



OPEN

## Identification of potentially common loci between childhood obesity and coronary artery disease using pleiotropic approaches

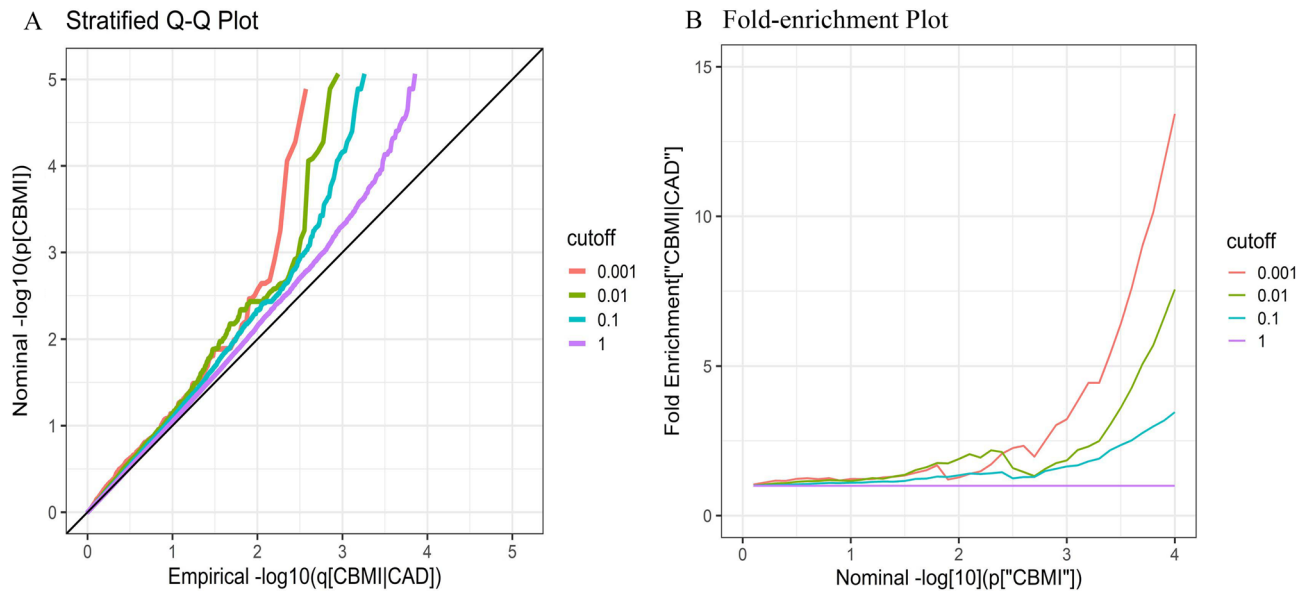
Lianke Wang<sup>1,2</sup>, Qiang Zhang<sup>2</sup>, Fei Xu<sup>1</sup>, Anna Brickell<sup>3</sup>, Qianyu Zhou<sup>1</sup>, Bin Yang<sup>1</sup> & Changqing Sun<sup>1,2</sup>✉

Childhood obesity remains one of the most important issues in global health, which is implicated in many chronic diseases. Converging evidence suggests that a higher body mass index during childhood (CBMI) is significantly associated with increased coronary artery disease (CAD) susceptibility in adulthood, which may partly arise from the shared genetic determination. Despite genome-wide association studies (GWASs) have successfully identified some loci associated with CBMI and CAD individually, the genetic overlap and common biological mechanism between them remains largely unexplored. Here, relying on the results from the two large-scale GWASs ( $n = 35,668$  for CBMI and  $n = 547,261$  for CAD), linkage disequilibrium score regression (LDSC) was used to estimate the genetic correlation of CBMI and CAD in the first step. Then, we applied different pleiotropy-informed methods including conditional false discovery rate (cFDR) and genetic analysis incorporating pleiotropy and annotation (GPA) to detect potentially common loci for childhood obesity and CAD. By integrating the genetic information from the existing GWASs summary statistics, we found a significant positive genetic correlation ( $r_g = 0.127$ ,  $p = 2E-4$ ) and strong pleiotropic enrichment between CBMI and CAD ( $LRT = 79.352$ ,  $p = 5.2E-19$ ). Importantly, 28 loci were simultaneously discovered to be associated with CBMI, and 13 of them were identified as potentially pleiotropic loci by cFDR and GPA. Those corresponding pleiotropic genes were enriched in trait-associated gene ontology (GO) terms “amino sugar catabolic process”, “regulation of fat cell differentiation” and “synaptic transmission”. Overall, the findings of the pleiotropic loci will help to further elucidate the common molecular mechanisms underlying the association of childhood obesity and CAD, and provide a theoretical direction for early disease prevention and potential therapeutic targets.

Childhood obesity has emerged as an important public health problem and threatens future health and longevity, including an increased risk of many chronic disorders and premature death<sup>1-3</sup>. In recent decades, the worldwide prevalence of childhood obesity is on the increase remarkably<sup>2</sup>, and more than 22% of children are overweight or obese in developed countries<sup>3</sup>. The data from the World Health Organization (WHO) shows that there are over 340 million overweight and 124 million obese children and adolescents in 2016<sup>4</sup>. BMI is an established and most commonly used index to quantitatively measure obesity and health. Numerous studies have demonstrated that a higher body mass index during childhood (CBMI) is associated with an increased risk of coronary artery disease (CAD) in later adult life<sup>5-7</sup>. Several lines of evidence suggested that this association may partially arise from the common genetic foundation between the two phenotypes<sup>8,9</sup>. However, the biological mechanisms underlying this relationship are still unclear.

