

Frequency and spectrum of actionable pathogenic secondary findings in 196 Korean exomes

Mi-Ae Jang, MD^{1,4}, Sang-Heon Lee, BS^{2,3}, Namshin Kim, PhD^{2,3} and Chang-Seok Ki, MD, PhD¹

Purpose: One of the biggest challenges of exome and genome sequencing in the era of genomic medicine is the identification and reporting of secondary findings. In this study we investigated the frequency and spectrum of actionable pathogenic secondary findings in Korean exomes.

Methods: Data from 196 Korean exomes were screened for variants from a list of 56 genes recommended by the American College of Medical Genetics and Genomics (ACMG) for return of secondary findings. Identified variants were classified according to the evidence-based guidelines reached by a joint consensus of the ACMG and the Association for Molecular Pathology.

Results: Among the 196 exomes, which were from 100 healthy controls and 96 patients with suspected genetic disorders, 11

variants in 13 individuals were found to be pathogenic or likely pathogenic. We estimated that the frequency of actionable pathogenic secondary findings was 7% for the control subjects (7/100) and 6% for the patients with disease (6/96). For one autosomal-recessive disease, four individuals exhibited either one pathogenic or one likely pathogenic variant of the *MUTYH* gene, leading to a carrier frequency of 2% (4/196).

Conclusion: Secondary findings are not uncommon in Korean exomes.

Genet Med advance online publication 9 April 2015

Key Words: exome; Korean; secondary finding; variant

INTRODUCTION

As whole-exome and -genome sequencing has become less expensive and bioinformatics has improved, the clinical use of genome sequencing has rapidly increased. One important concern related to this expanded clinical use is the identification and reporting of secondary findings that are unrelated to the indication of the sequencing but are of medical value for patient care.^{1,2} There has been much discussion and publication on the topic of secondary findings,^{3–5} and the American College of Medical Genetics and Genomics (ACMG) has released a set of guidelines.¹ The ACMG recommends restricting the return of secondary findings to a minimum set of 56 actionable genes and restricting the variants reported as secondary findings to those fitting one of two categories: “known pathogenic” or “expected pathogenic.”¹

However, classification of identified variants to the appropriate category is very difficult, even with the ACMG recommendations for interpretative categories of sequence variants.⁶ The ACMG and the Association for Molecular Pathology (AMP) recently came to a joint consensus regarding updated guidelines for variant classification during the 2013 ACMG Annual Clinical Genetics Meeting. The new recommendation is to classify variants into pathogenic or likely pathogenic categories according to evidence-based guidelines, in which multiple lines of evidence (e.g., population data, computational data, functional data, segregation data) are combined. The proposed

evidence-based guidelines are more stringent and objective than previous guidelines; however, their value to clinical exomes has not been elucidated.

A substantial number of secondary findings exist in the human genome. Each individual carries about 50–100 variants in disease-associated genes,⁷ and a screen of 1,000 exomes with actionable variants revealed an estimated frequency of 3.4% for those of European descent and 1.2% for those of African descent harboring the pathogenic or likely pathogenic variants.⁸ In Korea the number of individuals undergoing exome or genome sequencing is also rapidly increasing. However, no study has yet addressed the prevalence of secondary findings per individual in Korea. In this study we analyzed exome data from 196 Korean individuals to search for actionable pathogenic variants as secondary findings using the proposed evidence-based guidelines.

MATERIALS AND METHODS

Study population

This study was approved by the institutional review board of Samsung Medical Center, Seoul, South Korea. A total of 196 exomes, including 100 exomes from healthy controls (group A, control exomes) and 96 exomes from patients with any kind of underlying disorders (group B, disease exomes), were screened for variants from a list of 56 genes recommended by the ACMG for return of secondary findings.¹ Control exome data were

¹Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea; ²Epigenomics Research Center Genome Institute, Korea Research Institute of Bioscience and Biotechnology, Daejeon, South Korea; ³Department of Bioinformatics, Korea University of Science and Technology, Daejeon, South Korea; ⁴Current address: Department of Laboratory Medicine, Korea University College of Medicine, Seoul, South Korea. Correspondence: Chang-Seok Ki (changski@skku.edu) or Namshin Kim (deepreds@kribb.re.kr)

