

Original Research Article

Phenome-wide association study of genetically predicted B vitamins and homocysteine biomarkers with multiple health and disease outcomes: analysis of the UK Biobank

Lijuan Wang^{1,2,†}, Xue Li^{1,†}, Azita Montazeri³, Amanda J. MacFarlane^{4,§}, Franco Momoli³, Susan Duthie⁵, Marjanne Senekal⁶, Ines Mesa Eguiagaray², Ron Munger⁷, Derrick Bennett⁸, Harry Campbell², Michele Rubini⁹, Helene McNulty¹⁰, Julian Little^{3,‡}, Evropi Theodoratou^{2,11,*}

¹ School of Public Health and the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China; ² Centre for Global Health, Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom; ³ School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada; ⁴ Nutrition Research Division, Health Canada, Ottawa, Ontario, Canada; ⁵ School of Pharmacy and Life Sciences, Robert Gordon University, Aberdeen, United Kingdom; ⁶ Department of Human Biology, University of Cape Town, Cape Town, South Africa; ⁷ Department of Nutrition and Food Sciences and the Center for Epidemiologic Studies, Utah State University, Logan, UT, USA; ⁸ Medical Research Council Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom; ⁹ Department of Neuroscience and rehabilitation, University of Ferrara, Ferrara, Italy; ¹⁰ Nutrition Innovation Centre for Food and Health, Ulster University, Coleraine, Northern Ireland, United Kingdom; ¹¹ Cancer Research UK Edinburgh Centre, The University of Edinburgh MRC Institute of Genetics and Cancer, Edinburgh, United Kingdom

A B S T R A C T

Background: Although a number of health outcomes such as CVDs, metabolic-related outcomes, neurological disorders, pregnancy outcomes, and cancers have been identified in relation to B vitamins, evidence is of uneven quality and volume, and there is uncertainty about putative causal relationships.

Objectives: To explore the effects of B vitamins and homocysteine on a wide range of health outcomes based on a large biorepository linking biological samples and electronic medical records.

Methods: First, we performed a phenome-wide association study (PheWAS) to investigate the associations of genetically predicted plasma concentrations (genetic component of the circulating concentrations) of folate, vitamin B6, vitamin B12, and their metabolite homocysteine with a wide range of disease outcomes (including both prevalent and incident events) among 385,917 individuals in the UK Biobank. Second, 2-sample Mendelian randomization (MR) analysis was used to replicate any observed associations and detect causality. We considered MR $P < 0.05$ as significant for replication. Third, dose-response, mediation, and bioinformatics analyses were carried out to examine any nonlinear trends and to disentangle the underlying mediating biological mechanisms for the identified associations.

Results: In total, 1117 phenotypes were tested in each PheWAS analysis. After multiple corrections, 32 phenotypic associations of B vitamins and homocysteine were identified. Two-sample MR analysis supported that 3 of them were causal, including associations of higher plasma vitamin B6 with lower risk of calculus of kidney (OR: 0.64; 95% CI: 0.42, 0.97; $P = 0.033$), higher homocysteine concentration with higher risk of hypercholesterolemia (OR: 1.28, 95% CI: 1.04, 1.56; $P = 0.018$), and chronic kidney disease (OR: 1.32, 95% CI: 1.06, 1.63; $P = 0.012$). Significant nonlinear dose-response relationships were observed for the associations of folate with anemia, vitamin B12 with vitamin B-complex deficiencies, anemia and cholelithiasis, and homocysteine with cerebrovascular disease.

Conclusions: This study provides strong evidence for the associations of B vitamins and homocysteine with endocrine/metabolic and genitourinary disorders.

Keywords: B vitamins, homocysteine, phenome-wide association study, Mendelian randomization

* Corresponding author.

E-mail address: e.theodoratou@ed.ac.uk (E. Theodoratou).

† LW and XL contributed equally to this work.

‡ JL and ET share last authorship.

§ AJM is a member of the Journal's Editorial Board and played no role in the Journal's evaluation of this study.

Abbreviations

ACME	average causal mediation effect
ADE	average direct effect
CMA	causal mediation analysis
EHR	electronic health record
eQTL	expression quantitative trait loci
FDR	false discovery rate
GWAS	genome-wide association study
IVW	inverse variance weighted
KEGG	Kyoto Encyclopedia of Genes and Genomes
MR	Mendelian randomization
PheWAS	phenome-wide association study
PPI	protein–protein interaction
RCT	randomized clinical trial
SNP	single nucleotide polymorphism
TC	total cholesterol
wGRS	weighted genetic risk score

Introduction

The B vitamins, including folate (B9), vitamin B6, and vitamin B12, have been investigated to have effects on many health and disease outcomes. In a recent umbrella review of meta-analyses of observational studies and of randomized clinical trials (RCTs), folate (dietary intake, dietary supplementary, and plasma concentration) has been linked to >100 unique health outcomes [1]. In addition, vitamins B6 and B12 (dietary intake, dietary supplementary, and plasma concentration) have been identified to be associated with risk of CVD, metabolic-related outcomes, neurological disorders, pregnancy outcomes, and several types of cancers [2–4]. One-carbon metabolism is a process in which folate transfers 1-carbon units to support a wide range of biological processes including DNA synthesis and methylation [5]. Vitamins B6 and B12 interact with folate as methyl donors within this network, and deficiency of either folate, B6, or B12 can lead to increased circulating concentration of the related metabolite, homocysteine (Figure 1) [5], which has been suggested as an independent risk factor for CVD [6–8].

Based on a study of elderly twins, genetic polymorphisms account for a substantial component of the heritability of B vitamin concentrations, with estimates of 56%, 59%, and 66% for folate, vitamin B12, and homocysteine, respectively [9]. Thus, identification of genetic variants that affect circulating concentrations of B vitamins can give insights into the interplay of diet, genetics, and human health. Genome-wide association studies (GWASs) of serum folate, vitamin B6, vitamin B12, and homocysteine have identified and replicated several single nucleotide polymorphisms (SNPs) associated with these biomarkers [10–12]. The relevant genetic variants can be used as instruments to predict biomarker concentrations, and examination of their associations with disease outcomes can help strengthen the causal inference [13].

With the recently increased availability of dense electronic health records (EHRs), phenomes related to a participant's health conditions, lifestyle, and environmental exposures can be characterized [14]. By linking large biorepositories containing human DNA samples and EHRs, we can comprehensively evaluate associations between genetically predicted biomarkers and a wide range of phenotypes by using the phenome-wide association study design (PheWAS) [15]. PheWAS has been introduced as an approach to replicate associations previously identified in observational studies and detect novel

genotype-phenotype associations [16]. In this study, we aim to explore the phenotypic effects of B vitamins and homocysteine under the phenome-wide association framework, by using genetic proxies in the UK Biobank.

Methods**Study design**

The UK Biobank genetic data contains genotypes for 488,366 participants. After quality control, a total of 385,917 individuals were retained for the subsequent analyses including PheWAS, Mendelian randomization (MR), dose-response, and mediation analyses. We first performed a PheWAS analysis to investigate the associations between genetic proxies of 4 biomarkers (folate, vitamin B6, vitamin B12, and homocysteine) and a wide range of disease outcomes. Associations with false discovery rate (FDR) <0.05 were considered significant. Second, the observed significant associations were replicated and regarded as successful with a threshold of MR $P < 0.05$. Third, dose-response, mediation, and bioinformatics analyses were conducted to characterize and quantify the associations between the biomarkers and the associated outcomes. The study design is presented in Figure 2.

UK Biobank

The UK Biobank is a large-scale, population-based prospective cohort study, which recruited 502,490 adult participants aged between 40 and 69 y in 2006 to 2010 and combined extensive measurement of baseline and genotype data linked with wide-ranged health outcomes [17]. In total, 613,944 SNPs were genotyped by the Affymetrix UK BiLEVE Axiom array and Affymetrix UK Biobank Axiom array, and 30,798,054 SNPs were imputed on the basis of a merged reference panel of the Haplotype Reference Consortium and the UK10K haplotype resources. Various health outcomes from 3 main different types of national medical records (for example, in-patient hospital episode records, cancer registry, and death registry) were incorporated into the UK Biobank to follow up the disease diagnosis, cancer occurrence, and causes of death among the enrolled participants. For quality control of genotype data, a list of field variables was made available by the UK Biobank to indicate the genotype quality, population structure, and genetic relatedness. To minimize the influence of diverse population structure within UK Biobank, quality control of samples was also conducted, and those that were identified as outliers with high heterozygosity or with high missing rate, sex mismatch or putative aneuploidy in sex chromosome, non-European ancestry, or individuals with relatedness were excluded from the analysis. More details on the quality control of UK Biobank samples are given in Supplemental Figure 1.