Genome-wide association studies (GWASs) have been widely used to identify single nucleotide polymorphisms (SNPs) contributing to the variation of complex traits or diseases, including CBMI and CAD. Heritability studies indicated a strong genetic contribution to both CBMI ( $h^2$ : 67–93%)<sup>10</sup> and CAD ( $h^2$ : 40–60%)<sup>11</sup>. To date, GWASs have identified 161 loci for CAD and 25 loci for CBMI, accounting for 15.1% and 3.6% of the phenotypic variance of CAD and CBMI, respectively<sup>1,8,12</sup>. This indicated the heritability explained by the variants

<sup>1</sup>Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou 450001, Henan, China. <sup>2</sup>College of Nursing and Health, Zhengzhou University, Zhengzhou 450001, Henan, China. <sup>3</sup>College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA. ✉email: zzugwsy@163.com



**Figure 1.** Stratified Q–Q plot (A) and Fold-enrichment plot (B). Stratified Q–Q plot: Nominal versus empirical  $-\log_{10}(p)$  values in CBMI as a function of significance of association with CAD at different levels. Fold-enrichment plot: Enrichment versus nominal  $-\log_{10}(p)$  values for CBMI as a function of significance of the association with CAD.

previously reported was limited (especially for CBMI) and most of the heritability was probably not missing but undetected. Additionally, hundreds of SNPs associated with adult BMI or obesity-related traits had been identified, by contrast, less was well known about the genetic background of CBMI<sup>13</sup>. Meanwhile, evidence from previous researches suggested that common genetic variant existed in CBMI and CAD<sup>9</sup>, indicating shared genetic determinants or risk factors between the two phenotypes.

GWAS can uncover more genetic loci to further explain the heritability of complex phenotypes by increasing the sample size<sup>14,15</sup>, but it is commonly not feasible since recruiting additional participants and SNP genotyping is too costly. Moreover, GWASs are usually performed on a single trait or disease, rather than analyzing them concurrently. Therefore, pleiotropic methods without extensive additional subject recruitment requirements are effective to discover novel genetic variants associated with multiple diseases or traits. Pleiotropy is the phenomenon of a single locus influencing more than two distinct diseases or traits<sup>16</sup>. Now, with the active pleiotropic method development in this field, it is possible to comprehensively elucidate the genetic overlap and biological underpinnings between different traits or diseases<sup>17</sup>.

Considering the close relationship, high heritability, and potential pleiotropy between CBMI and CAD, we assume the two phenotypes are ideal for further analyses to improve the detection of associated loci and explore the common etiology. Additionally, the pleiotropy between adult BMI and CAD has been studied by our team before<sup>18</sup>. Owing to the genetic correlation of CBMI and adult BMI and the significance of early disease prevention, the shared genetic signals between CBMI and CAD are worthy of further study with larger sample size. In the present study, with summary statistics from the two large-scale existing GWAS datasets, we first utilized the linkage disequilibrium score regression (LDSC) method to estimate the overall genetic correlation between CBMI and CAD. Next, we applied the conditional false discovery rate (cFDR) and the genetic analysis incorporating pleiotropy and annotation (GPA) approaches to explore whether CBMI shares susceptibility loci with CAD<sup>19,20</sup> and discover some novel etiologic relationship between them. The current study will be a foundation of further researches about the two phenotypes and helpful for screening the shared genetic factor. Additionally, the potentially shared genetic influences and biological mechanisms can provide novel helpful strategies for preventing and treating CAD early.

## Results

**Genetic correlation.** The LDSC results showed that the genome-wide SNP-based heritability was estimated to be 24.21% ( $se=2.21\%$ ) for CBMI and 5.83% ( $se=0.33\%$ ) for CAD. Additionally, the LDSC intercept was 0.9971 ( $se=0.007$ ) for CBMI and 0.9951 ( $se=0.011$ ) for CAD, demonstrating that the results of our analysis were reliable and not affected by the confounding factors from population stratification or relatedness. Next, using cross-trait LDSC with 1,053,840 common SNPs from filtered summary statistics, we found that there existed a significant positive genetic correlation between CBMI and CAD ( $r_g = 0.127$ ,  $p = 2E-4$ ), indicating the genetic overlap between the two phenotypes. Therefore, further analyses were implemented to explore the specific genetic mechanisms shared by CBMI and CAD.

**Pleiotropic enrichment.** A stratified Q–Q plot for CBMI conditioned on different strengths of the association of CAD was shown in Fig. 1A. Strong pleiotropic enrichment between CBMI and CAD was revealed in the

	$\pi_{00}$	$\pi_{10}$	$\pi_{01}$	$\pi_{11}$	LRT	<i>p</i> -value
CBMI-CAD	0.857 (0.006)	0.048 (0.006)	0.090 (0.003)	0.005 (0.002)	79.352	5.2E-19

**Table 1.** Pleiotropy assessed between CBMI and CAD by GPA method. The values in the brackets are standard errors of the estimates. The last two columns provide the LRT statistics and *p*-values of hypothesis testing. The  $\pi_{00}$  represents the proportion of the SNPs not associated with CBMI or CAD;  $\pi_{10}$  and  $\pi_{01}$  represent the proportion of the SNP associated with CBMI and CAD, respectively;  $\pi_{11}$  represents the proportion of the SNP associated with CBMI and CAD.

