Folate Network Genetic Variation Predicts Cardiovascular Disease Risk in Non-Hispanic White Males


4Division of Nutritional Sciences, 5Department of Molecular Biology and Genetics, and 6Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 7Channing Laboratory, Brigham and Women’s Hospital, Boston, MA; 8Harvard Medical School, Boston, MA; 9VA Normative Aging Study, VA Boston Healthcare System, Boston, MA; 10Division of Aging, Brigham and Women’s Hospital, Boston, MA; 11Department of Medicine, Boston University School of Medicine, Boston MA; and 12School of Health Professions, Northeastern University, Boston, MA

Abstract

Genes functioning in folate-mediated 1-carbon metabolism are hypothesized to play a role in cardiovascular disease (CVD) risk beyond the current narrow focus on the MTHFR 677 C→T (rs1801133) polymorphism. Using a cohort study design, we investigated whether sequence variants in the network of folate-related genes, particularly in genes encoding proteins related to SHMT1, predict CVD risk in 1131 men from the Normative Aging Study. A total of 330 single nucleotide polymorphisms (SNPs) in 52 genes, selected for function and gene coverage, were assayed on the Illumina GoldenGate platform. Age- and smoking-adjusted genotype-phenotype associations were estimated in regression models. Using a nominal \( P = 5.00 \times 10^{-3} \) significance threshold, 8 SNPs were associated with CVD risk in single locus analyses. Using a false discovery rate (FDR) threshold \( \text{P\_adjusted} \leq 1.00 \times 10^{-3} \), a SNP in the GGH gene remained associated with reduced CVD risk, with a stronger association in early onset CVD cases \(<55\) y). A gene \( \times \) folate interaction \((\text{MAT2B} \text{ and } G \text{ and } B \text{ and } \text{SLC25A32})\) reached the FDR \( \text{P\_adjusted} \leq 2.00 \times 10^{-1} \) threshold. Three biological hypotheses related to SHMT1 were explored and significant gene \( \times \) gene interactions were identified for TYMS by UBE2N, FTH1 by CELF1, and TYMS by MTHFR. Variations in genes other than MTHFR and those directly involved in homocysteine metabolism are associated with CVD risk in non-Hispanic white males. This work supports a role for SHMT1-related genes and nuclear folate metabolism, including the thymidylate biosynthesis pathway, in mediating CVD risk. J. Nutr. 142: 1272–1279, 2012.

Introduction

Heart disease and stroke were responsible for about one-third of U.S. deaths in 2004 (1). There is extensive evidence for the contribution of dietary and lifestyle factors to risk; however, the origins of cardiovascular disease (CVD) are complex and the interactive effects of genetic and environmental factors, including nutrition, play an important role.

Folate and other B vitamins contribute to biological processes important to health, including DNA synthesis and repair, and the generation of cellular methylation potential for a variety of methylation reactions. Folate status is influenced by dietary intake and variation in the genes encoding folate-related enzymes; altered folate status is associated with adverse outcomes, including birth defects, CVD, and cancer (2). Elevated plasma homocysteine, a sulfur-containing amino acid, is a marker of disturbed folate-mediated 1-carbon metabolism, and is associated with an increased risk of CVD (3–6). Homocysteine concentrations are modulated by nutrition, particularly folate and vitamin B-12 (7), and by genetic variants, including a well-studied single nucleotide polymorphism (SNP) in the methyl-
enetetrahydrofolate reductase gene (MTHFR 677 C→T; rs1801133) (8). Thus, the association of sequence variation in the 1-carbon network with CVD may be mediated through altered homocysteine metabolism, cellular methylation potential, and/or DNA synthesis.

Although several candidate genes contributing to folate-mediated, 1-carbon metabolism have been extensively studied in relation to CVD risk, most notably the MTHFR rs1801133 SNP, other genes less proximal to homocysteine metabolism are understudied. A recent publication examined the association of genetic variation in the 1-carbon network and related nutritional factors with biological markers of CVD risk [plasma homocysteine and global genomic DNA methylation (9)], and a similar evaluation of CVD is needed. A more complete understanding of the role of the 1-carbon network in mediating CVD risk would support the development of novel prevention strategies. This study investigated 330 SNPs in 52 genes that play a role in folate-mediated, 1-carbon metabolism. The set of genes, SNPs, and biomarkers examined in this study was purposefully selected to represent the full functional variation of the folate-mediated, 1-carbon metabolic network and prospectively evaluated for association with CVD in a U.S. population studied prior to the initiation of mandatory folate fortification.

**Participants and Methods**

**Study population.** The longitudinal Normative Aging Study (NAS) was established by the Veterans Administration in 1961 and men were followed-up at 3- to 5-y intervals through 1998 (end date for this study); 2280 men aged 21–81 y (mean age 42 y at study entry) were enrolled based on health criteria (10).

