Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway

Aditi Hazra1,2,*, Peter Kraft1, Ross Lazarus2,3, Constance Chen1, Stephen J. Chanock4, Paul Jacques5, Jacob Selhub5 and David J. Hunter1,2

1Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA, 2Channing Laboratory, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA, 3Department of Ambulatory Care and Prevention, Harvard Medical School, Boston, MA, USA, 4Division of Cancer Epidemiology and Genetics, Department of Health and Human Services (DHHS), National Cancer Institute (NCI), US National Institutes of Health (NIH), Bethesda, MD 20892, USA and 5Vitamin Metabolism and Aging Laboratory and Nutritional Epidemiology Program, Jean Mayer US Department of Agriculture Human Nutrition Research Center on Aging, Tufts University, Boston, MA 02111, USA

Received May 21, 2009; Revised and Accepted September 4, 2009

Low plasma B-vitamin levels and elevated homocysteine have been associated with cancer, cardiovascular disease and neurodegenerative disorders. Common variants in FUT2 on chromosome 19q13 were associated with plasma vitamin B12 levels among women in a genome-wide association study in the Nurses’ Health Study (NHS) NCI-Cancer Genetic Markers of Susceptibility (CGEMS) project. To identify additional loci associated with plasma vitamin B12, homocysteine, folate and vitamin B6 (active form pyridoxal 5’-phosphate, PLP), we conducted a meta-analysis of three GWA scans (total n = 4763, consisting of 1658 women in NHS-CGEMS, 1647 women in Framingham-SNP-Health Association Resource (SHARE) and 1458 men in SHARE). On chromosome 19q13, we confirm the association of plasma vitamin B12 with rs602662 and rs492602 (P-value = 1.83 × 10^-15 and 1.30 × 10^-14, respectively) in strong linkage disequilibrium (LD) with rs601338 (P = 6.92 × 10^-15), the FUT2 W143X nonsense mutation. We identified additional genome-wide significant loci for plasma vitamin B12 on chromosomes 6p21 (P = 4.05 × 10^-08), 10p12 (P-value = 2.87 × 10^-9) and 11q11 (P-value = 2.25 × 10^-10) in genes with biological relevance. We confirm the association of the well-studied functional candidate SNP 5,10-methylene tetrahydrofolate reductase (MTHFR) Ala222Val (dbSNP ID: rs1801133; P-value = 1.27 × 10^-9), on chromosome 1p36 with plasma homocysteine and identify an additional genome-wide significant locus on chromosome 9q22 (P-value = 2.06 × 10^-8) associated with plasma homocysteine. We also identified genome-wide associations with variants on chromosome 1p36 with plasma PLP (P-value = 1.40 × 10^-15). Genome-wide significant loci were not identified for plasma folate. These data reveal new biological candidates and confirm prior candidate genes for plasma homocysteine, plasma vitamin B12 and plasma PLP.

INTRODUCTION

One-carbon metabolism comprises a network of biochemical reactions involved in the transfer of single-carbon moieties essential for purine and thymidylate synthesis necessary for maintenance of genomic integrity and remethylation of homocysteine for S-adenosylmethionine (SAM)-dependent DNA methylation reactions (1,2). Folate, a B-complex nutrient from diet and supplements, donates methyl groups to generate methionine from homocysteine. Plasma homocysteine (Hcy), an integrated marker of one-carbon metabolism, is inversely related to folate, vitamin B6 (3) and vitamin B12 and positively related to high alcohol consumption (4,5). Vitamin B12 is a necessary co-factor for the methionine synthase catalyzed
remethylation of Hcy into methionine. The biologically active form of vitamin B6, pyridoxal 5'-phosphate (PLP), catalyzes Hcy transsulfuration (6) and is involved in over 100 enzyme reactions in the synthesis and catabolism of neurotransmitters (7), amino acids, glycogen and lipids (8).

