A Comprehensive Follow-Up Study Identifying Diagnostic Evidences for Secondary Finding

shu li  
Maternal and Child Health Hospital of Hunan Province

li haoyu  
Xiangya Hospital, Central South University

xiao neng  
Chenzhou First People's Hospital

lin yixiang  
Children's Hospital of Fudan University

yu wenxian  
Maternal and Child Health Hospital of Hunan Province

li yingjing  
Maternal and Child Health Hospital of Hunan Province

xi hui  
Maternal and Child Health Hospital of Hunan Province

jia zhengjun  
Maternal and Child Health Hospital of Hunan Province

yu fuli  
Baylor College of Medicine

wang hua  
Maternal and Child Health Hospital of Hunan Province

mao xiao  
Maternal and Child Health Hospital of Hunan Province

Research

Keywords: follow-up study, secondary findings, secondary variants, whole-exome sequencing, family

DOI: https://doi.org/10.21203/rs.3.rs-51673/v1

License: ☇ ○ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background: There was a lacking of clinical diagnostic evidence in follow-up studies for reporting of secondary variants in 59 genes in American College of Medical Genetics and Genomics recommendations for reporting secondary findings and various strategies were applied to interpret the secondary variants.

Results: Out of 1330 participants performed whole-exome sequencing, we identified 15 families with convincing clinical evidence. After Sanger validation and a comprehensive clinical follow-up, 10 families with both convincing clinical evidence and convincing genetic evidence of hereditary variants were found. Detailed clinical presentations and related clinical evidence were collected.

Conclusions: Our research is a comprehensive follow-up study to identify secondary variants with convincing genetic and clinical evidence and it could help improve the strategy of screening actionable secondary variants and contribute to translation of genetic findings into medical practice.

Background

As genome and exome sequencing have widely been applied in clinical and research contexts and the interpretation of mutations with primary disease is becoming streamlined, the interpretative challenge of secondary findings is a heated debate. Secondary (Incidental) findings are defined as exome and genome sequencing results unrelated to the indication for ordering the sequencing but of medical value for patient care. In 2017, the American College of Medical Genetics and Genomics (ACMG) published recommendations for reporting secondary findings, including 59 medically actionable genes that supplement the previous 56 gene list. It has been suggested that the report of actionable secondary findings should be based at the variant level, not just the gene level. These secondary variants could help prevent treatable diseases, help in life planning for untreatable diseases, tailor effective treatments, etc. The incidence and clinical significance of the secondary variants have been explored by many cohort studies and have become a research hotspot.

The criteria for classification of known pathogenic (KP) variants and expected pathogenic (EP) variants in ACMG recommendations for reporting of secondary findings has differed in studies so far. Workflows composed of various filtering methods and different medically curated databases, including the Human Gene Mutation Database (HGMD), ClinVar databases, Locus Specific Databases (LSDBs) and ACMG guidelines for interpretation of genetic variants accompanied by manual assessments, were used in the identification of secondary variants. Due to these various interpretation strategies, the frequency of actionable secondary variants in cohorts could range from 1–9% in previous studies.

Because the clinical relevance of the secondary variants is not clear, there is a lack of actual evidence to guide the interpretation of secondary variants; thus, the return of the reports has an unknown impact on individuals carrying these variants. There were few studies with limited clinical evidence that would lead to the inability to judge the clinical significance. Moreover, without research on families with multiple generations, it is difficult to determine the true penetrance of the variants. The difference in penetrance of the deleterious secondary variants would result in specific medical recommendations for improving clinical outcomes. Therefore, due to the lack of long-term follow-up studies about the clinical significance of the secondary variants, researchers considered the different filtering strategies adopted by previous cohort studies to exert a high false-positive rate.

In our research, we performed a comprehensive follow-up study of 1330 participants and collected clinical and genetic convincing evidence for families with phenotypes associated with the incidental findings. Our study confirmed the clinical significance of secondary variants in 59 genes based on ACMG recommendations and is an exploration for the refinement of detection secondary variants in whole-exome sequencing (WES) data based on empirical data.

Methods

Subject Cohort

Patients were recruited from the outpatient department of medical genetics in Maternal and Child Health Hospital of Hunan Province from 2015.2 to 2019.10 who used whole exome sequencing (WES) as one of the tests for etiologic diagnosis. A total of 1330 participants were enrolled including 648 patients (48.7%) and 682 unaffected family members (51.3%). Among the patients, there were 211 patients with microcephaly (9.2%), 82 patients with developmental and epileptic encephalopathy (6.2%), 228 patients with developmental delay/intellectual disability (17.1%) and 127 patients with other diseases (metabolic diseases, epilepsy, skeletal dysplasia, etc.) (9.5%) (Figure 1). The average and median age of the patients and the unaffected family members at the time of drawing blood were 44.24±44.67 months and 32.77±6.59 years, respectively.

Leukocyte DNA was extracted from peripheral blood by the phenol chloroform method. All participants signed informed consent forms, and the study was approved by the ethics committee of the Maternal and Child Health Hospital of Hunan Province. All participants were fully informed about the nature of the research and the possibility of secondary findings being generated in the study. The study was performed in accordance with the principles of the Declaration of Helsinki.

WES analyses

Library preparation

The Qiagen DNA Blood Midi/Mini Kit (Qiagen GmbH, Hilden, Germany) was used to extract DNA from peripheral blood. DNA was then sheared to approximately 200 bp by the Biorupter UCD-200 (Diagenode). The DNA fragments were then repaired at the end, and one A base was added to the 3' end. The DNA fragments were connected with sequencing adaptors, and fragments of approximately 320 bp were collected by XP beads. After PCR amplification, the
DNA fragments were hybridized and captured by IDT's xGen Exome Research Panel (Integrated DNA Technologies, San Diego, USA) according to the manufacturer's protocol. The hybrid products were eluted and collected. Then, DNA was PCR amplified and purified.

The libraries were tested for enrichment by qPCR, and size distribution and concentration were determined using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). WES was performed on the NovaSeq6000 platform (Illumina, San Diego, USA), with 150 bp pair-end reads used for sequencing the genomic DNA of the family. Raw image files were processed using CASAVA v1.82 for base calling and generating raw data.

**Raw data processing**

We first used cutadapt1 (v1.15) to trim adaptor sequences at the tail of sequencing reads and then aligned sequencing reads to the human reference genome (UCSC hg19/GRCh37) with the Burrows–Wheeler Aligner tool BWA2 (v0.7.15). Duplicated reads were marked by Picard3 (v2.4.1). Qualimap4 (v2.2.1) was used to calculate base quality metrics, genome mapping rate, and coverage of targeted regions. Base quality score recalibration, indel realignment and variant (SNVs & InDels) calling were performed following the best practice protocol of the Genome Analysis Toolkit5 (GATK, v3.8). Variants were annotated using ANNOVAR11. Annotation human population databases mainly included the following: human population databases, such as gnomAD (http://gnomad.broadinstitute.org/) and ExAC (http://exac.broadinstitute.org/); and in silico prediction algorithms, such as SIFT (http://sift.jcvi.org), FATHMM (http://fathmm.biocompute.org.uk), MutationAssessor (http://mutationassessor.org), and CADD (http://cadd.gs.washington.edu).

After the annotation process, variants within the 59 genes recommended by the ACMG were extracted.

**Variant filtering**

Our goal was to screen out family members of participants with hereditary secondary variants and participants with de novo secondary variants who developed secondary variant related diseases to clarify the clinical significance of the secondary variants; therefore, sensitivity was sacrificed.

**Step 1. Filtering variants according to HGMD and ACMG guidelines**

Disease-causing mutations (DM) and probable/possible pathological mutation (DM?) in the HGMD (Professional version 2019.1) database, and pathogenic (P) and likely pathogenic (LP) variants interpreted by ACMG guidelines for interpretation of genetic variants were included12.