Submitted 17 October 2014; accepted 23 January 2015; advance online publication 9 April 2015. doi:[10.1038/gim.2015.26](https://doi.org/10.1038/gim.2015.26)

obtained from the Korean Genome and Epidemiology Study, which began in 2001 as an ongoing, population-based study of Korean adults in Ansan and Ansong, South Korea. We also investigated disease exome data from 96 consecutive, unrelated Korean patients with suspected Mendelian disorders who underwent whole-exome sequencing between 2011 and December 2013 at Samsung Medical Center, Seoul, South Korea. The underlying disease conditions for patients who underwent whole-exome sequencing were Alzheimer disease (3/96, 3%), amyotrophic lateral sclerosis (5/96, 5%), atelosteogenesis (1/96, 1%), autism spectrum disorder (1/96, 1%), frontotemporal dementia (3/96, 3%), inherited metabolic disease (4/96, 4%), nontuberculous mycobacterial infection (20/96, 21%), primary congenital glaucoma (12/96, 13%), subcortical vascular cognitive impairment (45/96, 47%), retinitis pigmentosa (1/96, 1%), and vasculopathy (1/96, 1%). The population structure of our sample was examined to confirm its genetic homogeneity and to assess its stratification using principal component analysis.⁹

Exome sequencing

Genomic DNA was obtained from the participants in group A with a SureSelect Human All Exon v2 (Agilent Technologies, Santa Clara, CA) and sequenced on an Illumina HiSeq2000 platform (Illumina, San Diego, CA). Genomic DNA from the participants in group B was enriched using Agilent SureSelect Human All Exon v3 (Agilent Technologies), with the exception of five samples from patients with amyotrophic lateral sclerosis that were obtained using an Illumina TruSeq kit (Illumina) and sequenced using the Illumina HiSeq2000 machine (Illumina).

Criteria for evidence-based classification

We adopted the evidence-based guidelines recently recommended at the 2013 joint consensus meeting of the ACMG and the Association for Molecular Pathology to classify the identified variants. Variants were classified into five categories: pathogenic, likely pathogenic, uncertain significance, likely benign, and benign (**Supplementary Table S1** online). The criteria were meant to be stringent and stratified based on the combination of many lines of evidence, including population data, computational and predictive data, functional data, segregation data, de novo data, allelic data, and other data. Each piece of evidence was weighted as very strong, strong, moderate, or supporting. A detailed definition of evidence for each case is given in **Supplementary Table S2** online.

A primary literature review to determine their potential pathogenicity of all identified variants was conducted using various sources cited in the Human Gene Mutation Database (HGMD; professional version for release 2014.1) and PubMed. Population data were assessed using the allele frequency of a variant in the general population as described in databases such as the 1000 Genomes Project; the National Heart, Lung, and Blood Institute Exome Sequencing Project database; and the single-nucleotide polymorphism database. If the allele frequency was absent from controls or was present only at an extremely low frequency (minor allele frequency <0.005)

in the databases, moderate evidence for a pathogenic variant could be assigned to the variant. Computational and predictive data were assessed by different prediction tools such as Sorting Intolerant From Tolerant, Polymorphism Phenotyping version 2, Combined Annotation Dependent Depletion, and Human Splicing Finder, along with a conservation score (PhyloP). Cosegregation of a variant also was considered a criterion in determining pathogenicity. If a variant was described in the literature as de novo, there was increased suspicion that the variant was pathogenic.

Selection of pathogenic or likely pathogenic variants

We first reduced all variants to the set that occurs within the 56 genes and manually curated this set to exclude nongenic or intronic variants. To reduce the list of variants to a manageable size, we prioritized variants previously listed as disease-causing in the HGMD, as well as novel or disruptive truncating variants expected to cause disease with no previous reports in the HGMD. All available lines of evidence were reviewed for each selected variant according to evidence-based guidelines. After deliberate variant classification, variants classified as

Table 1 Filtering and classification of variants by study group

Classification	Group A (control)	Group B (disease)	Total
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants ^c (D)	6	5	11
Sum of participants with pathogenic or likely pathogenic variants (C + D)	7	6	13
Participants (n)			
Participants with pathogenic variants ^c (C)	1	1	2
Participants with likely pathogenic variants<			

“pathogenic” or “likely pathogenic” were selected as actionable pathogenic secondary findings.

RESULTS

Population stratification and exome sequencing statistics

Principal component analysis demonstrated that the genetic variants exhibited by the Korean subjects in our study are clearly distinguished from those of other ethnic groups, indicating the genetic homogeneity of our study population (Supplementary Figure S1 online).

The targeted coding sequences of the 56 genes examined 286,381 bp (1,161 exons). The proportion of target bases covered by 10 or more sequence reads was 90% in group A and 93% in group B. Detailed statistical data are given in Supplementary Table S3 online.