Epidemiologic studies have shown that individuals with reduced intake of folate (found in fruits, vegetables and legumes) or vitamin B6 (found in poultry, fish, meat, legumes, nuts and potatoes) or vitamin B12 (found in liver, shellfish, fish, poultry, eggs and dairy products (9)) have an inverse relationship with plasma levels of folate, vitamin B12 and Hcy (10,11). In addition, individuals with inadequate intake of folate and vitamin B12 are at an elevated risk of uracil misincorporation and vitamin B6 are an established risk predictor of cardiovascular diseases (12–16). In an increase in S-adenosylhomocysteine (AdoHcy) and is an established marker of the one-carbon metabolism pathway, specifically the focused on genes in the one-carbon metabolism pathway, associated with vitamin B12 levels in women (17). In a recent genome-wide association between common variants in reduced conversion of methylene-THF to methyl-THF and plasma Hcy, individuals with inadequate intake of folate and vitamin B12 are at an elevated risk of uracil misincorporation and vitamin B6 are an established risk predictor of cardiovascular diseases (12–16). In addition, individuals with inadequate intake of folate and vitamin B12 are at an elevated risk of uracil misincorporation into DNA and aberrant DNA methylation (17–20). Vitamin B12 deficiency (21), related to poor intestinal B12 absorption (9) or dietary deficiency, can lead to inactivation of methionine synthase and is associated with pernicious anemia (22), cardiovascular disease (23–25), cancer (26–32) and neurodegenerative disorders (33–35). Plasma PLP levels are significantly inversely associated with risk of colorectal cancer in men (36) and women (37).

Plasma Hcy is the integrated marker of the one-carbon metabolism pathway, with a heritability estimate of 44% (38) [heritability estimate range 8–57% (38–43)]. Until recently, common genetic variations reported to influence plasma Hcy, vitamin B12 and PLP, vitamin B12 and Hcy were focused on genes in the one-carbon metabolism pathway, specifically the Ala222Val polymorphism (23,44) (45) in 5,10 methylene-THF reductase (MTHFR), associated with reduced conversion of methylene-THF to methyl-THF (46,47). We previously identified a genome-wide significant association between common variants in FUT2 and plasma vitamin B12 levels in women (48). In a recent genome-wide analysis of data pooled from three genome-wide association study (GWASs) of 2931 persons and a replication study of 687 participants, Tanaka et al. confirmed this loci and reported additional loci that met genome-wide significance criteria for PLP. In the current study, we report a meta-analysis of three GWASs with 4763 participants (1658 women in CGEMS, SHARe Women and SHARe Men study populations with GW data analysis.}

### RESULTS

Genome-wide significant associations of P-value < 5 × 10^{-8} [corresponding to P < 0.05 corrected P-value after adjusting for approximately 1 million independent loci (50)] were identified in meta-analyses for log-transformed plasma Hcy, vitamin B12 and PLP, Age and the distributions of plasma vitamin B12 and homocysteine in the study populations are described in Table 1. As expected, plasma Hcy was found to be inversely related to plasma levels of folate, vitamin B12 and PLP.

#### Plasma homocysteine

The well-studied functional polymorphism in MTHFR Ala222Val (also reported as MTHFR C677T), rs1801133, on chromosome 1 [position (bp): 11,790,644] was associated with plasma Hcy (P-value = 1.27 × 10^{-8}; Fig. 1A; Table 2). Participants with the Ala/Ala (G/G) variant had lower plasma Hcy levels than variant carriers. Two other SNPs, rs12085006 [position (bp): 11,892,989] and rs1999594 [position (bp): 11,893,482], located 102 kb from MTHFR, had stronger associations with plasma Hcy (5.81 × 10^{-10} and 6.24 × 10^{-10}, respectively; Fig. 1A; Table 2). Participants homozygous for the rs12085006 G and rs1999594 A alleles had higher Hcy levels compared with variant carriers for each SNP. The rs12085006 A allele is correlated with the MTHFR 222Val allele (composite r^2 = 0.53; P < 0.0001) (51). After mutual

#### Table 1. Mean and median plasma one-carbon metabolite levels in the NHS CGEMS, SHARe Women and SHARe Men study populations with GW data

<table>
<thead>
<tr>
<th></th>
<th>NHS CGEMS, n = 1,658</th>
<th>SHARe Women, Exam 6*, n = 1,647</th>
<th>SHARe Men, Exam 6*, n = 1,458</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Mean, SD</td>
<td>58.92, 6.20</td>
<td>58.60,9.71</td>
<td>58.57, 9.70</td>
</tr>
<tr>
<td>Plasma HCY*, Mean, SD</td>
<td>11.52, 5.41</td>
<td>9.06, 3.87</td>
<td>10.54, 3.99</td>
</tr>
<tr>
<td>Plasma HCY Median</td>
<td>10.62 (4.04)</td>
<td>8.39 (3.32)</td>
<td>9.82 (3.55)</td>
</tr>
<tr>
<td>Plasma Folate*</td>
<td>10.02, 9.03</td>
<td>7.85, 5.32</td>
<td>6.54, 4.23</td>
</tr>
<tr>
<td>Plasma Folate Median</td>
<td>7.69, (7.44)</td>
<td>6.45 (5.80)</td>
<td>5.68 (4.22)</td>
</tr>
<tr>
<td>Plasma B12*, Mean, SD</td>
<td>473.90, 249.53</td>
<td>435.10, 181.80</td>
<td>398.59, 148.26</td>
</tr>
<tr>
<td>Plasma B12 Median</td>
<td>436.31 (224.0)</td>
<td>405.96 (211.91)</td>
<td>377.34 (172.68)</td>
</tr>
<tr>
<td>Plasma PLP*, Mean, SD</td>
<td>76.39, 90.55</td>
<td>83.14, 82.76</td>
<td>86.89, 84.05</td>
</tr>
<tr>
<td>Plasma PLP Median</td>
<td>46.70 (47.77)</td>
<td>57.55 (56.87)</td>
<td>61.72 (52.61)</td>
</tr>
</tbody>
</table>

