npj | genomic medicine

Article

Published in partnership with CEGMR, King Abdulaziz University

https://doi.org/10.1038/s41525-026-00565-0

Population-scale genomic screening reveals high frequency of actionable secondary findings in Chinese newborns

Yushan Huang[1,2], Ya Gao[2,3], Zonghao Duan[1,4], Xiao Jia[5], Yue Sun[6], Chunhua Liu[7], Hui Huang[8], Junnian Liu[9], Silin Pan[6], Xin Jin[2,3] & Mingyan Fang[2,3] ✉

Secondary findings (SFs) from genome sequencing have significant implications for disease prevention and early intervention, yet their population-specific spectrum remains poorly characterized in non-European cohorts. We performed whole-genome sequencing of 6685 Chinese newborns and evaluated pathogenic variants in 84 genes from the American College of Medical Genetics and Genomics (ACMG) SF v3.3 list according to ACMG/Association for Molecular Pathology (AMP) classification guidelines, and cross-referenced against ClinVar. We identified 306 unique actionable variants, comprising 172 known pathogenic variants (KP) and 134 expected pathogenic variants (EP). When heterozygous carriers of autosomal recessive (AR) variants were included, 9.12% (610/6685) of newborns carried at least one pathogenic variant. Under ACMG SF criteria, clinically actionable variants were identified in 5.06% (338/6685) of newborns, predominantly affecting cardiovascular disease genes (3.49%) and cancer predisposition genes (1.26%), most commonly involving *LDLR*, *TTN*, and *BRCA2*. Importantly, 28 variants across 12 genes showed significant allele frequency divergence between Chinese and European ancestries, highlighting ancestry-specific genetic architecture. Our findings support the inclusion of high-penetrance genes prevalent in East Asian populations in population-tailored genomic screening panels, providing essential reference data for the equitable implementation of precision newborn genomics in underrepresented populations.

Over the past two decades, the rapid advancement of high-throughput sequencing (HTS) technologies has significantly reduced the costs of sequencing and genetic analysis, driving the widespread adoption of genetic sequencing in the clinical practice. Beyond enabling molecular diagnosis for specific clinical symptoms, HTS can also identify pathogenic genetic variants unrelated to the current symptoms but with potential medical significance—termed secondary findings (SFs)[1]. SFs, often associated with severe genetic disorders, play a critical role in early disease prevention, risk prediction, and precision healthcare management. According to guidelines from the American College of Medical Genetics and Genomics (ACMG)[2,3], known pathogenic (KP) or expected pathogenic (EP) variants within its recommended gene list should be clinically reported, regardless of their relevance to the primary diagnostic goal. The ACMG guidelines emphasize that early identification of SFs provides actionable insights for disease prevention and timely intervention, underscoring their value in improving clinical decision making and patient outcomes.

In 2013, the ACMG first recommended the reporting of SFs limited to known pathogenic (P) or likely pathogenic (LP) variants, initially comprising 56 medically actionable genes associated with 24 severe but clinically intervenable genetic disorders, including hereditary cancer syndromes, cardiomyopathies, and miscellaneous groups[2]. By 2017, the ACMG SF v2.0 expanded to 59 actionable genes, adding genes linked to arrhythmias and metabolic diseases[4]. In 2021, the ACMG updated to SF v3.0[5], incorporating 73 genes, including 14 newly added genes. Subsequently, it was further upgraded to SF v3.2 in 2023[6], encompassing 81 genes, reflecting significant advancements in human genetics and disease etiology over recent years.

The evolution of the ACMG guidelines has driven a decade of research into medically actionable variants across diverse ethnic populations,

[1]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [2]BGI Research, Shenzhen 518083, China. [3]State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Shenzhen 518083, China. [4]BGI Research, Wuhan 430074, China. [5]Huangdao Maternity and Child Health Care Hospital of Qingdao, Qingdao 266400, China. [6]Qingdao Women and Children's Hospital, Qingdao University, Qingdao 266034, China. [7]Huangdao Maternal and Child Health and Family Planning Service Center of Qingdao, Qingdao 266000, China. [8]BGI Genomics, Shenzhen 518083, China. [9]BGI Research, Qingdao 266555, China. ✉e-mail: fangmingyan@aliyun.com

revealing profound ethnic heterogeneity in the prevalence of SFs. Population-specific analyses have demonstrated striking disparities: European-ancestry populations exhibit a 2% prevalence of pathogenic SNVs, while African-ancestry populations show a lower prevalence of approximately 1.2%[7]. East Asian-ancestry populations display even greater complexity, with reported frequencies ranging from 1.67% to 21% across 196 Korean exomes (7%)[8], 2049 Japanese cohorts (21%)[9], 1559 Thai populations (11.9%)[10], 1496 Taiwanese samples (1.67%)[11] and 954 mixed East Asians (2.5%)[12]. This remarkable inter-population variability goes beyond mere epidemiological observation, fundamentally reshaping our conceptual framework for the implementation of clinical genomics, and underscoring the critical gap in comprehensive Chinese population data.

The 2025 release of ACMG SF v3.3 list, which includes 84 genes associated with 52 inherited disorders or phenotypes[13], has not yet been systematically evaluated in Chinese populations. Given the demonstrated population-specific variant architecture, empirical evaluation of the frequency of SFs in healthy Chinese populations is scientifically imperative. Here, we analyzed whole-genome sequencing data from 6,685 unselected newborns to systematically characterize the spectrum and population frequency of SFs under the updated ACMG SF v3.3 framework, thereby establishing a foundational evidence bases for the development of population-tailored genomic screening strategies in Chinese healthcare systems.

Results

Population characteristics

The 10 K Newborn Genome Screening Project (QDnewborn) is a large Chinese newborn cohort comprising 9992 neonates from Qingdao, Shandong Province, China. Of these, 6685 (66.90%) with high-quality genomic data (mean sequencing depth, 65.29×) and research consent were included in this analysis. The final cohort comprised 3392 males (50.74%) and 3293 females (49.26%), predominantly of Northern Han Chinese ancestry.

Actionable variants in Chinese newborns

Using the established QDnewborn cohort, we characterized pathogenic variants within the 84 ACMG-recommended SF genes in 6685 Chinese neonates. As shown in Fig. 1, a total of 399,037 variants within the 84 ACMG-reportable genes were identified in our cohort, including 365,090 single nucleotide variants (SNVs) and 33,947 insertions and deletions (InDels). Following variant annotation (see Methods), a subset of 221,422 variants across coding region, introns, splice sites and UTRs was selected for further pathogenicity assessment, with the coding region accounting for 5.56% (12,301/221,422) of the total. Among the coding region variants, nonsynonymous and synonymous variants comprised 96.84% (11,912/12,301) (Supplementary Table S1).

According to the ACMG/AMP standards and guidelines for variant classification (see Methods), 306 unique variants were defined as actionable pathogenic variants. Of these, 172 variants were classified as KP variants based on P/LP status (≥2 stars) in the ClinVar database, and the remaining 134 truncating variants were classified as EP variants (Supplementary Table S2). By inheritance pattern, 210 variants were associated with genes underlying autosomal dominant phenotypes, 94 with autosomal recessive phenotypes, and 2 with X-linked phenotypes. Detailed classification and supporting evidence for all P/LP variants are provided in Supplementary Table S2.

