A Common Polymorphism in \textit{HIBCH} Influences Methylmalonic Acid Concentrations in Blood Independently of Cobalamin


Methylmalonic acid (MMA) is a by-product of propionic acid metabolism through the vitamin B$_{12}$ (cobalamin)-dependent enzyme methylmalonyl CoA mutase. Elevated MMA concentrations are a hallmark of several inborn errors of metabolism and indicators of cobalamin deficiency in older persons. In a genome-wide analysis of 2,210 healthy young Irish adults (median age 22 years) we identified a strong association of plasma MMA with SNPs in 3-hydroxyisobutyryl-CoA hydrolase (\textit{HIBCH}), and found that these SNPs were associated with MMA concentrations in newborns. The association remained significant across the gestational trimesters and in newborns. HIBCH is unique to valine catabolism.

Introduction

Methylmalonic acid (MMA) is a by-product of the reaction catalyzed by methylmalonylCoA mutase (MUT), one of two enzymes in mammals that use vitamin B$_{12}$ (cobalamin) as a cofactor. This mitochondrial enzyme plays an important role in the metabolism of odd-chain fatty acids and the branched-chain amino acids valine, leucine, and isoleucine. Catabolism of these molecules produces methylmalonylCoA which is converted to succinylCoA via MUT activity. If MUT does not function effectively, the MMA concentration in blood increases. Rare high-penetration mutations in \textit{MUT} (MIM: 609058) or in enzymes affecting the transport or synthesis of 5’adenosylcobalamin cause a number of inborn errors of metabolism characterized by methylmalonic acidemia. However, there are currently no data to suggest that genetic variation in \textit{MUT} determines plasma MMA levels outside of rare inborn errors of metabolism. Furthermore, there are no data on the relative contributions of branched-chain amino acid or fatty acid oxidation pathways to circulating MMA concentrations in healthy individuals. Transient or persistently elevated circulating MMA levels are observed in individuals with severe mutations in \textit{ALDH6A1} (MIM: 603178), which encodes methylmalonate semialdehyde dehydrogenase, and \textit{ACSF3} (MIM: 614245), which encodes a mitochondrial AcylCoA synthetase considered to have malonylCoA and methylmalonyl CoA activities, and in mitochondrial depletion diseases such as succinate Co-A ligase deficiency due to mutations of mitochondrial \textit{SUCLG1} (MIM: 611224).

In contrast to diseases of newborns that result in elevated MMA concentration, cobalamin deficiency is relatively common among the elderly; estimates of deficiency range from 6% to over 40% in this population. Untreated cobalamin deficiency can lead to megaloblastic anemia or irreversible neurological damage. The MUT enzyme is extremely sensitive to cobalamin deficiency, and circulating MMA concentration is regarded as a sensitive and specific marker of cobalamin status, such that plasma concentrations within the normal reference range (<0.27 \muM) exclude cobalamin deficiency and...
concentrations >0.5 μmol/L almost invariably indicate deficiency.13,14 However, MMA concentration is influenced by age and renal impairment.15–17 Many older persons have marginally elevated MMA concentrations in the range of 0.27–0.37 μmol/L.18 This elevation is not associated with the status of other cobalamin biomarkers, such as the serum total cobalamin or plasma homocysteine. Circulating MMA concentrations in excess of 0.37 μmol/L are used to define inadequate cobalamin status;14,19–21 nevertheless, in one study of older subjects, known determinants of MMA levels (age, plasma creatinine, and plasma total cobalamin) accounted for less than 17% of the variation in this trait.16 Several reports have demonstrated transiently high MMA concentration in plasma and urine of neonates up to 6 months old,22 but it remains unclear whether this is related to low cobalamin status or whether other processes involved in MMA handling that are independent of vitamin B12 function are not fully developed.23 Identification of genetic factors that influence circulating MMA concentration in healthy individuals would greatly help in interpreting MMA levels that are elevated but of unknown clinical significance.