**Step 2. Filtering variants according to allele frequency, variant type and inheritance mode**

Variants with minor allele frequencies (MAF) <0.1%, variant depth of coverage >=20 and alteration base depth of coverage >=4 were taken on for further analyses. Filtering was conducted on the remaining variants according to variant type and inheritance model of the associated disease of ACMG 59 medically actionable genes2. Variants in the exonic region or with splicing impact were included. Among exonic variants, loss-of-function variants (frameshift variants, start lost, stopgain, etc.) and nonsynonymous single nucleotide variants (SNVs) were retained, and synonymous SNVs were excluded. Heterozygous variants in two genes, ATP7B and MUTYH, in autosomal recessive inheritance were excluded, and homozygous variants in the genes remained. Heterozygous variants in the other 57 genes that fit the dominant inheritance mode were retained. Pass-filter variants were selected for clinical and genetic follow-up studies (Supplementary Table 1-4).

**Clinical and genetic follow-up study based on families**

A comprehensive follow-up study was conducted. Family members of participants with hereditary secondary variants and participants with de novo secondary variants were enrolled. Primary medical evaluation: Current and past medical history was assessed by email, WeChat software or telephone to identify confirmed individuals according to diagnostic criteria or possible individuals fitting at least one clinical manifestation in OMIM database (https://omim.org/). Secondary medical evaluation: physical examination and necessary examinations by face-to-face evaluation in the clinic or home visits were also used to identify individuals with convincing clinical evidence. All clinical information was collected and evaluated by experienced clinicians, including pediatricians and neurosurgeons. The diagnostic criteria of the gene-associated diseases are listed in the detailed clinical manifestations of each family separately.

Individuals and families with convincing clinical evidence were performed Sanger sequencing was performed to validate the variant. Experienced geneticists analyzed the variants, including variant types and inheritance modes. All family members who took the test signed the informed consent form.

Published articles in the HGMD database related to secondary variants with convincing clinical and genetic evidence were reviewed to explore the pathogenicity of the variants.

**Results**

The flowchart showing research steps including variant screening and follow-up study is presented in Figure 2. A total of 8851 secondary variant counts classified as DM or DM? in the HGMD database or P or LP in the ACMG guidelines in 1330 participants were obtained after variant filtering step 1. These secondary variants included 2777 DM variant counts, 5986 DM? variant counts, 28 P variant counts and 537 LP variant counts. The total number of variants was 450, consisting of 186 DM variant counts, 208 DM? variants, 20 P variants and 203 LP variants (Figure 3a). After variants were filtered in step 2 according to allele frequency, variant type and inheritance mode, there were 998 variant counts including 493 DM variant counts, 422 DM? variant counts, 22 P variant counts and 399 LP variant counts. The total number of variants was 297, with 149 DM variants, 94 DM? variants, 16 P variants and 181 LP variants (Figure 3b).
Characteristics of the included variants

These 297 pass-filter variants in 638 individuals were included in the follow-up study. These variants were in 48 genes among 59 ACMG genes related to secondary findings (Table 1). The top six most frequently observed secondary variants in genes were BRCA2 (86 individuals, 13.5%), APOB (57 individuals, 8.9%), DSP (49 individuals, 7.7%), MYBPC3 (41 individuals, 6.4%), MYH7 (40 individuals, 6.3%) and RYR2 (40 individuals, 6.3%).

We classified these secondary variants into four categories according to gene-associated phenotypes: variants in genes associated with oncogenic diseases, variants in genes associated with cardiogenic diseases, variants in genes associated with connective tissue diseases, variants in genes associated with hypercholesterolemia, and others. The number of individuals carrying secondary variants from maximum to minimum were oncogenic disease (369 individuals, 57.8%), cardiogenic diseases (322 individuals, 50.5%), hypercholesterolemia (92 individuals, 14.4%), connective tissue diseases (74 individuals, 11.6%) and others (malignant hyperthermia susceptibility, ornithine transcarbamylase deficiency) (41 individuals, 6.4%) (Table 1, Figure 4a). The order is similar to the number of secondary variants identified in each category (Figure 4b). The top three/four most common variants in genes related to different disease categories separately were DSP, MYBPC3, MYH7 and RYR2 (cardiogenic diseases); FBN1, COL3A1, and MYH11 (connective tissue diseases); APOB, LDLR, and PCSK9 (hypercholesterolemia); BRCA2, TSC1, and APC (oncogenic diseases) and RYR1, CACNA1S, and OTC (others) (Table 1).

Follow-up study

A comprehensive follow-up study was performed in 638 individuals and their family members. Among these participants, 105 people were lost to follow-up (38 people could not be contacted, 67 people refused to participate in the study), and 532 people finally took part in the follow-up study including family members of 420 participants with hereditary secondary variants and 112 participants with de novo secondary variants. The detailed screening strategies to identify individuals with convincing clinical evidence in the follow-up study is shown in Figure 5. Examplified by follow-up study of participants carrying MPBPC3 variants, strategies including primary medical evaluation and secondary medical evaluation were performed to identify individuals with convincing clinical evidence (Figure 5). Based on the clinical features of different genes associated with secondary findings, 13 families with sufficient clinical evidence were found in the follow-up study while no participants with de novo variants presented convincing clinical manifestation (Table 1).

Sanger sequencing was conducted to validate variants in families with convincing clinical evidence. There was 1 family who refused Sanger sequencing (Table 2). After Sanger sequencing of the 12 families, 10 families with both convincing clinical and genetic evidence were detected (Table 3). 2 variants did not cosegregate with disease phenotypes in 2 families (Table 2). Among these final identified individuals having clinically and genetically confirmed secondary variants, 8 individuals were from the WES 1330 cohort. Therefore, the frequency of the secondary variants with both convincing clinical and genetic evidence in our cohort was 0.6% (8/1330).

Among the 10 families with convincing clinical and genetic evidence, there were 2 families carrying the same COL3A1 variant c.3133G>A. A two families carried two different APOB variants (c.1342G>A, c.35_39del), and 2 families carried two different BRCA1 variants (c.811G>A, c.427G>A). TNNT2 c.422G>A, MSH2 c.2197G>A, MYBPC3 c.2543C>T, and BRCA2 c.1832C>G were each observed in one family separately. In these 9 validated disease-associated variants, 6 variants (TNNT2 c.422G>A, APOB c.1342G>A, MSH2 c.2197G>A, MYBPC3 c.2543C>T, BRCA2 c.1832C>G, BRCA1 c.811G>A) were classified as DM/LP, 2 variants (COL3A1 c.3133G>A, BRCA1 c.427G>A) were classified as DM/?VUS, and 1 variant (APOB c.35_39del) was classified as LP. The gene-associated genotype categories were cardiogenic diseases (TNNT2-related dilated cardiomyopathy, MYBPC3-related hypertrophic cardiomyopathy); connective tissue diseases (COL3A1-related Ehlers-Danlos syndrome-vascular type); hypercholesterolemia (APOB-familial hypercholesterolemia); and oncogenic diseases (BRCA1 or BRCA2-related hereditary breast and ovarian cancer, MSH2-related Lynch syndrome).

Detailed clinical and genetic evidence of the 9 secondary variants in 10 families are shown as follows along with penetrance calculation of variants in families based on age at onset of gene-associated diseases.

Family 1

TNNT2 c.422G>A was identified in the mother of the proband with microcephaly (III:2) by trio-WES in family 1. The heterozygous variant was validated by Sanger sequencing in 5 family members who agreed to do the test. Among them, the variant was detected in 4 family members (II:1, II:2, III:2, III:3) and was not found in 1 family member (IV:1) (Figure 6-7). The diagnosis of dilated cardiomyopathy was made by first-line imaging test echocardiography13. The sister of the grandmother of the proband (II:1) was previously diagnosed with dilated cardiomyopathy by echocardiography (Figure 8a). The grandmother of the proband (II:3) often suffered from chest tightness and dizziness, especially after exercise. Echocardiography showed dilated cardiomyopathy (Figure 8b). The mother of the proband (III:2) had occasional chest distress and palpitation, but she did not seek medical advice from the doctors. After we reported the variant to her, she came to the hospital and underwent echocardiography examination. Dilated cardiomyopathy was presented as a result (Figure 8c) (Table 4). As a gene related to child/adult-onset disease, the penetrance of TNNT2 c.422G>A was 3/4 in family 1.