**DNA extraction, SNP selection, and genotyping.** DNA extraction, SNP selection, and genotyping were carried out as described elsewhere (9). Fifty-two genes contributing to folate-mediated, 1-carbon metabolism were identified (Supplemental Table 1). SNP selection encompassed 2 kb on either side of the gene to include promoter and/or regulatory region variants. SNPs within a gene were selected to minimize redundancy (9). No assumptions were made about a given SNP representing gain or loss of function and SNPs with a minor allele frequency (MAF) ≥5% in European-ancestry populations were selected where possible, although exceptions were made for SNPs with prior evidence of putative function or when no SNPs with a MAF ≥5% were available [details presented elsewhere (9)]. A total of 384 SNPs were submitted to the Center for Inherited Disease Research at Johns Hopkins University for genotyping via an Illumina GoldenGate custom panel; 54 SNPs were ultimately excluded (reasons included assay failure, monomorphic genotype data, MAF <1%, and genotype frequencies out of Hardy-Weinberg equilibrium), leaving 330 SNPs for analysis (Supplemental Table 2). SNP selection was completed in January 2008. A total of 1304 participant samples were submitted for genotyping: 54.4% were genomic and 45.5% were whole-genome amplified. This study included the 1131 successfully genotyped, non-Hispanic, white males.

**Covariates.** Extensive previously collected data on NAS participants included physical measurements, lifestyle factors, and blood biomarkers; the methods for covariate measurement were previously described (11–14). Since enrollment, participants have had clinical examinations at 3- to 5-y intervals, with a response rate of >90% for mailed questionnaires. Plasma biomarkers were assayed in an unselected subset of stored blood samples; CV were excellent [folate, 4.3%; vitamin B-6 (pyridoxal-5’-phosphate), 5.0%; vitamin B-12, 4.7%; total homocysteine, 4.0%] (11).

Up to 3 measurements of homocysteine, folate, vitamin B-6, and vitamin B-12 were available prior to December 31, 1998, the end of follow-up, which coincided approximately with the start of mandatory folate fortification in the US. The intracllass correlation coefficients for log-transformed homocysteine, folate, vitamin B-6, and vitamin B-12 confirmed that a high proportion of the total variance was between- (vs. within-) person; thus, the first available biomarker measurement was used to best represent nutrient status in the pre-fortification era and to maximize case number, because only cases with nutrients measured prior to disease onset were included. The biomarkers were frequently measured together and the date range corresponding to the first available measure of each biomarker was April 1993 to December 1998. All biomarkers were log-transformed to adjust for skewness and centered for best performance in regression models.

**Phenotype assessment.** In this longitudinal cohort study, follow-up time for study participants was estimated as time from study baseline to first occurrence of a CVD event (angina; fatal and nonfatal MI or other coronary heart disease; fatal and nonfatal stroke) or death from a competing cause. For all other men, follow-up time was time from baseline to December 31, 1998, the end of follow-up for all study analyses.

The ascertainment of CVD events within the NAS is described elsewhere (13,15). Briefly, the possible occurrence of nonfatal MI and angina was ascertained by participant report and confirmed by a cardiologist. Similarly, the occurrence of a nonfatal stroke was ascertained by participant report and confirmed by a neurologist. Fatal myocardial infarction and fatal stroke in men with no prior history of CVD were ascertained through death certificates, which were reviewed by a cardiologist who assigned cause of death using International Classification of Disease codes. CVD events comprised the following diagnosis codes (International Classification of Disease, version 8): 410–414.9 (heart disease including angina and myocardial infarction), and 436.9, 430, 431 (stroke). Ascertainment of fatal CVD events was nearly 100% (13,16). In all instances, the confirming physician was unaware of the participant’s exposure status (genotype). From 1964 to 1998, 381 cases of fatal or nonfatal CVD occurred among the 1131 men with successful Illumina genotyping (31,871 person-years of follow-up).

**Statistical analyses.** Time-to-event analysis evaluated the relation of SNP markers to CVD risk by using a Cox proportional hazards model. Genetic models of inheritance (additive, dominant, recessive, and over-dominant) were tested for each SNP and the model most significantly associated with CVD risk in the unadjusted single SNP analysis was chosen as the best model going forward. SNP main effects were captured in a single term; significance of the SNP was assessed by the P value for the SNP regression coefficient. Although the first selection step affects the P value distribution going forward, no additional multiple testing corrections were applied beyond the false discovery rate (FDR) given that the P values from the various models were not independent (e.g., between additive and dominant models) and because the work was hypothesis oriented and seeking to nominate candidate genes for further investigation. For the MTHFR rs1801133 SNP, a dummy-coded scheme was used for consistency to past work, which almost universally evaluates the CT and TT genotypes separately against the CC reference category (6,15,17), and the likelihood ratio test (LRT) P value was used to assess significance of the set of dummy variables. The potentially confounding factors considered included age, smoking, diabetes, hypertension, obesity, alcohol consumption, and education; however, only age, smoking, diabetes, and hypertension were associated with CVD and thus these variables were included in further models. All regression models were adjusted for age and smoking status (baseline smoking status and change in smoking status over follow-up) and sensitivity analyses were performed to evaluate the SNP-CVD association adjusted for diabetes and hypertension comorbidities. Extended models adjusted for biomarker residuals (i.e., variation in nutrient biomarker not predicted by SNP) and the MTHFR rs1801133 variant (using recessive coding to account for the pattern of association using the fewest model terms). Previous work in this cohort demonstrated no population substructure (18); thus, no adjustments were made.

