variants, more sophisticated software (e.g., neural networks) [34], and artificial intelligence programs (e.g., machine learning) [35], based on formal, probabilistic frameworks [36].

Reviewers in this study frequently noted that gaining a working familiarity with the ACMG/AMP guidelines was demanding. As a quality control procedure, we compared criterion scores (0 or 1) with the respective comments provided by the reviewer; we observed confusion related to multiple criteria. Despite these challenges and after correction of inconsistently scored criteria (based on the comments provided), our secondary review and consensus process showed a concordance rate of 83.3%, which is at the upper limit of previously reported concordance rates (34–79%) [11, 23, 37, 38]. In many cases, ambiguous words in the ACMG/AMP criteria such as “well-established” (e.g., in criteria PM1, PS3, BS3) and “multiple” (PP1, BP4) are subjective and unavoidably led to discrepancies between reviewers in the consensus process. These differences in criteria interpretation were resolvable with a discussion between the primary and secondary reviewers. Suggestions to refine the wording of the ACMG/AMP guidelines, as well as other practical improvements (e.g., specific cutoff

MAF for each disease, resources for which genes cause disease by loss of function, which functional assays are appropriate, and the quantitative threshold for segregation) have been promulgated [37]. To resolve this ambiguity, the Clinical Genome Resource (ClinGen) [39] is working with experts in the field to refine the guidelines. For hereditary cancer, there are five different working groups (breast and ovarian cancer, *CDH1*, colon cancer, *PTEN*, and *TP53*).

We acknowledge the limitations of our study. Since cases and controls were anonymized, there were restrictions on the depth and detail of clinical information. Thus, we were not able to assess *de novo* or *cis/trans* status or to assess segregation of a variant with phenotype. Availability of these data may have increased the number of variants that were definitively classifiable, which would have reduced the number of VUS. In addition, we did not review all VUS variants called by InterVar; we only considered the 115 variants that were reported by HGMD as DM or for which sufficient information was found in Google Scholar. Furthermore, this study only examined the white population, potentially limiting the applicability of these findings on P/LP prevalence to other ethnic groups and variant review time. Lastly, although the cases were drawn from a heterogeneous, convenient cohort of families assembled over multiple decades and protocols, the generalizability of our results may be limited, given the broad spectrum of cancer diagnoses.

Conclusions

We found a non-significant difference in the prevalence of P/LP variants from the ACMG SF v2.0 *cancer* genes in a cancer cohort (1.2%) and ethnically matched healthy controls (0.8%). Variant review, even with the help of sophisticated software tools, is time-consuming. Newer approaches, perhaps using artificial intelligence tools and neural networks, are needed to simplify and expedite this important task.

Additional files

Additional file 1: Supplemental Methods. Detailed methods on DNA preparation, hybridization, exome sequencing, variant calling, use of ethnicity-informative variation, quality control, and capture region matching. (PDF 233 kb)

Additional file 2: Figure S1. Population stratification of cancer cases and controls. Figure S2. Principal component analysis of cancer cases and controls. Table S1. Study names and predominant cancer types in DCEG Familial Exome cohort. (PDF 449 kb)

Additional file 3: Table S2. InterVar classification of ACMG SF v2.0 cancer genes after expert review. (XLSX 30 kb)

Abbreviations

ACMG: American College of Medical Genetics and Genomics; ACMG/AMP: American College of Medical Genetics and Genomics/Association for

Molecular Pathology; B: Benign; CEU: Northern and Western European ancestry; CGR, NCI: Cancer Genomics Research Laboratory, National Cancer Institute; CPISII: Cancer Prevention Study; DCEG: Division of Cancer Epidemiology and Genetics; DM: Disease mutation; EAGLE: Environment and Genes in Lung Cancer Etiology; HGMD: Human Gene Mutation Database; LB: Likely benign; LP: Likely pathogenic; P: Pathogenic; P/LP: Pathogenic/likely pathogenic; PLCO: Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; QC: Quality control; SF: Secondary finding; VUS: Variants of unknown significance

Acknowledgements

This work utilized the computational resources of the NIH High Performance Computing Biowulf Cluster.

Funding

This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics of the National Cancer Institute, Bethesda, MD. JJJ and LGB were supported by grants HG200359 09 and HG200387 04 by the Intramural Research Program National Human Genome Research Institute. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Availability of data and materials

The datasets generated and analyzed in this study are not publicly available due to the anonymization of the data. The only person with access to both the deidentified ID and original ID is the study "honest broker." If the genetic data from this study is posted, it may be feasible to identify individual participants using previously posted data. This would violate the NIH Office of Human Subjects Research Protection approval for this study, in which we certified that we would not in any way link the deidentified ID with the original ID. However, data from this study can be made available from the authors upon reasonable request.