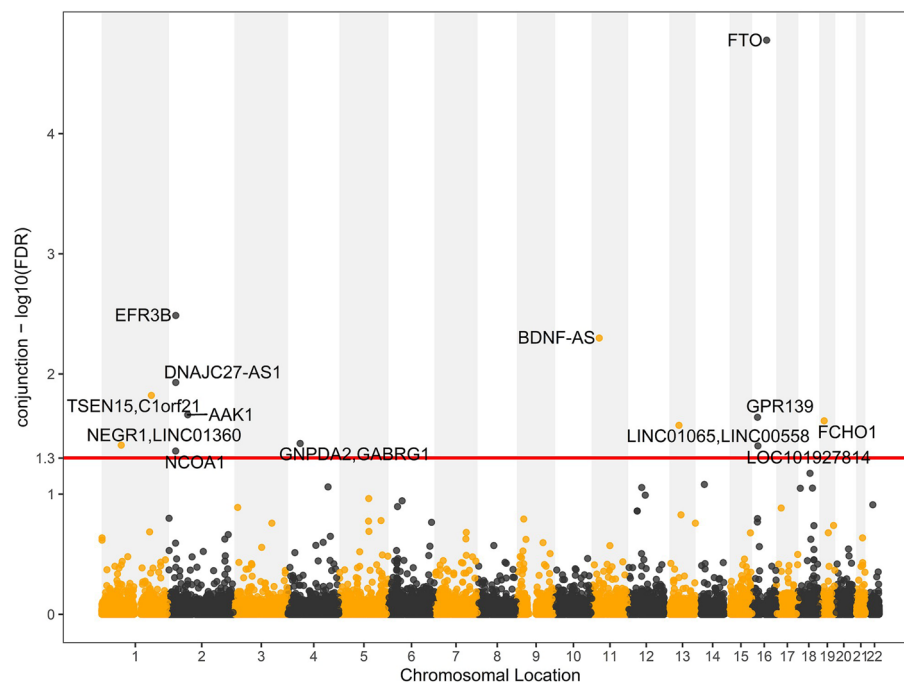
Chr	SNP	Allel	Role	Gene	<i>P</i> .valueA	<i>P</i> .valueB	cFDR.AcB	cFDR.BcA	ccFDR	fdr11.GPA
chr16	rs9940128	G/A	Intronic	<i>FTO</i>	1.67E-05	9.25E-14	1.67E-05	2.45E-11	1.67E-05	0.001527
chr2	rs1866146	G/A	UTR3	<i>EFR3B</i>	0.000594	2.04E-06	0.003264	0.000802	0.003264	0.031703
chr11	rs7127507	T/C	ncRNA_intronic	<i>BDNF-AS</i>	0.000754	1.29E-05	0.005024	0.003705	0.005024	0.053765
chr2	rs11125884	A/G	ncRNA_intronic	<i>DNAJC27-AS1</i>	0.01177	2.73E-13	0.01177	4.88E-10	0.01177	0.109489
chr1	rs7536226	C/T	Intergenic	<i>TSEN15, C1orf21</i>	0.000247	5.31E-05	0.004194	0.015092	0.015092	0.064655
chr2	rs7420531	G/A	UTR3	<i>AAK1</i>	0.000533	8.77E-05	0.007644	0.021845	0.021845	0.088304
chr16	rs16969473	A/G	Intronic	<i>GPR139</i>	0.002157	8.22E-05	0.018119	0.022945	0.022945	0.126129
chr19	rs13382133	C/T	Intronic	<i>FCHO1</i>	7.67E-12	0.000558	1.1E-09	0.024548	0.024548	0.12562
chr13	rs12429545	G/A	Intergenic	<i>LINC01065, LINC00558</i>	0.02676	3.66E-11	0.02676	7.68E-08	0.02676	0.175974
chr4	rs1996023	T/G	Intergenic	<i>GNPDA2, GABRG1</i>	0.007574	2.73E-07	0.03787	0.00037	0.03787	0.094976
chr1	rs7531118	T/C	Intergenic	<i>NEG1, LINC01360</i>	0.01305	1.29E-05	0.03915	0.007047	0.03915	0.182283
chr16	rs194546	G/A	ncRNA_intronic	<i>LOC101927814</i>	0.006605	8.62E-06	0.03963	0.007183	0.03963	0.123251
chr2	rs17736503	G/T	Intronic	<i>NCOA1</i>	0.007106	6.91E-05	0.04382	0.030049	0.04382	0.188942

**Table 2.** The pleiotropic loci identified by cFDR and GPA method (ccFDR < 0.05 and fdr11.GPA < 0.2). The SNPs are listed with their chromosomal, allele, role, annotated genes, original *p*-value for each phenotype, cFDR value for each phenotype, ccFDR and fdr.11.GPA values. A means CAD and B means CBMI. Chr: chromosom, ccFDR: conjunctive conditional false discovery rate, fdr.11.GPA: false discovery rate of GPA when the SNP was associated with both phenotypes, GPA: genetic analysis incorporating pleiotropy and annotation.

conditional Q-Q plot, as reflected by a large degree of leftward deflection from the null line under the successively decreasing thresholds. Earlier departures from the null line with higher levels of association with CAD demonstrate a greater proportion of true associations for a given nominal *p*-value. The fold-enrichment plot also displayed the genetic enrichment for CBMI conditioned on varying significance levels of CAD by observing the extent of upward shift from the reference line (Fig. 1B). Meanwhile, according to the LRT result (Table 1), we found that there was significantly pleiotropic enrichment between CBMI and CAD (the LRT *p*-value was 5.2E-19), which entirely supported the evidence provided from the above two different plots.

**CBMI genetic loci identified with two methods.** We identified 31 and 29 significant SNPs associated with CBMI according to the thresholds of cFDR < 0.05 and fdr.GPA < 0.2, respectively. As illustrated in the conditional Manhattan plot for CBMI (Fig. S1), a total of 28 significant SNPs was identified to be associated with CBMI conditioned on CAD, which was annotated to 41 genes and located on 14 different chromosomes (Table S1). Among the 28 SNPs, 9 SNPs had *p*-values smaller than the genome-wide significance level (5E-8) and were considered to be associated with CBMI<sup>21</sup>, and 2 SNPs rs7127507 and rs7531118 were in the same LD block with the previously identified CBMI-associated SNPs rs17309874 ( $r^2 = 0.6306$ ) and SNP rs3101336 ( $r^2 = 0.6805$ ), respectively. These genetic variants were regarded as a replication of the previous CBMI-associated GWASs. The rest of 17 independent SNPs with *p*-values larger than the genome-wide significance threshold were considered novel SNPs to be associated with CBMI. For the 41 genes annotated by these SNPs, 29 genes were newly discovered in comparison with the original CBMI GWAS and previous CBMI-related studies. Other information on these 28 SNPs was presented in Table S1. Among the 41 genes detected with CBMI, most of them were enriched in CBMI-related GO terms such as “positive regulation of cellular process”, “limb morphogenesis” and “positive regulation of cell proliferation”. More detailed information on enrichment analysis was shown in Table 3. As for PPI results (Fig. S2), 28 genes including 21 novel ones were annotated in this protein network. The proteins encoded by these genes indirectly interacted with many other proteins encoded by CBMI-associated genes, proteins including *FTO*, *SEC16B*, *GNPDA2*, and *TFAP2B* had very close contact, which was proven to have a confirmed correlation with CBMI in original GWAS.