Weighted genetic risk score

To generate weighted genetic risk scores (wGRSs) for plasma folate, vitamin B6, vitamin B12, and homocysteine, we applied a 2-sample study design (exposures and outcomes are measured in nonoverlapping populations) by retrieving the GWAS summary data from previous published studies, aiming to increase the statistical power by incorporating multiple sources from the same ancestry. SNPs associated with plasma folate and vitamin B12 were identified from GWASs with 37,341 and 45,576 individuals of Icelandic and Danish populations [10]. SNPs associated with plasma vitamin B6 and homocysteine were identified in GWASs of 4763 and 44,147 individuals

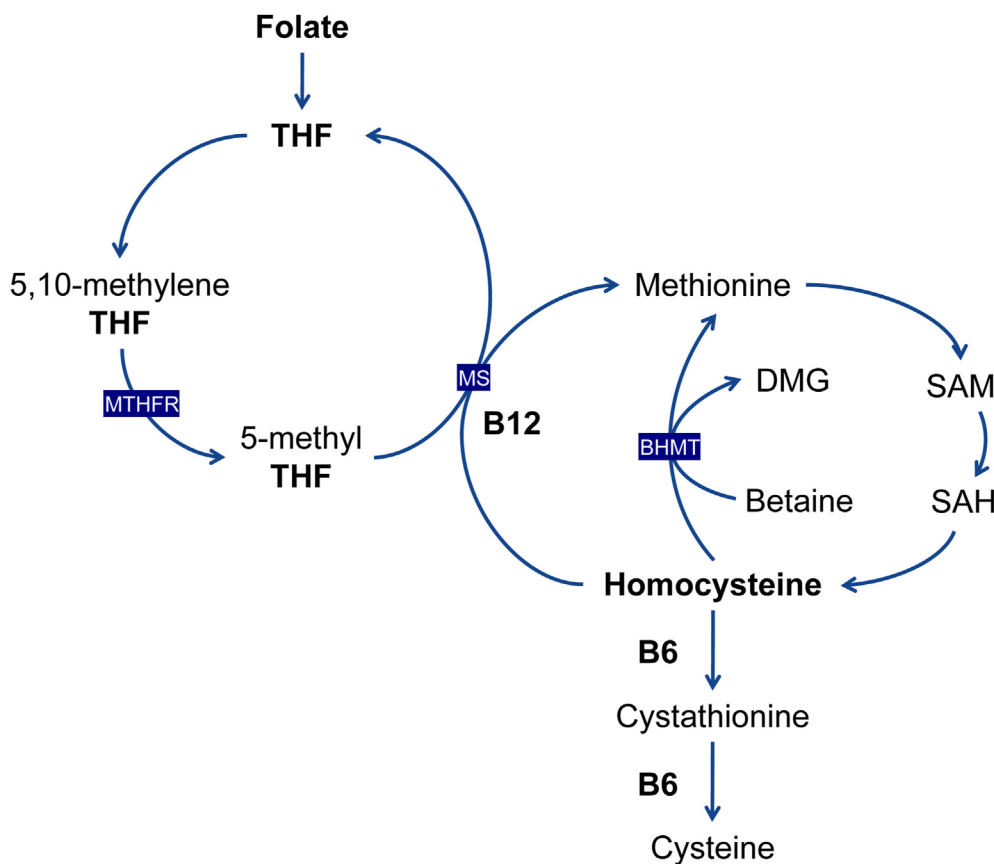


FIGURE 1. Interrelation between folate, vitamin B6, vitamin B12, and homocysteine metabolism. Folate is reduced to tetrahydrofolate (THF). A methyl group is then transferred to THF, forming 5,10-methylene-THF. 5,10-methylene-THF can be reduced by methylenetetrahydrofolate reductase (MTHFR) to 5-methyl-THF. The methyl group of 5-methyl-THF is transferred to homocysteine by methionine synthase (MS), generating methionine and regenerating THF. Vitamin B12 is a cofactor for MS. Methionine can also be generated independently of folate and B12, by the action of betaine homocysteine methyltransferase (BHMT), which transfers a methyl group from betaine to homocysteine. Methionine is then activated to form S-adenosyl methionine (SAM), which serves as a universal methyl donor for numerous reactions. S-adenosyl homocysteine (SAH) is 1 of the products of these methylation reactions and is subsequently hydrolyzed to generate homocysteine. Homocysteine is used either to regenerate methionine, or it is converted to cystathionine and then cysteine. BHMT, betaine homocysteine methyltransferase; DMG, dimethylglycine; MS, methionine synthase; MTHFR, methylenetetrahydrofolate reductase; SAH, S-adenosyl homocysteine; SAM, S-adenosyl methionine; THF, tetrahydrofolate.

of the European descent, respectively [11, 12]. We selected variants that were associated with biomarkers at genome-wide significance ($P < 5 \times 10^{-8}$) and clumped them by using a linkage disequilibrium threshold of $r^2 < 0.01$ according to the European reference panel of the 1000 Genomes project. As a result, we selected 2 SNPs for folate, 1 SNP for vitamin B6, 14 SNPs for vitamin B12, and 14 SNPs for homocysteine from the external GWAS summary data, which explained 0.55% of variance for folate, 0.65% of variance for vitamin B6, 5.59% of variance for vitamin B12, and 3.22% of variance for homocysteine, respectively (Supplemental Table 1). We also calculated the F-statistic of each instrument, and no weak instruments were identified (F-statistic > 10). For individuals in the UK Biobank, a wGRS for each biomarker was then calculated by adding up the number of biomarker-increasing alleles for each SNP weighted for the reported effect size of the variant.

Phenome framework and PheWAS analysis

The ontology of the phenome was defined on the base of the ICD codes in the EHRs. Records date back to 1997 for England, 1998 for Wales, and 1981 for Scotland. All of the current UK Biobank linked English, and most Welsh hospital data are coded in ICD10. However, because the collection of Scottish data collection began in 1981, the Scottish data collected before 1997 are coded in ICD-9, and small number of Welsh records are coded with ICD-9. Thus, we included both ICD-9 and ICD-10 codes in our PheWAS analysis. Individual ICD codes could not be directly used to define the phenome because they represent specific subphenotypes of a similar set of outcomes, instead of independent phenotypes. To account for the correlations between ICD codes,

we define the phenome framework by using the PheCODE schema that combines ≥ 1 related ICD codes into distinct outcome groups [18]. For a given phenotype, the case group included patients recorded as having the specific phecode that most closely related to the etiology of the disease, and the control group was defined on the basis of the absence of the phecode. Participants with a disease code that was related to 1 of the examined case group were also excluded from the control group [19]. As suggested by a simulation study, outcomes with >200 cases were included in the analysis [20]. In the PheWAS analysis, multivariable logistic regression was used to explore associations between wGRS and phecodes with adjustment for multiple covariates including sex, age, assessment center, and the first 10 genetic principal components (PCs). We also conducted sex-stratified analysis by dividing the population into male and female subgroups and reconducted PheWAS analysis in each subgroup. The Cochran Q test and the I square metric were calculated to detect potential heterogeneity among subgroups. The Benjamini-Hochberg method was applied for each biomarker to account for multiple testing and associations with FDR < 0.05 were considered statistically significant [21]. The PheWAS analysis was implemented by using the “PheWAS” [19] package (R version 4.0.3).

MR analysis

To replicate and inform potential causality for biomarker-outcome associations identified in the PheWAS analysis, we carried out 2-sample MR analysis in other independent European populations derived from the OpenGWAS database [22]. In addition, we also examined the causal relationships between the biomarkers and some intermediate phenotypes including total cholesterol (TC), TG, HDL cholesterol,