*NH blood collection was during 1989–1990.

| Exam 6* refers to Framingham Heart Study laboratory exam 6 measurements (1995–1998). For all plasma, one-carbon metabolites exam 6 measurements were log-transformed for GWAS analysis.

<table>
<thead>
<tr>
<th>IQR, interquartile range.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma Folate measure in ng/ml; data were log-transformed for GWAS analysis.</td>
</tr>
<tr>
<td>Plasma Vitamin B12 measured in pg/ml; data were log-transformed for GWAS analysis.</td>
</tr>
<tr>
<td>Plasma PLP measured in pmol/ml; data were log-transformed for GWAS analysis.</td>
</tr>
</tbody>
</table>

and vitamin B6 in our study population (Supplementary Material, Table S1). No significant heterogeneity by sex or by study (Table 2) was detected for the genome-wide significant associations.
Table 2. Genome-wide significant associations for plasma one-carbon metabolites

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Position (bp)</th>
<th>Gene</th>
<th>rs</th>
<th>A/G or C/T</th>
<th>Genotype carriers</th>
<th>Plasma folate</th>
<th>Plasma vitamin B12</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9473558</td>
<td>6p</td>
<td>49,517,467</td>
<td>MUT</td>
<td>C/G</td>
<td>0.35</td>
<td>homozygous</td>
<td>0.07</td>
<td>0.15</td>
</tr>
<tr>
<td>rs9473555</td>
<td>6p</td>
<td>49,520,392</td>
<td>MUT</td>
<td>G/C</td>
<td>0.35</td>
<td>homozygous</td>
<td>0.07</td>
<td>0.15</td>
</tr>
</tbody>
</table>

**Plasma folate**

Overall, there were no genome-wide significant associations for plasma folate in the meta-analysis (Fig. 1B). The strongest associations were in chromosome 2 in the *FIGN* gene, rs982393 (P-value = 8.36 × 10^{-8}; data not shown) and rs2119289 (P-value = 1.03 × 10^{-7}; data not shown). Although not genome-wide significant, a moderate signal for SNP rs112012222Val variant carriers had lower plasma folate compared with the Ala/Ala variant (MTFR Ala222Val P-value = 1.14 × 10^{-3} from meta-analysis, data not shown). As expected (53), the MTFR Ala222Val SNP in *CBS*, *CPS1*, *MUT* and *NOX4* reported in the Women’s Genome Health Study by Paré et al. (52) were not genome-wide significant in our meta-analysis data (rs6586282 in *CBS*: P-value = 6.30 × 10^{-5}; rs7422339 in *CPS1*: P-value = 4.18 × 10^{-4}; rs4267943 in *MUT*: P-value = 6.43 × 10^{-1}; rs11018628 in *NOX4*: P-value = 4.35 × 10^{-3}).

**Plasma vitamin B12**

Genome-wide significant associations were identified on chromosomes 6, 10, 11 and 19 for plasma vitamin B12 (Fig. 1C; Table 2). Two SNPs, in LD, in *Methylmalonyl-CoA mutase (MUT)* on chromosome 6p21—rs9473558 SNP and rs9473555 [chromosome position (bp): 49,520,392 and 49,517,466, respectively]—were genome-wide significant in the meta-analysis for plasma vitamin B12 (P-meta = 4.91 × 10^{-8}, 4.05 × 10^{-8} respectively, Fig. 1; Table 2). Participants homozygous for the rs9473558 C and rs9473555 G alleles had higher B12 levels compared with participants homozygous for the major allele. We identified an association (P = 2.87 × 10^{-9}; Fig. 1C; Table 2) between rs18012222 in *CUBN* (Cubilin) on chromosome 10p12 and plasma vitamin B12 levels in the meta-analysis of 4763 participants. The SNP is in exon 8 of *CUBN* and codes for an amino acid change Ser253Phe [chromosome position (bp): 17,196,157].
Participants homozygous for the rs1801222 G allele had higher B12 levels. The meta-analysis indicated several signals in the P < 10^-8 range for other SNPs in CUBN and on chromosome 10p12 in the neighboring TRDMT1 gene (DNMT2), which may be explained by LD. A genome-wide significant association was also observed for rs526934 in the transcobalamin 1 (TCN1) gene (P-value = 2.25 × 10^-8) located on chromosome 11q11 [position (bp): 59,390,069]. Variant carriers of the rs526934 G allele had lower vitamin B12 levels when compared with individuals with the homozygous AA variant.