**Pleiotropic loci identified with two methods.** Using the two distinct methods with the thresholds of ccFDR < 0.05 and fdr11.GPA < 0.2, we identified 13 and 14 potentially pleiotropic SNPs, respectively. Taken together, 13 significant SNPs were detected to be associated with CBMI and CAD, these SNPs were annotated to 17 genes and located on 7 different chromosomes (Table 2 and Fig. 2). Most of the pleiotropic SNPs were



**Figure 2.** Conjunction Manhattan plot of  $-\log_{10}(\text{ccFDR})$  values for CBMI and CAD. The red line marks the  $-\log_{10}(\text{ccFDR})$  value of 1.3 corresponds to ccFDR of 0.05. The figure shows the genomic locations of shared variants and the overlapped genes between CBMI and CAD identified by cFDR and GPA method.

Go term	Pathway description	P-value	Combined score	Genes
CBMI GO:0048522	Positive regulation of cellular process	6.17E-04	41.68	<i>PKHD1, TFAP2B, AKAP6, CACUL1, TBX3, DYNAP</i>
GO:0035108	Limb morphogenesis	8.40E-04	329.02	<i>TFAP2B, TBX3</i>
GO:0008284	Positive regulation of cell proliferation	0.001672	36.78	<i>PKHD1, TFAP2B, CACUL1, TBX3, DYNAP</i>
GO:0072001	Renal system development	0.006516	83.23	<i>PKHD1, TFAP2B</i>
GO:1903779	Regulation of cardiac conduction	0.008333	69.71	<i>FPGT-TNNI3K, TNNI3K</i>
GO:0030136	Clathrin-coated vesicle	0.001126	99.35	<i>FCHO1, AAK1, RAB27B</i>
GO:0035612	AP-2 adaptor complex binding	8.55E-05	1305.44	<i>FCHO1, AAK1</i>
GO:0001104	RNA polymerase II transcription cofactor activity	6.79E-04	127.09	<i>NCOA1, TFAP2B, MED13L</i>
GO:0008528	G-protein coupled peptide receptor activity	0.007628	74.33	<i>PRLHR, GPR139</i>
CBMI and CAD GO:0046348	Amino sugar catabolic process	0.00509	1035.40	<i>GNPDA2</i>
GO:0090335	Regulation of brown fat cell differentiation	0.006781	734.36	<i>FTO</i>
GO:1901072	Glucosamine-containing compound catabolic process	0.008469	561.33	<i>GNPDA2</i>
GO:0006054	N-acetylneuraminate metabolic process	0.008469	561.33	<i>GNPDA2</i>
GO:0051932	Synaptic transmission, GABAergic	0.009313	500.15	<i>GABRG1</i>
GO:0030135	Coated vesicle	0.001633	212.68	<i>FCHO1, AAK1</i>
GO:0035612	AP-2 adaptor complex binding	1.42E-05	3750.98	<i>FCHO1, AAK1</i>
GO:0017162	Aryl hydrocarbon receptor binding	0.006781	734.36	<i>NCOA1</i>
GO:0008503	Benzodiazepine receptor activity	0.008469	561.33	<i>GABRG1</i>

**Table 3.** Gene functional annotation and enrichment result.

situated in the intronic (54%) and intergenic (31%) regions while two were in the untranslated regions (15%). All of the 13 pleiotropic SNPs were not previously identified to be associated with both CBMI and CAD, only 3 SNPs (rs9940128, rs11125884, and rs12429545) and 1 SNP rs13382133 passed genome-wide significance level in the original CBMI and CAD GWASs<sup>12,21</sup>, respectively. There were 8 SNPs (rs1866146, rs1996023, rs17736503, rs9940128, rs11125884, rs7127507, rs12429545 and rs7531118) previously reported to be associated with adult BMI-related traits<sup>22,23</sup>, and one SNPs rs9940128 was previously reported to be related with diabetes and high-density lipoprotein-cholesterol (HDL-C)<sup>24,25</sup>. All of the 17 genes these SNPs annotated to were novel because none of them were previously confirmed to be associated with both CBMI and CAD. More information on these SNPs and genes was presented in Table 2. A large number of the genes were enriched in CBMI and CAD-related terms such as “amino sugar catabolic process”, “regulation of brown fat cell differentiation” and “AP-2 adaptor complex binding”. More information on enrichment analysis was shown in Table 3. All the pleiotropic genes were imported into the STRING database, and only 9 genes were annotated in the PPI network. Those proteins encoded by the 9 genes were enriched into three disparate clusters (Fig. S3). For example, four proteins like *FTO*, *NEGR1*, *GABRG1*, and *GPR139* had the interplay in the *GNPDA2* cluster and the protein *FCHO1* directly interacted with the other protein *AAK1*.

## Discussion

In this study, by integrating two GWASs summary statistics of CBMI and CAD into cFDR and GPA framework, we identified 28 SNPs associated with CBMI, and 13 pleiotropic SNPs between CBMI and CAD. Among those annotated genes, 17 novel genes were found to be pleiotropic. In the present study, given the background and the focus of the study, we highlighted the SNPs identified with CBMI and the common SNPs associated with the two phenotypes.

The findings of some genetic signals in present study were consistent with the evidence from previous studies and the pleiotropic enrichment indicated a genetic relationship between CBMI and CAD. Moreover, most loci were not identified in previous genetic researches, indicating the effectivity of the two pleiotropic approaches on the detection of overlapped genetic variants. Meanwhile, the statistical power was largely improved, it was reported that there were about 15–20 times of power increased to detect more variants through a comparison of conditional vs. unconditional FDR approach<sup>19</sup>, and the statistical power of the GPA method was comparable to that of cFDR method<sup>20</sup>. More importantly, through contrastive analysis, we quantitatively investigated the pleiotropic enrichment evaluation between CBMI and CAD based on the LRT, and robust results were obtained. The findings of polygenic pleiotropy between CBMI and CAD enabled us to further illustrate the overlapped biological mechanisms underlying the different traits or diseases.