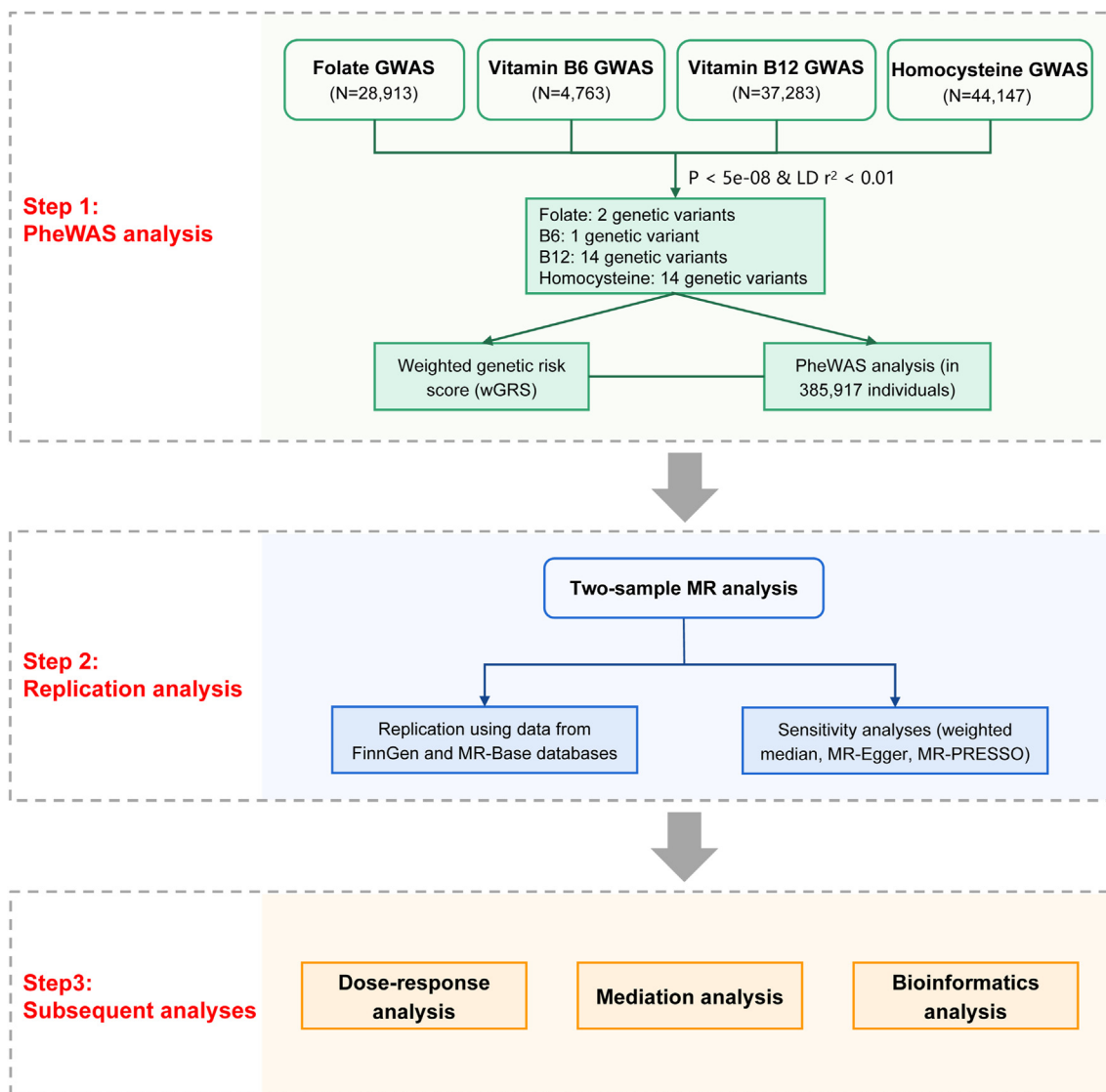


FIGURE 2. Flowchart for the study design. Step 1, PheWAS analysis was performed for each biomarker to investigate associations with a wide range of disease outcomes; Step 2, the observed significant associations were replicated by using several MR approaches (IVW, weighted median, MR-Egger, and MR-PRESSO); Step 3, dose-response, mediation, and bioinformatics analyses were conducted to quantify and characterize the associations between the biomarkers and the associated outcomes. GWAS, genome-wide association study; IVW, inverse variance weighted; MR, Mendelian randomization; wGRS, weighted genetic risk score.

LDL cholesterol, cystatin C, and creatinine. We applied the inverse variance weighted (IVW) approach as the main statistical method where ≥ 2 exposure SNPs were available, and the causal estimates were calculated by using the meta-analyzing SNP-specific Wald ratio estimates on the basis of the random-effects inverse variance method that weights each ratio by its SE [23]. Where 1 exposure SNP was available for analysis, we used the Wald ratio method. In addition, 3 sensitivity analyses, including the weighted median, MR-Egger, and MR-PRESSO, were applied to detect horizontal pleiotropic effects in the causal estimates when performing multivariable MR analysis [23–25]. Under the assumption that the association of each genetic variant with the exposure is independent of the pleiotropic effect of the variant (not through the exposure), the MR-Egger regression can be used to detect the bias owing to directional horizontal pleiotropy when conducting MR analysis by using multiple genetic instruments [23]. The weighted median model generates consistent estimates of causal effects if at least half of the weights come from valid SNPs [24]. The

MR-PRESSO method can detect outlying SNPs and provide causal estimates after removal of possible outliers under the assumption that the used SNPs are valid [25]. The ORs and corresponding 95% CIs of outcomes were scaled to 1-SD increase in genetically predicted circulating concentrations of B vitamins and homocysteine. The association with a $P < 0.05$ was deemed significant in the 2-sample MR analysis for replication. All tests were conducted by using the “Two-SampleMR” [22] and “MR-PRESSO” [25] packages in R Software 4.0.3.

Dose-response analysis

For significant PheWAS associations, we first conducted linear regression to test for any potential linear dose-response relationship; however, no significant linear association was identified. Then, we performed nonlinear dose-response analysis by using a restricted cubic spline function with 5 knots located at the fifth, 25th, 50th, 75th, and 95th percentiles of genetically predicted biomarker concentrations [26].

Multivariable logistic regression models were adopted to estimate ORs with 95% CIs for risk of outcomes, adjusted for age, sex, assessment center, and the first 10 PCs. Prediction regression implemented in “rms” R package [27] was applied to test for trends by assigning the median value for each category and modeling this variable as a continuous variable, and a $P < 0.05$ stands for the significance of nonlinearity. A dose-response curve was used to present the dose-response relationship between the biomarker and risk of outcome.

Mediation analysis

Given that B vitamins, including folate and vitamins B6 and B12, are important for the metabolism of homocysteine, abnormal concentrations of any of these B vitamins can lead to a change in homocysteine concentration. Therefore, to further explore any pleiotropic association derived from the correlations between B vitamins and homocysteine, we performed mediation analysis by using wGRSs as proxies for blood biomarker concentrations considering that data for B vitamin supplements in UKB are dietary estimates and only cover one-fifth of the population. We used genetically predicted plasma biomarker concentrations as the exposure and performed the mediation analysis with the “mediation” R package [28]. The package performs causal mediation analysis (CMA) under the assumption of sequential ignorability, which implies that there is no unmeasured confounding of the exposure-mediator, exposure-outcome, and mediator-outcome relationships while conditioning on covariates. More details regarding the rationale of mediation analysis can be found in Figure 3. The analysis reports an average causal mediation effect (ACME) that is transmitted through a mediator to the outcome and an average direct effect (ADE) that is explained by the exposure and the proportion of explained variance by the mediator. This approach was applied to disentangle the underlying mediating pathophysiological processes.

Bioinformatics analysis

For outcomes validated in the MR analysis, subsequent bioinformatics analysis was carried out to explore the underlying biological mechanisms. First, to identify whether the instrumental variants exert effects through regulating the expression of located genes, we conducted expression quantitative trait loci (eQTL) analysis based on the GTEx v8 release (<https://gtexportal.org/home/>), which consists of normalized gene expression data in whole blood tissue. Briefly, the cis-eQTL mapping window was defined as 1 MB up- and downstream of the transcription start site. The effect sizes were defined as the slopes of the linear regression computed as the effect of the alternate allele relative to the reference allele [29]. Then, to explore the interactions

between biomarker-associated genes and outcome-associated genes, we used STRING (<https://string-db.org/>) and Cytoscape (<https://cytoscape.org/>) to construct a protein–protein interaction (PPI) network by using the instrumental SNPs located genes and the top 10 associated genes of biomarker and outcome [30, 31]. The top 10 associated genes of biomarker and outcome were searched through the GeneCards website (<https://www.genecards.org/>) [32]. The biological processes in which these genes are involved were also explored by performing pathway enrichment analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome were used as sources [33, 34]. FDR correction was used for multiple testing, and an adjusted $P < 0.05$ was regarded as a cutoff threshold [21].

Results

UK Biobank participants

In this study, we used the UK Biobank genetic data of 488,366 participants. Genetic quality control was performed centrally by the UK Biobank. A total of 385,917 unrelated European individuals were included in our final analysis, consisting of 177,690 men (46.0% of the cohort) and 208,227 women (54.0% of the cohort). Basic characteristics including sociodemographic information and biomarker concentrations are present in Table 1.

PheWAS associations

The phenome defined by PheCODE schema consisted of 1804 distinct phecodes. After filtering out the phecodes with < 200 cases, PheWAS analysis was performed for 1117 phecodes that could be classified into 17 broadly related disease categories (Supplemental Table 2). Findings from the PheWAS analysis indicated that higher genetically predicted plasma folate was significantly associated with lower risk of megaloblastic anemia (OR: 0.91; 95% CI: 0.87, 0.95; $P = 2.50 \times 10^{-5}$), essential hypertension (OR: 0.98; 95% CI: 0.98, 0.99; $P = 5.01 \times 10^{-5}$), lipid metabolism disorders including hypercholesterolemia (OR: 0.98; 95% CI: 0.97, 0.99; $P = 9.92 \times 10^{-5}$) and hyperlipidemia (OR: 0.98; 95% CI: 0.97, 0.99; $P = 1.06 \times 10^{-4}$), and regional enteritis (OR: 0.92; 95% CI: 0.89, 0.96; $P = 3.02 \times 10^{-4}$) (Table 2 and Figure 4). Higher blood vitamin B6 was inversely associated with risk of urinary calculus (OR: 0.95; 95% CI: 0.93, 0.97; $P = 7.72 \times 10^{-7}$) and calculus of kidney (OR: 0.94; 95% CI: 0.92, 0.97; $P = 4.03 \times 10^{-5}$). Higher vitamin B12 concentration was related to lower risk of vitamin B–complex deficiencies (OR: 0.47; 95% CI: 0.40, 0.55; $P = 1.36 \times 10^{-21}$) and megaloblastic anemia (OR: 0.73; 95% CI: 0.62, 0.86; $P = 1.53 \times 10^{-4}$) and may increase risk of neoplasm of