Family 2

Trio-WES was performed, and APOB c.1342G>A was observed in the mother and the proband who presented with developmental and epileptic encephalopathy (III:3, IV:1). Sanger sequencing revealed the heterozygous variant in 6 family members (II:1, II:2, II:6, III:1, III:3, IV:1) who wanted to undergo the test (Figure 9-10). The diagnosis was made according to the Dutch Criteria for Familial Hypercholesterolemia14. Two family members (II:2, II:1) died of cerebral infarction accompanied by hyperlipidemia at the ages of 71 and 69. The grandfather of the proband had hyperlipidemia with serum total cholesterol (TC) 7.68 mmol/L and low-density lipoprotein cholesterol (LDL-C) 4.97 mmol/L. The younger brother of the grandfather of the proband (II:6) had facial paralysis after cerebral infarction at the age of 50. The TC was 9.22 mmol/L, and LDL-C was 5.04 mmol/L in the recent blood lipid test. The older brother of the mother of the proband (III:1) had hyperlipidemia (TC 7.91 mmol/L, LDL-C 5.17 mmol/L) and had fatty liver detected by abdominal ultrasound examination. The mother of...
the proband (Ill:3) also had hyperlipidemia with TC 8.18 mmol/L and LDL-C 5.46 mmol/L (Table 4). As the age at onset of APOB-related familial hyperlipidemia was child/adult, the penetrance of APOB c.1342G>A in family 2 was 5/6.

Family 3

APOB c.35_39del was identified in the mother of the proband with microcephaly (Ill:2) after trio-WES. The heterozygous variant was validated in four family members by Sanger sequencing. Among them, three family members carried the variant (Il:2, Il:3, Ill:2), and 1 family member (IV:1) did not (Figure 11-12). Familial hypercholesterolemia was diagnosed according to Dutch criteria14. The father of the grandmother of the proband (I:1) was diagnosed with coronary heart disease and hypercholesterolemia 20 years ago. The grandmother (Il:2), the brother of grandmother (Il:3) and the mother of the proband (Ill:2) all had hypercholesterolemia with high TC and LDL-C (TC 8.05 mmol/L LDL-C 5.39 mmol/L; TC 7.72 mmol/L LDL-C 4.98 mmol/L; TC 7.99 mmol/L LDL-C 5.51 mmol/L) (Table 4). According to the age at onset of APOB-related familial hyperlipidemia (child/adult), the penetrance of APOB c.35_39del was 3/3 in family 3.

Family 4

MSH2 c.2197G>A was detected in the father of the proband with microcephaly (Ill:1) by trio-WES. Three family members (Il:1, Il:4, Ill:1) agreed to undergo Sanger sequencing, and all of them were confirmed to carry this heterozygous variant (Figure 13-14). Lynch syndrome was diagnosed by Amsterdam criteria15. The grandmother of the proband died of rectal cancer at the age of 65. The brother of the grandmother (Il:4) and father of the proband (Ill:1) had rectal cancer at the ages of 62 and 41, respectively (Figure 15a-b) (Table 4). The penetrance of adult-onset MSH2-associated Lynch syndrome was 2/3.

Family 5

After WES sequencing of trios, the proband with microcephaly (IV:3) and mother (III:5) of family 1 carried the secondary heterozygous variant COL3A1 c.3133G>A. Sanger sequencing was applied to the family members who agreed to perform the test. Five people carried the variant (III:2, III:3, III:5, IV:2, IV:3), and 2 people (II:1, IV:1) did not (Figure 16-17). The diagnosis for Ehlers-Danlos syndrome-vascular type was made by clinical features and COL3A1 mutations16,17. The brother of the proband's grandfather (Il:1) reported a history of inguinal herna. The grandfather of the proband (Il:2) had an acute onset severe headache at the age of 47, and subarachnoid hemorrhage was detected by brain tomography angiography (CT) (Figure 18). Brain computed tomography angiography (CTA) was performed, and an aneurysm of the internal carotid system was identified (Figure 19a). Surgical interventions were conducted. The younger sister of the mother (III:3) had headache of unknown reason at the age of 28. Due to the cerebral aneurysm hemorrhage of her father, brain CTA was performed, and an internal carotid system aneurysm was found (Figure 19b). After being informed that he carried the same variant as his family members, the asymptomatic brother of the mother (III:2) was suggested to undergo brain CTA, and an aneurysm of the internal carotid system was also identified (Figure 19c). The mother and sister of the proband showed dermatopathological features of Ehlers-Danlos syndrome-vascular type. The mother of the proband (Ill:3) had a brown-black plaque above the left ankle manifesting atrophic skin and ulcers (Figure 20a) and ecchymosis on the left side of the knee (Figure 20b). The sister of the proband (IV:2) presented with scar hyperplasia surrounded by hypopigmented pityriasis versicolor-like lesions over the knee (Figure 20c) (Table 4). The penetrance was 4/5 based on child/adult age at onset of COL3A1-related Ehlers-Danlos syndrome-vascular type.

Family 6

The WES of trios in family 2 identified the same heterozygous variant COL3A1 c.3133G>A in the proband with developmental delay (IV:1) and her mother (III:2). Sanger sequencing validated the variant in 4 family members. The diagnosis of Ehlers-Danlos syndrome-vascular type was made as stated in family 5. Three family members (Il:5, Ill:2, IV:1) carried the variant, and 1 family member (Ill:3) did not (Figure 21-22). The younger brother of the grandfather of the proband died of acute myocardial infarction at the age of 22 (Il:4). The grandfather (Il:5) and mother (III:2) of the proband had a history of gastric perforation due to severe headache at the age of 47, and subarachnoid hemorrhage was detected by brain tomography angiography (CT) (Figure 18). Brain computed tomography angiography (CTA) was performed, and an aneurysm of the internal carotid system was identified (Figure 19a). Surgical interventions were conducted. The younger sister of the mother (III:3) had headache of unknown reason at the age of 28. Due to the cerebral aneurysm hemorrhage of her father, brain CTA was performed, and an internal carotid system aneurysm was found (Figure 19b). After being informed that he carried the same variant as his family members, the asymptomatic brother of the mother (III:2) was suggested to undergo brain CTA, and an aneurysm of the internal carotid system was also identified (Figure 19c). The mother and sister of the proband showed dermatopathological features of Ehlers-Danlos syndrome-vascular type. The mother of the proband (Ill:3) had a brown-black plaque above the left ankle manifesting atrophic skin and ulcers (Figure 20a) and ecchymosis on the left side of the knee (Figure 20b). The sister of the proband (IV:2) presented with scar hyperplasia surrounded by hypopigmented pityriasis versicolor-like lesions over the knee (Figure 20c) (Table 4). The penetrance was 2/3 based on child/adult age at onset of COL3A1-related Ehlers-Danlos syndrome-vascular type.

Family 7

Trio-WES results showed that the father of the proband with microcephaly (IV:1) carried the heterozygous MYBPC3 c.2543C>T in family 7. Three family members of family 7 underwent Sanger sequencing. Two family members were verified to carry the variant (III:1, IV:1), and 1 family member (IV:1) was not (Figure 23-24). The European Society of Cardiology (ESC) on the diagnosis and management of hypertrophic cardiomyopathy was used as a diagnostic criterion18. Four family members died suddenly. Among them, two family members (Il:2, Il:3) had sudden death before 40 with unknown cause. Two family members (Il:1 and Ill:3) died of cardiac arrest at the ages of 45 and 38, respectively. The grandfather of the proband (Ill:1) was admitted to the hospital due to shortness of breath and chest tightness, and echocardiography detected features of hypertrophic cardiomyopathy (Figure 25a). The asymptomatic father of the proband (IV:1) was advised to undergo echocardiography after she was informed that she was a variant carrier. Echocardiography also showed evidence of hypertrophic cardiomyopathy (Figure 25b) (Table 4). The penetrance of MYBPC3 c.2543C>T in family 7 was 2/2 (child/adult onset).

