Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease


ABSTRACT

Background: The strong observational association between total homocysteine (tHcy) concentrations and risk of coronary artery disease (CAD) and the null associations in the homocysteine-lowering trials have prompted the need to identify genetic variants associated with homocysteine concentrations and risk of CAD.

Objective: We tested whether common genetic polymorphisms associated with variation in tHcy are also associated with CAD.

Design: We conducted a meta-analysis of genome-wide association studies (GWAS) on tHcy concentrations in 44,147 individuals of European descent. Polymorphisms associated with tHcy (\(P < 10^{-8}\)) were tested for association with CAD in 31,400 cases and 92,927 controls.

Results: Common variants at 13 loci, explaining 5.9% of the variation in tHcy, were associated with tHcy concentrations, including 6 novel loci in or near MMACHC (2.1 \(\times 10^{-8}\)), SLC17A3 (1.0 \(\times 10^{-8}\)), GTPB10 (1.7 \(\times 10^{-8}\)), CUBN (7.5 \(\times 10^{-10}\)), HNF1A (1.2 \(\times 10^{-10}\)), and FUT2 (6.6 \(\times 10^{-9}\)), and variants previously reported at or near the MTHFR, MTR, CPS1, MUT, NOX4, DPEP1, and CBS genes. Individuals with the highest 10% of the genotype risk score (GRS) had 3-\(\mu\)mol/L higher mean tHcy concentrations than did those with the lowest 10% of the GRS (\(P = 1 \times 10^{-36}\)). The GRS was not associated with risk of CAD (OR: 1.01; 95% CI: 0.98, 1.04; \(P = 0.49\)).

Conclusions: We identified several novel loci that influence plasma tHcy concentrations. Overall, common genetic variants that influence plasma tHcy concentrations are not associated with risk of CAD in white populations, which further refutes the causal relevance of moderately elevated tHcy concentrations and tHcy-related pathways for CAD.

INTRODUCTION

Observational studies have consistently shown that elevated concentrations of total homocysteine (tHcy)\(^6\) are associated with an increased risk of coronary artery disease (CAD), stroke, and venous thrombosis in addition to other clinical conditions (1–6). However, it is uncertain whether these associations are causal, given the absence of protection against cardiovascular events in homocysteine-lowering trials (7). We therefore used a Mendelian randomization strategy to test whether long-term elevated concentrations of homocysteine are associated with an increased risk of CAD. Mendelian randomization is a method of using measured variation in genes of known function to examine the causal effect of a modifiable exposure on disease in

---

1 From the Departments of Internal Medicine (JBJvM, FR, and AGU), Epidemiology (AD, A Hofman, FR, JCMW, and AGU), and Clinical Chemistry (IL), Erasmus MC, Rotterdam, the Netherlands; the Netherlands Genomics Initiative, sponsored by the Netherlands Consortium for Healthy Aging, Rotterdam, Leiden the Netherlands (JBJvM and FR); the Department of Pathology, McMaster University, Hamilton, Canada (GP); the Cardiovascular Health Research Unit, Department of Epidemiology, University of Washington, Seattle, WA (SMS); the Channing Laboratory, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School and Department of Epidemiology, Harvard School of Public Health, Boston (AH, CC, DJH, and PK); the Clinical Research Branch, National Institute on Aging, Baltimore, MD (TT, and LF); the Department for Health Evidence, Radboud University Medical Centre, Nijmegen, the Netherlands (SHV and LALMK); the Department of Twin Research and Genetic Epidemiology, King’s College London, London, United Kingdom (IC and KRA); the Genetics Division, GlaxoSmithKline, King of Prussia, PA (XY, DMW, and VM); the Department of Medicine, Division of Internal Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland (XY, PV, GW, DMW, and VM); the Atherosclerosis Research Unit, Karolinska Institute, Stockholm, Sweden (AM and A Hamsten); the Geriatric Unit, Azienda Sanitaria Firenze, Florence, Italy (SB); the VU Medical Centre, Amsterdam, the Netherlands (HB); the Clinical Pharmacology Unit, University of Cambridge, Cambridge, United Kingdom (MJB); the Medical Genetics Research Institute, Cedars-Sinai Medical Centre, Los Angeles, CA (Y-DC); the Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, United Kingdom (RJC); Medizinische Klinik II, Universität zu Lübeck, Lübeck, Germany (JE and HS); the Department of Cardiovascular Medicine, University of Oxford, United Kingdom (AG and HW); the Vitamin Metabolism and Aging Laboratory and Nutritional Epidemiology Program, Tufts University, Boston, MA (JS and PJ); the Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA (ADJ, DPK, CJO, and PJ); the Cardiovascular Research Center and Cardiology Division, Boston, MA (SK); the Division of Human

---

nonexperimental studies. By examining whether common polymorphisms associated with variations in tHcy are also associated with CAD, we could explore the association between tHcy and CAD.