The most significant association in the meta-analysis data (P-value = 1.83 × 10^-15; Fig. 1C; Table 2) was for the rs602662 SNP in FUT2. The FUT2 nonsense SNP W143X, rs601338 [chromosome 19q13 position (bp): 53,898,486], the determinant of FUT2 secretor status, was the second most significant association with a P-value = 6.92 × 10^-15. Participants homozygous for non-secretor variants had higher B12 levels than carriers of the secretor genotype. The reported rs492602 SNP in NBPF3, upstream of ALPL, had a P-value = 4.30 × 10^-11 with plasma PLP in these data (composite r^2 for rs1256335 and rs4654748 is 0.16, P < 0.0001).

DISCUSSION

We identified genome-wide significant associations for plasma homocysteine, plasma vitamin B12 and plasma PLP. SNPs in chromosome 6p21, 10p12, 11q11 and 19q13 showed strong signals for plasma vitamin B12. SNPs in chromosomes 1p36 and 9q22 were strongly associated with plasma homocysteine. We identified genome-wide significant associations between SNPs in chromosome 1p36 and plasma PLP.

Plasma homocysteine

Chromosome 1p36. MTHFR is a key enzyme required for DNA synthesis that catalyzes the irreversible transformation...
of 5,10-MTHF into 5-MTHF (54). The TT variant alleles of the well-known MTHFR 677 (amino acid change: MTHFR Ala222Val) polymorphism are related to a 30% reduction in enzyme activity (55). A decrease in MTHFR activity associated with the 222 Val/Val (677TT) polymorphism shifts the folate pool, leading to an elevation in 5,10-methyleneTHF. The MTHFR Ala222Val SNP results in a moderate (up to 15–19%) increase in mean Hcy levels between the Ala/Ala and Val/Val variants (23), and this effect is greatly attenuated in those individuals with mid-range to high folate levels (56). Rs12085006 and rs1999594, located 102 kb from MTHFR, had an even stronger association with plasma Hcy in our meta-analysis. Rs1999594 had a genome-wide significant association ($P = 8.5 \times 10^{-12}$) with plasma Hcy in the Women’s Genome Health Study (WGHS), but only rs180133 showed evidence for non-additive effects (52). The rs1999594 SNP was reported by Tanaka et al. (49) to be associated with plasma folate in a meta-analysis of 2931 participants ($P = 1.12 \times 10^{-7}$), although the association did not reach genome-wide significance in their report on plasma folate and was not reported for plasma Hcy. Given the strong correlation between rs1999594 (or rs12085006) and rs180133 and the modest evidence for association between rs1999594 (or rs12085006) and plasma Hcy after adjusting for MTHFR Ala222Val, we cannot conclude that these two SNPs mark distinct loci associated with plasma Hcy.

Chromosome 9q22. We identified a genome-wide significant association between plasma Hcy and a variant in GPR51. Carriers of the rs10986018 T allele had lower plasma vitamin $B_{12}$ compared with carriers of the CC variant. Expression of GPR51 may be regulated by nicotine (57). Epidemiologic studies have reported increased homocysteine levels in smokers compared with non-smokers (58–62). The association with plasma Hcy and rs10986018 warrants future study on plasma Hcy and effect modification by smoking status.

Plasma folate

We did not identify any genome-wide significant association with plasma folate in this meta-analysis, although the MTHFR Ala222Val variant had the expected direction of association (53,56), with a modest P-value. FHS SHARe Exam 6 data available in dbGaP had the largest sample size for all four plasma one-carbon metabolites and laboratory assays comparable to the NHS CGEMS study population. All NHS CGEMS participants had plasma folate measured on blood drawn prior to folic acid fortification (1989–1990). However, the Framingham SHARe laboratory exam 6 measurements were taken during 1995–1998. While the majority of the SHARe participants included in this analysis had measurements prior to the US mandate for folic acid fortification in 1998, evidence suggests that the fortification may have started in 1996–1997. Pre-fortification plasma folate levels in the prior Framingham SHARe laboratory exam, available in dbGaP, were measured using a microassay and would not be equivalent to the HPLC assay used in the NHS CGEMS data. We also recognize that there may also be misclassification due to folic acid supplementation.