To identify loci that influence the plasma MMA concentration in healthy individuals, we conducted a genome-wide association study (GWAS) in a cohort of healthy young Irish adults aged between 18 and 28 years. Our study of this ethnically and environmentally homogeneous group identified variants that were also associated with circulating MMA levels in two Irish replication cohorts (older adults and pregnant women). We present evidence that the functional variant underlying the strongest association signal acts at the level of transcription.

Material and Methods

Recruitment of the Subjects for the GWAS
The Trinity Student Study (TSS) population has been described previously.24–26 The cohort was established over one academic year (2003–2004) for the purpose of exploring the heritability of quantitative traits. A total of 2,524 eligible students attending the University of Dublin, Trinity College (TCD) were invited to participate in the study. Eligibility criteria included being between 18 and 28 years old, having no current serious medical condition, and being of Irish ethnicity based on geographic origins of grandparents. All participants gave a non-fasting 30 mL venous blood sample and completed a health and lifestyle questionnaire in which information on age, gender, height, weight, medical history and medications, smoking, alcohol intake, and dietary habits was collected. Detailed supplement and fortified food intake data were additionally collected as part of the questionnaire. Participants were paid a modest honorarium for completing the study. Fifteen students did not return a questionnaire and were excluded. One student was found to have participated twice and the second blood sample and questionnaire were excluded. Samples from 2,508 participants with questionnaire data went forward for metabolite analysis and DNA extraction. Ethical approval was obtained from the Dublin Federated Hospitals Research Ethics Committee, which is affiliated with TCD, and was reviewed by the Office of Human Subjects Research at the NIH. Written informed consent was obtained from the participants when they enrolled.

Recruitment of the Older Adult Replication Cohort
The replication cohort consisted of 5,186 subjects recruited as part of the Trinity, Ulster, Department of Agriculture (TUDA) Study, a study of older Irish adults designed to investigate nutritional factors, related gene-nutrient interactions, and a range of health and lifestyle factors in the development of chronic diseases of aging in non-institutionalized adults. Recruitment was carried out in Dublin and in Northern Ireland between September 2008 and December 2012 from either hospital outpatient clinics or the community, and subjects were deemed eligible if they were aged over 60 years, of Irish parents, and without a prior diagnosis of dementia. Detailed health, lifestyle, and clinical information, including medication use, smoking status, physical activity, and vitamin supplement usage, was obtained for all subjects. Further details of the cohort have been published previously.27 Cobalamin status was assessed with both serum total cobalamin and serum holotranscobalamin (holoTC) measurements on all subjects. Plasma MMA concentration was measured on a subset of 1,511 individuals, based on their cobalamin status. All individuals who had either serum cobalamin ≤ 148 pmol/L or holoTC ≤ 30 pmol/L (n = 856) and a random (generated from subject ID) selection of individuals with both serum cobalamin and holoTC above the respective cut-offs (n = 655) were assigned to plasma MMA analysis in order to perform a quantitative trait analysis of plasma MMA concentration on the pooled samples. The study was conducted according to the guidelines in the Declaration of Helsinki, and ethical approval was granted by the relevant authorities in each jurisdiction: the Research Ethics Committee of St. James’s Hospital and The Adelaide and Meath Hospital, Dublin, and the Office for Research Ethics Committees, Northern Ireland (ORECNI; reference 08/NI/RO3113) with corresponding approvals from The Northern and Western Health and Social Care Trusts, Northern Ireland. Participants in all studies provided written informed consent at the time of enrollment.

Recruitment of the Longitudinal Pregnancy Cohort
A total of 201 women were recruited between 2004 and 2005 at their first ante-natal visit to the Coombe Women and Infants University Hospital in Dublin, Ireland. They agreed to participate in a longitudinal study involving collection of non-fasting blood samples at 14 weeks, 24 weeks, and 34 weeks of pregnancy and a cord blood sample upon delivery. Although time points are missing for some women, and cord blood samples were only available from 117 deliveries, all available samples were retained for analysis. All participants had uncomplicated pregnancies and gave birth to infants free of any major health issues. Participants’ blood samples were used to extract genomic DNA for genotyping and to measure the concentrations of circulating metabolites in plasma. All participants gave written informed consent, and ethical approval was granted by the Coombe Women and Infants University Hospital Research Ethics Committee and the Office of Human Research Subjects at the NIH.