Among genes associated with autosomal recessive (AR) conditions, *MUTYH* showed the highest proportion of actionable variants (2.40%), followed by *GAA* (2.18%), *CYP27A1* (1.27%) and *ATP7B* (1.05%). In autosomal dominant (AD) inheritance patterns, the highest proportion of actionable variants were observed in the *DES* (1.03%), *LDLR* (0.95%), *MEN1* (0.91%) and *BRCA2* (0.86%) genes (Supplementary Table S3). Among X-linked genes, 0.60% of the variants in *GLA* gene were classified as actionable variants.

Estimation of the population frequency of actionable variants

At the individual level, approximately 9.12% (610/6685) of Chinese newborns carried at least one pathogenic variant, with an average of 0.10 pathogenic variants per newborn (Fig. 2A). After excluding heterozygous carriers of autosomal recessive variants (not meeting ACMG-reportable criteria), the clinically reportable population frequency of actionable pathogenic variants was 5.06% (338/6685) (Fig. 2B). For the autosomal dominant conditions, such as cardiovascular and cancer diseases, 3.49% (233/6685) of newborns harbored actionable variants in 26 cardiovascular disease genes, while 1.26% (84/6685) had variants in 13 cancer predisposition genes (Fig. 2C). For inherited metabolic disorders (AR conditions), 1.85% (124/6685) of the cohort were identified as heterozygous carriers of actionable variants, and no individuals with biallelic pathogenic variants were detected (Supplementary Table S4).

Gene-level analysis revealed that 11 recessive or X-linked genes had pathogenic variants detected in multiple individuals, with *ATP7B* (1.33%, 89/6685), *CYP27A1* (1.06%, 71/6685), *GAA* (0.73%, 49/6685) and *TRDN* (0.49%, 33/6685) showing the highest carrier frequencies in Chinese newborns (Fig. 2D). Notably, one compound heterozygote was identified in *ATP7B*. Pathogenic variants in *ATP7B* cause Wilson's disease, a condition characterized by impaired excretion of copper ions, resulting in toxic accumulation in the liver, brain, and other tissues[14]. The *CYP27A1* gene, newly added to the SF v3.3 list, which is associated with cerebrotendinous xanthomatosis (CTX). The reported incidence of CTX ranges from 1:134,970 to 1:461,358 among Europeans, 1:71,677 to 1:148,914 in the Americans, and approximately 1:64,000 in East Asians. Due to insidious symptoms and delayed diagnosis, the actual prevalence of CTX may be underestimated, making early identification through genetic screening crucial for improving prognosis[15].

In contrast, gene-level analysis demonstrated that multiple autosomal dominant genes harbored pathogenic variants across distinct individuals (Fig. 2E). Specifically, *LDLR* exhibited the highest frequency of pathogenic variants in our cohort, followed by *TTN*. Mutations in the *LDLR* gene are the primary cause of familial hypercholesterolemia (FH), accounting for approximately 60% to 80% of cases[16].

TTN mutations, particularly truncating variants (TTNtv), represent the most prevalent genetic cause of dilated cardiomyopathy (DCM), the most common childhood cardiomyopathy, which is associated with significant morbidity and mortality[17]. A total of 67 TTNtv were initially identified in our cohort through the LOF prediction pipeline (Supplementary Table S5). These variants were distributed across sarcomeric regions: 34.33% (23/67) located in the A-band, 52.24% (35/67) in the I-band, 7.46% (5/67) in the M-band, and 5.97% (4/67) near or within the Z-disk. Applying the *TTN*-specific filtering criteria described in Methods (PSI ≥ 0.9 for constitutively expressed exons or A-band localization for splice variants), 33 variants were excluded and 34 high-confidence TTNtv were retained for pathogenicity assessment: 23 located in the A-band, 5 in the M-band, 5 in the I-band, and 1 near the Z-disk. All 34 retained variants were classified as pathogenic/likely pathogenic and included in the actionable variant set. Notably, 68% (23/34) were located in the A-band, the sarcomeric region most strongly associated with dilated cardiomyopathy[18].

BRCA2 was the most common AD cancer predisposition gene, which is associated with increased susceptibility to breast and ovarian cancer[19,20]. We also identified actionable variants in twelve other cancer-related genes, including *PALB2*, *PMS2*, *BRCA1*, *MSH6*, *MEN1*, *SDHB*, *VHL*, *TSC2*, *RET*, *TP53*, *SDHAF2*, and *APC*. Specifically, no homozygous individuals were observed for any of the analyzed genes.

Population-specific patterns of actionable variants

We further investigated population stratification in the distribution of actionable pathogenic variants across global populations. The results revealed significant differences in the allele frequencies of actionable pathogenic variants between Chinese and European populations (Table 1). A total of 28 P/LP variants across 12 genes showed marked population-specific differences, highlighting the importance of ancestry-specific considerations in SFs reporting.

Notably, four variants exhibited elevated minor allele frequencies in the Chinese cohort compared to non-East Asian populations: *APOB*

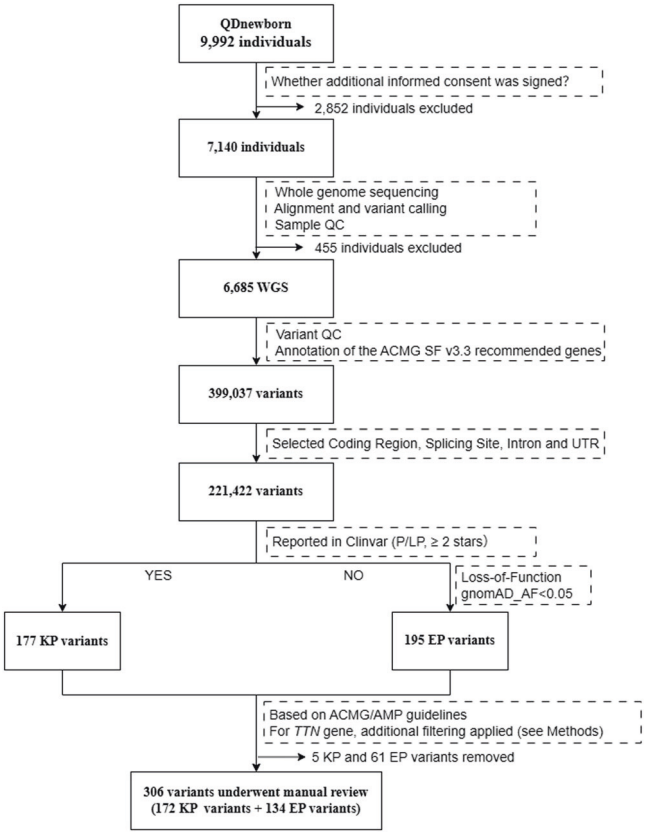

Fig. 1 | The workflow diagram of secondary findings (SFs).

(NM_000384.3: c.10579 C > T (p.Arg3527Trp)) (Fig. 3A), *ATP7B* (NM_000053.4:c.2975 C > T (p.Pro992Leu)) (Fig. 3B), *CYP27A1* (NM_000784.4:c.410 G > A (p.Arg137Gln)) (Fig. 3C), and *LDLR* (NM_000527.5:c.1747 C > T (p.His583Tyr)) (Fig. 3D). Conversely, European populations displayed significantly higher allele frequencies for *ATP7B* (NM_000053.4: c.122 A > G (p.Asn41Ser)) (Fig. 3E), *ATP7B* (NM_000053.4:c.2605 G > A (p.Gly869Arg)) (Fig. 3F), *GAA* (NM_000152.5:c.-32-13T > G) (Fig. 3G) and *HFE* (NM_000410.4:c.845 G > A (p.Cys282Tyr)) (Fig. 3H).