Plasma vitamin $B_{12}$

Chromosome 6. MUT, located on chromosome 6p21, encodes the mitochondrial enzyme methylmalonyl Coenzyme A (CoA) mutase. MUT requires 5-prime-deoxyadenosylcobalamin (AdoCb, a coenzyme form of vitamin $B_{12}$) to catalyze the isomerization of $L$-methylmalonyl–coenzyme A (CoA) to succinyl-CoA, in the mitochondria, independent of folate status. Mutations in methylmalonyl-CoA mutase, especially in the C-terminus, result in methylmalonic aciduria (63–67). A patient with the mut(-) MMA phenotype due to the MUT Gly717Val variant was also homozygous for the His532Arg (rs9473558) polymorphism (68). The T variant of rs9473558 has reduced vitamin $B_{12}$ levels in our data and may be associated with decreased conversion of methylmalonyl CoA to succinyl CoA. The rs9473558 SNP in MUT was associated with plasma homocysteine ($P = 2.1 \times 10^{-7}$) in the Women’s Genome Health Study (WGHS) participants (52). The rs9473558 SNP was positively associated with plasma homocysteine in the WGHS study ($\beta = 0.020$), but inversely associated with plasma vitamin $B_{12}$ in our meta-analysis ($\beta = -0.04$), which is consistent with the inverse relationship between homocysteine and vitamin $B_{12}$. The expression of MUT was associated with the rs9473555 SNP with a LOD score = 7.095 [enzyme activity for the $G$-allele is reduced by 0.42 standard deviation; mRNA by SNP browser 1-0-1 database (69)]. The preliminary association with rs9473558 or rs9473555 in the MUT gene and plasma vitamin $B_{12}$ should be replicated in other studies. Future studies on the relationship of rs9473555 or rs9473558 and MUT gene expression are warranted.

Chromosome 10. Cubilin (CUBN) is located on chromosome 10p12.3. The 460 kd multiligand hydrophobic CUBN protein binds to intrinsic factor-cobalamin (Cbl-IF) complexes with a high affinity and is expressed in both kidney and ileal epithelial cells. Mutations in CUBN are associated with megaloblastic anemia and characterized by selective intestinal vitamin $B_{12}$ malabsorption (70). The P1297L mutation in CUBN causes a 5-fold decrease of the binding domain affinity for Cbl-IF (the IF-Cbl binding region includes amino acids 928–1386). The rs1801222 (Ser253Phe) variant was associated with lower vitamin $B_{12}$ levels in this study and may decrease the binding and transport of vitamin $B_{12}$ in the ileum. Another SNP in CUBN reported by Tanaka et al. (49) with a P-value of $1.11 \times 10^{-5}$, rs11254363, had a P-value of $8.65 \times 10^{-5}$ in our meta-analysis. The neighboring gene, TRDMT1, located on 10p15.1 is a methyltransferase that methylates the aspartic acid transfer RNA molecule (tRNA(Asp)) (71). A recent study on spina bifida and folate-related genes reported an association for different SNPs in CUBN and TRDMT1 with increased serum RBC folate levels in controls (72). Further research is warranted to understand the LD pattern across the 305.3 kb CUBN gene and the cis relationships of the genome-wide significant SNPs with CUBN and TRDMT1.

Chromosome 11. TCN1, on chromosome 11q11, is a vitamin $B_{12}$ binding protein that promotes the cellular uptake of vitamin $B_{12}$ by receptor-mediated endocytosis. The protein
is a major constituent of secondary granules in neutrophils and facilitates the transport of cobalamin into cells (73). The meta-analysis and replication joint P-value for rs526934 from Tanaka et al., was $P$-value$=1.51 \times 10^{-6}$. The TCN1 rs526934 $G$ variant may reduce transport of cobalamin, resulting in lower plasma vitamin B$_{12}$ levels.