Samples for the TSS GWAS discovery cohort and for both replication cohorts were rendered anonymous prior to analysis.

Metabolite Analysis
Non-fasting blood samples were collected into EDTA, lithium heparin, and clotting tubes. Samples for MMA, cobalamin, holoTC,
and total homocysteine (tHcy) analysis were processed within 3 hr of collection and were stored below –80°C, until analyzed. Plasma MMA concentration was measured by Bevital, Bergen, with gas chromatography mass spectroscopy. 28 Plasma MMA measures were obtained for 2,483 of 2,508 participants. Missing plasma MMA values were due to an insufficient sample for the initial measure or its repeat in the case of a high coefficient of variation. Serum cobalamin was measured with colistin-resistant L. delbrueckii by microbiological assay. 29 HoloTC and plasma homocysteine were measured via automated AxSym-based immuno-fluorescence methods. 30,31 Inter-assay CVs for all methods were as follows: MMA, 8.1%; serum total cobalamin, 10.6%; serum holoTC, 9.4%; plasma homocysteine, 2.2%. Circulating MMA concentrations are reported to be stable during long-term storage below –40°C and are not affected by recent food intake. 32

**Blood DNA Extraction and Genotyping of the TSS**

We extracted genomic DNA from all samples by using QIAGEN QIAamp DNA Blood Mini Kits. In the TSS, eighteen samples with low quantity DNA were excluded and 2,490 samples were sent for genotyping. Genome-wide SNP genotyping was conducted at the Center for Inherited Disease Research (CIDR) in Baltimore with Illumina 1M HumanOmniI-Quad_v1-0_B chips. High-quality genotypes were obtained on 2,438 study samples, 14 randomly selected duplicates, and HapMap controls. The blind duplicates had a concordance rate of 99.997%, and the HapMap samples had a 99.71% concordance rate.

**Genotyping Quality Control and Data Preparation for GWAS**

In a preliminary analysis for anomalies, ten subjects were excluded because of gender discrepancy between self-report and genotypes (n = 7) and abnormal sex chromosomes (n = 3; one XXY male and two XX/XO mosaic females). 16 subjects with incomplete anthropometric data were removed. Siblings from 77 families reported entry into the study. These included two siblings from 74 families and three siblings from three families. Pair-wise tests for relatedness and population stratification within all the genotyped samples were performed with PLINK v.2.21 and a relatively independent but otherwise random subset of 25,718 SNPs (all selected SNPs had pair-wise intermarker linkage disequilibrium [LD] $R^2 < 0.04$ for both short and long-range blocks of SNPs, minor-allele frequency [MAF] > 0.1 and genotype call rates > 0.99). These analyses confirmed the self-reported sibling relationships, and identified additional pairs of individuals who appeared closely related, to the level of $1^{st}$ cousins. In order to reduce effects of relatedness, only one individual from each family was retained. In total, 175 individuals were removed (80 siblings and 95 $1^{st}$ cousins). Finally, principal components (PC) analysis was performed with the EIGENSOFT 3.0 software, 23 and this same low-LD subset of 25,718 SNPs to detect potential population stratification. Scores from the first two PCs (PC1 and PC2) were examined, and two samples were detected as outliers from the main-population cluster of samples, suggesting that their ancestry might be different than that of the remainder of the cohort (6 SD away from the mean). These two samples were excluded from further analyses.

Otherwise, the sample set was remarkably homogeneous for genetic variation, consistent with accurate self-report of participants and their Irish ethnicity. No significant evidence of population stratification (no significant principal components) was observed in these data after dropping the two outlier samples. As such, no adjustment was made for stratification in the association analyses. The full data preparation process resulted in a final dataset of 2,235 study samples in the GWAS association analysis.