Family 8

BRCA2 c.1832C>G was identified by trio-WES in the mother of the proband who presented with developmental delay (III:2). The heterozygous variant was validated by Sanger sequencing in 5 family members, including 4 carriers (Il:2, Il:3, Ill:2, Ill:3) and 1 noncarrier (Il:4) (Figure 26-27). One of the family members died of breast cancer at the age of 55. Pathological examination proved that two members had breast cancer in the left breast (Il:3, Ill:2) (Figure 28a-b) (Table 4). The penetrance of the variant was 2/4 in total and 2/3 in females in family 8 according to adult-onset BRCA2-associated breast cancer.

Family 9
**BRCA1** c.811G>A was detected in the proband with microcephaly by WES. Sanger sequencing was performed on 4 family members who agreed to the test. Three family members (I:2, II:1, II:2) were carriers of the variant, and 1 family member (II:4) was a noncarrier (Figure 29-30). Breast cancers were diagnosed histologically in two family members (I:2, II:2) (Figure 31a-b) (Table 4). Due to the adult onset of **BRCA1**-related breast cancer, the penetrance was 2/3 in total and 2/3 in females in family 9.

**Family 10**

**BRCA1** c.427G>A was identified in the mother (III:3) and the proband who presented with developmental and epileptic encephalopathy (IV:1) by trio-WES. Sanger sequencing validated the variant in four family members. Among them, 3 family members (II:2, III:1, III:3) carried the variant, and 1 family member (II:3) did not (Figure 32-33). Pathological findings proved that the three family members had breast cancer in the right breast (Figure 34a-b) (Table 4). The penetrance of the variant was 3/3 according to adult-onset **BRCA1**-related breast cancer.

**Discussion**

Our research is a comprehensive clinical and genetic follow-up study of individuals carrying secondary variants. A total of 9 variants in 10 families with actual clinical and genetic evidence were identified. These secondary variants cosegregated with previously reported breast or ovarian cancer, colon cancer, cardiomyopathy, and hypercholesterolemia, and we also first reported cosegregation of **COL4A1** variant and Ehlers-Danlos syndrome-vascular type in the secondary finding cohort. Additionally, we used these secondary variants shown by actual evidence as an indicator of developing screening strategies for secondary variants. We found that P/LP variants classified by ACMG guidelines could be reportable secondary variants, and DM/DM? variants in the HGMD database, even with VUS classifications, could be of clinical significance. In addition, we calculated the penetrance of these actual secondary variants based on families. Our results further confirmed the clinical significance and necessity of interpretation of secondary findings based on 59 genes recommended by ACMG SF v2.0.

In cohort studies on secondary findings, the frequencies of identified variants in secondary genes varied among different studies. In our research, 10 families with convincing clinical and genetic evidence were detected, and the frequency of the secondary variants with actual evidence in our cohort was 0.6%, which was in accordance with ACMG’s estimation of a 1% rate of reportable incidental findings. The most frequent secondary variants in genes after bioinformatic analyses were related to hereditary breast/ovarian cancer, familial hypercholesterolemia, arrhythmogenic right ventricular cardiomyopathy, hypertrophic cardiomyopathy, and catecholaminergic polymorphic ventricular tachycardia. The frequency of secondary variants in disease categories from the maximum to the minimum was ontogenetic disease, cardiogenetic diseases, hypercholesterolemia, connective tissue diseases and others. The disease categories were in accordance with ACMG recommendations for secondary findings that the disorders of 59 genes were primarily cancer-predisposition syndromes and inherited cardiovascular disorders such as hypertrophic cardiomyopathy.

In our research, we provided actual clinical and genetic evidence of the secondary variants that were reliable information on variant segregation. So far, there was limited actual clinical evidence that met the gold standard of secondary findings-related diseases or fulfilled the diagnostic criteria of the diseases used for identify secondary variants. If the compatible symptoms for secondary findings-related diseases were not specific, it was difficult to verify the diseases with the subjective reports of participants without convincing evidence. Additionally, there were possibilities that the variants did not cosegregate with clinical phenotypes similar to the 2 families in our follow-up study (Table 3). Calculation of reportable secondary variant frequency based on convincing clinical evidences combined with genetic evidence may provide a clue for identifying real secondary variants of clinical significance, which may benefit patients psychologically and help reduce the unnecessary costs of medical actions.

There is a lack of clinical follow-up studies in cohort studies to explore the clinical significance of secondary variants. In the few previous reports on relevant medical findings in individuals carrying secondary variants in 59 ACMG genes, clinical presentations of familial hypercholesterolemia reflected by TC and LDL-C levels were the most easily accessible clinical evidence reports. As for the two most common disease categories of secondary findings, oncogenic diseases and cardiogenic disease, carriers of variants in the cancer-susceptible genes **BRCA1**/2 and **MLH1** have been reported to have a family history of cancer in previous cohort study on secondary findings. Individuals carrying variants in genes such as **MYBPC3** and **KCNQ1** have also been reported to be related to cardiogenetic disease in researches about secondary findings. However, cardiovascular symptoms are very common and due to the lack of specificity of symptoms and difficulty in obtaining clinical evidence, it is impossible to determine whether cardiovascular symptoms are related to cardiac diseases such as cardiomyopathy and arrhythmia. In our follow-up study, the secondary variant-associated disease categories with clinical diagnostic evidence included oncogenic diseases (**BRCA1** variants c.811G>A, c.427G>A and **BRCA2** c.1832C>G-related breast cancer, **MSH2** c.2197G>A-related Lynch syndrome) in 4 families, hypercholesterolemia (**APOB** c.1342G>A, c.35_39del-related familial hypercholesterolemia), cardiogenic diseases (**TNNT2** c.422G>A-related dilated cardiomyopathy, **MYBPC3** c.2543C>T-related hypertrophic cardiomyopathy) and connective tissue diseases (**COL3A1** c.3133G>A-related Ehlers-Danlos syndrome-vascular type) in 2 families separately. Clinical and genetic evidence was collected to confirm the disease diagnosis and cosegregation of the variants. In addition to reporting of three most common disease categories-oncogenic diseases, hypercholesterolemia and cardiogenic diseases, our research first provided clinical diagnostic evidence for variants related to rare connective disease Ehlers-Danlos syndrome-vascular type in secondary variant research. In family members carrying **COL3A1** c.3133G>A in two unrelated families, typical cutaneous features of the disease, such as brown-black plaque, pityriasis versicolor-like lesions and cerebrovascular complications aneurysm accompanied by hernias, and perforation of organs were observed in family members carrying the variant.

Past studies have applied different criteria filtering procedures and classification of secondary variants. Interpreting secondary variants not related to primary disease requires a large amount of exploration work from researchers and clinicians and could be very challenging. Using ACMG guidelines alone or as the final screening step, as most published articles did, could indeed identify the majority of secondary variants with actual evidence reflected by our results.
However, restricting the variant classifications to P/LP in the guidelines might decrease the sensitivity of the identification of secondary variants such that the frequency of these variants was lower than 5% after variant filtrations. Some studies relied on the HGMD database accompanied by manual review of primary reviews to define the pathogenicity of secondary variants. The HGMD database is one of the most common medically curated databases used for determining the pathogenicity of variants. However, the variants in HGMD may be nonspecific and oversensitive. There were limited primary articles for variant curation, which might lead to inevitable misclassifications of variants. For example, one of the primary articles included in the curation of MSH2: c.2197G > A in the HGMD database explored the role of the variant in medulloblastoma. However, the disease is not well-established in Lynch syndrome-related tumors, such as cancers in the colon, rectum, and endometrium. Therefore, our preliminary exploration suggestions on the screening strategy of secondary variants included variants according to ACMG guidelines (P/LP) or the HGMD database (DM/DM?) to elevate the detection rate of clinical diagnostic variants and thus refine the screening strategies. Additionally, from the variants included in our research, we identified two DM?/VUS variants: COL3A1 c.3133G > A, BRCA1 c.427G > A in three families with confirmed clinical and genetic evidence indicating that VUS variants could also be disease-variant-related variants. Although the possibility of a given variant of solely classification of VUS as a disease-causing variant is extremely low, the "clinically significant" VUS subgroup variants such as the two DM7/VUS variants should be considered in future research to develop strategies of identifying secondary variants. Therefore, future studies are needed to explore further filtration strategies for secondary variants classified as DM7/VUS.