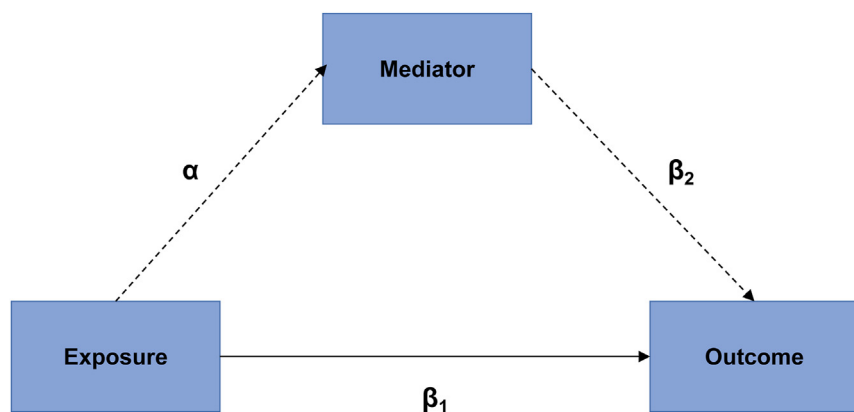


FIGURE 3. Schematic diagram illustrating the rationale of mediation analysis. Dotted lines ($\alpha\beta_2$) refer to the indirect effect of the exposure on the identified outcome through the mediator. Solid line (β_1) is the direct effect of the exposure on the identified outcome. The assumption under mediation analysis is that 1) the relationship between the exposure and the outcome is entirely mediated by the mediator, 2) the exposure is truly associated with the mediator, and 3) the exposure is not associated with any confounder of the relationship between the mediator and the outcome.

digestive system (OR: 1.50; 95% CI: 1.26, 1.77; $P = 2.54 \times 10^{-6}$) and cholelithiasis and cholecystitis (OR: 1.10; 95% CI: 1.05, 1.16; $P = 9.36 \times 10^{-5}$). As expected, higher circulating homocysteine was positively associated with risk of vitamin B-complex deficiencies (OR: 1.57; 95% CI: 1.25, 1.98; $P = 1.33 \times 10^{-4}$). This served as positive control for the genetic instrument of homocysteine. In addition, higher homocysteine was related to higher risk of lipid metabolism disorders including hyperlipidemia (OR: 1.14; 95% CI: 1.08, 1.20; $P = 2.17 \times 10^{-6}$) and hypercholesterolemia (OR: 1.13; 95% CI: 1.07, 1.19; $P = 1.34 \times 10^{-5}$), spondylosis with myelopathy (OR: 3.27; 95% CI: 1.80, 5.95; $P = 9.99 \times 10^{-5}$), kidney disease including chronic kidney disease (OR: 1.29; 95% CI: 1.13, 1.46; $P = 1.04 \times 10^{-4}$) and renal failure (OR: 1.21; 95% CI: 1.09, 1.34; $P = 1.90 \times 10^{-4}$), circulatory disease primarily cerebrovascular disease (OR: 1.21; 95% CI: 1.10, 1.33; $P = 1.12 \times 10^{-4}$) and coronary atherosclerosis (OR: 1.14; 95% CI: 1.06, 1.22; $P = 2.07 \times 10^{-4}$), celiac disease (OR: 1.50; 95% CI: 1.20, 1.87; $P = 3.27 \times 10^{-4}$), syncope and collapse (OR: 1.19; 95% CI: 1.09, 1.31; $P = 2.17 \times 10^{-4}$), celiac disease (OR: 1.50; 95% CI: 1.20, 1.87; $P = 3.27 \times 10^{-4}$), and migraine (OR: 1.32; 95% CI: 1.12, 1.54; $P = 5.90 \times 10^{-4}$). Sex-stratified analysis revealed considerable heterogeneity in the associations of vitamin B12 with vitamin deficiency ($P_{\text{heterogeneity}} = 0.022$, $I^2 = 81\%$) among men and women. More specifically, per 1SD high blood vitamin B12 was associated with vitamin deficiency in males only (OR: 0.64; 95% CI: 0.55, 0.76; $P = 6.50 \times 10^{-8}$) (Supplemental Table 3).

MR analysis

The MR IVW analysis revealed that genetically proxied high vitamin B6 was associated with lower risk of calculus of kidney (OR: 0.64; 95% CI: 0.42, 0.97; $P = 0.033$) (Table 3). For homocysteine, genetically proxied high concentration was associated with a reduction in HDL cholesterol concentrations (OR: 0.95; 95% CI: 0.90, 1.00; $P = 0.041$, $P_{\text{heterogeneity}} = 0.021$, $P_{\text{MR-Egger}} = 0.311$) and higher risk of hypercholesterolemia (OR: 1.28; 95% CI: 1.04, 1.56; $P = 0.018$, $P_{\text{heterogeneity}} = 0.046$, $P_{\text{MR-Egger}} = 0.695$). In addition, genetically predicted high homocysteine was related to elevated creatinine concentrations (OR: 1.11; 95% CI: 1.01, 1.21; $P = 0.027$, $P_{\text{heterogeneity}} < 0.001$, $P_{\text{MR-Egger}} = 0.437$) and higher risk of chronic kidney disease (OR: 1.32; 95% CI: 1.06, 1.63; $P = 0.012$, $P_{\text{heterogeneity}} = 0.276$, $P_{\text{MR-Egger}} = 0.784$).

TABLE 1

Baseline characteristics of participants in the UK Biobank

Baseline characteristic	All participants ($n = 385,917$)
Sex, n (%)	
Female	208,227 (54.0)
Male	177,690 (46.0)
Age (y), mean (SD)	56.7 (8.0)
BMI (kg/m^2), mean (SD)	27.4 (4.8)
Townsend deprivation index, mean (SD)	-1.5 (3.0)
Smoking status, n (%)	
Current	40,039 (10.4)
Former	136,651 (35.4)
Never	207,296 (53.7)
Unknown	1931 (0.5)
Blood pressure (mmHg), mean (SD)	
Diastolic blood pressure	82.2(10.7)
Systolic blood pressure	139.8 (19.7)
Biomarker concentration (mmol/L), mean (SD)	
LDL-cholesterol	3.6 (0.9)
HDL-cholesterol	1.5 (0.4)
Total cholesterol	5.7 (1.1)
TG	1.7 (1.0)

Egger = 0.784). All MR estimates for associations identified by PheWAS are presented in Supplemental Table 4.

Dose-response analysis

The dose-response results showed that risk of vitamin B-complex deficiencies ($P < 0.001$) and megaloblastic anemia ($P = 0.004$) reduced when vitamin B12 concentrations increase, whereas the estimated ORs of developing cerebrovascular disease increased in response to the increase of homocysteine ($P = 0.037$) (Figure 5A–C). In addition, significant U-shaped dose-response relationships were identified for the associations of folate with megaloblastic anemia ($P = 0.009$) and vitamin B12 with cholelithiasis and cholecystitis ($P = 0.013$) (Figure 5D, E).

Mediation analysis

Considering the limited variance of folate and vitamin B6 explained by the selected genetic instruments, we only examined the mediation effects of homocysteine on the associations between vitamin B12 and the identified outcomes. The results showed that homocysteine is a potential mediator in the associations of vitamin B12 with vitamin deficiency (ACME = -1.20×10^{-4} , $P < 0.001$, Supplemental Table 5), vitamin B-complex deficiencies (ACME = -9.01×10^{-5} , $P < 0.001$), megaloblastic anemia (ACME = -6.45×10^{-5} , $P < 0.001$), and pernicious anemia (ACME = -5.19×10^{-5} , $P = 0.020$), but the mediation only accounts for 3.4%, 2.2%, 4.2%, and 4.3% of the total effects, respectively.

Bioinformatics analysis

By using gene expression data from GTEx, we found that 3 homocysteine-associated genetic instruments including rs4660306, rs1801133, and rs9369898 were associated with the expression of *PRDX1*, *MTHFR*, and *MMUT* (Figure 6A). The protein-protein interaction analysis identified interactions between the homocysteine-associated genes (*MTHFR*, *MMUT*, but not *PRDX1*) and hyperlipidemia-associated genes (*APOB*, *APOE*, and *APOA5*) (Figure 6B). Pathway analysis revealed that these genes were significantly enriched in biological processes related to cholesterol metabolism, plasma lipoprotein remodeling, assembly of active lipoprotein lipase, and hepatic triacylglycerol lipase complexes (Supplemental Table 6), which play an important role in lipid metabolism; thus, alterations in these pathways may lead to the development of hyperlipidemia. The protein-protein interaction results also revealed interactions between the homocysteine-associated genes (*PRDX1*, *MTHFR*, and *MMUT*) and kidney disease-associated genes (*TP53*, *TNF*, *IL-6*, *REN*, *ACE*, and *ALB*) (Figure 6C, D), which are aggregated in the renin-angiotensin system, biological oxidations, and immune-related signaling (Supplemental Table 6). These pathways are biologically relevant to the development and progression of chronic kidney diseases. These data suggest possible mechanisms of action that may underlie the observed PheWAS associations between homocysteine and hyperlipidemia and chronic kidney disease.