Quality control assessment was performed based on genotyping data for 1,008,829 SNPs. SNPs and/or samples were excluded in two stages. In stage 1, SNPs that had less than 95% call rates were dropped and then samples with less than 97% call rates or that showed cryptic relatedness were dropped (57 samples total). In stage 2, SNPs were dropped that had (1) less than 98% call rate, (2) Mendelian errors when we used Hapmap trios, (3) discordant markers when we used Hapmap controls, and (4) discordant markers from one or more pairs based on study duplicates. SNPs with extreme deviation from the Hardy-Weinberg equilibrium ($p < 1 \times 10^{-8}$) were flagged for future reference and remained in the analysis. We also measured LD in our dataset and examined LD plots by using the HaploView program. 33 The data were checked for batch effects with several methods from published GWAS quality control guidelines. 34 No significant batch effects were detected. These methods included (1) descriptive statistics per batch (average MAF, average genotyping call rate across all SNPs for each plate) and (2) basic allelic chi-square association test (1 df) using the “–loop-assoc” command in PLINK; results compared each batch to the others and at random (median $p$ value, number of significant results at $p$ value thresholds of $< 1 \times 10^{-3}$ and $< 1 \times 10^{-4}$). The final dataset contained 758,443 genotyped SNPs (3,512 of these SNPs had Hardy-Weinberg equilibrium $p < 1 \times 10^{-4}$ after the exclusion of monomorphic and low MAF SNPs (MAF < 0.01) for 2,235 samples. TSS genotype and phenotype data have been deposited in dbGAP (dbGAP: phs000789.v1.p1) Collaborative Study of Genes, Nutrients and Metabolites (CSGNM).

**Imputation of SNPs in the TSS**

The MACH1.0 program 35 was used with a set of 757,533 post-quality-control autosomal SNPs to impute ungenotyped SNPs with the HapMap II CEU (CEPH; Utah residents with ancestry from northern and western Europe) reference panel, resulting in 2,350,751 imputed genotypes per person after quality control, such that imputed SNPs were only retained for analysis if the squared correlations between imputed and actual genotypes for a set of masked genotypes (R²) was greater than 0.5. We used 200 individuals for estimating model parameters with 100 iterations for the Monte Carlo procedure by using all available haplotypes for each update, and then we imputed all the SNPs in the HapMap II CEU reference panel.

**GWAS Statistical Analysis**

Plasma MMA results were available for 2,210 of the 2,235 samples that passed the SNP genotype-based quality control analyses, so the discovery analysis for plasma MMA concentration relates to these 2,210 individuals. Standard descriptive statistics, including frequency distribution of MMA values, median, mean, theoretical versus sample quantile Q-Q plots, and measurements of skewness and kurtosis, were performed on the observed MMA measures both with and without transformation. The distribution of the MMA measurements in the TSS study participants deviated significantly from a normal distribution, and a log₁₀ transformation was used to normalize the MMA values for the linear regression analyses. Therefore, for the GWAS, genotyped and imputed autosomal SNPs that passed quality control were tested for association with the log₁₀-transformed MMA values, adjusted for age and sex.
were calculated for each analysis (Table S2). Significant SNPs (Figures S1 and S3). Genomic-control lambda values were generated, to remove the effect on the plots of these extremely significant SNPs. The horizontal blue line indicates the genome-wide significance threshold (p = 5 × 10⁻⁵). The horizontal red line indicates the suggestive significance threshold (p = 1 × 10⁻⁴). The two strongest signals are found in HIBCH and ACSF3.

Genotyping in the Replication Cohorts and Statistical Analysis of Metabolites in the TSS, TUDA, and Longitudinal Pregnancy Cohorts

All 5,186 TUDA samples were genotyped for HIBCH (MIM: 610690) rs291466 (GenBank: NM_014362.3:c.*231G>C) and ACSF3 rs1054747 ([GenBank: NM_001127214.3] c.*231G>C) by LGC Genomics (Herts, UK) with KASP genotyping chemistry. The call rates were 98.9% and 98.0%, respectively. Duplicate samples (2%) were included, as well as the addition of a single control sample to every plate. Concordance for all duplicates was >99%.