Pathogenic or likely pathogenic variants by study group

As shown in Table 1, a total of 4,409 unique variants of the 56 ACMG-reportable genes were identified in the 196 exomes (2,309 and 2,100 in groups A and B, respectively). After exclusion of nongenic or intronic variants, 70 potentially pathogenic variants that were classified as disease-causing in the HGMD were identified (33 and 37 in groups A and B, respectively). Six novel truncating variants not listed as disease-causing in the

HGMD also were identified (two and four in groups A and B, respectively). These 76 variants, previously listed or not listed as disease-causing in the HGMD, were reviewed according to all available lines of evidence. This analysis identified 11 potentially pathogenic variants, including pathogenic or likely pathogenic variants, in 13 individuals (7 and 6 in groups A and B, respectively). These variants are listed in Table 2. The detailed classification of each pathogenic or likely pathogenic variant is described in Supplementary Table S4 online. All pathogenic or likely pathogenic variants seemed to be unrelated to the ascertainment of the disease group B.

Estimation of the population frequency of actionable pathogenic secondary findings

We identified 13 participants, 7 individuals in group A and 6 individuals in group B, with pathogenic or likely pathogenic variants as secondary findings in 56 medically-actionable genes, according to the ACMG (Table 1). Thus, ~7% (7/100 in group A; 6/96 in group B) of the participants analyzed, who were all Korean, had secondary findings. For 9 of the variants, these pathogenic or likely pathogenic variants were identified by their classification in the HGMD as disease-causing mutations, whereas two were identified via predicted premature truncations (Table 2). Among the 56 ACMG-reportable genes, only the *MUTYH* gene is an

Table 2 Pathogenic or likely pathogenic variants according to the evidence-based guidelines in ACMG reportable genes

Category	Gene	Variant	dbSNP	HGMD accession no.	Disease association per the ACMG	Frequency in A (control)	Frequency in B (disease)
P	<i>BRCA2</i>	NM_000059.3:c.1399A>T NP_000050.2:p.Lys467*	rs80358427	CM021955	Hereditary breast and ovarian cancer	1/100	1/96
P	<i>MUTYH</i>	NM_001128425.1:c.799C>T NP_001121897.1:p.Gln267*	NA	CM077607	MYH-associated polyposis	1/100	0/96
P	<i>MUTYH</i>	NM_001128425.1:c.934-2A>G	rs77542170	CS04822	MYH-associated polyposis	2/100	1/96
LP	<i>BRCA1</i>	NM_007294.3:c.3629dupA NP_009225.1:p.Ser1211Valfs*8	NA	NA	Hereditary breast and ovarian cancer	1/100	0/96
LP	<i>DSP</i>	NM_004415.2:c.269A>G NP_004406.2:p.Gln90Arg	rs188516326	CM063959	Arrhythmogenic right ventricular cardiomyopathy	1/100	0/96
LP	<i>MYBPC3</i>	NM_000256.3:c.1000G>A NP_000247.2:p.Glu334Lys	NA	CM073211	Hypertrophic cardiomyopathy, dilated cardiomyopathy	0/100	1/96
LP	<i>MYL3</i>	NM_000258.2:c.170C>G NP_000249.1:p.Ala57Gly	rs139794067	CM14210	Hypertrophic cardiomyopathy, dilated cardiomyopathy	0/100	2/96
LP	<i>RYR1</i>	NM_000540.2:c.5701C>T NP_000531.2:p.Gln1901*	NA	NA	Malignant hyperthermia susceptibility	0/100	1/96
LP	<i>RYR2</i>	NM_001035.2:c.1217C>T NP_001026.2:p.Ser406Leu	NA	CM126236	Catecholaminergic polymorphic ventricular tachycardia	0/100	1/96
LP	<i>SCN5A</i>	NM_198056.2:c.3250G>A NP_932173.1:p.Gly1084Ser	rs199473190	CM085685	Long QT syndrome	1/100	0/96
LP	<i>TP53</i>	NM_000546.5:c.91G>A NP_000537.3:p.Val31Ile	rs201753350	CM074603	Li-Fraumeni syndrome	3/100	0/96

ACMG, the American College of Medical Genetics and Genomics; HGMD, the Human Gene Mutation Database, professional version for release 2014.1; LP, likely pathogenic variant; NA, not available; P, pathogenic variant; dbSNP, single-nucleotide polymorphism database.

autosomal-recessive locus. These *MUTYH* variants were found in individuals who carry one copy of the variant.

DISCUSSION

In this study we found that ~7% (13/196) of the Korean individuals examined had a pathogenic or likely pathogenic variant as identified by exome sequencing. This study is the first to estimate the actionable pathogenic variant load in an Asian population and to apply evidence-based guidelines to variant classification.