The penetrance of secondary variants could range from 20–100% at different evaluation ages, and such variable variant penetrance may cause the genetic counseling of secondary variants to be very challenging. We addressed penetrance of specific variants in families with actual clinical and genetic evidence, and the penetration calculation was carried out based on more than 3 generations in families that had a broader age range than routine cohort studies collected at a certain age stage. The family-specific penetrance could decrease age-related bias and was more accurate in reflecting the actual variant penetrance, especially for variants in child/adult-onset genes. It has been reported that the evaluation age of a specific age might filter a number of high-penetrance cancer variants for mortality at younger ages. In our 10 families, the penetrance of secondary variants ranged from 67–100%, which was in accordance with the high-penetrance feature of secondary variants. Sexual penetrance differences of BRCA1-2 were also demonstrated in our research; female penetrance was higher than male penetrance.

The value of reporting secondary findings was shown in family members who did not exhibit overt disease phenotypes. These secondary findings may provide significant opportunities for asymptomatic individuals to prevent disease by surveillance for family members. It is crucial for the prediction and prevention of severe diseases by screening deleterious secondary variant carriers. For example, after learning about their grandfather's medical history of subarachnoid hemorrhage, two family members carrying COL3A1 c.3133G > A underwent CTA examination and were surgically treated by neurosurgeons. As significant symptoms do not usually appear before the rupture of aneurysms, the action has played a preventive role in this disease, which may be lethal and seriously affect the quality of life of patients. In the family of multiple sudden deaths and the grandfather with hypertrophic cardiomyopathy, the father of the proband was also suggested to undergo a relevant test after being reported as having the secondary variant MYBPC3 c.2543C > T and was also diagnosed with the disease. It is a life-changing action that leads to early detection of disease and early prevention of asymptomatic patients. Therefore, our research indicated that if a secondary variant with genetic and clinical evidence is identified in a patient, it is worth checking for the actual secondary variant in their asymptomatic family members to prevent future risk of actionable high-penetrance diseases by lifestyle changes or medical and/or surgical intervention.

Although our analysis is a comprehensive follow-up study of secondary findings, the research had some inevitable limitations. First, the included populations were from the clinics of our hospital. A total of 1330 participants could not compare to the sample size of multicenter WES studies, which may increase the generalizability of the variant interpretations. Second, the people who were lost to follow-up or refused the test would lead to lower final positive results of the follow-up study. Third, although our exploration of the filtration of actionable secondary variants may increase the sensitivity of the variants by incorporation of VUS/DM? variants, the specificity of the variants was compromised. Further bioinformatic studies based on clinical and genetic evidence need to be performed to seek appropriate procedures identifying reportable actionable secondary variants.

Conclusions

In conclusion, our comprehensive study provided a choice of refinement in the report of secondary findings so that the secondary variants could be more likely to be of actual clinical significance and reduce the omission of positive families. The penetrance calculations based on family studies were closer to the actual variant penetrance, which provided a valuable reference for clinical genetic counseling. Because overrated reportable secondary variants might cause anxiety and unnecessary examination of patients and family members and too stringent would filter positive variants and neglect true cause of secondary variant-related diseases, taking the actual clinical consequences into screening strategies of the variants highlights the areas of future research and contribute to translation of genetic findings into medical practice.

Declarations

Acknowledgements

We thank all the participated patients and their families for their cooperation, as well as all the cooperators and their team members.

Authors’ contributions

XM designed the research. LS interpreted the data and wrote the manuscript. HYL, NX, YXL, WXY, YJL, HX, ZJJ did the follow-up study and collected, evaluated the clinical and genetic evidence. FLY and HW revised the manuscript. All authors read and approved the final manuscript.
Funding

This work was supported by the National Natural Science Foundation of China (grant number 81801136), Hunan Provincial Major Science and Technology Project (grant number 2019SK1010) and the China Postdoctoral Science Foundation (2019M662804).

Availability of data and materials

The data supporting the conclusions of this article are included within the article and the supplementary materials.

Competing interests

The authors declare no conflicts of interest.

Ethics approval and consent to participate

The study was approved by the ethics committee of the Maternal and Child Health Hospital of Hunan Province. The study was performed in accordance with the principles of the Declaration of Helsinki.

Consent for publication

All participants signed informed consent forms for publication.

References

1. Berg JS, Amendola LM, Eng C, et al. Processes and preliminary outputs for identification of actionable genes as incidental findings in genomic sequence data in the Clinical Sequencing Exploratory Research Consortium. Genet Med. 2013;15:860–7.
2. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15:565–74.
3. Kalia SS, Adelman K, Bale SJ, 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;19:249–55.
4. Weck KE. Interpretation of genomic testing: variants should be considered uncertain until proven guilty. Genet Med. 2018;20:291–3.
5. Kaphingst KA, Ivanovich J, Biesecker BB, et al. Preferences for return of incidental findings from genome sequencing among women diagnosed with breast cancer at a young age. Clin Genet. 2016;89:378–84.
6. 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;16:741–50.
7. 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;91:97–108.
8. Chen W, Li W, Ma Y, et al. Secondary findings in 421 whole exome-sequenced Chinese children. Hum Genomics. 2018;12:42.
9. Thauvin-Robinet C, Thevenon J, Nambot S, et al. Secondary actionable findings identified by exome sequencing: expected impact on the organisation of care from the study of 700 consecutive tests. Eur J Hum Genet. 2019;27:1197–214.
10. Haer-Wigman L, van der Schoot V, Feenstra I, et al. 1 in 38 individuals at risk of a dominant medically actionable disease. Eur J Hum Genet. 2019;27:325–30.
11. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
12. Richards S, Aziz N, Bale S, 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;17:405–24.
13. Mathew T, Williams L, Navaratnam G, et al. Diagnosis and assessment of dilated cardiomyopathy: a guideline protocol from the British Society of Echocardiography. Echo Res Pract. 2017;4:G1–13.
14. Al-Rasadi K, Al-Waili K, Al-Sabti HA, et al. Criteria for Diagnosis of Familial Hypercholesterolemia: A Comprehensive Analysis of the Different Guidelines, Appraising their Suitability in the Omani Arab Population. Oman Med J. 2014;29:85–91.
15. Samadder NJ, Smith KR, Wong J, et al. Cancer Risk in Families Fulfiling the Amsterdam Criteria for Lynch Syndrome. JAMA Oncol. 2017;3:1697–701.
16. Malfait F, Francomano C, Byers P, et al. The 2017 international classification of the Ehlers-Danlos syndromes. Am J Med Genet C Semin Med Genet. 2017;175:8–26.
17. Pepin M, Schwarze U, Superti-Furga A, Byers PH. Clinical and genetic features of Ehlers-Danlos syndrome type IV, the vascular type. N Engl J Med. 2000;342:673–80.
18. Authors/Task Force m. Elliott PM, Anastasakis A, et al: 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). Eur Heart J 2014; 35: 2733–79.
19. Hart MR, Biesecker BB, Blout CL, et al. Secondary findings from clinical genomic sequencing: prevalence, patient perspectives, family history assessment, and health-care costs from a multisite study. Genet Med. 2019;21:1100–10.
Tables