Discussion

In this study, we comprehensively investigated the effects of B vitamins (folate, vitamin B6, and vitamin B12) and homocysteine on a wide range of disease outcomes. A total of 32 pairs of genotype-phenotype associations were identified, resulting in 25 unique health outcomes belonging to several disease groups including hematopoietic

TABLE 2
Significant PheWAS associations by using wGRS in the UK Biobank ($n = 385,917$)

Biomarker	Phecode	Description	Group	Cases	Participants	OR (95% CI)	<i>P</i> value	FDR adjusted <i>P</i> value
Folate	281.11	Pernicious anemia	Hematopoietic	1144	355,132	0.88 (0.83, 0.93)	2.37E-05	1.13E-02
	281.1	Megaloblastic anemia	Hematopoietic	1862	355,850	0.91 (0.87, 0.95)	2.50E-05	1.13E-02
	281	Other deficiency anemia	Hematopoietic	1972	355,960	0.91 (0.87, 0.95)	4.19E-05	1.13E-02
	401.1	Essential hypertension	Circulatory system	109,289	385,661	0.98 (0.98, 0.99)	5.01E-05	1.13E-02
	401	Hypertension	Circulatory system	109,545	385,917	0.98 (0.98, 0.99)	5.06E-05	1.13E-02
	272	Disorders of lipid metabolism	Endocrine/metabolic	54,333	385,917	0.98 (0.97, 0.99)	9.79E-05	1.48E-02
	272.11	Hypercholesterolemia	Endocrine/metabolic	50,143	381,727	0.98 (0.97, 0.99)	9.92E-05	1.48E-02
	272.1	Hyperlipidemia	Endocrine/metabolic	54,099	385,683	0.98 (0.97, 0.99)	1.06E-04	1.48E-02
	555.1	Regional enteritis	Digestive	2166	315,062	0.92 (0.89, 0.96)	3.02E-04	3.74E-02
Vitamin B6	594	Urinary calculus	Genitourinary	8724	385,126	0.95 (0.93, 0.97)	7.72E-07	8.62E-04
	594.1	Calculus of kidney	Genitourinary	4748	381,150	0.94 (0.92, 0.97)	4.03E-05	2.25E-02
Vitamin B12	261.2	Vitamin B–complex deficiencies	Endocrine/metabolic	2246	380,355	0.47 (0.40, 0.55)	1.36E-21	1.52E-18
	261	Vitamin deficiency	Endocrine/metabolic	5055	383,164	0.74 (0.67, 0.82)	5.51E-09	3.08E-06
	158	Neoplasm of unspecified nature of digestive system	Neoplasms	1540	362,470	1.50 (1.26, 1.77)	2.54E-06	9.44E-04
	574	Cholelithiasis and cholecystitis	Digestive	20,567	383,100	1.10 (1.05, 1.16)	9.36E-05	2.61E-02
	281.1	Megaloblastic anemia	Hematopoietic	1862	355,850	0.73 (0.62, 0.86)	1.53E-04	3.42E-02
	574.1	Cholelithiasis	Digestive	18,324	380,857	1.10 (1.05, 1.16)	2.37E-04	4.01E-02
	281.11	Pernicious anemia	Hematopoietic	1144	355,132	0.68 (0.55, 0.83)	2.51E-04	4.01E-02
Homocysteine	272.1	Hyperlipidemia	Endocrine/metabolic	54,099	385,683	1.14 (1.08, 1.20)	2.17E-06	1.53E-03
	272	Disorders of lipid metabolism	Endocrine/metabolic	54,333	385,917	1.14 (1.08, 1.20)	2.74E-06	1.53E-03
	272.11	Hypercholesterolemia	Endocrine/metabolic	50,143	381,727	1.13 (1.07, 1.19)	1.34E-05	4.98E-03
	721.2	Spondylosis with myelopathy	Musculoskeletal	335	374,722	3.27 (1.80, 5.95)	9.99E-05	2.08E-02
	585.33	Chronic Kidney Disease, stage III	Genitourinary	7753	362,882	1.29 (1.13, 1.46)	1.04E-04	2.08E-02
	433	Cerebrovascular disease	Circulatory system	14,058	384,206	1.21 (1.10, 1.33)	1.12E-04	2.08E-02
	261.2	Vitamin B–complex deficiencies	Endocrine/metabolic	2246	380,355	1.57 (1.25, 1.98)	1.33E-04	2.12E-02
	585.3	Chronic renal failure (CKD)	Genitourinary	12,777	367,906	1.21 (1.09, 1.34)	1.90E-04	2.20E-02
	261	Vitamin deficiency	Endocrine/metabolic	5055	383,164	1.34 (1.15, 1.57)	1.92E-04	2.20E-02
	411.4	Coronary atherosclerosis	Circulatory system	28,433	369,752	1.14 (1.06, 1.22)	2.07E-04	2.20E-02
	788	Syncope and collapse	Symptoms	14,378	385,917	1.19 (1.09, 1.31)	2.17E-04	2.20E-02
	557.1	Celiac disease	Digestive	2470	315,366	1.50 (1.20, 1.87)	3.27E-04	3.05E-02
	252.2	Hypoparathyroidism	Endocrine/metabolic	355	379,815	0.36 (0.20, 0.64)	5.89E-04	4.71E-02
	340	Migraine	Neurological	5026	376,143	1.32 (1.12, 1.54)	5.90E-04	4.71E-02

FDR, false discovery rate; wGRS, weighted genetic risk score.

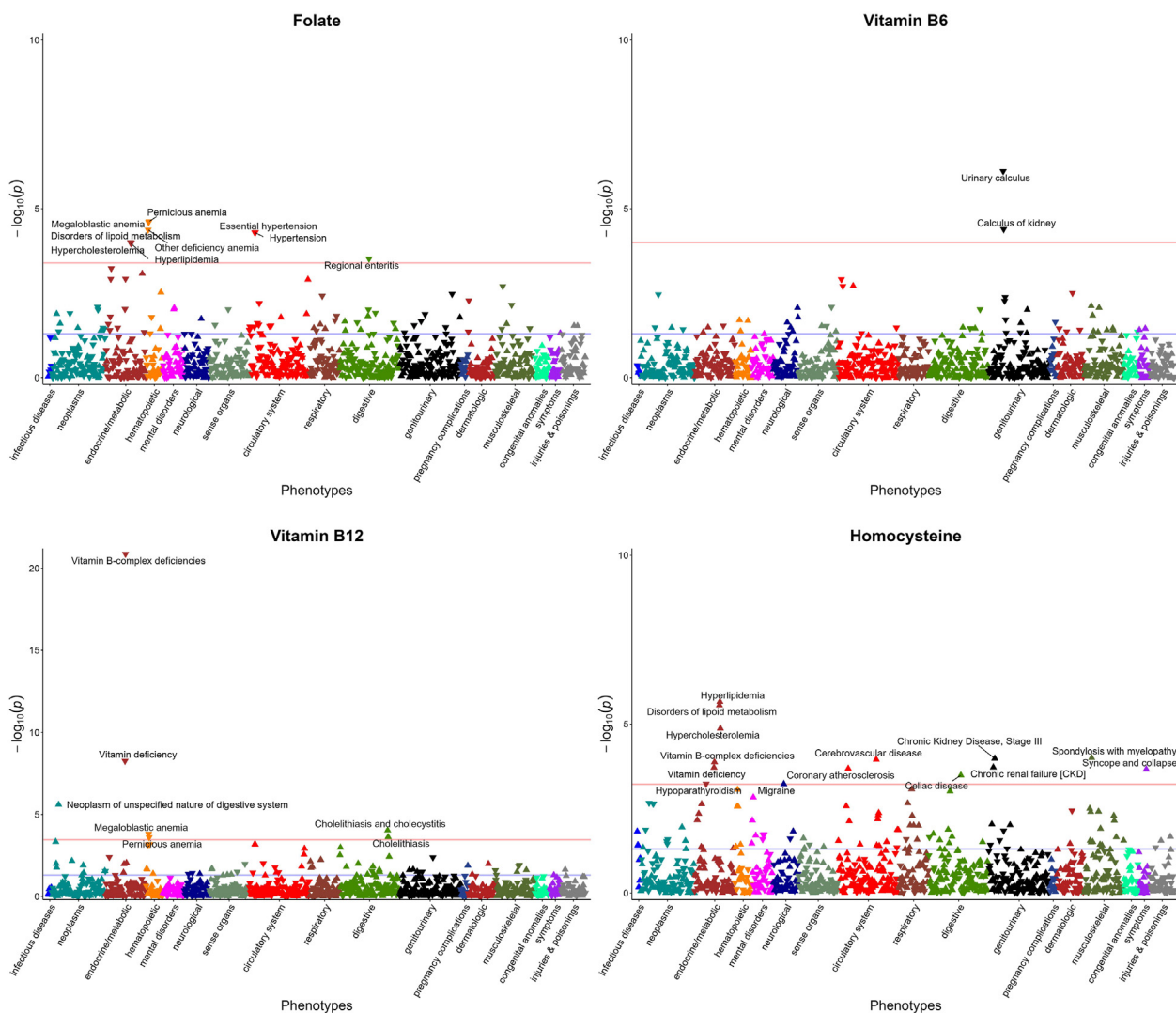


FIGURE 4. Plots for PheWAS associations of folate, vitamin B6, vitamin B12, and homocysteine ($n = 385,917$). The x axis represents distinct phenotypic groups by using different colors, and the y axis represents the P value for the phenotypic associations.

(for example, megaloblastic anemia), endocrine/metabolic (for example, hyperlipidemia), circulatory (for example, hypertension), genitourinary (for example, chronic kidney disease), digestive (for example, cholelithiasis and cholecystitis), musculoskeletal (for example, spondylosis with myelopathy), neurological disorders (for example migraine), neoplasm (neoplasm of digestive system), and symptoms (for example, syncope and collapse). Two-sample MR analysis further confirmed the potential causal effects of B vitamins and homocysteine on calculus of kidney, hypercholesterolemia, and chronic kidney disease. When incorporating evidence from published MR studies, the present study replicated the associated homocysteine and lipid metabolism disorders and identified a potentially new causal association between vitamin B6 and calculus of kidney and homocysteine and chronic kidney disease. Our findings provide evidence for further investigation on the clinical relevance of B vitamins and homocysteine with the identified disease outcomes.