**Chromosome 19.** We previously reported an association with variants in $FUT2$ on chromosome 19q13 and plasma vitamin B$_{12}$ [rs602662: $P$-value$=3.52 \times 10^{-15}$; rs492602: $P$-value$=5.36 \times 10^{-17}$ in the combined NHS CGEMS and colorectal neoplasia data sets; (48)]. In the SHARe women and the SHARe men data sets, we replicated our genome-wide significant finding for rs602662 (SHARE combined $P = 5.07 \times 10^{-8}$, Table 2 presents separate estimates and standard error, by SHARe Men and Women) and in this meta-analysis of CGEMS and SHARE data, we confirm the association of these SNPs with plasma vitamin B$_{12}$. In this study of plasma vitamin B$_{12}$, we confirm our previously reported hypothesis that $FUT2$ $W143X$, a nonsense mutation (74) and determinant of $FUT2$ secretor status, is the causal variant for the association with plasma B$_{12}$ levels (rs601338: $P$-value$=6.92 \times 10^{-15}$). Individuals with the non-secretor status variant had higher plasma vitamin B$_{12}$ levels compared with carriers of the secretor status genotypes. Reduced vitamin B$_{12}$ absorption (75–79) in carriers of the secretor genotype may be a consequence of susceptibility to $H. pylori$ infection compared with individuals with the non-secretor genotype. The association with rs602662 was replicated by an independent group (49) with a $P$-value of $2.43 \times 10^{-12}$ in a meta-analysis of three GWASs ($n = 2930$ participants).

**Plasma PLP**

Mutations in the $ALPL$ gene, on chromosome 1p36, are associated with hypophosphatasia. Increased circulating concentrations of PLP in all clinical forms of hypophosphatasia have been reported by Whyte et al. (80,81). The findings suggested that tissue-non-specific alkaline phosphatase is involved in vitamin B$_{6}$ metabolism. The $ALPL$ enzyme may function as an ectoenzyme to regulate extracellular concentrations of PLP (80). The rs1256335 $C$ allele had lower plasma PLP levels compared with the homozygous $A$ allele, possibly due to increased hydrolysis of PLP (82). The rs4654748 $C$ allele, in the neuroblastoma breakpoint family, member 3, (NBPF3, alternate name AE2) gene had lower plasma PLP levels compared with the homozygous major allele.

**Comparability of plasma phenotypes in the study population**

The NHS and FHS study populations have comparable metabolite levels for conducting a meta-analysis of GWA data of plasma one-carbon metabolites and the biomarker measurement in these studies were all done in a single laboratory. However, one limitation is that the NHS CGEMS participants were all postmenopausal women, whereas SHARE Women participants included of both pre- and postmenopausal women. Plasma Hey is lower in premenopausal women than men, but after menopause, homocysteine levels in women increased to approach those in men, consistent with our data. All NHS and the majority of the FHS blood samples for participants in this analysis were collected prior to the introduction of folic acid fortification of breads, cereals, flours, corn meal, pasta products, rice and other cereal grain products sold in the USA (83,84). Therefore, the genetic variability signal should not be attenuated by this recent environmental influence.

In summary, for plasma vitamin B$_{12}$, we replicated our finding for $FUT2$ (48) and identified genome-wide significant SNPs in biological candidate genes: $TCN1$, a vitamin B$_{12}$ binding protein; $MUT$, which converts methylmalonyl CoA to succinyl CoA; and $CUBN$, the receptor for intrinsic factor-vitamin B$_{12}$ complexes. For plasma homocysteine, we observed genome-wide significant associations with the $MTHFR$ functional SNP Ala222Val and a possible new independent locus 102 kb upstream of $MTHFR$. For plasma PLP, we noted genome-wide significant associations in $ALPL$. These data reveal new biological candidates and confirm prior candidate genes for plasma homocysteine, plasma vitamin B$_{12}$ and plasma PLP.

**MATERIALS AND METHODS**

**Ethics statement**

All participants provided written informed consent, and local institutional review boards approved the study protocols.

**Study populations**

We conducted a meta-analysis of genome-wide data in 1658 women genotyped with the HumanHap500 as part of the Cancer Genetic Markers of Susceptibility Project (CGEMS) and in 1682 women genotyped with the Affymetrix 500K mapping array and the Affymetrix 50K Supplementary array as a part of the SNP Health Association Resource (SHARE data set, part of NHLBI FHS) to identify novel loci that influence plasma one-carbon metabolite levels. All participants were of self-reported European ancestry.

**The Nurses’ Health Study (NHS)**

The NHS was initiated in 1976, when 121,700 US registered nurses between the ages of 30 and 55 returned an initial questionnaire reporting medical histories and baseline health-related exposures. Every 2 years, follow-up questionnaires are mailed to the participants (questionnaire response rate >90% for all follow-up cycles). Diet was assessed in 1986, 1990, 1994 and 1996 with a semiquantitative food frequency questionnaire. Between 1989 and 1990, blood samples and relevant sample collection information were collected from 32,826 women.