To investigate whether nutrient status modified the genotype-phenotype association, gene × nutrient interactions were evaluated. Product terms representing the interaction of each genotype with nutrient residuals (continuous variable) were included in models; significance was assessed by the P value for the interaction term. Where specific genetic models of inheritance led to sparse data for genotype × nutrient interaction, the analysis was repeated assessing nutrient remnants (the difference between the nutrient adjusted and the raw nutrient) instead of nutrient residuals as the predictor. These models were fit using the l1-lasso 0.1 penalty method to give the best results with an empirical Bayes model averaging approach (19).
nutrient interaction models, additive coding was used as the default. For the purposes of illustration, SNP-nutrient interactions are presented at 3 values of the centered, log-transformed nutrients: the 10th percentile (low), the 50th percentile (median), and the 90th percentile (high). To assess the joint role of multiple genetic variants, gene × gene interactions were tested using model-free dummy variable coding; significance was assessed using the LRT $P$ value for models with vs. models without interaction terms; sparse data prevented the evaluation of some interactions. Interactions affected by sparse data [<5 individuals in ≥1 cell of the SNP by SNP frequency table or improbably large regression estimates (likely indicates sparse data)] are presented in the online supplement only.

To account for multiple comparisons, the FDR multiple testing correction of Benjamini and Hochberg (19) was implemented (SAS, the online supplement only. Large regression estimates (likely indicates sparse data) are presented in Table 1).

For SNP main effects, regression coefficients with a nominal $P \leq 5.00 \times 10^{-3}$ were reported and a conditional FDR given the prior step identifying the best genetic model. In assessing interactions, a less stringent significance threshold was used: regression coefficients with a nominal $P \leq 2.00 \times 10^{-2}$ were reported. However, for the network-wide gene × gene interaction analysis, a nominal $P \leq 1.00 \times 10^{-3}$ threshold was used to limit the number of results reported.

All statistical analyses were conducted in SAS v. 9.2 unless otherwise specified. The study was approved by the Brigham and Women’s Hospital Human Subjects committee, the Veterans Administration Research and Development Committee, and the Cornell University Committee on Human Subjects.

Results

Among 1131 men with genotype data, 381 CVD cases occurred: 14% myocardial infarction only, 17% angina only, 4% other coronary heart disease only, and 8% stroke only. The remainder had multiple diagnoses. A subset of 735 men had measurements for plasma nutrient biomarkers (folate, vitamins B-6 and B-12, and homocysteine). Among the study participants, the $MTHFR$ rs1801133 TT genotype frequency was 12.8%, representing >2200 chromosomes (Table 1); this frequency is higher than a sample of 120 chromosomes from the HapMap CEPH population (6.7% TT) but similar to a sample of 5064 chromosomes from North American coronary heart disease controls in a meta-analysis (12.2% TT) (8). In exploratory regression models, prior to assessing the genotype-phenotype associations, baseline age was associated with CVD risk ($P \leq 1.00 \times 10^{-4}$), as were baseline smoking status and change in smoking status over follow-up ($P \leq 2.00 \times 10^{-2}$ for each term). When modeled individually, plasma folate ($P = 9.00 \times 10^{-1}$) and vitamin B-12 ($P = 7.00 \times 10^{-1}$) had little or no association with CVD, although vitamin B-6 ($P = 7.70 \times 10^{-2}$) and homocysteine ($P = 5.40 \times 10^{-2}$) had some association in unadjusted models; the associations were attenuated in age- and smoking-adjusted models.

Single locus marker associations. The 8 SNPs with nominal $P \leq 5.00 \times 10^{-3}$ for the CVD phenotype (Fig. 1) were generally common variants with a MAF ≥13% (Table 2). γ-Glutamyl hydrodase (GGH) rs12544405, a 5’ region marker, had a nominal $P = 5.50 \times 10^{-4}$, $P$-adjusted $= 9.10 \times 10^{-2}$. Compared with the AA/GG genotype group, the AG genotype was associated with a 32% reduction in CVD risk [OR = 0.68 (95% CI: 0.5, 0.8)]. A SNP in 5, 10-methylenetetrahydrofolate synthetase ($MTHFS$), rs7177659, was associated with reduced CVD risk; men in the AC genotype group had a 29% lower risk compared with the AA/CC genotype group [OR = 0.71 (95% CI: 0.6, 0.9)]. The well-known $MTHFR$ rs1801133 SNP was associated with risk: CT (vs. CC) was associated with a 6% increase in CVD risk [OR = 1.06 (95% CI: 0.8, 1.3)], and TT (vs. CC) was associated with a 71% increase in CVD risk [OR = 1.71 (95% CI: 1.3, 2.3)]. Further adjusting for diabetes or hypertension (baseline status and change over follow-up) made little difference overall and identified the same set of SNPs, although regression coefficients and $P$ values showed minor changes.