The shared genetic determination between CBMI and CAD emphasized the important role of CBMI in CAD development and supported the hypothesis that a common genetic basis existed between them. The two pleiotropy-informed methods jointly identified 13 shared SNPs annotated to 17 genes, these SNPs and corresponding genes were novel for the reason that no previous researches confirmed them to be related to both CBMI and CAD. These new findings could provide us a better understanding of the overlapped etiology between CBMI and CAD. Here, we will discuss some pleiotropic genetic signals for their potential biological function to further elucidate the phenotypic relevance.

The SNP rs9940128 (16q12.2) is the most statistically significant pleiotropic variant identified in our study, which is located in the intron region of the *FTO* gene, and was documented to be associated with CBMI in original GWAS<sup>21</sup>. Meanwhile, this variant also influencing adult BMI has been repeatedly confirmed in the earlier GWASs<sup>23,26</sup>. The *FTO* is a well-known gene involved in weight gain and obesity. Research based on the *FTO* mouse model further demonstrated that perturbing the *FTO* enzymatic activity could dysregulate genes related to energy metabolism, resulting in the imbalance of adipose tissue homeostasis in mice<sup>27</sup>. It provided clear evidence that *FTO* played important role in regulating food intake or energy metabolism. Furthermore, evidence from recent studies has emphasized the importance of *FTO* gene variation, revealing that they were associated with neuropsychiatric diseases<sup>28</sup> besides metabolic disorders and human adiposity. Additionally, the *FTO* gene also played a potential role in cardiovascular diseases (CVD). A meta-analysis comprising 10 studies reported a significant association between the *FTO* gene and CVD risk, and this association was independent of BMI and other conventional CVD risk factors<sup>29</sup>. Subsequently, one study including a total of 970 samples that came from Pakistan demonstrated the variant in *FTO* gene was associated with CAD through affecting plasma glucose metabolism<sup>30</sup>. In addition, common variants in this gene were also associated with diabetes-related risk metabolic traits, such as raised fasting glucose, insulin and triglycerides, and lower HDL-C<sup>31</sup>. As described above, *FTO* gene was associated with CBMI and CAD may through these biological processes and molecular pathways, but further studies involving the exact functional characterization of this genetic signal are required to elucidate the precise mechanism behind their relationship.

The pleiotropic SNP rs1996023 (4p12) is located in the intergenic region between the *GNPDA2* gene and the *GABRG1* gene. The protein encoded by the *GNPDA2* gene belongs to an allosteric enzyme that catalyzes the deamination of glucosamine 6-phosphate and participates in the hexosamine signaling pathway<sup>32,33</sup>. SNPs in or close to the *GNPDA2* gene had been documented to be associated with adult BMI or childhood obesity<sup>21</sup>. One animal study also found that the *GNPDA2* gene played an important role in regulating body fat, weight, and energy metabolism<sup>33</sup>. Additionally, another study investigating the possible mechanism of the *GNPDA2* in adipogenesis discovered that overexpression of *GNPDA2* enhanced the accumulation of lipid droplets and knockout of the gene resulted in the reduction of accumulation of lipid droplets, reporting that *GNPDA2* may be a key gene in lipid and glucose metabolism<sup>32</sup>. Furthermore, through assessing the impact of the obesity-related loci on the known obesity complications found that the *GNPDA2* gene was potentially associated with diabetes<sup>34</sup>. It is widely known that the abnormal metabolism of lipid and glucose is a causal risk factor for cardiovascular



diseases. A previous study also reported that the *GNPDA2* gene was highly expressed in the hypothalamus and brain, and played a vital role in the central nervous system processes of weight regulation<sup>34</sup>. The *GABRG1* gene is an integral membrane protein that encodes a protein belonging to the ligand-gated ionic channel family and participates in the inhibition of neural transmission<sup>35</sup>. *GABRG1* gene was reported to be implicated in neurological development conditions, including autism spectrum disorder<sup>36</sup>. The pathogenesis may involve the early development of the brain, affecting synaptic plasticity, neural transmission, and functional connectivity. There was evidence that children with neurodevelopmental disorders were at higher risk of developing depression<sup>37</sup>. Meanwhile, growing evidence shows depression is a risk factor for the development of CAD<sup>38</sup>. The *GNPDA2* and *GABRG1* genes increase the risk of susceptibility for CBMI and CAD may be through the same biological processes, but further experimentations are needed to determine the exact mechanism of the action of these genes.

The implementation of cFDR and GPA approaches on two large GWAS datasets let us successfully detect a large number of novel phenotype-associated genetic signals. Especially, an alternative pleiotropic method based on an independent algorithm was used simultaneously to ensure the reliability of the results. However, there are still a few limitations. First, the information about the precise effect size of the loci on the phenotype could not be provided, so it is impossible to determine which phenotype the locus has the greater influence on, but it can be inferred from the results of the original GWAS study. Second, the dataset of CAD is a meta-analysis of overwhelming European ancestry and a small fraction of other ancestries, we could not ensure uniformity of the ancestry due to the lack of data at the individual level, this may have an impact on the results. Nevertheless, the large sample size can contribute to the successful discovery of potential novel genetic variants. Third, independent replication is needed in this research. However, our CAD dataset is the largest to date which includes the dataset from the CARDIoGRAMplusC4D consortium and UKB. Moreover, we used the GWAS dataset of CAD from CARDIoGRAMplusC4D consortium for replication<sup>39</sup>, we observed a similar pleiotropic enrichment pattern and replicated 19 SNPs associate with CBMI and 3 pleiotropic SNPs associated with CBMI and CAD. These results illustrate that there is strong pleiotropic enrichment between CBMI and CAD and some common SNPs can be replicated across studies although with small sample size. Fourth, both pleiotropic methods cannot identify the causal variants for the phenotypes of interest, but the question can be partially addressed in the future through the integration with other types of data<sup>40</sup>. Fifth, most of our findings are statistically significant and require further biological functional experiments and clinical replication studies to support our findings.