Table 1. Characteristics of included secondary variants in 48 genes classified by disease categories and the follow-up steps
| Genes | Number of variants | Associated disease | Classifications | Number of de novo patients and families remained after primary medical evaluation | Number of families remained after secondary medical evaluation |
|-------|-------------------|--------------------|-----------------|---------------------------------|---------------------------------|
| DSP   | 8                 | Arrhythmogenic right ventricular cardiomyopathy | cardiogenic diseases | 49 | 2 | 0 |
| MYBPC3 | 19               | Hypertrophic cardiomyopathy, dilated | cardiogenic diseases | 41 | 7 | 1 |
| MYH7  | 14                | Hypertrophic cardiomyopathy, dilated | cardiogenic diseases | 40 | 3 | 0 |
| RYR2  | 13                | Catecholaminergic polymorphic ventricular tachycardia | cardiogenic diseases | 40 | 1 | 0 |
| TMEM43 | 3                | Arrhythmogenic right ventricular cardiomyopathy | cardiogenic diseases | 37 | 0 | 0 |
| DSG2  | 9                 | Arrhythmogenic right ventricular cardiomyopathy | cardiogenic diseases | 27 | 3 | 0 |
| TNNI3 | 2                 | Hypertrophic cardiomyopathy, dilated | cardiogenic diseases | 23 | 0 | 0 |
| SCN5A | 8                 | Romano-Ward long-QT syndrome types 1, 2, and 3, Brugada syndrome | cardiogenic diseases | 17 | 0 | 0 |
| KCNH2 | 7                 | Romano-Ward long-QT syndrome types 1, 2, and 3, Brugada syndrome | cardiogenic diseases | 11 | 0 | 0 |
| PKP2  | 3                 | Arrhythmogenic right ventricular Cardiomyopathy | cardiogenic diseases | 8 | 0 | 0 |
| KCNQ1 | 5                 | Romano-Ward long-QT syndrome types 1, 2, and 3, Brugada syndrome | cardiogenic diseases | 6 | 0 | 0 |
| TNNT2 | 4                 | Hypertrophic cardiomyopathy, dilated | cardiogenic diseases | 5 | 3 | 1 |
| DSC2  | 3                 | Arrhythmogenic right ventricular cardiomyopathy | cardiogenic diseases | 5 | 1 | 0 |
| MYL3  | 3                 | Hypertrophic cardiomyopathy, dilated | cardiogenic diseases | 4 | 0 | 0 |
| GLA   | 2                 | Hypertrophic cardiomyopathy, dilated | cardiogenic diseases | 4 | 0 | 0 |
| LMNA  | 2                 | Hypertrophic cardiomyopathy, dilated | cardiogenic diseases | 3 | 0 | 0 |
| PRKAG2 | 1                | Hypertrophic cardiomyopathy, dilated cardiomyopathy | cardiogenic diseases | 2 | 0 | 0 |
| FBN1  | 7                 | Marfan syndrome, Loeys-Dietz syndromes, and familial thoracic aortic | connective tissue diseases | 31 | 0 | 0 |
| Gene   | Mutations | Syndrome/Condition                                                                 |
|--------|-----------|------------------------------------------------------------------------------------|
| COL3A1 | 2         | Ehlers-Danlos syndrome, vascular type connective tissue diseases                    |
| MYH11  | 5         | Marfan syndrome, Loeys-Dietz syndromes, connective tissue diseases and familial thoracic aortic aneurysms and dissections |
| TGFBR1 | 2         | Marfan syndrome, Loeys-Dietz syndromes, connective tissue diseases and familial thoracic aortic aneurysms and dissections |
| TGFBR2 | 3         | Marfan syndrome, Loeys-Dietz syndromes, connective tissue diseases and familial thoracic aortic aneurysms and dissections |
| ACTA2  | 1         | Marfan syndrome, Loeys-Dietz syndromes, connective tissue diseases and familial thoracic aortic aneurysms and dissections |
| APOB   | 23        | Familial hypercholesterolemia hypercholesterolemia                                  |
| LDLR   | 11        | Familial hypercholesterolemia hypercholesterolemia                                  |
| PCSK9  | 6         | Familial hypercholesterolemia hypercholesterolemia                                  |
| BRCA2  | 26        | Hereditary breast and ovarian cancer ontogenetic diseases                            |
| TSC1   | 4         | Tuberous sclerosis complex ontogenetic diseases                                      |
| APC    | 10        | Familial adenomatous polyposis ontogenetic diseases                                  |
| MSH2   | 8         | Lynch syndrome ontogenetic diseases                                                 |
| RET    | 5         | Multiple endocrine neoplasia type 2 ontogenetic diseases                             |
| TSC2   | 9         | Tuberous sclerosis complex ontogenetic diseases                                      |
| BRCA1  | 11        | Hereditary breast and ovarian cancer ontogenetic diseases                            |
| MSH6   | 10        | Lynch syndrome ontogenetic diseases                                                 |
| MLH1   | 5         | Lynch syndrome ontogenetic diseases                                                 |
| MEN1   | 1         | Multiple endocrine neoplasia type 1 ontogenetic diseases                             |
| TP53   | 4         | Li-Fraumeni syndrome ontogenetic diseases                                           |
| RB1    | 3         | Retinoblastoma ontogenetic diseases                                                 |
| PMS2   | 3         | Lynch syndrome ontogenetic diseases                                                 |
| SDHD   | 2         | Hereditary paragangliomapheochromocytoma syndrome ontogenetic diseases               |
| PTEN   | 2         | PTEN hamartoma tumor syndrome ontogenetic diseases                                  |
| STK11  | 2         | Peutz-Jeghers syndrome ontogenetic diseases                                         |
| BMPR1A | 1         | Juvenile polyposis ontogenetic diseases                                             |
| SDHB   | 1         | Hereditary paragangliomapheochromocytoma syndrome ontogenetic diseases               |
| Gene.refGene | Chr  | Start    | End      | Ref  | Alt  | Genotype | Func.refGene | ExonicFunc.refGene | AAChange.refGene          | Classifications |
|-------------|------|----------|----------|------|------|----------|--------------|-------------------|-------------------------|-----------------|
| **NF2**     | 1    | Neurofibromatosis type 2 | 1 | 0 | 0 | 0 | 0 |
| **RYR1**    | 16   | Malignant hyperthermia susceptibility | others | 31 | 0 | 0 | 0 |
| **CACNA1S** | 4    | Malignant hyperthermia susceptibility | others | 6 | 0 | 0 | 0 |
| **OTC**     | 3    | Ornithine transcarbamylase deficiency | others | 4 | 0 | 0 | 0 |

Table 2. Families with sufficient clinical evidence, however, were not confirmed by Sanger sequencing

| Gene.refGene | Chr | Start | End | Ref  | Alt | Genotype | Func.refGene | ExonicFunc.refGene | AAChange.refGene | Classifications |
|--------------|-----|-------|-----|------|-----|----------|--------------|-------------------|-----------------|-----------------|
| APOB         | chr2 | 21263905 | 21263905 | C   | A   | het      | exonic       | nonsynonymous SNV | NM_000384.2:exon4: c.288G>T:p.Q96H | DM/LP |
| MLH1         | chr3 | 37053562 | 37053562 | C   | T   | het      | exonic       | nonsynonymous SNV | NM_000249.3:exon8: c.649C>T:p.R217C | DMII/VUS |
| BRCA2        | chr13 | 32929042 | 32929042 | C   | G   | het      | exonic       | nonsynonymous SNV | NM_000059.3:exon14: c.7052G>T:p.A2351G | DMII/VUS |