When analyses were restricted to the subset of men with biomarker data ($n = 735$), the set of most significant SNPs, and

### Table 1

**Descriptive characteristics of NAS participants at study baseline**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>40.2 ± 7.8</td>
</tr>
<tr>
<td>Education, % college graduate or higher</td>
<td>28.1</td>
</tr>
<tr>
<td>White, %</td>
<td>100</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>25.9 ± 2.9</td>
</tr>
<tr>
<td>Obese, % BMI ≥30 kg/m²</td>
<td>8.5</td>
</tr>
<tr>
<td>Cigarette smoking, %</td>
<td>30.1</td>
</tr>
<tr>
<td>Never</td>
<td>37.5</td>
</tr>
<tr>
<td>Former</td>
<td>32.4</td>
</tr>
<tr>
<td>Alcohol intake, % ≥2 drinks/d</td>
<td>13.4</td>
</tr>
<tr>
<td>Serum fasting glucose, mmol/L</td>
<td>5.5 ± 0.6</td>
</tr>
<tr>
<td>Diabetes diagnosis, %</td>
<td>0.18</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>123 ± 12.7</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>76.6 ± 8.6</td>
</tr>
<tr>
<td>Hypertension diagnosis, %</td>
<td>2.5</td>
</tr>
<tr>
<td>Serum total cholesterol, mmol/L</td>
<td>5.3 ± 1.1</td>
</tr>
<tr>
<td>$MTHFR$ 677 C→T (rs1801133) TT genotype, %</td>
<td>12.8</td>
</tr>
<tr>
<td>Plasma folate, μmol/l</td>
<td>23.6 ± 12.9</td>
</tr>
<tr>
<td>Plasma vitamin B-6, μmol/l</td>
<td>84.8 ± 85.0</td>
</tr>
<tr>
<td>Plasma vitamin B-12, μmol/l</td>
<td>339 ± 140</td>
</tr>
<tr>
<td>Plasma homocysteine, μmol/l</td>
<td>10.6 ± 3.7</td>
</tr>
<tr>
<td>Mean follow-up time, y</td>
<td>28.2 ± 6.7</td>
</tr>
</tbody>
</table>

1 Values are percent or mean ± SD. NAS, Normative Aging Study.
2 Plasma measures of folate, vitamin B-6, vitamin B-12, and homocysteine were collected prior to the initiation of folate fortification.

### Figure 1

Manhattan plot: folate-related SNPs as predictors of CVD risk in men from the NAS. Models adjusted for age and smoking status. The box indicates SNP that reached the FDR significance threshold. Eight SNPs are shown above the dotted line, including 2 SNPs on chromosomes 16 and 21. No SNPs were located on chromosomes 4 or 13 or the sex chromosomes. CVD, cardiovascular disease; FDR, false discovery rate; NAS, Normative Aging Study; SNP, single nucleotide polymorphism.
were consistent in direction of effect for early onset (P = 2.00 \times 10^{-2}) (Supplemental Table 5). Compared with the C/T/T genotype group, men with the CC genotype had 91% lower, 42% lower, and 7-fold greater CVD risk at high (39.1 nmol/L), median (21.8 nmol/L), and low (9.9 nmol/L) plasma folate concentrations, respectively. Two vitamin B-12 interaction terms were significant at FDR thresholds: for the rs585800 SNP in betaine-homocysteine S-methyltransferase (BHMT), compared with the AA/TT genotype group, men with the TA genotype had 36% lower, 19% higher, and 2.2-fold higher CVD risk at high (501 pmol/L), median (309 pmol/L), and low (191 pmol/L) plasma vitamin B-12 concentrations, respectively. For the rs1061196 SNP in solute carrier family 25, member 32 (SLC25A32), compared with the AA/CC genotype group, men with the AG genotype had 1.6-fold higher, 17% lower, and 55% lower CVD risk at high, median, and low plasma B-12 concentrations, respectively. No vitamin B-6 interaction terms reached the FDR significance level of P-adjusted = 2.00 \times 10^{-2}.

**Gene-gene interactions.** In models investigating interactions between MTHFR rs1801133 and each SNP marker, 6 interactions reached the nominal level of P = 2.00 \times 10^{-2}; no interactions were significant at the a priori FDR significance threshold of P-adjusted = 2.00 \times 10^{-2} (Supplemental Table 6).

Given the interconnectedness of folate-mediated, 1-carbon metabolism, gene-gene interactions within the network were hypothesized to predict CVD risk. Thus, an analysis of all pairwise gene-gene interactions within the network was conducted. Of 52,832 unique pairwise interactions tested, 55 were associated with CVD risk at P = 1.00 \times 10^{-3}; no interactions were significant at an FDR P-adjusted = 2.00 \times 10^{-1} (Supplemental Table 7). The interaction between CELF1 (formerly CUGBP1) rs7933019 and MTR rs16834388 reached nominal P = 1.20 \times 10^{-4}. In men with the CELF1 rs7933019 GG genotype, MTR rs16834388 GT and TT genotypes were associated with an increased risk of CVD (by 30 and 73%, respectively, vs. MTR rs16834388 GG genotype), whereas in men with the CELF1 rs7933019 CC genotype, the MTR GT and TT genotypes were inversely associated with CVD risk (Table 3).

Several hypotheses arising from animal and cell studies and prior population-level studies were investigated. Prior studies (15,20) reported a gene \times gene interaction for MTHFR by