Authors' contributions

JK analyzed and interpreted data and was a major contributor in writing the manuscript. WL, MW, AL, BA, NEC, SMG, LRG, JL, XRY, NDF, SG, MTL, BH, BZ, SJC, LM, LM, MAT, and SAS analyzed and interpreted the data. TW, MNF, JJJ, MLN, MR, MIA, SAB, MD, KCA, FF, MG, PK, MLM, LJM, AP, MP, and KS analyzed the original classification. JS consulted the statistical analysis. MHG, MMG, and LGB analyzed and interpreted the data and were a major contributors in editing the manuscript. AMG and DRS conceptualized, analyzed, and interpreted the data and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All participants provided written consent and were recruited through IRB-approved protocols. For these analyses, cases and controls underwent irrevocable anonymization. The project was reviewed and approved by the NIH Office of Human Subjects Research Protection, which granted a waiver of the IRB review requirement. The research was conducted in accordance with the declaration of Helsinki.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA. ²Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Leidos Biomedical Research, Inc., Frederick, MD 21701, USA. ³División de Investigación, Instituto Nacional de Cancerología, 14080 Mexico City, Mexico. ⁴Medical Genomics and Metabolic Genetics Branch, National

Human Genome Research Institute, NIH, Bethesda, MD 20892, USA.

⁵Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Gaithersburg, MD 20877, USA.

⁶Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH, Rockville, MD 20850, USA. ⁷International Research Center, A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil. ⁸Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA. ⁹Integrative Tumor Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA. ¹⁰Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA. ¹¹Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, USA. ¹²Department of Dermatology, Maurizio Bufalini Hospital, Cesena, Italy. ¹³Department of Internal Medicine and Medical Specialties, University of Genoa and Genetics of Rare Cancers, IRCCS Ospedale Policlinico San Martino, Genoa, Italy. ¹⁴Department of Dermatology, University of L'Aquila, L'Aquila, Italy. ¹⁵Department of Dermatology, Instituto Valenciano de Oncología, Valencia, Spain. ¹⁶Institute of Dermatology, Catholic University - Fondazione Policlinico Universitario A. Gemelli, IRCCS, Rome, Italy. ¹⁷Dermatology Department, Melanoma Unit, Hospital Clinic de Barcelona, IDIBAPS, Universitat de Barcelona, Barcelona, Spain & Centro de Investigación Biomedica en Red en Enfermedades Raras (CIBERER), Valencia, Spain. ¹⁸Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA. ¹⁹Office of the Director, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA. ²⁰Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA. ²¹Division of Cancer Epidemiology and Genetics, Human Genetics Program National Cancer Institute, NIH, Rockville, MD 20850, USA. ²²Centro de Oncologia, Hospital Sirio-Libanés, Sao Paulo, SP 013050-050, Brazil.