To date, 5 separate genome-wide association studies (GWAS) of blood tHcy concentrations have been published, including 4 in whites and 1 in Filipinos, which have collectively identified 7 independent loci definitively associated with tHcy concentrations (8–12). These include the MTHFR C677T variant (rs1801133), the most widely studied consistently associated polymorphism with tHcy, and variants at or near the CBS, MTR, DEPPE1, NOX4, CPS1, and MUT genes. Of these genes, the effect of the MTHFR functional variant on CAD was studied previously. In a meta-analysis of published results based on 26,000 CAD cases and >31,000 controls, Lewis et al (13) reported a significant association between the C677T variant with CAD in Asian populations only; the association between this variant and CAD in whites was close to null and not statistically significant. In contrast, the MTHFR Studies Collaboration showed no association of MTHFR C677T with risk of CAD in unpublished studies, irrespective of folate status or ethnic group. The latter meta-analysis compared the associations of MTHFR C677T with CAD in a meta-analysis of 19 unpublished data sets involving 48,175 CAD cases with a meta-analysis of 86 published studies involving 28,617 CAD cases and showed that the discrepant results of the unpublished and published studies were explained by publication bias and other methodologic problems in the published studies of Asian populations (14).

To identify other common variants associated with risk of CAD through effects on tHcy concentrations, we extended previous efforts by performing a meta-analysis of 2.7 million genotyped and imputed single nucleotide polymorphisms (SNPs) from 10 GWAS, with data on a total of 44,147 individuals of European ancestry (including new data on 26,669 individuals) with measured blood tHcy concentrations, and tested whether common polymorphisms associated with variations in tHcy were also associated with CAD. We tested each individual tHcy-associated polymorphism and a weighted genotype risk score (GRS), composed of the top (highest P value) independent SNPs (as assessed by \( r^2 < 0.2 \) between associated SNPs) consistently associated with tHcy concentrations, for their association with CAD in 31,400 cases of CAD and 92,927 controls.

**SUBJECTS AND METHODS**

**Homocysteine meta-analysis populations**

The current meta-analysis included data from a total of 44,147 white individuals of European ancestry derived from 10 GWAS on tHcy concentrations (see supplementary materials under “Supplemental data” in the online issue for details of each cohort). Data from Invecchiare in Chianti (n = 1208), the Baltimore Longitudinal Study of Aging (BLSA) (n = 638), and the Nurses’ Health Study (NHS) (n = 1658) and part of the data (n = 13,974) from the Women’s Genome Health Study (WGHS) were published previously (8–10). Thus, the current meta-analysis includes new data on 26,669 individuals from 6 new cohorts. Each study was approved by the institutional ethics review committees at the relevant organizations, and all subjects provided informed consent for the collection of DNA and its use for genetic analyses.

**Coronary artery disease cohorts**

We tested the association of the tHcy-associated variants with risk of CAD in 6 independent cohorts and in the Coronary Artery Disease Genome-wide Replication And Meta-Analysis (CARDiOGRAM) Consortium. A full list of consortium members is listed under “Supplemental data” in the online issue.
criteria used for the recruitment of cases and controls in each of these studies is summarized elsewhere (see Supplementary Table 1 under “Supplemental data” in the online issue).

Blood tHcy measurements

THcy was measured in each cohort by using one of the following methods: isotope-dilution liquid chromatography–tandem mass spectrometry, gas chromatography–coupled mass spectrometry, HPLC, or enzymatic, immune, or chemiluminescence (see Supplementary Table 2 under “Supplemental data”).

Genotyping and quality control

The participants from each cohort were genotyped by using the Illumina Infinium HumanHap550, the Illumina Infinium HumanHap300, the Illumina HumanCNV370 BeadChips, or the Affymetrix 5.0 GeneChip 500K according to the manufacturers’ protocols and quality-control standards. The exclusion/filtering criteria for SNPs in each of the cohorts are described elsewhere (see Supplementary Table 2 under “Supplemental data” in the online issue).

Genotype imputation

Genotypes were imputed for all polymorphic SNPs based on HapMap CEU (haplotype map for people living in Utah with ancestry from northern and western Europe) phase II data (release 22, build 36) oriented to the positive strand. Hidden Markov model–based algorithms were used to infer unobserved genotypes probabilistically as implemented in either MACH (18), IMPUTE (19), or BimBam (20). Imputation quality-control metrics included the ratio of observed/expected variance of the allele dose ≤0.01 for MACH, proper information ≤0.40 for IMPUTE, or r² ≥ 0.3 for BimBam. Detailed descriptions of the quality-control and imputation procedures are summarized for all studies elsewhere (see Supplementary Table 3 under “Supplemental data” in the online issue).

Genome-wide association analysis

All analyses were performed on natural log-transformed tHcy concentrations by using sex-specific, age-adjusted standardized residuals (z scores). Each study performed a genome-wide association analysis for tHcy concentrations by using either MACH2QTL (www.sph.umich.edu/csg/abecasis/MACH/index.html) or SNPTTEST (www.stats.ox.ac.uk/~marchini/software/gwas/snpTest.html), which use the genotype dosage value as continuous additive predictors of tHcy concentrations in a linear regression framework or GenABEL (21) with the use of an additive genetic model while accounting for relatedness between the members of each family (see Supplementary Table 3 under “Supplemental data” in the online issue). Analysis of imputed genotype data accounted for uncertainty in each genotype prediction by using either the dosage information from MACH or BimBam or the genotype probabilities from IMPUTE. Genomic control was used to correct any inflation of test statistics due to population structure in each cohort before the meta-analysis.