A particularly striking example was observed in the p.C282Y variant in the *HFE* gene. Its allele frequency exceeds 6.50% in European (EUR), Admixed American populations (AMR) and African/African American (AFR) ($AF_{gnomAD\_EUR} = 0.071$, $AF_{gnomAD\_AMR} = 0.015$ and $AF_{gnomAD\_AFR} = 0.011$), but is rarely detected in the East Asian population ($AF_{QDnewborn} = 0.0007$ and $AF_{gnomAD\_EAS} = 0.0002$). This genetic difference aligns with epidemiological data showing a lower burden of hereditary hemochromatosis in East Asians compared to other ethnic groups (*e.g.*, 12.5% in Ireland)[21]. *HFE*-related hemochromatosis (HH) represents the most prevalent genetic iron overload condition among European populations, especially those with northern European heritage. The p.C282Y variant in the *HFE* gene is the most common mutation, occurring in approximately 1 out of every 10 individuals of northern European ancestry, resulting in a homozygous frequency of 1 in 200 individuals[22,23]. In contrast, the *HFE* p.C282Y mutation is almost entirely absent in East Asian populations, such as Chinese, Koreans, and Japanese.

Given the ancestry-related variation in disease prevalence, the recommendation for secondary findings should prioritize phenotypes and genes with high penetrance in the East Asian population. For example, Pei-Kuan Cong et al. suggested that the highly penetrant *SLC25A13* gene in East Asian populations[24], which is associated with Citrin deficiency, should be included in the recommendation list. Citrin deficiency is an autosomal recessive inherited metabolic disorder initially reported and predominantly identified in individuals of East Asian ancestry. In our cohort, nine heterozygous P/LP variants in the *SLC25A13* gene (Supplementary Table S6), (NM_014251.3: c.852_855del (p.Met285Profs*2)) ($AF_{QDnewborn} = 0.0021$), (NM_014251.3.6: c.1177+1 G > A) ($AF_{QDnewborn} = 0.00075$), (NM_014251.3:c.615+5 G > A) ($AF_{QDnewborn} = 0.0007$), (NM_014251.3: c.1399 C > T (p.Arg467*)) ($AF_{QDnewborn} = 0.0005$), (NM_014251.3: c.1048 G > A (p.Asp350Asn)) ($AF_{QDnewborn}$=0.0002), (NM_014251.3: c.550 C > T (p.Arg184*)) ($AF_{QDnewborn} = 0.00015$), (NM_014251.3: c.1078 C > T (p.Arg360*)) ($AF_{QDnewborn} = 0.000075$), (NM_014251.3: c.955 C > T (p.Arg319*)) ($AF_{QDnewborn} = 0.000075$) and (NM_014251.3: c.775 C > T (p.Gln259*)) ($AF_{QDnewborn}$ =0.000075) were detected in 1.06% (61/6,685) of individuals. Notably, the NM_014251.3: c.852_855del variant was enriched among East Asians ($AF_{QDnewborn} = 0.0021$ and $AF_{gnomAD\_EAS} = 0.0040$) but very rarely detected in other populations.

Similarly, we have found that combined methylmalonic aciduria and homocystinuria, cblC Type (MAHCC, OMIM277400) also exemplifies these principles. Methylmalonic aciduria and homocystinuria (MAHC) comprises a group of autosomal recessive disorders caused by pathogenic variants in multiple genes, resulting in distinct subtypes. Among these, the cblC type represents the most prevalent congenital disorder affecting vitamin B12 (cobalamin, Cbl) metabolism. This form is caused by mutations in the *MMACHC* gene, which is located on chromosome 1p34.1[25]. Estimates indicate that the incidence in Western populations ranges from approximately 1:46,000 to 1:200,000, whereas East Asian populations exhibit a higher prevalence, with rates between roughly 1:3,220 and 1:21,488[26]. In our cohort, 11 heterozygous P/LP variants in the *MMACHC* gene (Supplementary Table S7), (NM_015506.3:c.609 G > A (p.Trp203*)) ($AF_{QDnewborn} = 0.0032$), (NM_015506.3: c.658_660del (p.Lys220del)) ($AF_{QDnewborn} = 0.0018$), (NM_015506.3: c.80 A > G (p.Gln27Arg)) ($AF_{QDnewborn} = 0.0007$), (NM_015506.3: c.565 C > T (p.Arg189Cys)) ($AF_{QDnewborn} = 0.0005$), (NM_015506.3: c.617 G > A (p.Arg206Gln)) ($AF_{QDnewborn} = 0.0003$), (NM_015506.3: c.445_446del (p.Cys149Hisfs*32)) ($AF_{QDnewborn} = 0.0003$), (NM_015506.3: c.481 C > T (p.Arg161Ter)) ($AF_{QDnewborn} = 0.0002$), (NM_015506.3: c.541_548del (p.Asp181Thrfs*5)) ($AF_{QDnewborn} = 0.0001$), (NM_015506.3: c.228_231del (p.Asp77Glnfs*22)) ($AF_{QDnewborn} = 0.0001$), (NM_015506.3: c.389 A > G (p.Tyr130Cys)) ($AF_{QDnewborn} = 0.0001$) and (NM_015506.3: c.452 A > G (p.His151Arg)) ($AF_{QDnewborn} = 0.0001$) were detected in 1.45% (97/6,685) of individuals. The most common mutation was NM_015506.3: c.609 G > A ($AF_{QDnewborn} = 0.0032$, and $AF_{gnomAD\_EAS} = 0.0003$), which was reported as the most frequent mutation in the Chinese population with cblC disease, followed by NM_015506.3:c.658_660del and NM_015506.3:c.80 A > G[26].

Discussion

This study represents the first evaluation of frequencies of secondary findings in a cohort of healthy Chinese newborns using the ACMG SF 3.3 framework. We analyzed genome sequencing data from 6685 newborns and interpreted variants according to ACMG/AMP guidelines, identifying 306 unique actionable pathogenic variants across 84 ACMG-recommended genes. Overall, 9.12% (610/6685) of newborns carried at least one pathogenic variant when heterozygous carriers of autosomal recessive variants were included. After excluding carriers of AR variants, the frequency of clinically reportable findings was 5.06% (338/6685).

Our results align with previous studies of secondary findings in Chinese populations. Huang et al. (2022)[27] reported a reportable SF variant frequency of 5.3% using the ACMG SF v3.0 framework in a clinically referred cohort, providing an early baseline for the Chinese population. Using the updated ACMG SF v3.3 gene list in an unselected newborn population, we observed a comparable reportable frequency of 5.06%. Beyond this consistency, our study extends prior work by applying whole-genome sequencing to a large, population-based cohort of healthy newborns and incorporating newly added actionable genes, offering updated and