### Table 2: SNPs associated with CVD risk in men from the NAS at nominal P ≤ 5.00 \times 10^{-3,1-3}

<table>
<thead>
<tr>
<th>Gene</th>
<th>rs number</th>
<th>( \beta )</th>
<th>SE</th>
<th>OR (95% CI)</th>
<th>Nominal ( P )</th>
<th>Chr</th>
<th>Coded allele frequency</th>
<th>Coded allele frequency</th>
<th>Genetic model (number)</th>
<th>SNP type</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGH</td>
<td>rs12544045</td>
<td>−0.38</td>
<td>0.11</td>
<td>0.68 (0.6, 0.9)</td>
<td>5.51 \times 10^{-4}</td>
<td>8</td>
<td>A</td>
<td>29</td>
<td>O (AG: 464, AA + GG: 660)</td>
<td>5</td>
</tr>
<tr>
<td>MTHFS</td>
<td>rs7177659</td>
<td>−0.34</td>
<td>0.11</td>
<td>0.71 (0.6, 0.9)</td>
<td>1.18 \times 10^{-3}</td>
<td>15</td>
<td>A</td>
<td>47</td>
<td>O (AC: 457, AA + CC: 556)</td>
<td>1</td>
</tr>
<tr>
<td>MTHFR</td>
<td>rs1801133</td>
<td>0.06</td>
<td>0.12</td>
<td>1.06 (0.8, 1.3)</td>
<td>1.48 \times 10^{-3}</td>
<td>1</td>
<td>T</td>
<td>35</td>
<td>(CC: 482, CT: 502, TT: 145)</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>rs1801133 TT</td>
<td>0.54</td>
<td>0.15</td>
<td>1.71 (1.3, 2.3)</td>
<td></td>
<td>11</td>
<td>C</td>
<td>5</td>
<td>(CT: 103, CC + TT: 1027)</td>
<td>3</td>
</tr>
<tr>
<td>TCN1</td>
<td>rs17154234</td>
<td>0.49</td>
<td>0.16</td>
<td>1.62 (1.2, 2.2)</td>
<td>2.40 \times 10^{-3}</td>
<td>11</td>
<td>C</td>
<td>5</td>
<td>(CT: 70, GG + TT: 858)</td>
<td>1</td>
</tr>
<tr>
<td>UBEZI</td>
<td>rs11248868</td>
<td>0.34</td>
<td>0.11</td>
<td>1.40 (1.1, 1.8)</td>
<td>3.33 \times 10^{-3}</td>
<td>16</td>
<td>G</td>
<td>13</td>
<td>(GGH: 39, CTG: 102, TTG: 31)</td>
<td>1</td>
</tr>
<tr>
<td>CBS</td>
<td>rs6858282</td>
<td>−0.32</td>
<td>0.11</td>
<td>0.73 (0.6, 0.9)</td>
<td>3.35 \times 10^{-3}</td>
<td>21</td>
<td>T</td>
<td>18</td>
<td>A (CC: 759, CT: 335, TT: 31)</td>
<td>1</td>
</tr>
<tr>
<td>CBS</td>
<td>rs6858281</td>
<td>−0.35</td>
<td>0.12</td>
<td>0.71 (0.6, 0.9)</td>
<td>3.55 \times 10^{-3}</td>
<td>21</td>
<td>A</td>
<td>18</td>
<td>(OGM: AA + G: 565, GG: 756)</td>
<td>1</td>
</tr>
<tr>
<td>UBEZI</td>
<td>rs909915</td>
<td>0.33</td>
<td>0.11</td>
<td>1.39 (1.1, 1.7)</td>
<td>4.02 \times 10^{-3}</td>
<td>16</td>
<td>T</td>
<td>13</td>
<td>(CTC: 268, TT + CC: 860)</td>
<td>1</td>
</tr>
</tbody>
</table>

1. Model adjusted for age and smoking, forward strand allele shown. A, additive; N, coding nonsynonomous; Chr, chromosome; CVD, cardiovascular disease; DOM, dominant; FDR, false discovery rate, \( i \), inter; LRT, likelihood ratio test; NAS, Normative Aging Study; O, overdominant; SNP, single nucleotide polymorphism.
2. No SNPs map to more than one gene.
3. Sensitivity analyses considered only those men with complete follow-up data (i.e., never missed a scheduled NAS follow-up visit by >1 y), and regression coefficients and SE were similar to models including all participants. When the outcome was limited to early-onset CVD (onset age <55 y), associations for some of the SNPs were strengthened, whereas other associations were attenuated (Supplemental Table 4). The GGH rs12544045 SNP association with CVD increased by 215% in the analysis limited to early-onset cases (nominal P = 2.90 \times 10^{-4}, P-adjusted = 9.50 \times 10^{-2}). Compared with the AA/CC genotype group, the AG genotype was associated with a 70% reduction in CVD risk for early onset cases (OR = 0.3 (95% CI: 0.2, 0.6)). MTHFR rs1801133 associations were consistent in direction of effect for early onset cases (P = 3.40 \times 10^{-1}) and all CVD (P = 1.50 \times 10^{-3}). Effect sizes for markers in the UBE21 gene, which encodes a small ubiquitin-like modifier (SUMO)-conjugating enzyme (rs11248868, rs909915, and rs9926094), were approximately doubled in the early onset CVD analysis.