Meta-analysis

The minor allele from HapMap CEU genotypes was used to define the coded allele in all analyses, regardless of frequency in individual cohorts. A total of 2,090,256 SNPs with a minor allele frequency ≥1% and directly genotyped or imputed with sufficient quality (see Supplementary Table 3 under “Supplemental data” in the online issue) were included in the meta-analysis. We performed an inverse-variance, weighted, meta-analysis of the summary data from the 10 studies to obtain a pooled estimate of the overall β coefficient and SE by using METAL (www.sph.umich.edu/csg/abecasis/metal/index.html). We considered the results to be genome-wide significant at α = 5 × 10⁻⁸.

Conditional analysis to discover independent signals

To test for the presence of additional, independent signals at each of the top loci we carried out secondary GWAS conditioning on the top SNPs—referred to as the lead SNP—that had reached genome-wide significance. We re-ran each GWAS under the same conditions used in the primary analysis but included the maximum likelihood allele dose of each of the top 18 SNPs as covariates in the model. Parameters from the conditional analyses were then meta-analyzed by using a fixed-effects model. The regions within one mega-base upstream and downstream of the SNP conditioned on were examined for additional hits by using the significance criterion used for the original meta-analysis (α = 5 × 10⁻⁸). Any genome-wide significant SNP in these regions was considered as a second independent association in the region.

Risk of hyperhomocysteinemia

Using the method outlined by Horne et al (22), we generated a GRS equal to the sum of the expected number of risk alleles at each SNP weighted by their effect sizes (β coefficient from the tHcy meta-analysis). We used logistic regression to test the combined effect of the 18 loci associated with tHcy on the risk of hyperhomocysteinemia (defined as ≥15 μmol/L), with adjustments for age and sex, in the Rotterdam Study cohorts. In addition, we estimated the explained variance of the identified SNPs in the Rotterdam cohorts by calculating the R² in a linear regression analysis, in which tHcy concentration was the dependent variable and the SNPs were the independent variable.

Association of tHcy loci with coronary artery disease

We sought to assess whether common polymorphisms associated with tHcy identified through this meta-analysis are also associated with CAD. The association between tHcy SNPs on CAD were tested in 2 ways:

1) Each of the 18 SNPs that were associated with tHcy concentrations was tested for association with risk of CAD in a case-control study with the use of data from the CARDIoGRAM consortium (22,233 CAD cases and 64,762 controls) and in 3 separate CAD cohorts: the WGHS (23) (758 CAD cases and 22130 controls), which was also used in the discovery stage; the LOLIPOP (24) study, which contributed 2793 CAD cases and 3756 controls from South Asia; and the PROCARDIS study (25), which contributed a total of 5616 CAD cases and 2279 controls.

2) We estimated the combined effect of the 18 loci associated with tHcy in 6 studies (WGHS, PROCARDIS, The Wellcome...
trust case control consortium, pennsylvania catheterization study, lolipop, and rotterdam) by testing the association between an additive GRS for elevated tHcy concentrations and increased risk of CAD.

we conducted a logistic regression analysis treating either the SNPs or the GRSs as a continuous explanatory variable, including sex, age, and BMI as covariates. Output parameters were then meta-analyzed by using a fixed-effects model.

mendelian randomization study of tHcy and CAD

To quantify the expected causal relation between tHcy variation and risk of CAD, we used the methods outlined by minelli et al (26). Under the assumptions required for Mendelian randomization, the OR of CAD resulting from a unit change in tHcy (CAD$_{tHcy}$) is equivalent to the OR of CAD resulting from a unit change in the GRS (CADGRS) raised to the power of the inverse of the mean difference in tHcy per unit of the GRS ($\Delta$tHcyGRS):

$$\text{OR(CAD$_{tHcy}$)} = \text{OR(CADGRS)}^{\Delta\text{tHcy}_{\text{GRS}}}$$

We used the CADGRS from the meta-analysis and the ($\Delta$tHcyGRS) from the Rotterdam Study. We used the SE of the OR(CADGRS) to construct the 95% CIs for OR(CADGRS).

Expression quantitative trait loci analysis

For each of the 18 lead genome-wide significant tHcy SNPs, all proxy SNP identifiers were identified by using SNAP (27). All lead SNPs and their proxies ($r^2 > 0.7$) were then searched against a collected database of expression SNP results including the following tissues: fresh lymphocytes (28), fresh leukocytes (29), leukocyte samples in individuals with celiac disease (30), lymphoblastoid cell lines (LCLs) derived from asthmatic children (31), HapMap LCLs from 3 populations (32), a separate study on HapMap CEU LCLs (33), peripheral blood monocytes (34), adipose and blood samples (35, 36), 2 studies on brain cortex (34, 37), and 3 large studies of brain regions, including the prefrontal cortex, visual cortex, and cerebellum (V Emlisson, personal communication, 2012); liver (38); osteoblasts (39); skin (40); and additional fibroblast, T cell, and LCL samples (41). The collected expression SNP results met criteria for the statistical significance for association with gene transcript levels as described in the original articles.

RESULTS

Homocysteine cohort characteristics

The mean age of each cohort varied from 47.5 to 74.4 y, and the cohort participants ranged in age from 17 to 79 y. Four cohorts—including TwinsUK, WGHS, and NHS—were composed predominantly of females, whereas the remainder of the cohorts had roughly the same number of males and females (women: 45–58%). Additional characteristics of the study populations are presented elsewhere (see Supplementary Table 2 under “Supplemental data” in the online issue).