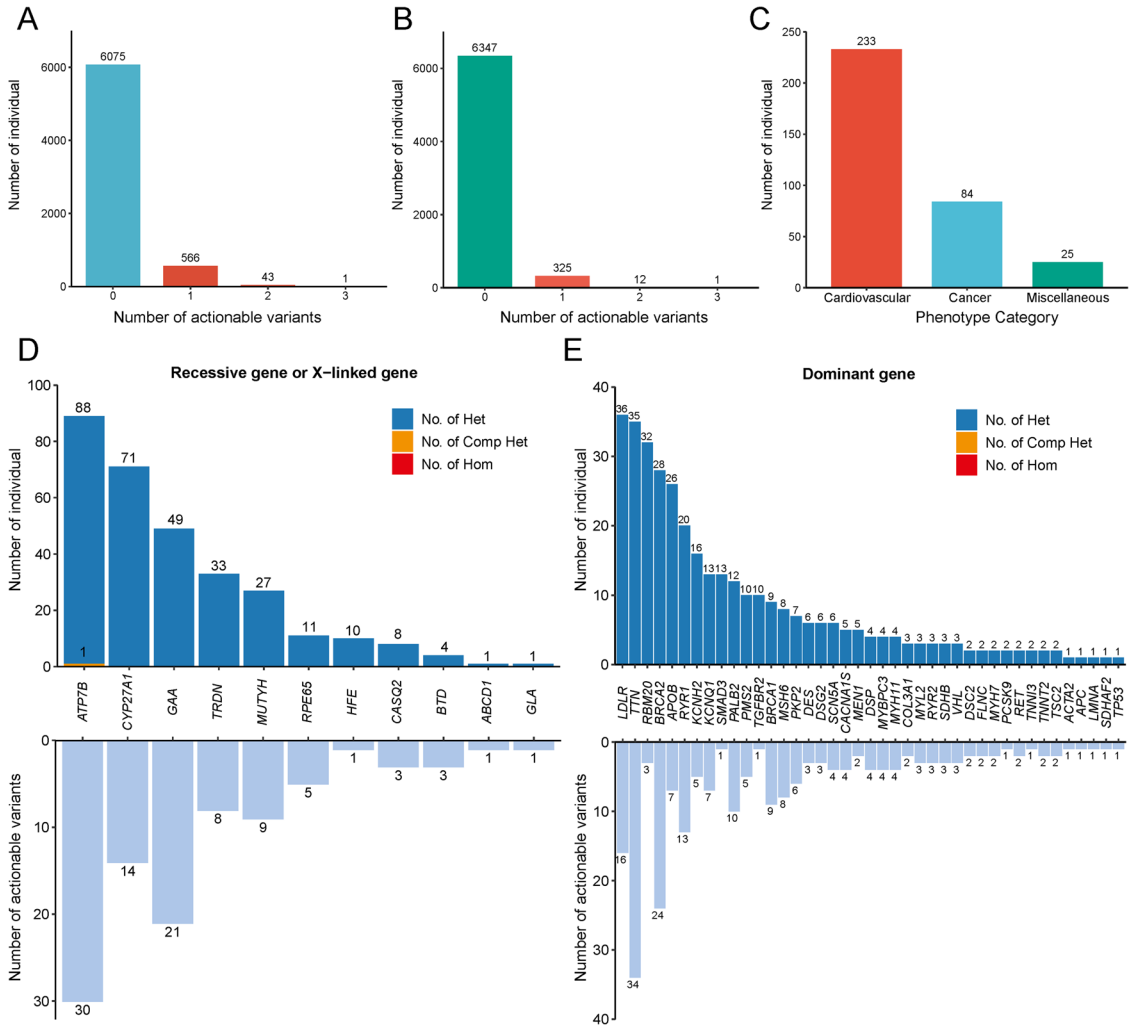

Fig. 2 | Landscape of actionable variants in the QDnewborn cohort. **A** Distribution of individuals based on the number of actionable variants in the QDnewborn cohort. **B** Distribution of individuals based on the number of actionable variants in the QDnewborn cohort, excluding heterozygous carriers of autosomal recessive variants. **C** Distribution of individuals with actionable variants in autosomal dominant genetic disorders, categorized by phenotype in the QDnewborn cohort. **D** Distribution of individuals with actionable variants in recessive or X-linked genes (top panel) and the corresponding number of actionable variants per gene (bottom panel) in the QDnewborn cohort. The top panel shows the count of individuals by genotype (Het: heterozygous; Comp Het: compound heterozygous; Hom: homozygous). **E** Number of individuals with actionable variants in dominant genes (top panel) and the associated number of actionable variants per gene (bottom panel) in the QDnewborn cohort.

complementary insights into the spectrum and population frequency of SFs at the earliest stage of life.

Previous studies on SFs from exome sequencing and genome sequencing have reported variable frequencies of P/LP variants. The rate of SFs variants we observed falls within the range of previously published studies[8,9,28]. The highest reported frequency 21% came from a study of 2049 Japanese individuals[9], followed by 8.5% in a cross-population analysis of 1092 asymptomatic individuals[28], 8% in 1254 Saudi population[29], and 7% in 196 Korean population[8]. Lower rates were documented in European (2%)[7], African (1.2%)[7], Chinese (5.3%)[27], Taiwanese (1.86%)[30], and broader East Asian populations (data not specified)[12]. The differences in SFs variants rates between different populations may arise from multiple factors, including differences in sequencing methodologies, variant classification criteria, reference databases, candidate gene panels, different filtering strategies and study populations. Among these, discrepancies in reference databases, variant classification standards and population background are the main causes of differences in SFs frequencies. Therefore, to enable more accurate and comprehensive reporting of SFs variants, we propose that collaborative efforts among research groups across countries are imperative to establish unified variant classification strategies.

Notably, previous studies on SFs had primarily focused on patient populations spanning a wide age range[8,27,31]—a likely reason for substantial variability in the previously reported SFs rates, given their focus on individuals with specific phenotypes. In our study, our cohort consisted solely of unselected newborns, and phenotypes associated with P/LP variants may not manifest at this age, supporting their classification as SFs—i.e., unrelated to primary clinical presentations in neonates. Based on our observations, we found that actionable pathogenic secondary findings are relatively common in the Chinese newborn population. Relevant studies indicated that individuals carrying pathogenic variants in ACMG actionable genes faced a significantly elevated overall risk of developing related diseases or clinical features[32,33]. We emphasize that SFs hold substantial practical value in the early monitoring and intervention of late-onset diseases.

ACMG secondary findings are typically associated with conditions that manifest later in life but may benefit from early risk assessment and preventive strategies. In contrast to congenital disorders presenting at birth, these conditions often remain asymptomatic during childhood but can progress to severe complications in adulthood if unrecognized. The identification of such variants in newborns may enable earlier risk stratification

https://doi.org/10.1038/s41525-026-00565-0

Table. 1 | Actionable variants showing significant ethnic differences in allele frequency between Chinese and European populations