Our PheWAS and MR analyses consistently demonstrated that genetically proxied high blood vitamin B6 was significantly associated with decreased risk of urinary calculus and calculus of kidney. Studies on vitamin B6 in relation to calculus are scarce and conflicting. Previous cohort studies found that the intake of vitamin B6

was not correlated with risk of kidney stone formation among men in the Health Professionals Follow-up Study, whereas a large dose of vitamin B6 seemed to be protective among women in the NHS [35, 36]. A recent prospective study conducted in the same cohorts with larger number of incident events (6308 incident kidney stones) and additional follow-up time (3,108,264 person-years) showed that there was no association between vitamin B6 intake and risk of kidney stones in either men or women [37]. The present study confirmed the previous findings of a lack of association between vitamin B6 and risk of kidney stones among women and identified that higher vitamin B6 appeared to exert a protective effect on calculus in men. In addition, our study adopted a MR approach to mitigate the impacts of residual confounding and reverse causality and provided further supportive evidence showing that high plasma vitamin B6 was associated with lower risk of calculus of kidney. From a biological perspective, it may be related to the role of vitamin B6 in reducing urinary excretion of oxalate, and thus, risk of developing calcium oxalate kidney stones [38].

For vitamin B12, this study found that higher genetically proxied plasma vitamin B12 was significantly associated with decreased risk of vitamin B-complex deficiencies and megaloblastic anemia in a dose-

TABLE 3
Replication of MR effect estimates in the OpenGWAS database

Exposure	Outcome and data source	MR method	Beta	SE	OR (95% CI)	<i>P</i> value ¹	Intercept ²	<i>P</i> value ²	Cases	Participants	
Vitamin B6	Calculus of kidney and ureter UKBB	PheWAS	−0.058	0.014	0.94 (0.92, 0.97)	4.03E-05	.	.	4748	381,150	
		Wald ratio ³	−0.453	0.213	0.64 (0.42, 0.97)	0.033	.	.	3856	176,613	
Homocysteine	HDL cholesterol concentrations OpenGWAS: ieu-4844	IVW	−0.055	0.027	0.95 (0.90, 1.00)	0.041	.	.	.	77,409	
		Weighted median ⁴	−0.061	0.027	0.94 (0.89, 0.99)	0.022	
		MR-Egger ⁴	−0.065	0.061	0.94 (0.83, 1.06)	0.311	0.001	0.853	.	.	
		MR-PRESSO ⁴	−0.059	0.026	0.94 (0.90, 0.99)	0.044	
	Hypercholesterolemia UKBB	PheWAS	0.122	0.028	1.13 (1.07, 1.19)	1.34E-05	.	.	50,143	381,727	
		IVW	0.245	0.103	1.28 (1.04, 1.56)	0.018	.	.	6840	167,301	
		Weighted median	0.210	0.106	1.23 (1.00, 1.52)	0.047	
		MR-Egger	0.093	0.230	1.10 (0.70, 1.72)	0.695	0.013	0.471	.	.	
	MR-PRESSO	0.245	0.103	1.28 (1.04, 1.56)	0.035	
		Creatinine concentrations OpenGWAS: met-Creatinine	IVW	0.103	0.046	1.11 (1.01, 1.21)	0.027	.	.	.	110,058
			Weighted median	0.013	0.022	1.01 (0.97, 1.06)	0.553
			MR-Egger	0.087	0.108	1.09 (0.88, 1.35)	0.437	0.001	0.875	.	.
MR-PRESSO	0.071		0.031	1.07 (1.01, 1.14)	0.490		
Chronic kidney disease UKBB	PheWAS	0.252	0.065	1.29 (1.13, 1.46)	1.04E-04	.	.	7753	362,882		
	IVW	0.276	0.110	1.32 (1.06, 1.63)	0.012	.	.	1533	20,920		
	Weighted median	0.134	0.149	1.14 (0.85, 1.53)	0.368		
	MR-Egger	0.070	0.250	1.07 (0.66, 1.75)	0.784	0.017	0.376	.	.		
MR-PRESSO	0.276	0.110	1.32 (1.06, 1.63)	0.026			

MR, Mendelian randomization; IVW, inverse variance weighted; UKBB, UK Biobank. ¹The *P* value of the MR effect estimates from different methods. ²The Egger regression intercept and horizontal pleiotropy *P* value of MR-Egger analysis. ³The Wald ratio method was applied to examine MR estimates for vitamin B6 as only 1 SNP was used as instrumental variable for vitamin B6. ⁴These 3 sensitivity methods were applied to detect any pleiotropic effect when performing multivariable MR analysis.

response manner, consistent with its known causal relationship with megaloblastic anemia [39, 40]. The mediation analysis revealed that homocysteine may mediate the associations of vitamin B12 with vitamin deficiency and anemia with small effects. A possible explanation may be that low vitamin B12 results in elevated homocysteine. As 1 of the major pathways of homocysteine metabolism, remethylation of homocysteine to methionine, requires vitamin B12 as the essential cofactor [41], vitamin B12 deficiency inhibits homocysteine to be methylated, thus leading to the accumulation of homocysteine, which also exerted positive effects on the development of vitamin deficiency and anemia observed in our study.

Homocysteine, a by-product of the 1-carbon metabolism, has been previously associated with lipid metabolism [42, 43]. In the current study, we found that higher genetically predicted plasma homocysteine was associated with a reduction in HDL cholesterol concentrations and increased risk of hypercholesterolemia. In addition, we conducted an overview of MR studies on the health effects of homocysteine and identified that genetically predicted high homocysteine has been reported to be related to lower HDL cholesterol and higher oxidized LDL cholesterol [44], providing support for our findings. Because homocysteine is reported to reduce the concentration of HDL cholesterol in plasma by inhibiting the hepatic synthesis of apoA-I [45, 46], this is a

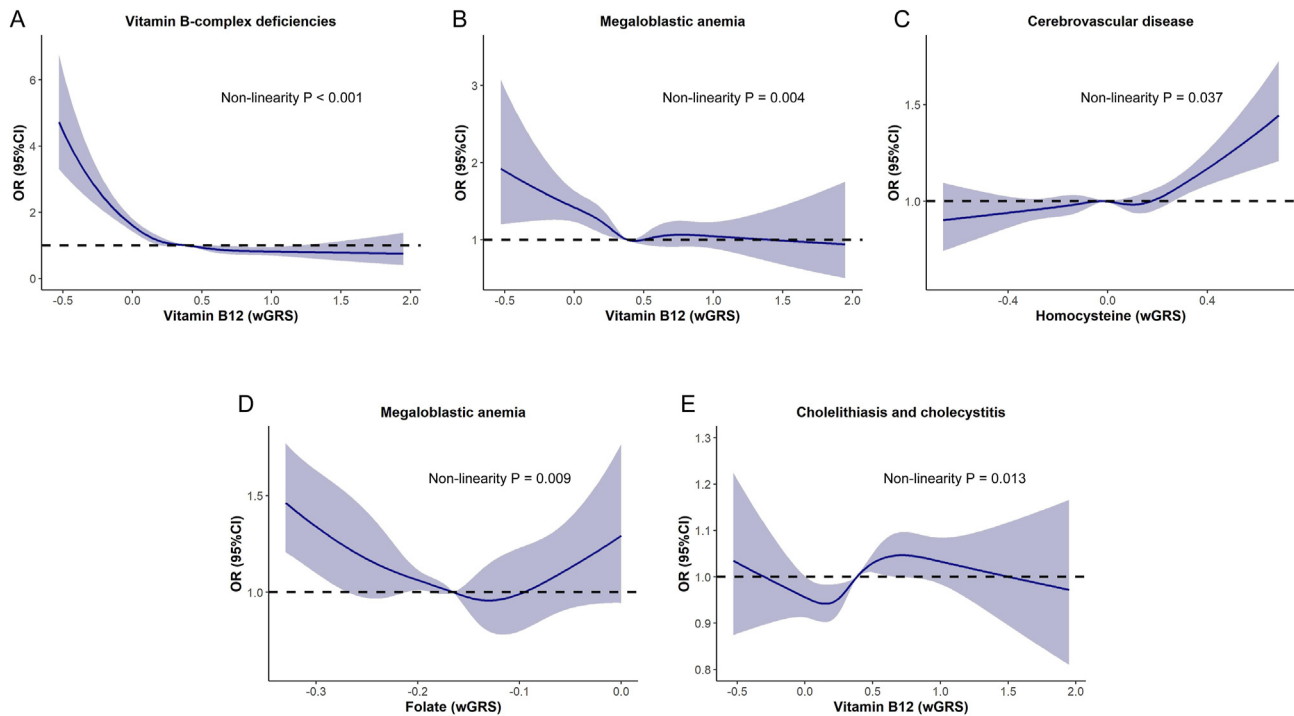


FIGURE 5. Dose-response relationships between genetically predicted biomarker concentrations and risk of outcomes identified by PheWAS ($n = 385,917$). (A) Nonlinear relationship between genetically predicted vitamin B12 concentration and vitamin B-complex deficiencies risk. (B) Nonlinear relationship between genetically predicted vitamin B12 concentration and megaloblastic anemia risk. (C) Nonlinear relationship between genetically predicted homocysteine concentration and cerebrovascular disease risk; (D) Nonlinear relationship between genetically predicted folate concentration and megaloblastic anemia risk. (E) Nonlinear relationship between genetically predicted vitamin B12 concentration and cholelithiasis and cholecystitis risk. wGRS, weighted genetic risk score.

possible mechanism linking homocysteine to the development of hypercholesterolemia. In addition, the present study revealed significant causal associations of higher genetically proxied homocysteine concentration with elevated creatinine concentrations (a biomarker for the evaluation of kidney function) and increased risk of chronic kidney disease. Consistently, a recent MR study revealed that higher blood homocysteine decreased estimated glomerular filtration rate (eGFR) and thus reduced kidney function [47], which may partly explain the associations between homocysteine and chronic kidney disease. Moreover, we identified that the underlying potential biological mechanisms may be related to the effects of homocysteine-associated genetic variants on the expression of genes within important pathways (for example, biological oxidations and immune-related signaling) that are involved in the development of chronic kidney disease [48].