Meta-analysis of genome-wide association scans of tHcy concentrations

The meta-analysis identified SNPs from 13 independent genomic loci exceeding the GWAS threshold ($P < 5 \times 10^{-8}$) (see Supplementary Figure 1 under “Supplemental data” in the online issue, Figure 1, and Table 1). These included 6 previously unreported loci in or near MMACHC ($P = 2.1 \times 10^{-9}$), SLC17A3 ($1.0 \times 10^{-8}$), GTPB10 ($1.7 \times 10^{-8}$), CUBN ($P = 7.5 \times 10^{-10}$), HNF1A ($1.2 \times 10^{-12}$), and FUT2 ($6.6 \times 10^{-9}$). In addition, we confirmed 7 loci previously reported to be associated with tHcy concentrations at or near the MTHFR, MTR, CPS1, MUT, NOX4, DPEP1, and CBS genes. We failed to replicate previous associations with variants at or near the ZNF366.
Additional SNPs associated with total plasma homocysteine concentrations

TABLE 1

<table>
<thead>
<tr>
<th>SNPs associated with total plasma homocysteine concentrations</th>
<th>Nearest gene</th>
<th>Chromosome</th>
<th>Coded allele</th>
<th>Frequency</th>
<th>$\beta$ (SD/allele)</th>
<th>SE</th>
<th>$P$(meta)</th>
<th>$P$(het)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known loci $^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1801133 MTHFR</td>
<td>1</td>
<td>A</td>
<td>0.34</td>
<td>0.1583</td>
<td>0.007</td>
<td></td>
<td>4.34 $\times 10^{-14}$</td>
<td>6.90 $\times 10^{-8}$</td>
</tr>
<tr>
<td>rs2275565 MTR</td>
<td>1</td>
<td>T</td>
<td>0.21</td>
<td>$-0.0542$</td>
<td>0.009</td>
<td></td>
<td>1.96 $\times 10^{-10}$</td>
<td>0.05</td>
</tr>
<tr>
<td>rs7422339 CPS1</td>
<td>2</td>
<td>A</td>
<td>0.33</td>
<td>0.0864</td>
<td>0.008</td>
<td></td>
<td>4.58 $\times 10^{-8}$</td>
<td>0.61</td>
</tr>
<tr>
<td>rs9369898 MUT</td>
<td>6</td>
<td>A</td>
<td>0.62</td>
<td>0.0449</td>
<td>0.007</td>
<td></td>
<td>2.17 $\times 10^{-10}$</td>
<td>0.66</td>
</tr>
<tr>
<td>rs7130284 NOX4</td>
<td>11</td>
<td>T</td>
<td>0.07</td>
<td>$-0.1242$</td>
<td>0.013</td>
<td></td>
<td>1.88 $\times 10^{-20}$</td>
<td>0.35</td>
</tr>
<tr>
<td>rs1546675 DPEP1</td>
<td>16</td>
<td>A</td>
<td>0.47</td>
<td>0.0963</td>
<td>0.007</td>
<td></td>
<td>1.74 $\times 10^{-43}$</td>
<td>0.09</td>
</tr>
<tr>
<td>rs234709 CBS</td>
<td>21</td>
<td>T</td>
<td>0.45</td>
<td>$-0.0718$</td>
<td>0.007</td>
<td></td>
<td>3.90 $\times 10^{-24}$</td>
<td>0.44</td>
</tr>
<tr>
<td>Novel loci</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4603096 MACHC</td>
<td>1</td>
<td>T</td>
<td>0.33</td>
<td>0.0435</td>
<td>0.007</td>
<td></td>
<td>2.33 $\times 10^{-09}$</td>
<td>0.18</td>
</tr>
<tr>
<td>rs549878 SLC17A3</td>
<td>6</td>
<td>C</td>
<td>0.13</td>
<td>0.0597</td>
<td>0.010</td>
<td></td>
<td>1.12 $\times 10^{-8}$</td>
<td>0.45</td>
</tr>
<tr>
<td>rs24684 GTPB10</td>
<td>7</td>
<td>A</td>
<td>0.40</td>
<td>$-0.0395$</td>
<td>0.007</td>
<td></td>
<td>1.97 $\times 10^{-8}$</td>
<td>0.57</td>
</tr>
<tr>
<td>rs1801222 CUBN</td>
<td>10</td>
<td>A</td>
<td>0.34</td>
<td>0.0453</td>
<td>0.007</td>
<td></td>
<td>8.43 $\times 10^{-10}$</td>
<td>0.05</td>
</tr>
<tr>
<td>rs2251468 HNF1A</td>
<td>12</td>
<td>A</td>
<td>0.65</td>
<td>$-0.0512$</td>
<td>0.007</td>
<td></td>
<td>1.28 $\times 10^{-12}$</td>
<td>0.43</td>
</tr>
<tr>
<td>rs838133 FUT2</td>
<td>19</td>
<td>A</td>
<td>0.45</td>
<td>0.0422</td>
<td>0.007</td>
<td></td>
<td>7.48 $\times 10^{-9}$</td>
<td>0.31</td>
</tr>
<tr>
<td>Additional SNPs associated after conditional analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12134663 MTHFR</td>
<td>1</td>
<td>A</td>
<td>0.80</td>
<td>$-0.101$</td>
<td>0.011</td>
<td></td>
<td>2.54 $\times 10^{-21}$</td>
<td>0.43</td>
</tr>
<tr>
<td>rs12780845 CUBN</td>
<td>10</td>
<td>A</td>
<td>0.65</td>
<td>0.0529</td>
<td>0.009</td>
<td></td>
<td>7.8 $\times 10^{-10}$</td>
<td>0.73</td>
</tr>
<tr>
<td>rs957140 NOX4</td>
<td>11</td>
<td>A</td>
<td>0.45</td>
<td>$-0.045$</td>
<td>0.008</td>
<td></td>
<td>2.43 $\times 10^{-8}$</td>
<td>0.97</td>
</tr>
<tr>
<td>rs12921383 DPEP1/LOC126661</td>
<td>16</td>
<td>T</td>
<td>0.87</td>
<td>$-0.090$</td>
<td>0.014</td>
<td></td>
<td>8.22 $\times 10^{-11}$</td>
<td>0.96</td>
</tr>
<tr>
<td>rs2851391 CBS</td>
<td>21</td>
<td>T</td>
<td>0.47</td>
<td>0.056</td>
<td>0.008</td>
<td></td>
<td>1.70 $\times 10^{-12}$</td>
<td>0.10</td>
</tr>
</tbody>
</table>