Gene	Variant name	rsID	AF (QDnewborn)[a]	AF (gnomAD Overall)[b]	AF (gnomAD EUR)[c]	AF (gnomAD AFR)[d]	AF (gnomAD AMR)[e]	AF (gnomAD EAS)[f]	AF (gnomAD SAS)[g]	ACMG Classification[h]	Mode of Inheritance[i]	Disease / Phenotype
APOB	NM_000384.3: c.10579 C > T (p.Arg3527Trp)	rs144467873	0.001346	0.000058	0.000028	0.000013	0	0.000513	0.000384	P	AD	Familial hypercholesterolemia
ATP7B	NM_000053.4: c.2975 C > T (p.Pro992Leu)	rs201038679	0.001272	0.000010	0.000002	0	0	0.000312	0	P	AR	Wilson disease
ATP7B	NM_000053.4: c.2755 C > G (p.Arg919Gly)	rs121907993	0.000823	0.000014	0	0	0	0.000468	0	P	AR	Wilson disease
ATP7B	NM_000053.4: c.2621 C > T (p.Ala874Val)	rs121907994	0.000748	0.000054	0.000041	0.000013	0.000033	0.000735	0.000022	P	AR	Wilson disease
ATP7B	NM_000053.4: c.3859 G > A (p.Gly1287Ser)	rs762866453	0.000374	0.000060	0.000051	0.000254	0.000033	0.000245	0	P	AR	Wilson disease
ATP7B	NM_000053.4: c.3646 G > A (p.Val1216Met)	rs776280797	0.000374	0.000032	0.000004	0	0.000033	0.000089	0.000022	P	AR	Wilson disease
ATP7B	NM_000053.4: c.122 A > G (p.Asn41Ser)	rs201738967	0.000224	0.000614	0.000788	0.000067	0.000017	0.000045	0.000011	P	AR	Wilson disease
ATP7B	NM_000053.4: c.2605 G > A (p.Gly869Arg)	rs191312027	0.000224	0.001290	0.001595	0.000360	0.000616	0.000178	0.000022	LP	AR	Wilson disease
ATP7B	NM_000053.4: c.3955 C > T (p.Arg1319*)	rs193922109	0.000075	0.000140	0.000170	0.000040	0.000117	0.000022	0.000033	P	AR	Wilson disease
CASQ2	NM_001232.4: c.381 C > T (p.Gly127 =)	rs775663612	0.000449	0.000009	0.000008	0	0.000050	0.000022	0.000011	LP	AR	Catecholaminergic polymorphic ventricular tachycardia
CYP27A1	NM_000784.4: c.410 G > A (p.Arg137Gln)	rs587778818	0.001272	0.000027	0.000015	0.000107	0.000067	0.000156	0.000066	LP	AR	Cerebrotendinous xanthomatosis
CYP27A1	NM_000784.4: c.1263+1 G > A	rs397515355	0.001047	0.000064	0.000076	0.000080	0.000017	0.000045	0.000022	P	AR	Cerebrotendinous xanthomatosis
CYP27A1	NM_000784.4: c.435 G > T (p.Gly145 =)	rs587778796	0.000673	0.000009	0	0	0	0.000334	0	LP	AR	Cerebrotendinous xanthomatosis
CYP27A1	NM_000784.4: c.379 C > T (p.Arg127Trp)	rs201114717	0.000449	0.000051	0.000047	0.000173	0.000067	0.000089	0.000011	LP	AR	Cerebrotendinous xanthomatosis
CYP27A1	NM_000784.4: c.1420 C > T (p.Arg474Trp)	rs121908098	0.000449	0.000012	0.000008	0.000013	0.000017	0.000156	0.000011	P	AR	Cerebrotendinous xanthomatosis
CYP27A1	NM_000784.4: c.1214 G > A (p.Arg405Gln)	rs121908099	0.000374	0.000025	0.000008	0.000013	0.000033	0.000535	0.000011	P	AR	Cerebrotendinous xanthomatosis

https://doi.org/10.1038/s41525-026-00565-0

Table 1 (continued) | Actionable variants showing significant ethnic differences in allele frequency between Chinese and European populations

Gene	Variant name	rsID	AF (QDnewborn)[a]	AF (gnomAD Overall)[b]	AF (gnomAD EUR)[c]	AF (gnomAD AFR)[d]	AF (gnomAD AMR)[e]	AF (gnomAD EAS)[f]	AF (gnomAD SAS)[g]	ACMG Classification[h]	Mode of Inheritance[i]	Disease / Phenotype
GAA	NM_000152.4: c.2662 G > T (p.Glu888*)	rs765718882	0.000898	0.000006	0.000003	0	0	0.000134	0	P	AR	Pompe disease
GAA	NM_000152.5: c.2431del (p.Leu811Trpfs*37)	rs2143925402	0.000449	0	0	0	0	0	0	P	AR	Pompe disease
GAA	NM_000152.5: c.2214 G > A (p.Trp738*)	rs1057516328	0.000374	0.000002	0	0.000013	0	0.000045	0	P	AR	Pompe disease
GAA	NM_000152.5: c.-32-13T > G	rs386834236	0.000075	0.005255	0.006355	0.000957	0.003936	0.000113	0.002123	P	AR	Pompe disease
HFE	NM_000410.4: c.845 G > A (p.Cys282Tyr)	rs1800562	0.000748	0.056980	0.071020	0.010530	0.015060	0.000201	0.002152	LP	AR	Hereditary hemochromatosis
KCNQ1	NM_000218.3: c.176del (p.Pro59Glnfs*27)	rs1565023136	0.000301	0	0	0	0	0	0	LP	AD	Long-QT syndrome type 1
LDLR	NM_000527.5: c.1747 C > T (p.His583Tyr)	rs730882109	0.001122	0.000028	0	0	0	0.000490	0.000176	P	AD	Familial hypercholesterolemia
MUTYH	NM_001048174.2: c.13 C > T (p.Arg5*)	rs587780088	0.000598	0.000011	0.000002	0	0.000017	0.000201	0.000055	P	AR	MUTYH-associated polyposis
MUTYH	NM_001048174.2: c.715 C > T (p.Gln239*)	rs786203115	0.000524	0.000004	0	0	0	0.000134	0	P	AR	MUTYH-associated polyposis
PMS2	NM_000535.7: c.164-1 G > C	rs763308607	0.000374	0.000001	0	0	0	0.000023	0	P	AD	Lynch syndrome
RBM20	NM_001134363.3: c.3267del (p.Ile1090Serfs*14)	rs760334885	0.001722	0.000015	0.000019	0	0	0	0	LP	AD	Dliated cardiomyopathy
TRDN	NM_006073.4: c.1186del (p.Lys396Asnfs*27)	rs748084011	0.001720	0.000009	0	0	0	0.000333	0	LP	AR	Long QT syndrome

a. AF (QDnewborn): The allele frequency observed in our study cohort of 6,685 newborns.
b. AF (gnomAD Overall): The allele frequency of reported in the gnomAD v4.1 total population.
c. AF (gnomAD EUR): The allele frequency reported in the gnomAD v4.1 European (non-Finnish) population.
d. AF (gnomAD AFR): The allele frequency reported in the gnomAD v4.1 African/African American population.
e. AF (gnomAD AMR): The allele frequency reported in the gnomAD v4.1 Admixed American population.
f. AF (gnomAD EAS): The allele frequency reported in the gnomAD v4.1 East Asian population.
g. AF (gnomAD SAS): The allele frequency reported in the gnomAD v4.1 South Asian population.
h. ACMG Classification: *P* pathogenic, *LP* likely pathogenic.
i. Mode of Inheritance: *AD* autosomal dominant, *AR* autosomal recessive.