In summary, our results highlight the feasibility and improved power of the cFDR and GPA methods in detecting novel genetic pleiotropic loci between CBMI and CAD. The novel findings may facilitate the discovery of the shared genetic mechanisms underlying CBMI and CAD, and provide us novel insights into early disease prevention and treatment strategies.

## Materials and methods

**GWAS datasets.** Our analyses were conducted on summary statistics from two independent GWASs for CBMI and CAD<sup>12,21</sup>. The CBMI GWAS dataset was derived from a GWAS meta-analysis of European descent involving 35,668 individuals performed by the EGG Consortium (<http://www.egg-consortium.org/childhood-bmi.html>). The CAD GWAS dataset was derived from a meta-analysis of 547,261 individuals (122,733 cases and 424,528 controls) of mainly European descent, which combined a GWAS dataset from the CARDIoGRAMplusC4D consortium with another GWAS dataset from the UK Biobank (UKB). The aggregate dataset of CAD was available through the CardiOmics website (<https://www.cardiomics.net/download-data>). The two datasets consist of summary statistics for each SNP, and their alleles, *p*-value, effect size, and direction of effect. More information such as study design, inclusion criteria of subjects, and quality control during the analysis is provided in the original study. There are no reduplicating subjects between the two datasets, avoiding the result of increased Type I error and biased effect estimates<sup>41</sup>.

We used summary-level statistics from the two GWA studies. All participants in the original study received informed consent, and the ethical approval had been obtained from their respective institutional review boards. The details can be found in the two GWASs<sup>12,21</sup>. This study was approved by the Ethical Committee of the Life Sciences of Zhengzhou University.

**Data processing.** The two GWAS datasets were used here containing 2,499,691 SNPs for CBMI and 7,947,837 SNPs for CAD. Before conducting the following analysis, we extracted the SNPs and their corresponding variables (eg, alleles, effect sizes, *p*-values, sample sizes) from the two datasets for each phenotype. The HapMap3 CEU genotypes were used as the reference panel to label allele frequencies and calculate LD values between pairs of SNPs. For the LDSC analysis, the summary statistics were dealt with default SNP quality control (QC) filters (INFO > 0.9 and MAF > 0.01). After QC, a total of 1,054,231 and 1,181,375 SNPs were reserved for CBMI and CAD, respectively. Using cross-disease LDSC, 1,053,840 common SNPs were remained after merging. Before cFDR and GPA analysis, we first combined the common SNPs in the two datasets, then the SNPs were processed using the linkage disequilibrium (LD)-based pruning method by PLINK1.9. If the pairs of SNPs with LD values ( $r^2$ ) larger than 0.2, one of the SNPs with smaller minor allele frequency (MAF) was discarded<sup>19</sup>. Lastly, the remaining 128,749 common SNPs were prepared for the pleiotropic analysis. Genomic control (GC) corrected genomic inflation resulting from the potential population structure in the individual studies and the GWAS meta-analysis<sup>42</sup>. Therefore, there was no need for us to reuse the GC correction in this analysis.

**Statistical analyses.** The Linux operating system and R 4.0.5 were used for data processing and statistical analysis. The LDSC, cFDR, and GPA analyses were performed by “LDSC”, “GWAScFDR” and “GPA” packages, respectively. The quantile–quantile (Q-Q) plot, fold-enrichment plot, and Manhattan plot were mainly generated by the “ggplot2” package.

**Estimation of genetic correlation.** With the summary statistics from the two GWAS datasets, we first estimated the SNP heritability for each phenotype and evaluated the genetic correlation between CBMI and CAD using LDSC analysis. LDSC is a method of quantifying the contribution of polygenic heritability and confounding bias by examining the regression relationship between GWAS test statistics and LD to determine whether the inflation of test statistics is caused by polygenic effects or confounding bias<sup>43,44</sup>. The intercept of LDSC regression analysis can be used to determine whether there are confounding factors in the results. If the intercept is near 1, it means there is no confounding factor. For analysis of a single phenotype, LDSC can identify confounding factors and estimate heritability. For multiple phenotypes, the genetic correlation between phenotypes can be calculated according to the corresponding chi-square statistics<sup>44</sup>.

**Pleiotropic enrichment assessment.** To assess the pleiotropic enrichment between the two phenotypes (CBMI and CAD), a stratified Q-Q plot was conducted by successively conditioning the principal phenotype on SNPs across varying levels of significance threshold for the given phenotype. The enrichment of the pleiotropic effect can be decided by the degree of deflection from the null line which is conditioned on the second phenotype. Meanwhile, we constructed a fold-enrichment plot to further confirm the pleiotropic enrichment between CBMI and CAD. The enrichment can be reflected by the larger degree of the upward shift from the null line when the p-values become smaller. Furthermore, in addition to the visual display of the pleiotropic enrichment above, we also used the likelihood ratio test (LRT) to quantitatively estimate the statistical significance of pleiotropic enrichment<sup>20</sup>.

**Calculation of cFDR, ccFDR and GPA.** Both cFDR and GPA are well-established pleiotropic methods for incorporating GWAS data at the summary statistic level, which have been widely applied to many diseases or traits<sup>45–48</sup>. We briefly summarized both approaches in the following section. The cFDR is based on the Bayesian formula, identified as a posterior probability of an SNP being null for the principal phenotype given that the p-values for the two phenotypes (principal and conditional) are as small as or smaller than the observed ones. In the following formula,  $p_i$  and  $p_j$  are the observed p-values of a SNP for the principal and conditional phenotypes, respectively.  $H_0^{(i)}$  means a null hypothesis that the SNP was not associated with the principal phenotype.

$$\text{cFDR}(p_i|p_j) = \Pr(H_0^{(i)}|P_i \leq p_i, P_j \leq p_j) \quad (1)$$

The cFDR threshold was set to 0.05, which meant the SNP was considered to be associated with the principal phenotype if the value is less than 0.05<sup>19</sup>. To identify an SNP to be pleiotropic, conjunction cFDR (ccFDR) is defined based on cFDR and is calculated as the maximum value of the cFDRs for the two phenotypes<sup>19</sup>. Finally, the conditional and conjunction Manhattan plots were applied to present the location of the significant SNPs associated with CBMI, as well as pleiotropic SNP associated with CBMI and CAD.