$^1$ $\beta$, $\beta$ coefficient presented as the number of SD differences in homocysteine concentrations per allele; $P$(het), $P$ value for the Cochran’s heterogeneity statistic; $P$(meta), $P$ value for the overall meta-analysis; SNP, single nucleotide polymorphism.

$^2$ Those loci identified previously by Pare et al (9), Tanaka et al (8), Hazra et al (10), and Lange et al (12). The total number of individuals used to identify the genetic loci was 44,147.

(rs7445013) and the PTPRD (rs973117) genes (11) ($P = 0.29$ and 0.39, respectively).

**Conditional independence of variants associated with tHcy concentrations**

Visual inspection of the regional association plots of the 13 loci suggested that, for a number of loci, there was evidence to support the presence of additional SNPs that showed an association with variation in tHcy concentrations independently of the lead variant. This was supported by the high level of significance despite the low correlation ($r^2 < 0.2$) with the lead SNP and its position with respect to recombination hot spots.

Regional plots of the results from this conditional analysis are shown in Supplementary Figure 2. Our analysis identified $\geq 5$ other SNPs—at or near MTHFR, CUBN, NOX4, DPEP1, and CBS—which showed a significant association with variation in tHcy concentrations ($P < 5 \times 10^{-8}$) that were conditionally independent of the lead variant at these loci. In addition, there was evidence of a secondary signal at the SLC17A3 locus on chromosome 6, although this did not reach genome-wide significance (results not shown). In summary, we identified a total of 18 independent SNPs from 13 loci that are significantly associated with tHcy concentrations (see Supplementary Table 4 under “Supplemental data” in the online issue).

**Functional genomics of identified variants**

We tested for an association of the 18 lead SNPs [or SNPs in high linkage disequilibrium ($r^2 > 0.8$) with the lead SNP; ie, proxies] with cis-allelic expression of gene transcripts by using publicly available databases (see Supplementary Table 5 under “Supplemental data” in the online issue). Associations were seen for transcripts of LOC126661, RASIP1, CDK10, and CBS. The most significant correlation with expression in several tissues was seen with CBS for rs2851391—the SNP that was identified after conditional analysis. Notably, the lead variant on HNF1A was not associated with differential expression of the gene in any of the tissues tested, including the liver.

**Effect of tHcy loci and the tHcy risk score on hyperhomocysteinemia**

We examined the combined effect of the top SNPs arising from the 13 associated tHcy loci on the risk of tHcy concentrations and hyperhomocysteinemia (defined as tHcy $\geq 15 \mu mol/L$) in 5528 individuals from the prospective population-based Rotterdam Study I and II. Weighted risk allele scores were generated, and the population was divided into deciles according to risk score. The combined risk score explained 4.6% (Rotterdam Study I) to 5.9% (Rotterdam Study II) of the variance in tHcy concentrations. A highly significant linear increase in the mean tHcy concentrations of individuals with increasing values of GRS ($P$-trend: $5 \times 10^{-57}$) is shown elsewhere (see Supplementary Figure 3 under “Supplemental data” in the online issue). Individuals within the highest 10% of the GRS ($n = 552$) had 3-µmol/L higher tHcy concentrations than did those in the lowest 10% of GRS ($n = 552$) ($P = 1 \times 10^{-36}$; ANCOVA). In addition, the GRS was significantly associated with the risk of hyperhomocysteinemia.
Effect of tHcy loci and the tHcy risk score on CAD

Given the strong observational association between tHcy concentrations and risk of cardiovascular disease (42), we tested \( I \) each individual tHcy-associated SNP, and \( 2 \) the combined effect of the 18 SNPs from the 13 tHcy loci, with CAD risk in up to 31,400 CAD cases and 92,927 controls. We did not have Hcy concentrations available for most of the cases and controls who were used to assess the association between the GRS and CAD. The variants rs2251468 (OR: 0.93; 95% CI: 0.91, 0.95; \( P = 2.5 \times 10^{-10} \)) and rs7422339 (OR: 0.96; 95% CI: 0.93, 0.99; \( P = 0.01 \)) (Table 2), located in the HNF1A and CPS1 genes, respectively, are associated with an increased risk of CAD, although the latter failed to pass the Bonferroni correction criteria of \( P < 0.004 \).