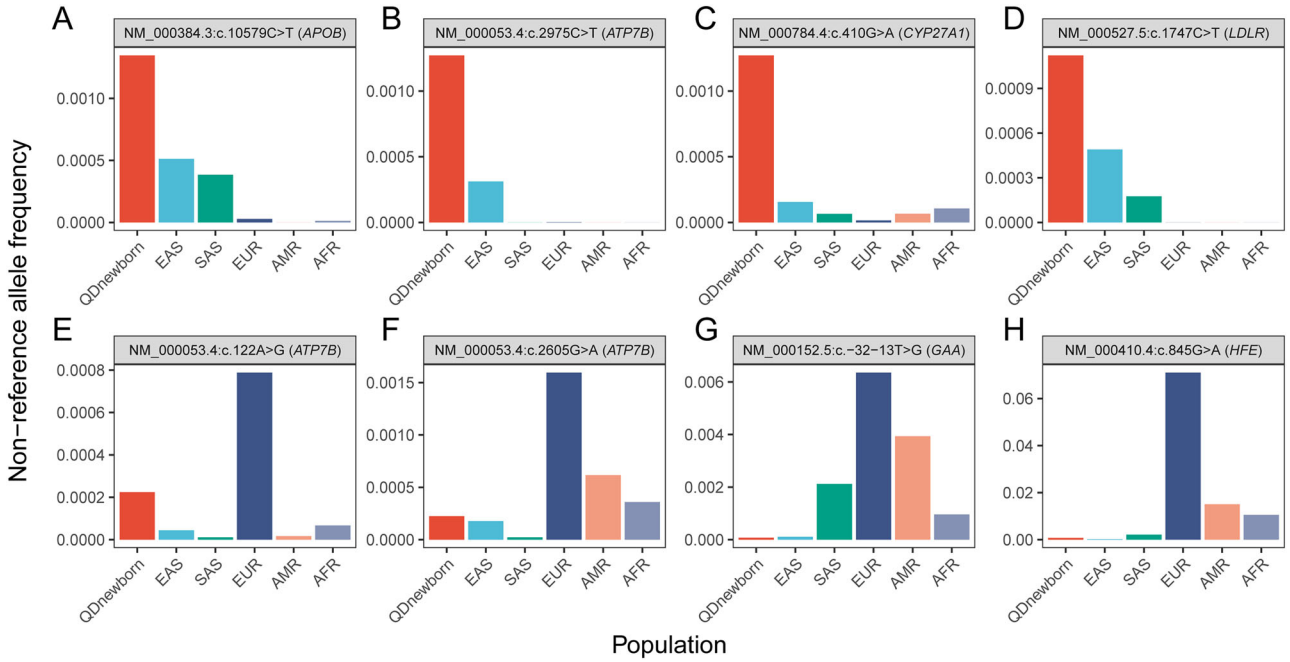

Fig. 3 | Comparison of the non-reference allele frequency of eight population-specific actionable variants across global populations. **A** NM_000384.3:c.10579C>T in *APOB*. **B** NM_000053.4:c.2975C>T in *ATP7B*. **C** NM_000784.4:c.410G>A in *CYP27A1*. **D** NM_000527.5:c.1747C>T in *LDLR*. **E** NM_000053.4:c.122A>G in *ATP7B*. **F** NM_000053.4:c.2605G>A in *ATP7B*. **G** NM_000152.5:c.-32-13T>G in *GAA*. **H** NM_000410.4:c.845G>A in *HFE*. Bars represent non-reference allele frequencies in QDnewborn cohort (Chinese newborns, $n = 6685$) and the Genome Aggregation Database (gnomAD) reference populations (EUR, European; EAS, East Asian; AMR, Admixed American; SAS, South Asian; AFR, African/African American).

and longitudinal monitoring, potentially shifting clinical management toward preventive care when appropriate.

In our cohort, *LDLR* was the most frequently detected gene associated with familial hypercholesterolemia. Previous studies have shown that FH patients with pathogenic variants frequently exhibit elevated LDL-C levels and face an increased risk of cardiovascular disease[34]. For newborns identified with *LDLR* SFs, long-term follow-up and family-based cardiovascular risk assessment may be warranted. Consistent with current pediatric FH management guidelines[35,36], early recognition enables timely lipid monitoring and age-appropriate interventions, including childhood statin therapy when indicated[31,37,38].

The identification of ACMG SFs in newborns may also facilitate cascade genetic testing among first-degree relatives. A follow-up survey from the BabySeq project[39] revealed that a substantial proportion of families (76.5%) pursued additional clinical evaluations or genetic testing after receiving unanticipated monogenic findings. For instance, three mothers of infants with *BRCA2* or *MSH2* variants elected to undergo cancer risk-reducing surgeries after learning of their genetic risks through cascade testing. These findings suggest that identifying ACMG SFs in newborns may contribute to earlier detection of at-risk relatives and enable family-wide preventive care.

Nevertheless, the return of adult-onset secondary findings in newborns remains ethically and clinically debated, particularly regarding autonomy and long-term psychosocial impact[40,41]. Careful consideration of reporting criteria, informed consent processes, and population-specific risk–benefit balance is essential.

Overall, early identification allows for timely intervention, long-term disease prevention, and improved family planning decisions, ultimately supporting better health outcomes across the lifespan.

Our analysis centered on the ACMG-recommended 84-gene panel, which, while broadly applicable, may exclude genes with higher prevalence or penetrance in Asian populations. For example, *SLC25A13*—the causative gene for citrin deficiency, a disorder with high penetrance in East Asia[24]—carried nine heterozygous pathogenic variants in 1.06% (61/6685) of our cohort, a frequency rarely reported in non-Asian populations. *MMAHCC*—the causative gene for combined methylmalonic aciduria and homocystinuria, cblC Type, which was reported to have a higher prevalence in East Asian populations[26]. In our cohort, 11 heterozygous P/LP variants in the *MMACHC* gene were detected in 1.45% (97/6685) of individuals. Therefore, we suggested that high-penetrance genes in East Asian populations, like *SLC25A13 and MMAHCC*, should be included in the recommendation list.

Interpreting SFs inherently involves ethical dilemmas, particularly regarding neonatal and pediatric testing. Some argue against predictive testing in young children due to potential psychological burdens and ethical complexities, but the ACMG recommends that age should not restrict the search for or reporting of actionable SFs[2]. The goal of such reporting is to identify clinically actionable pathogenic mutations, enabling healthcare providers and individuals to proactively implement preventive or therapeutic strategies. From this perspective, respecting participants' autonomy and facilitating their access to beneficial genomic information represents a responsible approach to balancing scientific utility with ethical integrity.

While this study represents the first report on the frequency and spectrum of SFs in Chinese normal newborn population, this study has several limitations. First, focusing solely on newborns may introduce ethical considerations regarding predictive testing in this population. Second, the absence of family member data/samples hindered our ability to characterize de novo mutations and assess their pathogenicity. Third, as participants were predominantly from Qingdao, China, results may not generalize to the entire Chinese newborn population; thus, further randomized sampling is required to determine the true SFs rate in this demographic.

In conclusion, based on the latest ACMG SF v3.3 gene list, this study focused on investigating the spectrum and frequency of SFs in the Chinese newborn population. Our findings reveal that clinically actionable SFs are not rare in the Chinese newborn population, and further follow-up and monitoring may facilitate the early diagnosis and precise treatment of specific human diseases.

Methods

Subject and consent

We recruited 9992 newborns from Qingdao, Shandong province, China. All participating newborns were enrolled in the WGS-based newborn screening program, with genome sequencing provided as a free service. Of these, 7140 families signed additional informed consent forms granting permission for the use of their children's genomic data in research. All research protocols were approved by the Ethics Committee of the Qingdao Women and Children's Hospital (Approval No. QFELL-KY-2020–29) and BGI Research (Approval No. NO.BGI-IRB20065). Written informed consent was obtained from the parents or legal guardians of all neonates participating. Participants were free to withdraw from the study at any time. All procedures involving human participants were conducted in accordance with the ethical standards of the institutional research committees and with the Declaration of Helsinki and its later amendments.

Genome sequencing

Heel prick spots were collected from all subjects three days after birth, and genomic DNA was extracted using the MagPure Tissue & Blood DNA LQ Kit (Magen, Guangzhou, China). DNA quality was assessed using Qubit 3.0 fluorometer and agarose gel electrophoresis. Genome sequencing was performed using libraries constructed according to MGI guidelines (Shenzhen, China) and sequenced on the MGI DNBSEQ platform[42], to achieve 30X coverage depth with 100 bp paired-end reads.