The CGEMS project (http://www.cgems.cancer.gov) is an NCI initiative to conduct genome-wide association studies (GWAS) to identify genes involved in breast and prostate cancer. The initial CGEMS scan, designed and funded only to study the main effect of SNP variants on breast cancer risk in postmenopausal women, has been completed (85).
A subset of the participants in the CGEMS project also has been assayed for plasma vitamin B₁₂. The NHS CGEMS participants had similar laboratory measurement of the plasma one-carbon metabolites as well as host characteristics and blood draw features (fasting status and year and month of blood draw).

**Framingham Heart Study (FHS)**

The FHS began in 1948 with the enrollment of two-thirds of the adult population of Framingham, Massachusetts (86), including 2873 women aged 28–62 years. In 1971, 5124 offspring of the original cohort members and offspring spouses were enrolled in the Framingham Heart Study, including 2641 women ranging in age from 12 to 60 years. A data set consisting of repeated and highly accurate measurements of many different phenotypes collected over multiple exam biennial cycles is available for each participant. The NHLBI Framingham SNP-Health Association Resource (SHArE) is a large-scale GWAS on the FHS original, offspring and third-generation cohorts with data accessible to the research community (87). In the Framingham cohort, the SNP Health Association Resource (SHArE) project genotyped 9274 Caucasian participants with the Affymetrix 500K mapping array and the Affymetrix 50K supplemental array (Affymetrix, Santa Clara, CA, USA). The 8508 samples from these participants had a sample call rate ≥97% (88,89). Of the participants from the Framingham original and offspring cohorts, with measured plasma one-carbon metabolites at laboratory exam 6, between 1995 and 1998, the SHArEs scan included 1647 (16) Caucasian women and 1458 Caucasian men. Exam 6 had the largest sample size for all four plasma one-carbon metabolites and laboratory assays comparable to the NHS CGEMS study population. Information on Framingham SHArE can be found at [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v1.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v1.p1).

**Laboratory assays**

Plasma levels of homocysteine for both the NHS CGEMS and FHS SHArE studies were measured by high-performance liquid chromatography, with fluorescence detection (90). Plasma levels of folate and vitamin B₁₂ for the NHS CGEMS and FHS SHArE participants were measured using a radioassay kit (Bio-Rad, Richmond, CA, USA) according to the manufacturer’s instructions. Plasma PLP levels were measured by an enzymatic procedure using radioactive tyrosine and the apo-enzyme tyrosine decarboxylase.

All assays were conducted at the Jean Mayer U.S. Department of Agriculture Human Nutrition Research Center on Aging at Tufts University. Plasma samples for matched case–control sets were always positioned next to each other, in random order, in boxes sent to the lab and were assessed in the same batch to minimize the impact of laboratory error due to batch drift. Quality control (QC) samples were also submitted with each batch and were randomly placed throughout the boxes (two assay batches were used for NHS CGEMS participants). Laboratory technicians were blinded to case, control and QC status of the samples. QC samples consisted of replicates of two pools of plasma. One QC sample was assayed per 10 study samples (27,37). No significant differences in mean levels of plasma homocysteine, folate, vitamin B₁₂ or PLP were detected by breast cancer case–control status in the NHS. The mean coefficient of variation for 75 pairs of replicate 6.5% for folate, 7.9% for homocysteine, 7.3% for B₁₂ and 7.2% for PLP (27).

**Reliability of laboratory assays of plasma levels**

In an analysis of 188 NHS participants, the correlation coefficients between folate intake calculated from the 1980 food frequency questionnaire and erythrocyte folate concentrations in 1987 were 0.55 for folate from foods and supplementation and 0.38 for folate from food only (18). Similar correlation coefficients were noted for plasma folate (0.55 for total folate; 0.35 for dietary folate) (18,27). The intraclass correlation coefficient over a 3-year period on 82 participants was 0.63 for plasma folate, suggesting reproducibility of these measurements over time.