Folate status during pregnancy is essential for adequate fetal development and for the long-term health of the individual. Low maternal folate status may induce adverse birth/pregnancy outcomes including neural tube defects (NTD) [49, 50]. Therefore, several countries (that is, Canada and the US) have implemented public policies to fortify foods with folic acid since 1998, which resulted in the decrease of NTD prevalence [51,52]. However, owing to limited data on pregnancy-related outcomes in the UK Biobank, we did not identify any relevant outcome associated with the instrumental variable for genetically predicted folate.

The present study has several strengths and limitations. We applied a rigorous study design, starting with a PheWAS analysis, based on which we investigated the relationship of genetically predicted concentrations of B vitamins and homocysteine with a wide spectrum of phenotypes. We applied additional analyses including MR, dose-

response analysis, mediation analysis, and pathway enrichment analysis to further elucidate the identified associations. The study also has some limitations. Our study was confined to individuals of European ancestry to minimize population structure bias, which hinders the generalizability of our findings to other populations. Moreover, considering that participants in the UK Biobank are exclusively from the UK, our results may not be applicable to the wider Caucasian population. Another limitation is the small variance of exposure explained by the genetic instruments, which may lead to inadequate statistical power in detecting phenotypic associations and underestimated association and mediation effects. In addition, we did not use the sex-specific GWAS SNPs as instrumental variables when performing stratified analysis owing to the lack of sex-specific GWAS data. Finally, we only incorporated EHR data for case ascertainment, and survival bias may occur because the medical records start before the time of cohort participation. More advanced criteria should be developed to help improve the coverage and validity of case definition for future PheWAS studies.

In conclusion, this study suggests robust associations between genetically predicted B vitamins and homocysteine concentrations with a group of disease outcomes, including hematopoietic, metabolic, circulatory, genitourinary, digestive, musculoskeletal, neurological disorders, neoplasm, and symptoms. In particular, abnormal concentrations of B vitamins and homocysteine could be associated with calculus of kidney, hypercholesterolemia, and chronic kidney disease. The biomarker-associated genetic variants exert their effects on the expression of genes that play important role in the development of the outcomes, which may be potential biological mechanisms underlying the identified associations.

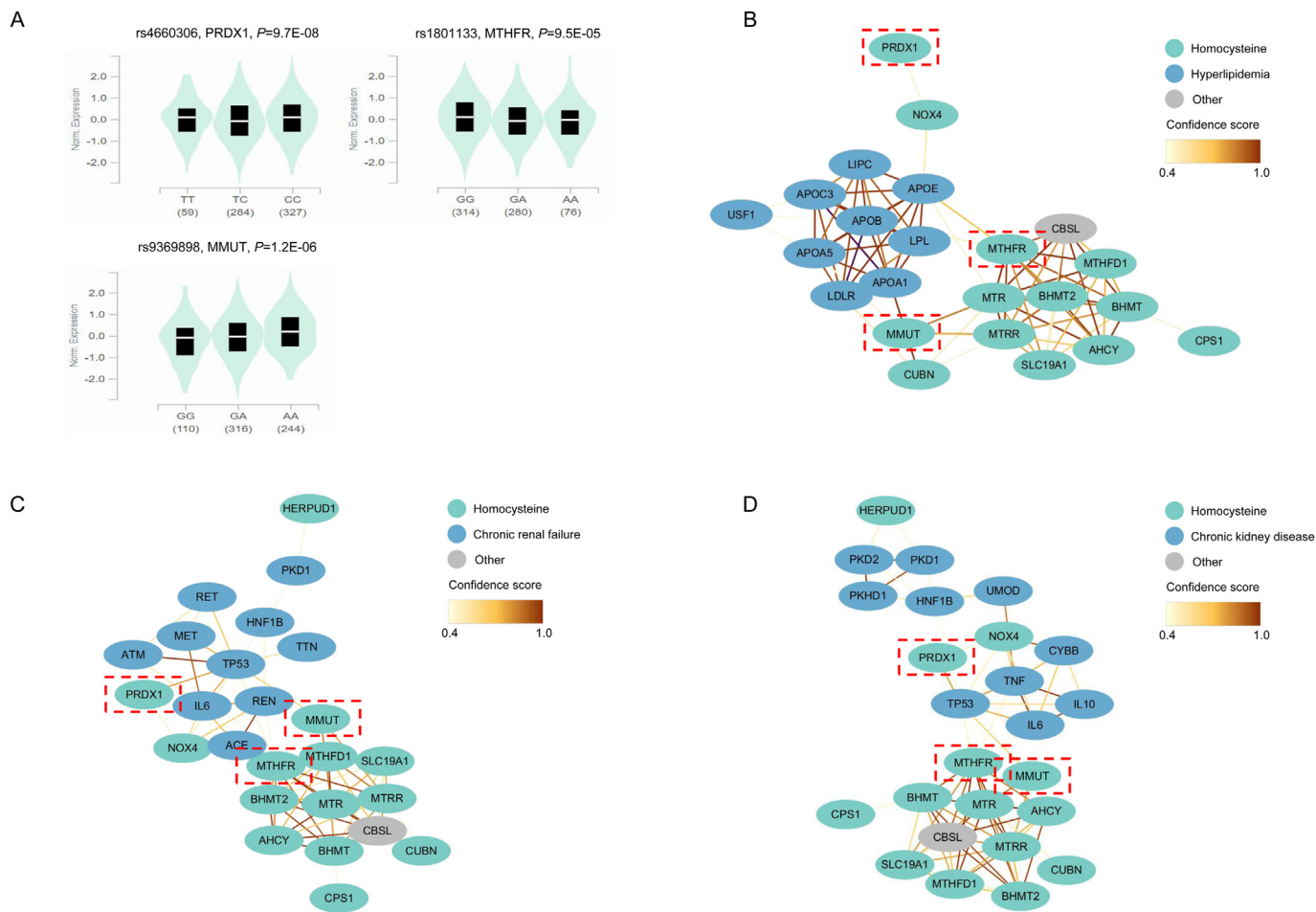


FIGURE 6. Bioinformatics analysis for homocysteine-associated instrumental variants. (A) eQTL violin plots of the associations between homocysteine-associated SNPs and expression of located genes. (B) Diagram of the interactions between homocysteine-associated genes and hyperlipidemia-associated genes. (C) Diagram of the interactions between homocysteine-associated genes and chronic renal failure-associated genes. (D) Diagram of the interactions between homocysteine-associated genes and chronic kidney disease-associated genes. eQTL, expression quantitative trait loci; MMUT, methylmalonyl-CoA mutase; MTHFR, methylenetetrahydrofolate reductase; PRDX1, peroxiredoxin 1; SNP, single nucleotide polymorphism.

Conflict of interest

All authors declare no competing interest.

Acknowledgements

We thank UK Biobank for their help in providing the data.

Author contribution

The authors’ responsibilities were as follows: ET, JL, and XL conceived and designed the study; LW, ET, and XL designed the methodology; LW conducted data analysis, made figures, and drafted the manuscript. All authors advised on statistical analyses and made critical revisions of the manuscript for important intellectual content. All authors have read and approved the final version of the manuscript.

Funding

LW is supported by a Darwin Trust PhD studentship. ET is supported by a Cancer Research UK Career Development Fellowship

(C31250/A22804). This work was in part supported by the Canadian Institutes of Health Research (CIHR)-Funding Reference Number: 175263.

Data Availability

This research has been conducted using the UK Biobank Resource under Application Number 10775. Data are available from the UK Biobank (<https://www.ukbiobank.ac.uk/>) for researchers who meet the criteria and gain approvals to access the research database from the UK Biobank.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ajcnut.2023.01.005>.

References

[1] Y. Bo, Y. Zhu, Y. Tao, X. Li, D. Zhai, Y. Bu, et al., Association between folate and health outcomes: an umbrella review of meta-analyses, *Front Public Health* 8 (2020) 550753.