Alignment and variant calling

Variant detection and joint genotype calling analyses were performed using the ZTRON and ZBOLT[43] pipeline (MGI, Shenzhen, China), following the GATK's best practices recommendations[44]. The Burrows-Wheeler Aligner (BWA)[45] was used to align sequencing reads from each sample to the GRCh38 reference genome, followed by the post-processing steps, including duplicate marking, InDel realignment, and base quality score recalibration (BQSR). The alignment sequences were stored in BAM format and underwent variant detection using GATK[46] HaplotypeCaller. Joint variant calling was performed using ZBOLT Joint-Calling (v1.0) after generating of single-sample GVCF file. Variant Quality Score Recalibration (VQSR) was applied using GATK (v4.1.7.0) with specific training sets and annotations for SNPs and INDELs. Finally, variants that passed the VQSR filtration were retained for downstream analysis. Multi-allelic sites were normalized into bi-allelic variants using BCFtools (v1.9)[47,48].

Sample and variant quality control (QC)

Rigorous sample and variant QC steps were applied to ensure data integrity. For sample QC, the following criteria were applied: base quality (Q20) > 90% and Q30 > 80%; GC content between 39% and 44%; mapping rate > 98%; N content rate < 1%; mean sequence depth > 15X; 20X coverage > 90%; total autosomal bi-allelic SNVs ranging from 2.4 M to 3.75 M; total singletons < 100 K; ratio of heterozygous to homozygous variants < 3.3; samples with a call rate > 90% were included. Sample contamination was assessed using VerifyBamID2 (v 2.0.1)[49], excluding samples with a freemix > 0.05. Sex determination employed SAMtools[47] (v1.3) to calculate the X-to-autosome (X/A) ratio of sequencing depths, classifying samples with X/A ratio <0.75 as male and >0.75 as female. Samples showing discrepancies between inferred genetic sex and documented sex at birth were excluded.

Variant QC involved multiple filtering steps. Variants with ExcessHet >54.69 or an inbreeding coefficient < -0.3 were excluded. Variants that failed to achieve high-quality genotypes in all samples or showed Hardy-Weinberg Equilibrium (HWE) p value < 1E−06 (determined by VCFtools (v 0.1.13))[50] were excluded. Individual genotypes with GQ < 20 or DP < 10 were designated as missing data, and variants with >15% missing genotypes across individuals were excluded. KING (v 2.2.7)[51] was used to infer the kinship and identify unrelated individuals.

Variant annotation

The function effects of variants were annotate using ANNOVAR[52] (version 20200607) with various databases, including RefGene, ESP, ExAC, ClinVar and so on. The ACMG SF v3.3 list includes 84 genes, comprising 42 genes related to 19 cardiovascular phenotypes, 28 genes related to 19 cancer phenotypes, 9 genes related to 7 miscellaneous phenotypes and 5 genes related to 5 inborn errors of metabolism phenotypes. Variant annotation was limited to the 84 genes recommended by the ACMG SF v3.3 list. Variants were classified into functional categories, such as nonsynonymous, synonymous, intronic, splicing, 5'-untranslated region (UTR), and 3'-UTR. Variants located in the coding region, splice sites, intronic and UTRs were selected for further pathogenicity evaluation. The pathogenicity of variants was predicted using SIFT[53], PolyPhen[54,55] and MutationTaster[56].

To identify the candidate P/LP variants, we retrieved the NCBI ClinVar database[57] (release 20250615, the raw VCF files was downloaded from https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/archive_2.0/2025/clinvar_20250615.vcf.gz). In the ClinVar database, a variant is classified as high-confidence P/LP variant if it meets the following criterion: 1) the variant was annotated with a review status of two stars or above (≥2 stars); and 2) the variant was annotated as "Pathogenic" or "Likely Pathogenic" or "Pathogenic/Likely Pathogenic" with no conflicting interpretations. Following the criteria above, we retrieved a total of 82,811 high-confidence P/LP variant with a review status of 2 stars or higher from the ClinVar database. Additionally, for the *HFE* gene, only *HFE* (NM_000410.3: c.845 G > A (p.Cys282Tyr)) was retained. Allele frequency for each variant was obtained from our cohort and the Genome Aggregation Database (gnomAD, v4.1.0)[58].

Variant classification

To reduce the variants list to a manageable size, we prioritized variants reported as P/LP (≥ 2 stars) in the ClinVar database. Variants were identified by matching genomic coordinates, reference allele, and alternative allele to P/LP variants reported in ClinVar. These variants were categorized as KP (known pathogenic) variants. The remaining variants, which were previously unreported but predicted to result in loss-of-function (LOF) (including nonsense, frameshift, and splice site disrupting variants), were classified as EP (expected pathogenic) variants. We used the Variant Effect Predictor (v103)[59] and the LOFTEE plugin[60] (https://github.com/konradjk/loftee) to identify LOF variants, including stop gain, splice-site, or frameshift variants, which were predicted with "high confidence" to significantly disrupt gene function.

Given the large size and complex structural characteristics of the *TTN* gene, we followed the management strategy for TTN-truncating variants (TTNtv) proposed by Roberts et al[18]. for further gene-specific filtering. TTNtv identified by the LOF prediction pipeline were annotated according to sarcomeric band location (A-band, I-band, M-band, or Z-disk) using reference coordinates from the *TTN* transcript annotations database (http://cardiodb.org/titin)[18]. Variants were retained as expected pathogenic variants if they met one of the following criteria: 1) the variant is located in exons shared by N2BA and N2B isoforms, with a percentage spliced in (PSI) ≥ 0.9; or 2) for splice site variants, they must be within the A-band region.

In accordance with the standards and guidelines jointly recommended by the ACMG/AMP, all candidate variants were manually reviewed, and supporting evidence for each variant was provide to assess its pathogenicity. After careful variant classification, variants classified as P/LP were selected as actionable pathogenic variants.

Data availability

The data supporting the finding of this study, including the allele frequency for all 399,037 genetic variants across the 84 ACMG-recommended genes, have been deposited in the Genome Variation Map (https://ngdc.cncb.ac.cn/gvm/) at the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, under the accession number that can be publicly accessible at https://ngdc.cncb.ac.cn/gvm/getProjectDetail?project=GVM001119. The release

of the data was approved by the Ministry of Science and Technology of China (Project ID: PRJCA043552). To prevent the disclosure of individuals' genetic identity, the raw sequencing data and information of the research participants are not publicly available. Further analysis of sequencing data will be made available for collaborating researchers upon request, dependent of the HGRAC's approval.

Received: 19 November 2025; Accepted: 18 March 2026;
Published online: 02 April 2026