Table 3. Nine secondary variants in 10 families with convincing clinical and genetic evidence

| Gene.refGene | Chr  | Start    | End      | Ref  | Alt  | Genotype | Func.refGene | ExonicFunc.refGene | AAChange.refGene | Classifications |
|-------------|------|----------|----------|------|------|----------|--------------|-------------------|-----------------|-----------------|
| TNNT2       | chr1 | 20133463 | 20133463 | C   | T   | het      | exonic       | nonsynonymous SNV | NM_001001430.2:exon10: c.422G>A:p.R141Q | DM/LP |
| APOB        | chr2 | 21255236 | 21255236 | C   | T   | het      | exonic       | nonsynonymous SNV | NM_000384.2:exon10: c.1342G>T:p.A448T | DM/ |
| APOB        | chr2 | 21266779 | 21266783 | C   | T   | het      | exonic       | frameshift deletion | NM_000384.2:exon1: c.35_39del:p.L12Pfs*44 | LP |
| MSH2        | chr2 | 47703697 | 47703697 | G   | A   | het      | exonic       | nonsynonymous SNV | NM_000251.2:exon13: c.2197G>A:p.A733T | DM/ |
| COL3A1      | chr2 | 18987110 | 18987110 | G   | A   | het      | exonic       | nonsynonymous SNV | NM_000090.3:exon43: c.3135G>A:p.A1045T | DM/ |
| MYBPC3      | chr11 | 47359001 | 47359001 | G   | A   | het      | exonic       | nonsynonymous SNV | NM_000256.3:exon25: c.2543C>G:p.A848V | DM/ |
| BRCA2       | chr13 | 32907447 | 32907447 | C   | G   | het      | exonic       | stopgain          | NM_000059.3:exon10: c.1832C>G:p.S611* | DM/ |
| BRCA1       | chr17 | 41246737 | 41246737 | C   | T   | het      | exonic       | nonsynonymous SNV | NM_007294.3:exon10: c.811G>C:p.A271T | DM/ |
| BRCA1       | chr17 | 41256153 | 41256153 | C   | T   | het      | exonic       | nonsynonymous SNV | NM_007294.3:exon6: c.427G>A:p.E143K | DM/ |

Table 4. Secondary findings-related disease phenotypes in family members
| Family | Gene     | Variant | Individual | Genotype         | Age  | Detailed clinical features                                      |
|--------|----------|---------|------------|------------------|------|----------------------------------------------------------------|
| 1      | TNNT2    | c.452G>A| II:1        | Heterozygous variant | 54   | dilated cardiomyopathy                                         |
| 1      | TNNT2    | c.452G>A| II:3        | Heterozygous variant | 53   | dilated cardiomyopathy                                         |
| 1      | TNNT2    | c.452G>A| III:2       | Heterozygous variant | 31   | dilated cardiomyopathy                                         |
| 2      | APOB     | c.1342G>A| I:2         | NA*              | 71   | 71y died of cerebral infarction, hyperlipidemia               |
| 2      | APOB     | c.1342G>A| II:1        | NA*              | 69   | 69y died of cerebral infarction, hyperlipidemia               |
| 2      | APOB     | c.1342G>A| II:2        | Heterozygous variant | 62   | TC 7.68mmol/L LDL-C 4.97mmol/L                               |
| 2      | APOB     | c.1342G>A| II:6        | Heterozygous variant | 51   | Facial paralysis after cerebral infarction at the age of 50; TC 9.22mmol/L LDL-C 5.04mmol/L |
| 2      | APOB     | c.1342G>A| III:1       | Heterozygous variant | 35   | Fatty liver; TC 7.91mmol/L LDL-C 5.17mmol/L               |
| 2      | APOB     | c.1342G>A| III:3       | Heterozygous variant | 34   | TC 8.18mmol/L LDL-C 5.46mmol/L                               |
| 3      | APOB     | c.35_39del| I:1         | Patient refused the test | 68   | coronary heart disease for at least 20 years               |
| 3      | APOB     | c.35_39del| II:2        | Heterozygous variant | 51   | TC 8.05mmol/L LDL-C 5.39mmol/L                               |
| 3      | APOB     | c.35_39del| II:3        | Heterozygous variant | 46   | TC 7.72mmol/L LDL-C 4.98mmol/L                               |
| 3      | APOB     | c.35_39del| III:2       | Heterozygous variant | 28   | TC 7.99mmol/L LDL-C 5.51mmol/L                               |
| 4      | MSH2     | c.2197G>A| II:3        | NA*              | 65   | 65y died of rectal cancer                                     |
| 4      | MSH2     | c.2197G>A| II:4        | Heterozygous variant | 62   | rectal cancer                                                 |
| 4      | MSH2     | c.2197G>A| II:1        | Heterozygous variant | 41   | rectal cancer                                                 |
| 5      | COL3A1   | c.3133G>A| II:1        | NA*              | 62   | inguinal hernia                                               |
| 5      | COL3A1   | c.3133G>A| II:2        | NA#              | 50   | subarachnoid hemorrhage due to intracranial aneurysm          |
| 5      | COL3A1   | c.3133G>A| III:2       | Heterozygous variant | 31   | Intracranial aneurysm                                        |
| 5      | COL3A1   | c.3133G>A| III:3       | Heterozygous variant | 28   | Intracranial aneurysm                                        |
| 5      | COL3A1   | c.3133G>A| III:5       | Heterozygous variant | 25   | skin brown-black plaque and ecchymosis                       |
| 5      | COL3A1   | c.3133G>A| IV:2        | Heterozygous variant | 4    | skin scar hyperplasia and pityriasis versicolor-like lesions |
| 6      | COL3A1   | c.3133G>A| II:4        | NA*              | 22   | acute myocardial infarction                                  |
| 6      | COL3A1   | c.3133G>A| II:5        | Heterozygous variant | 49   | Spontaneous gastric perforation                               |
| 6      | COL3A1   | c.3133G>A| III:2       | Heterozygous variant | 27   | Spontaneous gastric perforation                               |
| 7      | MYBPC3   | c.2543C>T| I:2         | NA*              | before 40 | sudden death before 40                                       |
| 7      | MYBPC3   | c.2543C>T| II:1        | NA*              | 45   | 45y died of cardiac arrest                                    |
| 7      | MYBPC3   | c.2543C>T| II:3        | NA*              | before 40 | sudden death before 40                                       |
| 7      | MYBPC3   | c.2543C>T| III:1       | Heterozygous variant | 47   | hypertrophic cardiomyopathy                                  |
| Family | Gene   | Variant        | Generation | Status       | Age | Cause of Death                        |
|--------|--------|----------------|------------|--------------|-----|---------------------------------------|
| 7      | MYBPC3 | c.2543C>T      | III:3      | NA*          | 38  | 38y died of cardiac arrest            |
| 7      | MYBPC3 | c.2543C>T      | IV:1       | Heterozygous variant | 25  | hypertrophic cardiomyopathy           |
| 8      | BRCA2  | c.1832C>G      | I:2        | NA*          | 55  | 55y died of breast cancer             |
| 8      | BRCA2  | c.1832C>G      | II:3       | Heterozygous variant | 52  | breast cancer in the left breast      |
| 8      | BRCA2  | c.1832C>G      | III:2      | Heterozygous variant | 31  | breast cancer in the left breast      |
| 9      | BRCA1  | c.811G>A       | I:2        | Heterozygous variant | 52  | bilateral breast cancer               |
| 9      | BRCA1  | c.811G>A       | II:2       | Heterozygous variant | 28  | breast cancer in the left breast      |
| 10     | BRCA1  | c.427G>A       | II:2       | Heterozygous variant | 51  | breast cancer in the right breast     |
| 10     | BRCA1  | c.427G>A       | III:1      | Heterozygous variant | 31  | breast cancer in the right breast     |
| 10     | BRCA1  | c.427G>A       | III:3      | Heterozygous variant | 29  | breast cancer in the right breast     |

Abbreviations: *, Dead; #, Patient refused the test.

**Figures**

- Microcephaly
- Developmental and epileptic encephalopathy
- Developmental delay/Intellectual disability
- Others (Metabolic diseases, Epilepsy, Skeletal dysplasia etc.)
- unaffected family members of patients

**Figure 1**

Pie chart of the frequency distribution of primary diseases in the population.
Flowchart

**step 1**
- A total of 1330 participants were enrolled.
- WES was conducted and variants within the 59 genes were called and annotated.