**Gene-nutrient interactions.** Plasma folate and vitamins B-6 and B-12 are cofactors for enzymes in the 1-carbon metabolic network and may modify the SNP-phenotype association. In further analyses assessing SNP by nutrient interactions, one folate interaction term, with the rs6882306 variant in the methionine adenosyltransferase II, beta (MAT2B) gene, reached FDR significance levels (P-adjusted = 2.00 \times 10^{-1}) (Supplemental Table 5).
SHMT1; the interaction was replicated here and mediation by folate, vitamins B-6 and B-12, and homocysteine was explored. Adjusting for homocysteine strengthened the interaction (LRT $P = 1.00 \times 10^{-2}$). Thus, in men with the SHMT1 rs1979277 TT genotype, the MTHFR rs1801133 CT genotype was associated with a 2.4-fold increased risk of CVD (vs. the rs1801133 CC genotype) and the MTHFR rs1801133 CT genotype was associated with a 10.4-fold increased risk of CVD (vs. the rs1801133 CC genotype) in models adjusted for homocysteine. Similarly, in men with the SHMT1 rs1979277 TT genotype, the MTHFR rs1801133 TT genotype (vs. rs1801133 CC genotype) was associated with a 3-fold increased risk of CVD without adjusting for homocysteine and a 15-fold increased risk of CVD with adjustment for homocysteine. There was little or no association of the MTHFR genotype with CVD risk in the subgroup of men with the SHMT1 rs1979277 CT/CC genotype either with or without adjusting for homocysteine.

The SHMT1 enzyme governs the competing flow of 1-carbon units to thymidylate biosynthesis or remethylation reactions (21); thus, modifications in protein expression or subcellular location may affect CVD risk by influencing the distribution of 1-carbon units to competing reactions. Three hypotheses related to SHMT1 were investigated. First, a nuclear 1-carbon metabolism hypothesis was evaluated by considering all pairwise interactions among SNPs in the SHMT1, TYMS, UBE2I, and UBE2N genes in relation to CVD risk. Of 451 unique pairwise interactions tested, 8 reached nominal ($P \leq 2.00 \times 10^{-2}$) threshold (Supplemental Table 8). The most significant finding was the interaction of UBE2N rs7300607 and TYMS rs502396 (Table 4); among men with the TYMS rs502396 CC genotype, the UBE2N rs7300607 CC genotype increased CVD risk 3-fold (vs. rs7300607 TT reference group). Second, an SHMT1 expression hypothesis (22) was explored and all pairwise interactions among SNPs in the FTHI, CELFI, and SHMT1 genes were investigated in relation to CVD risk. Of 92 unique pairwise interactions tested, 12 reached nominal ($P \leq 2.00 \times 10^{-2}$) (Supplemental Table 9). The interaction of FTHI rs17156609 and CELFI rs7933019 (Table 4) was significant; among men with the FTHI rs17156609 GA genotype, the CELFI rs7933019 CC genotype increased CVD risk 3-fold (vs. rs7933019 GG reference group). One category, FTHI rs17156609 AA / CELFI rs7933019 CC, could not be evaluated due to sparse data. Finally, because mitochondrial folate metabolism produces formate, a major source of cytoplasmic 1-carbon units, (2), interactions may exist among MTHFDD1, SHMT1, [influences 1-carbon unit distribution among the remethylation and thymidylate synthesis pathways (21)], and 2 downstream enzymes in each pathway, MTHFR and TYMS (the mitochondrial 1-carbon metabolism hypothesis). Of 377 unique pairwise interactions tested, 2 reached the nominal $P \leq 2.00 \times 10^{-2}$ threshold (Supplemental Table 10). The interaction of TYMS rs502396 and MTHFR rs12121543 (Table 4) exemplifies the finding that in men with the TYMS rs502396 CC genotype, the MTHFR rs12121543 AA genotype increased CVD risk 2-fold (vs. rs12121543 CC reference group).

A landmark GWAS, the Wellcome Trust Case Control Consortium, previously identified an intronic SNP (rs6922269) in MTHFDD1, a mitochondrial folate metabolism gene, which was associated with coronary artery disease risk (23). Based on HapMap linkage plots, this SNP is located in a tightly linked region HapMap linkage plots, this SNP is located in a tightly linked region.

### Table 4

Investigation of 3 hypotheses for folate-related genes in relation to risk in men from the NAS.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Gene</th>
<th>Genotype</th>
<th>OR (95% CI)</th>
<th>n</th>
<th>OR (95% CI)</th>
<th>n</th>
<th>OR (95% CI)</th>
<th>n</th>
<th>LRT $P^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear 1-carbon metabolism</td>
<td>TYMS rs502396</td>
<td>TT</td>
<td>1.00 (0.6, 1.6)</td>
<td>300</td>
<td>0.59 (0.3, 0.9)</td>
<td>200</td>
<td>0.80 (0.5, 1.3)</td>
<td>100</td>
<td>3.00 x 10^{-2}</td>
</tr>
<tr>
<td></td>
<td>UBE2N rs7300607</td>
<td>TT vs. GG</td>
<td>1.00 (0.8, 1.2)</td>
<td>200</td>
<td>0.80 (0.6, 1.3)</td>
<td>100</td>
<td>1.20 (0.9, 1.6)</td>
<td>50</td>
<td>7.00 x 10^{-3}</td>
</tr>
<tr>
<td>SHMT1 expression</td>
<td>FTHI rs17156609</td>
<td>GG</td>
<td>1.00 (0.8, 1.2)</td>
<td>200</td>
<td>0.80 (0.6, 1.3)</td>
<td>100</td>
<td>1.20 (0.9, 1.6)</td>
<td>50</td>
<td>7.00 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>CELFI rs7933019</td>
<td>GG vs. CT</td>
<td>1.00 (0.8, 1.2)</td>
<td>200</td>
<td>0.80 (0.6, 1.3)</td>
<td>100</td>
<td>1.20 (0.9, 1.6)</td>
<td>50</td>
<td>7.00 x 10^{-3}</td>
</tr>
<tr>
<td>Mitochondrial 1-carbon metabolism</td>
<td>TYMS rs502396</td>
<td>GG</td>
<td>1.00 (0.8, 1.2)</td>
<td>200</td>
<td>0.80 (0.6, 1.3)</td>
<td>100</td>
<td>1.20 (0.9, 1.6)</td>
<td>50</td>
<td>7.00 x 10^{-3}</td>
</tr>
</tbody>
</table>