References

1. Saeidian, A. H. et al. Secondary ACMG and non-ACMG genetic findings in a multiethnic cohort of 16,713 pediatric participants. *Genet. Med.* **26**, 101225 (2024).
2. Green, R. C. et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15**, 565–574 (2013).
3. Miller, D. T. et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **23**, 1391–1398 (2021).
4. Kalia, S. S. 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*. **19**, 249–255 (2017).
5. Miller, D. T. et al. ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **23**, 1381–1390 (2021).
6. Miller, D. T. et al. ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **25**, 100866 (2023).
7. Amendola, L. M. et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res.* **25**, 305–315 (2015).
8. Jang, M. A., Lee, S. H., Kim, N. & Ki, C. S. Frequency and spectrum of actionable pathogenic secondary findings in 196 Korean exomes. *Genet. Med.* **17**, 1007–1011 (2015).
9. Yamaguchi-Kabata, Y. et al. Evaluation of reported pathogenic variants and their frequencies in a Japanese population based on a whole-genome reference panel of 2049 individuals. *J. Hum. Genet.* **63**, 213–230 (2018).
10. Chetruengchai, W. & Shotelersuk, V. Actionable secondary findings in the 73 ACMG-recommended genes in 1559 Thai exomes. *J. Hum. Genet.* **67**, 137–142 (2022).
11. Hsu, J. S. et al. Complete genomic profiles of 1496 Taiwanese reveal curated medical insights. *J. Adv. Res.* **66**, 197–207 (2024).
12. Tang, C. S. et al. Actionable secondary findings from whole-genome sequencing of 954 East Asians. *Hum. Genet.* **137**, 31–37 (2018).
13. Lee, K. et al. ACMG SF v3.3 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med*. **27**, https://doi.org/10.1016/j.gim.2025.101454 (2025).
14. Ovchinnikova, E. V., Garbuz, M. M., Ovchinnikova, A. A. & Kumeiko, V. V. Epidemiology of Wilson's disease and pathogenic variants of the ATP7B gene leading to diversified protein disfunctions. *Int. J. Mol. Sci.* **25**, https://doi.org/10.3390/ijms25042402 (2024).
15. Appadurai, V. et al. Apparent underdiagnosis of Cerebrotendinous Xanthomatosis revealed by analysis of ~60,000 human exomes. *Mol. Genet. Metab.* **116**, 298–304 (2015).
16. Santos, R. D. Advancing prediction of pathogenicity of familial hypercholesterolemia LDL receptor commonest variants with machine learning models. *JACC Basic Transl. Sci.* **6**, 828–830 (2021).
17. Jammal Addin, M. B., Young, D., McCarrison, S. & Hunter, L. Dilated cardiomyopathy in a national paediatric population. *Eur. J. Pediatr.* **178**, 1229–1235 (2019).
18. Roberts, A. M. et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci. Transl. Med.* **7**, 270ra276 (2015).
19. Jayson, G. C., Kohn, E. C., Kitchener, H. C. & Ledermann, J. A. Ovarian cancer. *Lancet* **384**, 1376–1388 (2014).
20. Loibl, S., Poortmans, P., Morrow, M., Denkert, C. & Curigliano, G. Breast cancer. *Lancet* **397**, 1750–1769 (2021).
21. Pietrangelo, A. Hereditary hemochromatosis: pathogenesis, diagnosis, and treatment. *Gastroenterology* **139**, 393–408.e392 (2010).
22. Burke, W. et al. Hereditary hemochromatosisgene discovery and its implications for population-based screening. *JAMA* **280**, 172–178 (1998).
23. Allen, K. J. et al. HFE Cys282Tyr homozygotes with serum ferritin concentrations below 1000 microg/L are at low risk of hemochromatosis. *Hepatology* **52**, 925–933 (2010).
24. Cong, P. K. et al. Identification of clinically actionable secondary genetic variants from whole-genome sequencing in a large-scale Chinese population. *Clin. Transl. Med.* **12**, https://doi.org/10.1002/ctm2.866 (2022).
25. Lerner-Ellis, J. P. et al. Identification of the gene responsible for methylmalonic aciduria and homocystinuria, cblC type. *Nat. Genet.* **38**, 93–100 (2006).
26. Wang, C. et al. Mutation spectrum of MMACHC in Chinese pediatric patients with cobalamin C disease: a case series and literature review. *Eur. J. Med. Genet.* **62**, 103713 (2019).
27. Huang, Y. et al. Landscape of secondary findings in Chinese population: a practice of ACMG SF v3.0 list. *J. Pers. Med.* **12**, https://doi.org/10.3390/jpm12091503 (2022).
28. Cassa, C. A., Tong, M. Y. & Jordan, D. M. Large numbers of genetic variants considered to be pathogenic are common in asymptomatic individuals. *Hum. Mutat.* **34**, 1216–1220 (2013).
29. Aloraini, T. et al. The rate of secondary genomic findings in the Saudi population. *Am. J. Med. Genet. A* **188**, 83–88 (2022).
30. Kuo, C. W. et al. Frequency and spectrum of actionable pathogenic secondary findings in Taiwanese exomes. *Mol. Genet. Genom. Med* **8**, e1455 (2020).
31. Xiao, H. et al. Secondary genomic findings in the 2020 China Neonatal Genomes Project participants. *World J. Pediatr.* **18**, 687–694 (2022).
32. Natarajan, P. et al. Aggregate penetrance of genomic variants for actionable disorders in European and African Americans. *Sci. Transl. Med.* **8**, 364ra151 (2016).
33. Dewey, F. E. et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**, https://doi.org/10.1126/science.aaf6814 (2016).
34. Khera, A. V. et al. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J. Am. Coll. Cardiol.* **67**, 2578–2589 (2016).
35. Nordestgaard, B. G. et al. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *Eur. Heart J.* **34**, 3478–3490 (2013).
36. Myśliwiec, M. et al. 2024 Polish recommendations for the management of familial hypercholesterolemia in children and adolescents. *Arch. Med. Sci.* **20**, 1741–1753 (2024).
37. Bogsrud, M. P. et al. Treatment goal attainment in children with familial hypercholesterolemia: a cohort study of 302 children in Norway. *J. Clin. Lipido.* **12**, 375–382 (2018).
38. Luirink, I. K. et al. 20-year follow-up of statins in children with familial hypercholesterolemia. *N. Engl. J. Med.* **381**, 1547–1556 (2019).

39. Green, R. C. et al. Actionability of unanticipated monogenic disease risks in newborn genomic screening: findings from the BabySeq Project. *Am. J. Hum. Genet.* **110**, 1034–1045 (2023).
40. Wade, C. H., Wilfond, B. S. & McBride, C. M. Effects of genetic risk information on children's psychosocial wellbeing: a systematic review of the literature. *Genet. Med.* **12**, 317–326 (2010).
41. Botkin, J. R. et al. Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. *Am. J. Hum. Genet.* **97**, 6–21 (2015).
42. Huang, C. et al. An integrated Asian human SNV and indel benchmark established using multiple sequencing methods. *Sci. Rep.* **10**, 9821 (2020).
43. Li, Z. et al. An efficient large-scale whole-genome sequencing analyses practice with an average daily analysis of 100Tbp: ZBOLT. *Clin. Transl. Discov.* **3**, e252 (2023).
44. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.1–11.10.33 (2013).
45. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
46. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
47. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, https://doi.org/10.1093/gigascience/giab008 (2021).
48. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
49. Zhang, F. et al. Ancestry-agnostic estimation of DNA sample contamination from sequence reads. *Genome Res.* **30**, 185–194 (2020).
50. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
51. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
52. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
53. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
54. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
55. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7–20 (2013).
56. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
57. Landrum, M. J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
58. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
59. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
60. Gudmundsson, S. et al. Addendum: the mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **597**, E3–E4 (2021).

Acknowledgements

We are especially grateful to the participation of the volunteers and their families. We sincerely thank Yuhua Ye for his valuable assistance in the review and interpretation of genetic variants. This work was supported by the National Natural Science Foundation of China (2022YFC2703102), and the China-Serbia Science and Technology Cooperation Committee Exchange Program, Sixth Session (Project 6-3 to M. F.).

Author contributions

M.F. and X.J. conceived and designed the study. Y.H. performed data analysis, quality control, variant annotation and data interpretation. Y.G., Z.D., X.J., Y.S., and C.L. were responsible for participants recruitment and the collection of samples and associated information. H.H., J.L., S.P., and X.J. contributed to data generation. M.F. and Y.H. drafted the manuscript. M.F., Y.H., and X.J. revised the manuscript. All authors have read and approved the final manuscript.

Competing interests

Authors Y.H., Y.G., Z.D., H.H., J.L., X.J., and M.F. were employed by Beijing Genomics Institution (BGI) in Shenzhen. The other authors do not have a competing interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-026-00565-0.

Correspondence and requests for materials should be addressed to Mingyan Fang.