In this sample, measures of plasma MMA data were available for 1,511 individuals; 1,481 of these had genotype data for HIBCH rs291466 and 1,458 of these had genotype data for ACSF3 rs1054747. The replication analysis therefore relates to these 1,481 individuals. All 201 pregnant mothers and 117 cord bloods were genotyped for HIBCH rs291466 and ACSF3 rs1054747 by detection of allele-specific extension products via matrix-assisted laser desorption/ionization and time-of-flight (MALDI-TOF) mass spectrometry (Sequenom). For both SNPs, the call rates were 94% for the mothers and >96% for the newborns, and genotypes were 100% concordant for the >12% of samples that were re-plated and independently typed. The final replication analyses involved 185 mothers and 110 newborns with valid HIBCH rs291466 genotype and MMA data, and 182 mothers and 113 newborns with valid ACSF3 rs1054747 genotype and MMA data.

In the TSS, direct genotyping of HIBCH rs291466 yielded calls for 2,208 of the 2,210 subjects with MMA data who were included in the GWAS. Similar to the TSS data, distribution of the MMA measurements in the TUDA study participants and in the longitudinal pregnancy cohort did not approximate normal distributions, and a log₁₀ transformation was used to approximately normalize the MMA values for the linear regression analyses. HIBCH rs291466 was tested for association with the transformed unadjusted MMA values in the TUDA cohort and in the longitudinal pregnancy cohort. These association tests were performed with PLINK v.2 by simple linear regression and under the assumption of an additive genetic model as described above for the TSS GWAS.

Data for cobalamin biomarkers in the TSS and the replication datasets were not normally distributed. For Tables 1–3, all data are presented as medians with 25th and 75th percentile values (interquartile intervals; IQI). Comparisons of data across the three HIBCH rs291466 genotype groups were carried out with the non-parametric Kruskal-Wallis one-way ANOVA by ranks for three or more independent groups. p values < 0.05 were considered
significant and represent a difference across the three genotype groups, without specifying exactly which groups are different.

GWAS Catalog Query
To determine whether the lead SNPs identified in this study might have been previously associated with any disease phenotype or other trait, we searched the National Human Genome Research Institute and the European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog. No entries were found for HIBCH rs291466 or ACSF3 rs1054747, or for SNPs in LD (\( r^2 > 0.8 \), 1000 Genomes EUR super population) with them (n = 40 and n = 84, respectively).

Analysis of mRNA Abundance
Because HIBCH rs291466 and ACSF3 rs1054747 are coding SNPs, RNA-seq data can be used to quantitate transcripts carrying each allele. Two approaches were used to evaluate whether mRNA abundance is influenced by the SNPs associated with plasma MMA concentration. First, RNA-seq data from heterozygotes were used to determine how much HIBCH or ACSF3 transcript was expressed from each allele. Second, individuals were divided into groups based on HIBCH rs291466 or ACSF3 rs1054747 genotypes to compare the overall abundance of these transcripts.

First, to calculate the relative ratios of the T (Met) and C (?) alleles in the Geuvadis RNA Sequencing Project were downloaded. Samtools’ mpileup was used to find the heterozygous individuals in the Geuvadis RNA Sequencing Project. No entries were found for rs1054747, or for SNPs in LD (\( r^2 > 0.8 \), 1000 Genomes EUR super population). With them (n = 40 and n = 84, respectively).

Results

GWAS Discovery Cohort
We identified two regions of genome-wide significant association with \( \log_{10} \)-transformed MMA values (Figures 1A and 1B). The strongest association was observed with HIBCH rs291466 (\( p = 8.42 \times 10^{-8} \)), a directly typed SNP on chromosome 2 accounting for 9.9% of the total variance. This SNP is a non-synonymous polymorphism (c.2T>C) resulting in a methionine to a presumed threonine change in the translational start codon of the encoded protein, p.Met1? (note: the coding impact of this SNP in the initiator methionine is unknown and we therefore refer to this isoform as “?”). Although T (Met) is the ancestral allele (based on the allele present in the great apes), the C (?) allele is more common in the studied population. The median plasma MMA concentration for the corresponding CC (?) homozygotes was 46% higher than the TT (Met/Met) homozygotes (Table 1). This effect was independent of biochemical markers of cobalamin status, including serum total cobalamin, holotranscobalamin, and plasma homocysteine (Figures 1C and 1D and Table 1).