1. Model adjusted for age and smoking; forward strand allele shown. CVD, cardiovascular disease; FDR, false discovery rate; LRT, likelihood ratio test; NAS, Normative Aging Study; SNP, single nucleotide polymorphism.
2. No lower quality SNPs.
3. P value represents LRT for models with vs. without interaction terms.
4. SNP maps to more than one gene: rs17156609 also maps to bestrophin 1 (BEST1).
5. The FTHI rs17156609 AA genotype did not exist in this dataset and therefore the associated OR are not estimated.
14-kb block in the MTHFD1L gene. Although the intronic rs6922269 SNP, along with the nearby nonsynonymous rs9767752 SNP, were not assayed (both failed in the genotyping), 6 other tag SNPs successfully cover this region. Joint genotypes involving 2 SNPs within the 14-kb region (rs1474787 and rs803447) best captured the association with CVD risk. In the rs1474787 GA genotype group, men with the rs803447 TT genotype had a 3.6-fold increased CVD risk compared with the rs803447 CC reference group (P-interaction = 0.077). One category, the rs1474787 AA / rs803447 TT genotype group, could not be evaluated due to sparse data. In models adjusted for homocysteine, the interaction was strengthened; thus, in the rs1474787 GA genotype group, men with the rs803447 TT genotype had a 6-fold increased risk of CVD compared with the rs803447 CC reference group (P-interaction = 0.077) (data not shown).

Discussion

We investigated sequence variation in a network of candidate genes involved in 1-carbon metabolism in relation to CVD risk and several specific hypotheses based on animal and cell studies were confirmed in this population-level study. Overall, there was little evidence that plasma nutrients or homocysteine mediated the SNP-CVD risk associations. Among the genes considered, an FDR-significant association was identified for a GGH 5’ region SNP; the rs12544045 AG genotype was associated with a 32% reduction in CVD risk. The GGH enzyme facilitates cellular folate export and previous work identified polymorphisms in the GGH 5’ region that were associated with increased promoter activity (24) and DNA uracil content (25). After plasma folate adjustment, the GGH-CVD association decreased only slightly, suggesting that plasma folate does not fully mediate the association of GGH rs12544045 with CVD or that our understanding of the role of GGH in contributing to CVD risk is incomplete. When restricted to early onset cases (onset age <55 y), main effects for GGH rs12544045 were stronger, consistent with a true effect on risk. In the MTHFS gene, the rs7177659 AC genotype was associated with a 29% reduction in CVD risk, although the functional significance of the variant allele is not known. MTHFS is an essential enzyme in mice that regulates de novo purine biosynthesis and catalyzes the conversion of 5-formyltetrahydrofolate to methenyltetrahydrofolate (26). Increased expression of MTHFS enhances purine biosynthesis (27,28), potentially by delivering the substrate 10-formyltetrahydrofolate to the purinosine, a multienzyme complex (29).

Focusing on the SHMT1 intersection in the folate network, several specific hypotheses based on animal and cell studies were confirmed in this population-level study. In vitro studies demonstrated protein interactions between SHMT1 and UBC9 (yeast homolog of human UBE2I), resulting in the SUMOylation and nuclear localization of SHMT1, along with the other 2 components of thymidylate synthesis, TYMS and DHFR (30–32). UBC13 (yeast homolog of human UBE2N) is also a SHMT1 binding partner, suggesting that SHMT1 may be a target for ubiquitination (30). We could not directly assess SHMT1 by UBE2I interactions due to sparse data. However, UBE2N by TYMS interaction was nominally significant; the protein products of these 2 genes physically interact during the sumoylation of TYMS. Other studies suggest that CELF1 (formerly CUGBP1) and FTH1 interact with the 3′-untranslated region and 5′-untranslated region of SHMT1 mRNA to upregulate translation (22). CELF1 and FTH1 were shown to physically interact, consistent with our findings demonstrating significant interactions for the gene pair FTH1-CELF1. Dysregulation of SHMT1 expression may be fundamental to folate-associated pathologies; a 50% reduction in expression sensitizes mice to intestinal carcinogenesis (33) and neural tube defects (34).

Gene-nutrient interactions. The association of genetic variants with CVD risk may be modified by nutrient status. The previously identified MTHFR rs1801133-folate interaction (8,35) was not significant in these data (P = 9.00 × 10−13); the relatively high folate status in this cohort, compared with NHANES data (36), may explain the lack of interaction. In analyses restricted to men with folate concentrations below the 25th percentile (<14.0 nmol/L; 187 men, 59 CVD events), the interaction was evident, marginally significant, and effect sizes were consistent with previous studies (5,6,8,37,38).