The GPA method is another available statistical tool applied to detect the pleiotropic genetic effects between CBMI and CAD in the expectation of providing robust statistical evidence. This approach can use the GWAS summary results without annotation data, which has better comparability with the cFDR analysis results in our study. We took the intersection of the p-values of corresponding phenotypes as input, then a GPA model was fitted based on the provided p-value matrix data. The four-group model for two GWAS data sets was presented to implement the GPA analysis.

$$\begin{aligned} \pi_{00} &= \Pr(Z_{j00} = 1) : (P_{j1}|Z_{j00} = 1) \sim U[0, 1], (P_{j2}|Z_{j00} = 1) \sim U[0, 1], \\ \pi_{10} &= \Pr(Z_{j10} = 1) : (P_{j1}|Z_{j10} = 1) \sim \text{Beta}[\alpha_1, 1], (P_{j2}|Z_{j10} = 1) \sim U[0, 1], \\ \pi_{01} &= \Pr(Z_{j01} = 1) : (P_{j1}|Z_{j01} = 1) \sim U[0, 1], (P_{j2}|Z_{j01} = 1) \sim \text{Beta}[\alpha_2, 1], \\ \pi_{11} &= \Pr(Z_{j11} = 1) : (P_{j1}|Z_{j11} = 1) \sim \text{Beta}[\alpha_1, 1], (P_{j2}|Z_{j11} = 1) \sim \text{Beta}[\alpha_2, 1] \end{aligned} \quad (2)$$

The latent variables  $Z_j$  represent the association between the  $j$ -th SNP and the two phenotypes,  $Z_{j00} = 1$  means the SNP is associated with neither of them,  $Z_{j10} = 1$  and  $Z_{j01} = 1$  means the SNP is only associated with one of the two phenotypes, and  $Z_{j11} = 1$  means the SNP is associated with both. The  $U[0, 1]$  and  $\text{Beta}[\alpha_k, 1]$  means the p-values from the null group (Uniform distribution) and non-null group (Beta distribution, where  $0 < \alpha_k < 1$ ,  $k=1,2$ ), respectively. The Expectation–Maximization (EM) algorithm was used to estimate the statistical inference of the model parameters and determine the SNP ranking<sup>20</sup>. Meanwhile, using the LRT-based hypothesis testing for pleiotropy, we calculated the value of  $\text{fdr.GPA}$  (false discovery rate of GPA) to decide which variant to be associated with CBMI and the value of  $\text{fdr1.GPA}$  to determine the common variants between CBMI and CAD. We used the same the GPA significance threshold of 0.2 according to the criterion set in the original paper, which means this SNP is associated with the corresponding phenotype if the value is less than 0.2<sup>20</sup>.

**Definition of potentially novel SNP.** The SNP was defined as potentially novel one should not have been reported in the previous GWASs and was not in linkage disequilibrium (LD) ( $r^2 \geq 0.6$ ) with the previously GWASs confirmed SNPs. We mainly searched GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) and PhenoScanner (<http://www.phenoscanter.medschl.cam.ac.uk/>) to determine the SNP and LD value ( $r^2$ ). If the value of  $r^2$  was greater than 0.6, it means the two SNPs were in high LD and the SNP identified was considered to be reported or replicated.

**Functional analysis of the identified genes.** To explore the functional role of the identified genes, we performed a comprehensive gene set enrichment analysis using Enrichr (<http://amp.pharm.mssm.edu/Enrichr>)

hr/)<sup>49</sup>. The significant genes were annotated and clustered in three main categories (biological processes, cellular components, and molecular functions) to clarify polygenic associations and determine whether the implicated genes were involved in a biological process. To further reveal and visualize the functional partnership and interaction of identified genes, protein–protein interaction (PPI) network was established by utilizing the STRING database (<http://string-db.org/>), which could construct the corresponding protein association networks by comprising known and predicted associations<sup>50</sup>. The above analyses enabled us to get a systematical evaluation of the underlying biology and relevance between the genes enriched in clusters, especially enhanced our understanding of the biological underpinnings of potential associations between different phenotypes.

## Data availability

The data that support the findings of the current study are openly available in the Early Growth Genetics Consortium at <http://www.egg-consortium.org/childhood-bmi.html> and CardiOmics.net at <https://www.cardiomics.net/download-data>.