We failed to show an association between the tHcy GRS and an increased risk of CAD (OR: 1.01; 95% CI: 0.98, 1.04; \( P = 0.49 \)). We quantified the expected causal relation between tHcy variation and risk of CAD with the use of a Mendelian randomization framework. Using the output from the GRS analysis for hyperhomocysteinemia and CAD—ie, OR(CADGRS) = 1.01 or \( \beta = 0.01, \Delta tHcyGRS = 0.18 \), and SE of OR(CADGRS) = 0.016—
the OR(CADtHcy) was calculated at 1.05 (95% CI: 0.85, 1.2).

DISCUSSION

Given the strong observational link between tHcy concentrations and risk of CVD (1–3, 6), we tested whether common polymorphisms associated with variation in tHcy are also associated with CAD. To this end, we carried out a meta-analysis of 10 independent genome-wide association studies comprising a total of 44,147 individuals and identified 6 novel loci in or near genes MMACHC, SLC17A3, GTPB10, CUBN, HNF1A, and FUT2 and replicating variants previously reported at or near the MTHFR, MTR, CPS1, MUT, NOX4, DPEP1, and CBS loci.

Furthermore, we provide statistical evidence pointing to the fact that there are \( \geq 2 \) significant, independent loci in MTHFR, CUBN, NOX4, DPEP1, CBS, and SLC17A3. We then tested for an association between each lead SNP at these loci and a GRS combining these variants with an increased risk of hyperhomocysteinemia and CAD. Our results showed that genetic variants at the HNF1A locus—which are associated with reduced concentrations of tHcy—and previously associated with Maturity Onset Diabetes of the Young increased the risk of type 2 diabetes, and plasma concentrations of C-reactive protein, total cholesterol, and low-density lipoprotein are also significantly associated with a reduced risk of CAD. It is therefore plausible that this association is not caused by Hcy concentrations but rather to the pleiotropic role of this DNA variant, possibly through lipid metabolism. Although the aggregate GRS was significantly associated with hyperhomocysteinemia (OR: 1.40), we failed to detect an association between the GRS and CAD.

We used tHcy genetic variants to examine whether Hcy concentrations are causally related to CAD. Previous meta-analyses of published studies of MTHFR C667T and CAD had reported associations between the MTHFR variant and CAD (13, 43, 44). However, the MTHFR Studies Collaboration reported no association of MTHFR C677T with CAD in unpublished studies and showed that the discrepant results of unpublished and published studies were a result of publication bias or other methodologic problems in the Asian studies. The current meta-analysis extends the findings for MTHFR C677T and CAD in European populations by studying the association of CAD with all available genetic variants that influence Hcy concentrations. With the possible exception of the HNF1A locus, our results did not provide any support for an association of the tHcy-associated variants with CAD, including the most widely studied MTHFR C677T variant.

A formal assessment of the expected causal association under the assumptions of Mendelian randomization, with the GRS used as the instrumental variable, is consistent with the conclusion that tHcy does not cause CAD. Notably, our study \( I \) did

**Table 2**

Association between total homocysteine–associated loci and risk of coronary artery disease in 31,400 cases and 92,927 controls