Considering vitamin B-6, no interactions between studied SNPs and vitamin B-6 reached the FDR threshold for significance. However, FDR-significant interactions were identified for plasma folate (MAT2B rs6882306) and vitamin B-12 (BHMT rs585800 and SLC25A32 rs1061196). The few gene-nutrient interactions identified may reflect limited power or the lack of men with low B vitamin status, even within this pre–folate fortification cohort.

The MAT2B gene product regulates S-adenosylmethionine synthesis. rs6882306 is an intronic SNP, but HapMap plots, although incomplete, suggest high linkage across the gene. Although MAT2B polymorphisms are not known to interact with folate, a recent study identified variants in a closely related gene, methionine adenosyltransferase I, alpha (MATIA), that decreased risk of hypertension and DNA damage and increased risk of stroke independently of the MTHFR rs1801133 polymorphism (39). MATIA interactions with folate predicted homocysteine and were consistent with the direction of effect reported here for the MAT2B-folate interaction and CVD risk.

The BHMT gene encodes a cystosolic enzyme that remethylates betaine to homocysteine. The rs585800 SNP is a 3′ region polymorphism and thus may have a regulatory function. Although BHMT is not known to bind vitamin B-12, previous work has identified regions homologous to bacterial vitamin B-12-dependent methionine syntheses (40). In rats, hepatic BHMT activity was not sensitive to vitamin B-12 deficiency (41); however, epidemiologic work in humans suggests that enhancing methionine-synthase–dependent remethylation down-regulates BHMT reactions (42), suggesting a mechanism for B-12 sensitivity.

The SLC25A32 gene encodes a folate transporter that shuttles folate into the mitochondria. The rs1061196 polymorphism is a 3′ region variant and thus may have a regulatory function. No prior reports link SLC25A32 to biochemical or disease phenotypes and a biological basis for the interaction of this gene with vitamin B-12 metabolism could not be identified.

The MTHFR rs1801133 TT-CVD association identified here [OR = 1.71 (95% CI: 1.3, 2.3)] is somewhat larger than previously reported despite the relatively folate-replete status of this population. Prior reports identified variability in the effect of the TT genotype (8), which may be due in part to differences in folate status between populations, but gene-gene interactions involving MTHFR rs1801133 (15,20) may also play a role (Supplemental Table 6). Thus, interpretation of a MTHFR “main effect” association with CVD is complex, because differences in the interacting genotype frequency are expected to affect the risk estimates for the MTHFR genotype. Furthermore, the association of the MTHFR rs1801133 genotype with CVD risk was not entirely mediated through homocysteine, because...
the risks associated with MTHFR rs1801133 considered alone and in an interaction with SHMT1 rs1979277 both increased after adjusting for plasma nutrients and/or homocysteine.

**Gene-gene interactions.** Considering the studied SNPs and MTHFR rs1801133, no interactions were significant at the FDR threshold. Similarly, of all network pairwise interactions evaluated, none exceeded the FDR significance threshold (P-adjusted ≤ 2.00 × 10⁻¹⁰).

Pleiotropic effects were examined across outcomes. Previously, SNPs associated with plasma homocysteine and global genomic DNA methylation were identified (9). Of 8 nominally significant SNPs identified in the CVD analysis, one pleiotropic SNP, CBS rs6586282, was strongly associated with the homocysteine phenotype (nominal P-CVD = 3.30 × 10⁻³, nominal P-homocysteine = 4.06 × 10⁻³); this SNP was also identified in a GWAS of the homocysteine phenotype (43). The CBS rs6586282 CT (vs. combined CC/TT) genotype was associated with a 5.8% reduction in plasma homocysteine; under an additive model, the T allele was associated with a 29% reduction in CVD risk. The CBS rs6586282 SNP was the only polymorphism in the CVD analysis mediated by nutrition, consistent with the association of this SNP with both homocysteine and CVD. Similarly, GGH was nominally significant for homocysteine and FDR significant for CVD; MTHFR was nominally significant for CVD and LINE-1 methylation.

The strengths of this study include a pathway-wide investigation of folate network genetic variation, use of a large cohort with CVD outcomes collected prior to the introduction of mandatory folate fortification, and a systematic approach to data analysis. The findings were corrected for multiple comparisons and hypotheses derived from animal and cell studies were explored. Despite restricting data to the pre–folate fortification period, the folate status of the population was relatively high, which may have attenuated some associations, and the study represents data from men only. Therefore, associations may differ in females.

In conclusion, variation in genes other than MTHFR and those directly involved in homocysteine metabolism are associated with CVD risk in 1131 non-Hispanic white males from the NAS. The epidemiologic findings presented herein replicated molecular biology studies of interactions among SHMT1-related proteins, supporting the importance of the SHMT1 intersection in folate metabolism in relation to CVD risk. These findings may help explain the limited success of randomized controlled trials focused on CVD risk reduction through homocysteine lowering (44,45). A thorough understanding of the role of folate network genetic and nutritional variation in relation to CVD is important, particularly in the context of the numerous unfortified populations around the world.

**Acknowledgments**

P.A.C. and S.M.W. designed the research; S.M.W., P.A.C., A.G.C., A.A.L., S.T.W., J.M.G., P.S.V., and K.L.T. conducted the research; S.M.W. and P.A.C. analyzed the data; S.M.W., P.A.C., P.J.S., A.G.C., and M.T.W. wrote the paper; and P.A.C. had primary responsibility for all work and final content. All authors read and approved the final manuscript.

**Literature Cited**

23. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661–78.