Received: 17 April 2018 Accepted: 5 December 2018

Published online: 24 December 2018

References

- Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med.* 2013;15(7):565–74.
- ACMG Board of Directors. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genet Med.* 2015;17(1):68–9.
- Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med.* 2017;19(2):249–55.
- Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE. Molecular subcommittee of the ALQAC: ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet Med.* 2008;10(4):294–300.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. 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.
- Biesecker LG. ACMG secondary findings 2.0. *Genet Med.* 2017;19(5):604.
- Natarajan P, Gold NB, Bick AG, McLaughlin H, Kraft P, Rehm HL, Peloso GM, Wilson JG, Correa A, Seidman JG, et al. Aggregate penetrance of genomic variants for actionable disorders in European and African Americans. *Sci Transl Med.* 2016;8(364):364ra151.
- Olfson E, Cottrell CE, Davidson NO, Gurnett CA, Heusel JW, Stitzel NO, Chen LS, Hartz S, Nagarajan R, Saccone NL, et al. Identification of medically actionable secondary findings in the 1000 genomes. *PLoS One.* 2015;10(9):e0135193.
- Gambin T, Jhangiani SN, Below JE, Campbell IM, Wiszniewski W, Muzny DM, Staples J, Morrison AC, Bainbridge MN, Penney S, et al. Secondary findings and carrier test frequencies in a large multiethnic sample. *Genome Med.* 2015;7(1):54.
- Jang MA, Lee SH, Kim N, Ki CS. Frequency and spectrum of actionable pathogenic secondary findings in 196 Korean exomes. *Genet Med.* 2015;17(12):1007–11.
- Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, Murray ML, Tokita MJ, Gallego CJ, Kim DS, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res.* 2015;25(3):305–15.
- Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am J Hum Genet.* 2017;100(2):267–80.
- Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, Fogel R, Gelmann EP, Gilbert F, Hasson MA, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials.* 2000;21(6 Suppl):273S–309S.
- Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, Feigelson HS, Thun MJ. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer.* 2002;94(2):500–11.
- Landi MT, Consonni D, Rotunno M, Bergen AW, Goldstein AM, Lubin JH, Goldin L, Alavanja M, Morgan G, Subar AF, et al. Environment And Genetics in Lung cancer Etiology (EAGLE) study: an integrative population-based case-control study of lung cancer. *BMC Public Health.* 2008;8:203.
- Shi J, Yang XR, Ballew B, Rotunno M, Calista D, Fargnoli MC, Ghiorzo P, Bressac-de Paillerets B, Nagore E, Avril MF, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat Genet.* 2014;46(5):482–6.
- Yang XR, Rotunno M, Xiao Y, Ingvar C, Helgadottir H, Pastorino L, van Doorn R, Bennett H, Graham C, Sampson JN, et al. Multiple rare variants in high-risk pancreatic cancer-related genes may increase risk for pancreatic cancer in a subset of patients with and without germline CDKN2A mutations. *Hum Genet.* 2016;135(11):1241–9.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945–59.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
- Kleinberger J, Maloney KA, Pollin TI, Jeng LJ. An openly available online tool for implementing the ACMG/AMP standards and guidelines for the interpretation of sequence variants. *Genet Med.* 2016;18(11):1165.
- Nielsen M, Lynch H, Infante E, Brand R. MUTYH-associated polyposis. In: *GeneReviews*(®). Edited by Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LH, Stephens K, Amemiya A. Seattle; 1993–2018.
- Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, Bennett RL, Jones KL, Tokita MJ, Bennett JT, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet.* 2013;93(4):631–40.
- Jurgens J, Ling H, Hetrick K, Pugh E, Schietecatte F, Doheny K, Hamosh A, Avramopoulos D, Valle D, Sobreira N. Assessment of incidental findings in 232 whole-exome sequences from the Baylor-Hopkins Center for Mendelian Genomics. *Genet Med.* 2015;17(10):782–8.
- Lawrence L, Sincan M, Markello T, Adams DR, Gill F, Godfrey R, Golas G, Groden C, Landis D, Nehrebecky M, et al. The implications of familial incidental findings from exome sequencing: the NIH Undiagnosed Diseases Program experience. *Genet Med.* 2014;16(10):741–50.
- Seifert BA, O'Daniel JM, Amin K, Marchuk DS, Patel NM, Parker JS, Hoyle AP, Mose LE, Marron A, Hayward MC, et al. Germline analysis from tumor-germline sequencing dyads to identify clinically actionable secondary findings. *Clin Cancer Res.* 2016;22(16):4087–94.
- Johns AL, McKay SH, Humphris JL, Pinesse M, Chantrill LA, Mead RS, Tucker K, Andrews L, Goodwin A, Leonard C, et al. Lost in translation: returning germline genetic results in genome-scale cancer research. *Genome Med.* 2017;9(1):41.
- Kim J, Field A, Schultz KAP, Hill DA, Stewart DR. The prevalence of DICER1 pathogenic variation in population databases. *Int J Cancer.* 2017;141(10):2030–6.
- de Andrade KC, Mirabello L, Stewart DR, Karlins E, Koster R, Wang M, Gapstur SM, Gaudet MM, Freedman ND, Landi MT, et al. Higher-than-expected population prevalence of potentially pathogenic germline TP53 variants in individuals unselected for cancer history. *Hum Mutat.* 2017;38(12):1723–30.
- Thompson ER, Rowley SM, Li N, McInerney S, Devereux L, Wong-Brown MW, Trainer AH, Mitchell G, Scott RJ, James PA, et al. Panel testing for familial

- breast cancer: calibrating the tension between research and clinical care. *J Clin Oncol.* 2016;34(13):1455–9.
30. Maxwell KN, Domchek SM, Nathanson KL, Robson ME. Population frequency of germline BRCA1/2 mutations. *J Clin Oncol.* 2016;34(34):4183–5.
 31. Kohlmann W, Gruber SB: Lynch syndrome. In: *GeneReviews*(R). Edited by Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A. Seattle; 1993-2018.
 32. Dewey FE, Grove ME, Pan C, Goldstein BA, Bernstein JA, Chaib H, Merker JD, Goldfeder RL, Enns GM, David SP, et al. Clinical interpretation and implications of whole-genome sequencing. *JAMA.* 2014;311(10):1035–45.
 33. Caswell-Jin JL, Gupta T, Hall E, Petrovchich IM, Mills MA, Kingham KE, Koff R, Chun NM, Levonian P, Lebensohn AP, et al. Racial/ethnic differences in multiple-gene sequencing results for hereditary cancer risk. *Genet Med.* 2018;20(2):234–9.
 34. Peixoto LA, Bhering LL, Cruz CD. Artificial neural networks reveal efficiency in genetic value prediction. *Genet Mol Res.* 2015;14(2):6796–807.
 35. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet.* 2015;16(6):321–32.
 36. Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, Biesecker LG. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med.* 2018; 20(9):1054–60.
 37. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, Berg JS, Biswas S, Bowling KM, Conlin LK, et al. Performance of ACMG-AMP variant-interpretation guidelines among nine laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet.* 2016;99(1):247.
 38. Gradishar W, Johnson K, Brown K, Mundt E, Manley S. Clinical variant classification: a comparison of public databases and a commercial testing laboratory. *Oncologist.* 2017;22(7):797–803.
 39. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, et al. ClinGen—the clinical genome resource. *N Engl J Med.* 2015;372(23):2235–42.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