**step 2**
- Step 1: Filtering variants according to HGMD and ACMG guidelines
- 8851 secondary variant counts and 450 variants were identified

**step 3**
- Step 2: Filtering variants according to allele frequency, variant type and inheritance mode
- 998 secondary variant counts and 297 variants were identified

**step 4**
- 638 individuals were included for follow-up study
- Clinical and genetic evidence were collected

**step 5**
- 10 families with convincing clinical and genetic evidence were identified
- Variant specific penetrance was calculated and literature review was done

---

Figure 2

Flowchart of variant screening and follow-up study.
Figure 3

3a. Venn diagram of variants identified after variant filtering step 1. 3b. Venn diagram of variants identified after variant filtering step 2.

Figure 4

4a. Number of individuals carrying secondary variants in different categories. 4b. Number of secondary variants in different categories.

Family members of 420 participants with hereditary secondary variants and 112 participants with de novo secondary variants

Primary medical evaluation: current and past medical history assessment to identify
- Confirmed individuals fitting diagnostic criteria
- Possible individuals fitting at least one clinical manifestation in OMIM database (https://omim.org/)

Secondary medical evaluation: physical examination or necessary examinations assessment to identify individuals with convincing clinical evidence

Participants with MYBPC3 variants

Confirmed individuals fitting diagnostic criteria:
The European Society of Cardiology (ESC) on the diagnosis and management of hypertrophic cardiomyopathy

Possible individuals fitting at least one clinical manifestation in OMIM database:
Cardiomyopathy, hypertrophic, 4 (MIM 115197) showing chest pain, cardiomegaly, bilateral ventricular hypertrophy, cardiac arrest etc.

Echocardiography (ECG), cardiac magnetic resonance imaging (CMR)

Figure 5

The detailed screening strategies to identify individuals with convincing clinical evidence in the follow-up study exemplified by participants carrying MPBPC3 variants.
Figure 6

Pedigrees of family 1 with TNNT2 c.422G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.

Figure 7

Pedigrees of family 1 with TNNT2 c.422G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.
Sanger sequencing results of TNNT2 c.422G>A in family members.

Figure 8

Pedigrees of family 2 with APOB c.1342G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.

8a

8b

8c
Echocardiographic findings in TNNT2 c.422G>A family members. Echocardiography showed that all cases demonstrated left ventricular enlargement, decreased cardiac contractility, and variant degrees of mitral valve regurgitation, with or without right ventricular dysfunction. 8a. LVIDd 74.1 mm, LVIDs 60.2 mm, LAD 52.5 mm, LVEF 37.6%. 8b. LVIDd 64.8 mm, LVIDs 54.2 mm, LAD 36.4 mm, LVEF 33.2%. 7c. LVIDd 63.4 mm, LAD 37 mm. LVIDd/LVIDs: left ventricular internal dimension-diastole/systolic, LAD: left atrium diameter, LVEF: left ventricular ejection fraction.

Sanger sequencing results of APOB c.1342G>A in family members.
Figure 11

Pedigrees of family 3 with APOB c.35_39del. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.

Figure 12

Sanger sequencing results of APOB c.35_39del in family members.
Figure 13
Pedigrees of family 4 with MSH2 c.2197G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.

Figure 14
Sanger sequencing results of MSH2 c.2197G>A in family members.
Figure 15
Histologic features of MSH2 c.2197G>A family members. 15a. Histologic image showed middle to low differentiated adenocarcinoma, part of which was mucinous adenocarcinoma. The tumor invaded the deep myometrium (ulcerative mass, 3.5 × 2 × 1 cm in size); tumor thrombus and perineural invasion were seen in the vessel; no cancer metastasis was found in the lymph nodes (0 / 7). 15b. Histologic image showed low differentiated adenocarcinoma (tumor size 6 × 4 × 2 cm). The tumor invaded the whole layer, and cancer tissue could be seen in the abdominal wall. Mesenteric lymph nodes showed cancer metastasis (1 / 40); Peri-intestinal lymph nodes showed no cancer metastasis (0 / 8).

Figure 16
Pedigrees of family 5 with COL3A1 c.3133G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.

Figure 17

Brain CT showed subarachnoid hemorrhage.
Figure 18

Sanger sequencing results of COL3A1 c.3133G>A in family members.
19a-c. Brain CTA showed an aneurysm of the internal carotid system.

19a-c. Brain CTA showed an aneurysm of the internal carotid system.
Skin manifestations in COL3A1 c.3133G>A family members. 20a. Brown-black plaque above the left ankle manifesting atrophic skin and ulcers. 20b. Ecchymosis on the left side of the knee. 20c. Scar hyperplasia surrounded by hypopigmented pityriasis versicolor-like lesions over the knee.

Figure 21

Pedigrees of family 6 with COL3A1 c.3133G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and --, respectively.

- myocardial infarction
- gastric perforation
Figure 22

Sanger sequencing results of COL3A1 c.3133G>A in family members.

Figure 23

Pedigrees of family 7 with MYBPC3 c.2543C>T. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.

hypertrophic cardiomyopathy
Figure 24

Sanger sequencing results of MYBPC3 c.2543C>T in family members.

Figure 25

Echocardiographic findings in MYBPC3 c.2543C>T family members. 25a. Asymmetrical hypertrophy of IVS and LVPW and moderate obstruction of LVOT, with obvious ventricular diastolic dysfunction. IVSd 24.4 mm, LVPWd 11.6 mm, EPSS 0 mm, LVOT velocity 2.48 m/s. 25b. Asymmetrical hypertrophy of IVS and LVPW and no obstruction of LVOT. IVSd 34.7 mm, LVPWd 11.5 mm, LAD 40.1 mm, EPSS 9.7 mm, LVOT velocity 1.23 m/s. IVSd: Interventricular septal-diastole, LVPWd: Left ventricular posterior wall-diastole, LAD: Left atrium diameter, EPSS: E-point septal separation, LVOT: left ventricular outflow tract.
Figure 26

Pedigrees of family 8 with BRCA2 c.1832C>G. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.
Figure 27

Sanger sequencing results of BRCA2 c.1832C>G in family members.

Figure 28

Histologic features of BRCA2 c.1832C>G family members. 28a. Histologic image showed (left breast) invasive carcinoma, nonspecial type, grade II histology, no clear tumor thrombus in the vessel, no nerve invasion; (base, skin margin and nipple) no cancer; (left armpit) lymph node with cancer metastasis (5 / 6).
28b. Histologic image showed (left breast) grade I invasive ductal carcinoma. No cancer was found in the upper, lower, inner, outer and bottom cutting edges, and no cancer metastasis was found in the lymph nodes (0 / 14).

Figure 29
Pedigrees of family 9 with BRCA1 c.811G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.

Figure 30
Pedigrees of family 9 with BRCA1 c.811G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.
Sanger sequencing results of BRCA1 c.811G>A in family members.

Figure 31
Histologic features of BRCA1 c.811G>A family members. 31a. Histologic image showed (left breast) invasive carcinoma, inclined to tubular carcinoma, grade I histology and (right breast) invasive cancer; the focus tends to be apocrine adenoid type, histology grade II. 31b. Histologic image showed (left breast) invasive carcinoma, nonspecial type, grade II histology, with more lymphocytic infiltration, 5 × 4 × 3 cm in size, no tumor thrombus in vessel, no definite nerve invasion, (basal, skin margin and papilla) no cancer invasion, (left axillary) lymph node no cancer metastasis (0 / 18).
Figure 32

Pedigrees of family 10 with BRCA1 c.427G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.

Figure 33

Pedigrees of family 10 with BRCA1 c.427G>A. Individuals with heterozygous variants and without the variant are indicated by +/- and -/-, respectively.
Sanger sequencing results of BRCA1 c.427G>A in family members.

Figure 34

Histologic features of BRCA1 c.427G>A family members. 34a. Histologic image of II:2 showed (right breast) nonspecific invasive carcinoma, grade II. 34b. Histologic image of III:1 showed (right breast) invasive carcinoma, nonspecific type, grade III. 34c. Histologic image of III:3 showed (right breast) nonspecific invasive carcinoma of grade II, part of which is high-grade intraductal carcinoma. There was no tumor thrombus in the vessel, no definite nerve invasion, intraductal carcinoma change under the nipple, and no cancer metastasis (0 / 3) in the sentinel lymph node.