This work provides several important findings. First, the frequency of secondary findings in our Korean study population falls within the range of previously published studies.^{8,10,11} A prior report of 1,000 clinical exomes from European/African subjects estimated that 1.2–3.4% of its subjects had secondary findings.⁸ Another study, by Xue *et al.*,¹¹ of 179 healthy individuals from the 1000 Genomes Project found frequencies of up to 11%. A screen of 104 exomes for 448 severe recessive diseases found an average carrier burden of 2.8 variants per person.¹² Another study found a high proportion of pathogenic variants (8.5%) in 1,092 asymptomatic individuals.¹⁰ The variation in these rates of secondary findings between different groups could be due to a number of reasons. Differences in database annotations and interpretations are a major source of differences in apparent secondary findings.¹³ Moreover, our interpretation criteria—evidence guidelines—are different from those used in earlier reports and might be an additional source of variation in our observed secondary findings.

The second important observation is that the detection rate of secondary findings was similar across study groups. Approximately 7% (7/100) of the individuals in the control group exhibited pathogenic or likely pathogenic variants; similarly, 6% (6/96) of the individuals suspected to have genetic disease exhibited pathogenic or likely pathogenic variants. Thus the burden of actionable pathogenic secondary findings in the Korean population is estimated to be ~7% (13/196), although additional data are needed to extrapolate our findings to a wider population. Third, because they provide detailed guidance for how to classify a variant based on multiple lines of evidence, the evidence-based guidelines effectively classified the identified variants in this study.

This study was limited by the following: (i) it included only 196 Korean participants, (ii) phenotype data were not available, (iii) the data may have been skewed by missense variants that are common among Koreans and are not found in the current exome databases, and (iv) the gene list did not consider disease prevalence by ethnic group. We focused our analysis on the 56 genes recommended by the ACMG. Although this gene list may be applied to any ethnicity, it might not include more prevalent and highly penetrant genes in the Asian population. For example, the *CDH1* gene, which is the causative gene for hereditary diffuse gastric cancer, with a lifetime risk of over 80%, was not included in the ACMG list of minimum genes.¹ In this study, we found two likely pathogenic variants of *CDH1*, c.1018A>G (p.Thr340Ala) and c.2494G>A (p.Val832Met), in

0.5% (1/196) and 1.0% (2/196) of all patients, respectively. The prevalence of gastric cancer in Korea is high, with crude incidence rates of 92.7 and 43.5 per 100,000 people per year among men and women, respectively, as estimated in 2014.¹⁴ Moreover, the frequency of this disease is expected to continue to increase.

Taken together, the results from our work demonstrate that actionable pathogenic secondary findings are frequently identified in the Korean population.

SUPPLEMENTARY MATERIAL

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

ACKNOWLEDGMENTS

This study was supported by grants from the Korea Healthcare Technology R&D Project, Ministry for Health & Welfare, Republic of Korea (HI10C1673, HI12C0135, and HI12C0713) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (2013R1A1A2009756). Biospecimens and data for this study were provided by the Korean Genome Analysis Project (4845-301), the Korean Genome and Epidemiology Study (4851-302), and the Korea Biobank Project (4851-307, KBP-2014-031), supported by the Korea Center for Disease Control and Prevention, Republic of Korea.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

- Green RC, Berg JS, Grody WW, *et al.*; American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013; 15:565–574.
- Presidential Commission for the Study of Bioethical Issues. *Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts*. Presidential Commission for the Study of Bioethical Issues: Washington, DC, 2013.
- Knoppers BM, Deschênes M, Zawati MH, Tassé AM. Population studies: return of research results and incidental findings Policy Statement. *Eur J Hum Genet* 2013;21:245–247.
- Wolf SM, Crock BN, Van Ness B, *et al.* Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genet Med* 2012;14:361–384.
- Green RC, Berg JS, Berry GT, *et al.* Exploring concordance and discordance for return of incidental findings from clinical sequencing. *Genet Med* 2012;14: 405–410.
- Richards CS, Bale S, Bellissimo DB, *et al.*; Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 2008;10:294–300.
- Abecasis GR, Altshuler D, Auton A, *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–1073.
- Dorschner MO, Amendola LM, Turner EH, *et al.*; National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet* 2013;93:631–640.
- Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 2012;28:3326–3328.
- Cassa CA, Tong MY, Jordan DM. Large numbers of genetic variants considered to be pathogenic are common in asymptomatic individuals. *Hum Mutat* 2013;34:1216–1220.

11. Xue Y, Chen Y, Ayub Q, et al.; 1000 Genomes Project Consortium. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet* 2012;91:1022–1032.
12. Bell CJ, Dinwiddie DL, Miller NA, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 2011;3:65ra4.
13. Ding LE, Burnett L, Chesher D. The impact of reporting incidental findings from exome and whole-genome sequencing: predicted frequencies based on modeling. *Genet Med* 2014;17:197–204.
14. Jung KW, Won YJ, Kong HJ, Oh CM, Lee DH, Lee JS. Prediction of cancer incidence and mortality in Korea, 2014. *Cancer Res Treat* 2014;46:124–130.