**Illumina SNP genotyping and quality-control analysis**

Genotyping in the CGEMS project was carried out using the Illumina Infinium Sentrix HumanHap550 bead chip, which contains 556 566 SNPs derived from the HapMap phase I and II data (85). The polymorphisms on the bead chip are selected to be direct surrogates (R², a measure of pair-wise SNP correlation, >0.8) of more than 75% of the common polymorphisms genotyped as part of the HapMap Phase II project. Overall, genotyping success was 98.45% of the possible genotypes in the first attempt and 99.27% after failed samples were repeated. Quality control of genotyping reproducibility was 99.99%. SNPs with significant deviations from Hardy–Weinberg (HWE) proportions (5% of SNPs at the level of P = 0.05 and 1.28% at P = 0.01) were not excluded from the analysis, as the tests for association applied to these data are valid in the presence of departure from HW proportions. The NHS cases and controls included in CGEMS are of self-reported European ancestry (85). Four samples (three cases and one control) were removed because of intercontinental admixture. In the CGEMS data, we saw no evidence for systematic bias in the distribution of P-values for analyses with and without further adjustment for residual population structure using four principal components of genetic variation compatible with no confounding of SNP-metabolite associations due to population stratification (85,91). A total of 1658 NHS participants had GWAS and plasma vitamin B₁₂ data.

**Affymetrix SNP genotyping and quality-control analysis**

Genotyping in the SHARe study (88,89) was carried out using the Affymetrix (Santa Clara, CA, USA) GeneChip Human Mapping 500K mapping array and supplemental 50K mapping array (HuGeneFocused50K) with gene-centric and coding SNPs. The Human Mapping 500K Array Set is comprised of two arrays, each capable of genotyping on average 250 000 SNPs (approximately 262 000 for Nsp arrays and 238 000 for Sty arrays).
Association analyses were restricted to individuals with ≥97% genotyping call-rate. SNP results were filtered on Hardy–Weinberg equilibrium $P$-value of $1 \times 10^{-6}$. SNP call-rate of 95% and minor allele frequency (MAF) of 0.01. After cleaning the genotype data, 347 K SNPs were available for the analysis.

Statistical methods

We tested the association between log-transformed plasma one-carbon metabolite levels and each SNP using linear regression adjusted for age. SNP genotypes were coded as counts of minor alleles (additive coding) unless otherwise specified. We adopted an additive coding because it has nearly optimal power for a range of realistic genetic models, with the exception of rare recessive effects, which we have little power to detect (92). In CGEMS, we further adjusted for principal components of genetic variation, calculated with the EIGENSTRAT software (91).

The SHARe data were analyzed using linear mixed effects models to account for residual familial correlation; aside from fixed effects (age, SNP genotype), this model includes individual random effects with a correlation matrix equal to the kinship matrix times a scalar variance parameter. These analyses were implemented using the lmkin function in the R package kinship.

We imputed SNP genotypes using the hidden Markov model implemented in MACH 1.0 (http://www.sph.umich.edu/csg/abecasis/MACH/index.html) for each CGEMS and SHARe separately. For the SHARe data, the hidden Markov model was trained using unrelated founders (922 women and 692 biologically unrelated men); the fitted model was then used to infer missing genotypes for the remaining subjects. The certainty of the imputation is dependent on the stretches of haplotypes shared between the reference panel (HapMap Phase II CEU population) and the study population. The imputation result at each of the imputed SNPs, per study participant, can be expressed as a set of three genotype probabilities, that vary between 0 and 1 per genotype and sum to 1, or as an ‘allele dosage’, which is defined as the expected number of minor alleles (additive coding) unless otherwise specified. We counted the minor alleles (additive coding) unless otherwise specified. We saw no evidence for systematic bias in the distribution of $P$-values from the meta-analyses (Supplementary Material, Fig. S1). The $P$-for-heterogeneity was reported in Table 2.

GLM was used to test the independence of the estimates of multiple SNPs in a model. Composite LD was calculated according to the notation of Weir (51).

Our analyses do not explicitly account for the ascertainment of subjects on the basis of primary disease (breast cancer for the NHS CGEMS; Framingham SHARe data are from a cohort study and not ascertained on disease). This approach is appropriate, since our loci of interest are not associated with primary disease evaluated in the CGEMS NHS participants (85,95). All $P$-values are based on two-sided tests. All statistical analyses were performed with SAS (version 9.1; SAS Institute, Cary, NC, USA), R project (http://www.r-project.org) and PLINK (96).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

In this study, all plasma assays for CGEMS and SHARe were conducted at the Jean Mayer U.S. Department of Agriculture Human Nutrition Research Center on Aging at Tufts University. This manuscript does not necessarily reflect the opinions or views of Boston University or the NHLBI.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Institutes of Health [research grant numbers. U54 CA100971, P01 CA87969, P01 CA55075, U01 CA098233, R01 CA 065725, R01 CA070817 and R03 CA133937].

The Framingham Heart Study is conducted and supported by the National Heart, Lung and Blood Institute (NHLBI) in collaboration with Boston University.

REFERENCES