Received: 17 November 2021; Accepted: 8 November 2022

Published online: 14 November 2022

## References

- Franks, P. W. *et al.* Childhood obesity, other cardiovascular risk factors, and premature death. *N. Engl. J. Med.* **362**, 485–493 (2010).
- Di Cesare, M. *et al.* The epidemiological burden of obesity in childhood: A worldwide epidemic requiring urgent action. *BMC Med.* **17**, 212 (2019).
- Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 766–781 (2014).
- Obesity and Overweight. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> (Accessed 15 July 2022).
- Ajala, O., Mold, F., Boughton, C., Cooke, D. & Whyte, M. Childhood predictors of cardiovascular disease in adulthood. A systematic review and meta-analysis. *Obes. Rev.* **18**, 1061–1070 (2017).
- Baker, J. L., Olsen, L. W. & Sørensen, T. I. Childhood body-mass index and the risk of coronary heart disease in adulthood. *N. Engl. J. Med.* **357**, 2329–2337 (2007).
- Umer, A. *et al.* Childhood obesity and adult cardiovascular disease risk factors: A systematic review with meta-analysis. *BMC Public Health* **17**, 683 (2017).
- Vogelezang, S. *et al.* Novel loci for childhood body mass index and shared heritability with adult cardiometabolic traits. *PLoS Genet.* **16**, e1008718 (2020).
- Tekola-Ayele, F., Lee, A., Workalemahu, T. & Sánchez-Pozos, K. Shared genetic underpinnings of childhood obesity and adult cardiometabolic diseases. *Hum. Genomics* **13**, 17 (2019).
- Maes, H. H., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).
- Watkins, H. & Farrall, M. Genetic susceptibility to coronary artery disease: From promise to progress. *Nat. Rev. Genet.* **7**, 163–173 (2006).
- van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443 (2018).
- Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
- Wray, N. R. *et al.* Research review: Polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* **55**, 1068–1087 (2014).
- Schwabe, I. *et al.* Unraveling the genetic architecture of major depressive disorder: Merits and pitfalls of the approaches used in genome-wide association studies. *Psychol. Med.* **49**, 2646–2656 (2019).
- Stearns, F. W. One hundred years of pleiotropy: A retrospective. *Genetics* **186**, 767–773 (2010).
- Hackinger, S. & Zeggini, E. Statistical methods to detect pleiotropy in human complex traits. *Open Biol.* **7**, 170125 (2017).
- Ly, W. Q. *et al.* Novel common variants associated with body mass index and coronary artery disease detected using a pleiotropic cFDR method. *J. Mol. Cell Cardiol.* **112**, 1–7 (2017).
- Andreassen, O. A. *et al.* Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet.* **9**, e1003455 (2013).
- Chung, D., Yang, C., Li, C., Gelernter, J. & Zhao, H. GPA: A statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLoS Genet.* **10**, e1004787 (2014).
- Felix, J. F. *et al.* Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum. Mol. Genet.* **25**, 389–403 (2016).
- Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
- Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
- Scott, R. A. *et al.* An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
- Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
- Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
- Fischer, J. *et al.* Inactivation of the Fto gene protects from obesity. *Nature* **458**, 894–898 (2009).
- Annapoorna, P. K. *et al.* FTO: An emerging molecular player in neuropsychiatric diseases. *Neuroscience* **418**, 15–24 (2019).
- Liu, C., Mou, S. & Pan, C. The FTO gene rs9939609 polymorphism predicts risk of cardiovascular disease: A systematic review and meta-analysis. *PLoS ONE* **8**, e71901 (2013).
- Shahid, S. U., Rehman, A. & Hasnain, S. Role of a common variant of Fat Mass and Obesity associated (FTO) gene in obesity and coronary artery disease in subjects from Punjab, Pakistan: A case control study. *Lipids Health Dis.* **15**, 29 (2016).
- Freathy, R. M. *et al.* Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. *Diabetes* **57**, 1419–1426 (2008).
- Wu, L., Ma, F., Zhao, X., Zhang, M. X. & Wu, J. GNPDA2 gene affects adipogenesis and alters the transcriptome profile of human adipose-derived mesenchymal stem cells. *Int. J. Endocrinol.* **2019**, 9145452 (2019).
- Ouyang, H. *et al.* Identification, expression and variation of the GNPDA2 gene, and its association with body weight and fatness traits in chicken. *PeerJ* **4**, e2129 (2016).
- Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).



35. Lo-Castro, A. & Curatolo, P. Epilepsy associated with autism and attention deficit hyperactivity disorder: Is there a genetic link?. *Brain Dev.* **36**, 185–193 (2014).
36. Lionel, A. C. *et al.* Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci. Transl. Med.* **3**, 95ra75 (2011).
37. Eyre, O. *et al.* Childhood neurodevelopmental difficulties and risk of adolescent depression: The role of irritability. *J. Child Psychol. Psychiatry* **60**, 866–874 (2019).
38. Pinto, V. L., Brunini, T. M., Ferraz, M. R., Okinga, A. & Mendes-Ribeiro, A. C. Depression and cardiovascular disease: Role of nitric oxide. *Cardiovasc. Hematol. Agents Med. Chem.* **6**, 142–149 (2008).
39. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
40. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
41. LeBlanc, M. *et al.* A correction for sample overlap in genome-wide association studies in a polygenic pleiotropy-informed framework. *BMC Genomics* **19**, 494 (2018).
42. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* **60**, 155–166 (2001).
43. Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S. & Yang, J. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
44. Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A. & Day, F. R. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
45. Andreassen, O. A. *et al.* Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes. *Hypertension* **63**, 819–826 (2014).
46. Yin, X. *et al.* Integration of expression quantitative trait loci and pleiotropy identifies a novel psoriasis susceptibility gene, PTPN1. *J. Gene Med.* **19**, e2939 (2017).
47. Johnston, K. J. A. & Adams, M. J. Identification of novel common variants associated with chronic pain using conditional false discovery rate analysis with major depressive disorder and assessment of pleiotropic effects of LRFN5. *Transl. Psychiatry* **9**, 310 (2019).
48. Chen, H., Wang, T., Yang, J., Huang, S. & Zeng, P. Improved detection of potentially pleiotropic genes in coronary artery disease and chronic kidney disease using GWAS summary statistics. *Front. Genet.* **11**, 592461 (2020).
49. Kuleshov, M. V. *et al.* Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
50. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–d613 (2019).

## Acknowledgements

This work was partially supported by grants from the Key Research & Development and Promotion Project of Henan Province (192102310191), Key Scientific Research Projects of Colleges and Universities in Henan Province (19A330005), National Social Science Foundation of China (20BRK041) and Special Discipline Cultivation Fund of Zhengzhou University (32410604). The funders had no role in the study design, data collection, analysis, writing the paper, and decision to submit the manuscript for publication. Meanwhile, we acknowledged the CARDIoGRAMplusC4D Consortium, UK Biobank, and EGG consortium for their GWAS datasets.

## Author contributions

C.Q.S. conceived the study and provided advice on study design. L.K.W. performed data analysis and drafted the manuscript as the first author. Q.Z. and F.X. gave help in analyzing data and provided constructive suggestions. A.B. provided revisions for this manuscript. Q.Y.Z. and B.Y. gave suggestions during the whole process. Appreciate all study participants for their contributions.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-24009-8>.

**Correspondence** and requests for materials should be addressed to C.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022