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>Chromosome</th>
<th>Position</th>
<th>Allele</th>
<th>OR (95L)</th>
<th>OR (95U)</th>
<th>P(meta)</th>
<th>Q</th>
<th>P(het)</th>
<th>P</th>
<th>No. of cases</th>
<th>No. of controls</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs154657</td>
<td>DPEP1</td>
<td>16</td>
<td>88235597</td>
<td>A</td>
<td>1.01</td>
<td>0.99</td>
<td>1.04</td>
<td>0.40</td>
<td>2.14</td>
<td>0.54</td>
<td>0.00</td>
<td>29,314, 91,599</td>
<td>++++</td>
</tr>
<tr>
<td>rs1801133</td>
<td>MTHFR</td>
<td>11</td>
<td>11778965</td>
<td>A</td>
<td>1.01</td>
<td>0.98</td>
<td>1.04</td>
<td>0.50</td>
<td>9.21</td>
<td>0.03</td>
<td>0.67</td>
<td>15,974, 78,259</td>
<td>+——</td>
</tr>
<tr>
<td>rs1801222</td>
<td>CUBN</td>
<td>10</td>
<td>17196157</td>
<td>A</td>
<td>0.98</td>
<td>0.96</td>
<td>1.01</td>
<td>0.29</td>
<td>10.04</td>
<td>0.02</td>
<td>0.70</td>
<td>18,817, 81,102</td>
<td>——</td>
</tr>
<tr>
<td>rs2251468</td>
<td>HNF1A</td>
<td>12</td>
<td>119889509</td>
<td>A</td>
<td>0.94</td>
<td>0.91</td>
<td>0.96</td>
<td>2.5E-07</td>
<td>2.85</td>
<td>0.42</td>
<td>0.00</td>
<td>29,014, 91,299</td>
<td>———</td>
</tr>
<tr>
<td>rs2275565</td>
<td>MTR</td>
<td>1</td>
<td>235115299</td>
<td>G</td>
<td>1.01</td>
<td>0.98</td>
<td>1.04</td>
<td>0.37</td>
<td>3.40</td>
<td>0.33</td>
<td>0.12</td>
<td>29,319, 91,604</td>
<td>———</td>
</tr>
<tr>
<td>rs234709</td>
<td>CBS</td>
<td>21</td>
<td>43360033</td>
<td>C</td>
<td>1.00</td>
<td>0.97</td>
<td>1.02</td>
<td>0.81</td>
<td>1.43</td>
<td>0.49</td>
<td>0.00</td>
<td>23,110, 84,432</td>
<td>—/—</td>
</tr>
<tr>
<td>rs42658</td>
<td>GTPBP10</td>
<td>7</td>
<td>89815696</td>
<td>A</td>
<td>1.01</td>
<td>0.98</td>
<td>1.03</td>
<td>0.68</td>
<td>4.33</td>
<td>0.23</td>
<td>0.31</td>
<td>28,384, 90,669</td>
<td>++</td>
</tr>
<tr>
<td>rs4663036</td>
<td>MMACHC</td>
<td>1</td>
<td>4575126</td>
<td>C</td>
<td>1.00</td>
<td>0.98</td>
<td>1.03</td>
<td>0.84</td>
<td>2.96</td>
<td>0.40</td>
<td>0.00</td>
<td>29,591, 91,876</td>
<td>———</td>
</tr>
<tr>
<td>rs548997</td>
<td>SLC17A3</td>
<td>6</td>
<td>25977350</td>
<td>C</td>
<td>1.00</td>
<td>0.96</td>
<td>1.04</td>
<td>0.92</td>
<td>0.92</td>
<td>0.82</td>
<td>0.00</td>
<td>18,612, 80,897</td>
<td>++</td>
</tr>
<tr>
<td>rs7130284</td>
<td>NOX4</td>
<td>11</td>
<td>88788020</td>
<td>C</td>
<td>1.00</td>
<td>0.96</td>
<td>1.05</td>
<td>0.75</td>
<td>8.32</td>
<td>0.04</td>
<td>0.64</td>
<td>28,330, 90,615</td>
<td>++</td>
</tr>
<tr>
<td>rs7422339</td>
<td>CPS1</td>
<td>2</td>
<td>211246752</td>
<td>A</td>
<td>0.96</td>
<td>0.93</td>
<td>0.99</td>
<td>0.01</td>
<td>1.34</td>
<td>0.72</td>
<td>0.00</td>
<td>20,900, 83,185</td>
<td>++</td>
</tr>
<tr>
<td>rs838133</td>
<td>FUT2</td>
<td>19</td>
<td>53951341</td>
<td>A</td>
<td>1.02</td>
<td>0.98</td>
<td>1.05</td>
<td>0.31</td>
<td>3.79</td>
<td>0.28</td>
<td>0.21</td>
<td>14,060, 76,345</td>
<td>—+++</td>
</tr>
<tr>
<td>rs9369988</td>
<td>MUT</td>
<td>6</td>
<td>49490152</td>
<td>A</td>
<td>1.02</td>
<td>0.99</td>
<td>1.04</td>
<td>2.86</td>
<td>6.65</td>
<td>0.03</td>
<td>0.65</td>
<td>28,492, 90,777</td>
<td>—+++</td>
</tr>
</tbody>
</table>

1. \( P \), Higgins heterogeneity index; \( P(meta) \), \( P \) value for the overall meta-analysis; \( P(het) \), \( P \) value for the Cochran’s heterogeneity statistic; \( Q \), Cochran’s heterogeneity statistic; SNP, single nucleotide polymorphism; 95L, lower 95th percentile; 95U, upper 95th percentile.
2. Allele denotes the modeled or coded allele.
3. Number with marker present.
4. Effects denotes the summary of effect directions of all the single studies used for the meta-analysis, where “+” indicates a positive (increased) effect of the coded allele on risk of disease, “−” indicates a negative (decreased disease risk) effect of the coded allele on risk of disease, and “/” indicates missing.
not have the potential publication bias that a meta-analysis on published studies might have had, and 2) had sufficient power to pick up the previously reported effect size for MTHFR (30% increased risk) and the other SNPs. Power calculations showed that, for CAD, we had ~90% power to detect an OR as low as 1.07 for SNPs with an allele frequency of ≥20% (with an α = 0.004). Because the GRS was associated with a 3-μmol/L average elevation of tHcy, an 11% increased risk might plausibly have been anticipated from the observational studies if the association of Hcy with coronary artery disease seen in those studies was causal. With 12,138 such events among 45,705 individuals, this collaboration had >85% power to detect that difference. Therefore, these data indicate that a genetically raised concentration of Hcy itself is unlikely to be even a modest causal factor in CAD. Because the association of the MTHFR variant with tHcy concentrations varies by folate status, the effect of other variants for tHcy-related pathways on tHcy concentrations may also vary by folate status because of differences in national policies on folate fortification (14, 44, 46). Despite differences in the effect of MTHFR C677T on tHcy concentrations by folate status, a previous meta-analysis has shown that such differences do not explain the null associations of MTHFR C677T with risk of CAD (14), and there is no evidence that differences in folate status would influence the effects of other tHcy-related variants on risk of CAD.