- [2] R. Green, L.H. Allen, A.L. Björke-Monsen, A. Brito, J.L. Guéant, J.W. Miller, et al., Correction: vitamin B12 deficiency, *Nat Rev Dis Primers* 3 (2017) 17054.
- [3] C.T. Peterson, D.A. Rodionov, A.L. Osterman, S.N. Peterson, B vitamins and their role in immune regulation and cancer, *Nutrients* 12 (11) (2020) 3380.
- [4] A.D. Smith, H. Refsum, Homocysteine, B vitamins, and cognitive impairment, *Annu Rev Nutr* 36 (2016) 211–239.
- [5] G.S. Ducker, J.D. Rabinowitz, One-carbon metabolism in health and disease, *Cell Metab* 25 (1) (2017) 27–42.
- [6] H. Jakubowski, Homocysteine modification in protein structure/function and human disease, *Physiol Rev* 99 (1) (2019) 555–604.
- [7] J.D. Spence, Q. Yi, G.J. Hankey, B vitamins in stroke prevention: time to reconsider, *Lancet Neurol* 16 (9) (2017) 750–760.
- [8] M.J. Jardine, A. Kang, S. Zoungas, S.D. Navaneethan, T. Ninomiya, S.U. Nigwekar, et al., The effect of folic acid based homocysteine lowering on cardiovascular events in people with kidney disease: systematic review and meta-analysis, *BMJ* 344 (2012) e3533.
- [9] S.E. Nilsson, S. Read, S. Berg, B. Johansson, Heritabilities for fifteen routine biochemical values: findings in 215 Swedish twin pairs 82 years of age or older, *Scand J Clin Lab Invest* 69 (5) (2009) 562–569.
- [10] N. Grarup, P. Sulem, C.H. Sandholt, G. Thorleifsson, T.S. Ahluwalia, V. Steinthorsdottir, et al., Genetic architecture of vitamin B12 and folate levels uncovered applying deeply sequenced large datasets, *PLOS Genet* 9 (6) (2013), e1003530.
- [11] A. Hazra, P. Kraft, R. Lazarus, C. Chen, S.J. Chanock, P. Jacques, et al., Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway, *Hum Mol Genet* 18 (23) (2009) 4677–4687.
- [12] J.B. van Meurs, G. Pare, S.M. Schwartz, A. Hazra, T. Tanaka, S.H. Vermeulen, et al., Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease, *Am J Clin Nutr* 98 (3) (2013) 668–676.
- [13] C.A. Emdin, A.V. Khera, S. Kathiresan, Mendelian randomization, *JAMA* 318 (19) (2017) 1925–1926.
- [14] P.B. Jensen, L.J. Jensen, S. Brunak, Mining electronic health records: towards better research applications and clinical care, *Nat Rev Genet* 13 (6) (2012) 395–405.
- [15] W.S. Bush, M.T. Oetjens, D.C. Crawford, Unravelling the human genome-phenome relationship using phenome-wide association studies, *Nat Rev Genet* 17 (3) (2016) 129–145.
- [16] J.C. Denny, L. Bastarache, M.D. Ritchie, R.J. Carroll, R. Zink, J.D. Mosley, et al., Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data, *Nat Biotechnol* 31 (12) (2013) 1102, 145.
- [17] C. Bycroft, C. Freeman, D. Petkova, G. Band, L.T. Elliott, K. Sharp, et al., The UK Biobank resource with deep phenotyping and genomic data, *Nature* 562 (7726) (2018) 203–209.
- [18] J.C. Denny, M.D. Ritchie, M.A. Basford, J.M. Pulley, L. Bastarache, K. Brown-Gentry, et al., PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations, *Bioinformatics* 26 (9) (2010) 1205–1210.
- [19] R.J. Carroll, L. Bastarache, J.C. Denny, R. PheWAS, data analysis and plotting tools for phenome-wide association studies in the R environment, *Bioinformatics* 30 (16) (2014) 2375–2376.
- [20] A. Verma, Y. Bradford, S. Dudek, A.M. Lucas, S.S. Verma, S.A. Pendergrass, et al., A simulation study investigating power estimates in phenome-wide association studies, *BMC Bioinformatics* 19 (1) (2018) 120.
- [21] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J R Stat Soc B* 57 (1) (1995) 289–300.
- [22] G. Hemani, J. Zheng, B. Elsworth, K.H. Wade, V. Haberland, D. Baird, et al., The MR-base platform supports systematic causal inference across the human phenome, *Elife* 7 (2018), e34408.
- [23] J. Bowden, G. Davey Smith, S. Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression, *Int J Epidemiol* 44 (2) (2015) 512–525.
- [24] J. Bowden, G. Davey Smith, P.C. Haycock, S. Burgess, Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator, *Genet Epidemiol* 40 (4) (2016) 304–314.
- [25] M. Verbanck, C.Y. Chen, B. Neale, R. Do, Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases, *Nat Genet* 50 (5) (2018) 693–698.
- [26] L. Desquilbet, F. Mariotti, Dose-response analyses using restricted cubic spline functions in public health research, *Stat Med* 29 (9) (2010) 1037–1057.
- [27] F.E. Harrell Jr., R.M. Califf, D.B. Pryor, K.L. Lee, R.A. Rosati, Evaluating the yield of medical tests, *JAMA* 247 (18) (1982) 2543–2546.
- [28] Z. Zhang, C. Zheng, C. Kim, S. Van Poucke, S. Lin, P. Lan, Causal mediation analysis in the context of clinical research, *Ann Transl Med* 4 (21) (2016) 425.
- [29] GTEx Consortium, The Genotype-Tissue Expression (GTEx) project, *Nat Genet* 45 (6) (2013) 580–585.
- [30] D. Szklarczyk, A.L. Gable, K.C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N.T. Doncheva, et al., The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets, *Nucleic Acids Res* 49 (D1) (2021) D605–D612.
- [31] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res* 13 (11) (2003) 2498–2504.
- [32] G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, et al., The GeneCards suite: from gene data mining to disease genome sequence analyses, *Curr Protoc Bioinformatics* 54 (2016) 1301–1313.
- [33] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res* 28 (1) (2000) 27–30.
- [34] M. Gillespie, B. Jassal, R. Stephan, M. Milacic, K. Rothfels, A. Senff-Ribeiro, et al., The reactome pathway knowledgebase 2022, *Nucleic Acids Res* 50 (D1) (2022) D687–D692.
- [35] G.C. Curhan, W.C. Willett, F.E. Speizer, M.J. Stampfer, Intake of vitamins B6 and C and the risk of kidney stones in women, *J Am Soc Nephrol* 10 (4) (1999) 840–845.
- [36] G.C. Curhan, W.C. Willett, E.B. Rimm, M.J. Stampfer, A prospective study of the intake of vitamins C and B6, and the risk of kidney stones in men, *J Urol* 155 (6) (1996) 1847–1851.
- [37] P.M. Ferraro, E.N. Taylor, G. Gambaro, G.C. Curhan, Vitamin B6 intake and the risk of incident kidney stones, *Urolithiasis* 46 (3) (2018) 265–270.
- [38] G.C. Curhan, E.N. Taylor, 24-h uric acid excretion and the risk of kidney stones, *Kidney Int* 73 (4) (2008) 489–496.
- [39] S.P. Stabler, Clinical practice. Vitamin B12 deficiency, *N Engl J Med* 368 (2) (2013) 149–160.
- [40] R. Green, Vitamin B12 deficiency from the perspective of a practicing hematologist, *Blood* 129 (19) (2017) 2603–2611.
- [41] N. Sukumar, P. Saravanan, Investigating vitamin B12 deficiency, *BMJ* 365 (2019) 11865.
- [42] A.K. Elshorbagy, E. Nurk, C.G. Gjesdal, G.S. Tell, P.M. Ueland, O. Nygård, et al., Homocysteine, cysteine, and body composition in the Hordaland Homocysteine Study: does cysteine link amino acid and lipid metabolism? *Am J Clin Nutr* 88 (3) (2008) 738–746.
- [43] B. Vanizor Kural, A. Orem, G. Cimşit, H.A. Uydu, Y.E. Yandi, A. Alver, Plasma homocysteine and its relationships with atherothrombotic markers in psoriatic patients, *Clin Chim Acta* 332 (1-2) (2003) 23–30.
- [44] J.V. Zhao, C.M. Schooling, Homocysteine-reducing B vitamins and ischemic heart disease: a separate-sample Mendelian randomization analysis, *Eur J Clin Nutr* 71 (2) (2017) 267–273.
- [45] P.J. Barter, K.A. Rye, Homocysteine and cardiovascular disease: is HDL the link? *Circ Res* 99 (6) (2006) 565–566.
- [46] A.M. Devlin, S.R. Lentz, ApoA-I: a missing link between homocysteine and lipid metabolism? *Circ Res* 98 (4) (2006) 431–433.
- [47] S. Park, S. Lee, Y. Kim, S. Cho, K. Kim, Y.C. Kim, et al., Causal effects of homocysteine, folate, and cobalamin on kidney function: a Mendelian randomization study, *Nutrients* 13 (3) (2021) 906.
- [48] K. Kalantar-Zadeh, T.H. Jafar, D. Nitsch, B.L. Neuen, V. Perkovic, Chronic kidney disease, *Lancet* 398 (10302) (2021) 786–802.
- [49] H.J. Blom, G.M. Shaw, M. den Heijer, R.H. Finnell, Neural tube defects and folate: case far from closed, *Nat Rev Neurosci* 7 (9) (2006) 724–731.
- [50] K.S. Crider, O. Devine, L. Hao, N.F. Dowling, S. Li, A.M. Molloy, et al., Population red blood cell folate concentrations for prevention of neural tube defects: Bayesian model, *BMJ* 349 (2014) g4554.
- [51] Dietary reference intakes for thiamin, riboflavin, niacin, vitamin B6, folate, vitamin B12, pantothenic acid, biotin, and choline, 1998 The National Academies Collection.
- [52] G.S. Garrett, L.B. Bailey, A public health approach for preventing neural tube defects: folic acid fortification and beyond, *Ann N Y Acad Sci* 1414 (1) (2018) 47–58.