It is important to highlight that our study had little power to detect associations with common variants of smaller effects or to rarer (2–5%) and/or rare (<1%) variants. Collectively, the identified polymorphisms explained only a small proportion of the total variance in tHcy concentrations, and the effect sizes attributable to the associated variants were modest, ranging from SD differences of 0.09 to 0.19 in tHcy per allele. This suggests that either a significant proportion of the genetic variants are yet to be discovered or that the true heritability of tHcy is substantially lower than the often quoted 63%.

A second potential limitation pertained to the study design. We included data from 10 different GWAS studies on tHcy concentrations in our meta-analysis. Five of the 10 cohorts included participants from the United States, and 1 of these—BLSA—recruited subjects after mandatory food fortification. For the other 5 cohorts, participants were predominantly from northwestern European ancestry. Note that tHcy concentrations were qualitatively very similar across the US cohorts, although folate concentrations were notably different, ranging from 7.9 to 23.6 pg/mL across the 4 cohorts (FHS, CHS, BLSA, and NHS), which had folate concentrations available (see Supplementary Table 2 under “Supplemental data” in the online issue). Total Hcy concentrations seem to vary significantly across the European cohorts. These differences might be explained in part by differences in age, sex distribution, medical and lifestyle factors, and technical differences in the measurement techniques. However, because we used normalized tHcy values for each population separately, and did not compare crude mean concentrations, these differences should not have affected the overall outcome of the meta-analysis.

The other important limitations pertain to the sex and ethnic distribution of the cohorts studied. More than 82% of the individuals from the 10 cohorts were female, and participants from 9 of the cohorts had a predominantly white ancestry, which meant that we had insufficient power to assess differences in gene associations by sex and our conclusions are restricted to populations of northwestern European ancestry.

Finally, results from the conditional analyses based on different assumptions about the penetrance model of the lead (eg, MTHFR) loci may be warranted in future studies. Similarly, a thorough evaluation of the potential effect of other confounding factors (eg, lack of adjustment for B vitamin status), which can only be speculated about currently, needs to be tested in future studies.

Interestingly, 3 of our novel associations in genes MMACHC, CUBN, FUT2, and in the confirmed association with MUT, have either been previously associated with variation in constitutive vitamin B-12 concentrations—MUT, FUT2, and CUBN—or in the case of MMACHC, are excellent candidates through association with Mendelian inborn errors of vitamin B-12 transport. Data from TwinsUK show that tHcy and vitamin B-12 concentrations are correlated (r = −0.09, P < 10−8). Thus, the simplest explanations for this could be that, because vitamin B-12 is an important cofactor in Hcy metabolism, any genetic perturbation of vitamin B-12 concentrations could have a pleiotropic effect downstream on circulatory concentrations of Hcy.

Finally, previous studies have shown that tHcy concentrations increase with decreasing glomerular filtration rate, although the precise mechanism by which tHcy concentrations is related to glomerular filtration rate and consequently to CAD has not been established (46, 47). Our results support the idea that it is not tHcy itself, but something that tHcy is a marker for, that is a causal risk factor for CAD. One possible candidate is renal impairment or perhaps more likely, a particular molecule that, such as tHcy, is increased in the setting of renal impairment. Two of our associations with tHcy, namely with variants in or in the vicinity of DPEP1 and SLC17A3, were previously shown to be associated with renal function. The gene DPEP1, which encodes for an enzyme that is highly expressed and active in the proximal convoluted tubules, is implicated in the renal metabolism of glutathione and its conjugates. Because DPEP1 deficiency leads to increased urinary excretion of cysteine—a precursor of tHcy—it has been hypothesized that genetic variation at DPEP1 affects renal activity by altering the renal handling of amino acids. We used data from the Rotterdam Study cohorts and adjusted our analyses for creatinine concentrations to test whether our association was driven by individuals with altered renal function, but this did not affect the results of the GWAS (data not shown). However, because creatinine concentrations serve as only a rough proxy for kidney function, we cannot completely rule out that this association is driven by kidney function.

The results of the current study extend the findings of previous studies of MTHFR C677T and CAD (14) by assessing the relevance for CAD of all available genetic determinants of tHcy concentrations, alone and in combination, and provide further refutation of the causal relevance of moderately elevated tHcy concentrations to CAD in white populations.

Study-specific acknowledgments are provided as part of the Supplementary Material.

The authors’ responsibilities were as follows—JBJvM and KRA: conceived and designed the study; JBJvM, GP, SMS, AH, KRA: researched the data, contributed to the discussion, wrote the manuscript, and reviewed and edited the manuscript; JBJvM, TT, SHV, IC, XY, AM, CN, LMR, KRA: contributed to the analysis; SB, JCB, HB, MBJ, CC, LF, A Hamsten, A Hofman, DJH, AG, ADJ, SK, EK, DPK, AALMK, JCC, FK, JL, BM, CN,
The sponsors of the study had no role in the study design, data collection, analysis, data interpretation, or writing of the report. Am J Hum Genet 2006;78:778–92.