The second region of genome-wide significance was in the ACSF3 region on chromosome 16, consistent with previous studies showing that severe mutations in ACSF3 affect MMA concentrations. Patients with these mutations were found to have combined malonic and methylmalonic aciduria (CMAMMA [MIM: 614265]). The described patients are homozygous or compound heterozygous for rare coding changes in ACSF3 and have plasma MMA levels an order of magnitude higher than normal. The strongest ACSF3 signal among the genotyped SNPs (rs1054747, MAF = 0.05, \( p = 3.48 \times 10^{-19} \)) fell in the 3’ UTR of the gene and accounted for 2.1% of the total variance. Just three homozygotes carrying the minor A allele were
detected, but GA heterozygotes had a higher plasma MMA concentration than GG homozygotes (0.22 μM versus 0.18 μM). This SNP is linked ($D^2 = 1$, $r^2 > 0.92$) to SNPs in two adjacent non-coding RNAs (LINC00304, LOC400558) and the nearby cadherin 15 (CDH15) gene, but ACSF3 remained the best candidate gene for containing the causal variant because of its biological relevance. The 40 SNPs in high LD ($r^2 \geq 0.95$) with ACSF3 rs1054747 are intronic or intergenic, and the functional variant responsible for influencing MMA concentration remains unidentified. No regions of significance were observed in chromosome 6, the location of MUT. Next, we focused our efforts on performing replication analyses in two independent populations and undertaking functional characterization studies.

**Older Adult Replication Cohort**

Because MMA concentration is known to increase significantly with age, it was possible that the impact of this genotype would disappear in the elderly. To test this hypothesis, we genotyped HIBCH rs291466 in 1,481 older individuals in whom we had measured plasma MMA levels and other biochemical markers of cobalamin status (Table 2). As in our young adults, the genotype effect for HIBCH rs291466 on the transformed MMA data was highly significant ($p = 4.01 \times 10^{-26}$), and the genotype effect was discernible even when these subjects were categorized into three groups by age or by glomerular filtration rate (GFR, a measure of renal function; Figures 2A and 2B). Of note, this effect was independent of cobalamin status using either holoTC or total cobalamin as the serum indicators (Figures 2C and 2D) and was also independent of plasma homocysteine, a metabolic biomarker of cobalamin deficiency and a metabolite that, like MMA, is sensitive to reduced renal function.

The effect of ACSF3 rs1054747 on log10-transformed MMA levels was similarly assessed in this population of older Irish adults, resulting in replication of the originally observed effect ($p = 0.0003$).

**Longitudinal Pregnancy Replication Cohort**

MMA concentrations are altered during pregnancy and in newborns; therefore, replication of the influence of HIBCH rs291466 on MMA concentration was tested in samples from a longitudinal study of pregnant women and their newborn offspring (Table 3). MMA concentration distributions by HIBCH rs291466 genotype show significantly different effects for each of three trimesters of pregnancy and for cord bloods ($p < 8.0 \times 10^{-05}$ for each maternal time point; $p = 0.0005$ for cord bloods, Table 4). TT (Met/Met) homozygotes have the lowest median MMA concentration, whereas CC (?/??) homozygotes have the highest median MMA concentration, consistent with the direction of effect originally observed in the discovery cohort of young Irish adults.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All*</th>
<th>CC (/??)</th>
<th>CT (/?/C)</th>
<th>TT (C/C)</th>
<th>Kruskal-Wallis Test*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Individuals</td>
<td>2,208</td>
<td>721</td>
<td>1,065</td>
<td>422</td>
<td>–</td>
</tr>
<tr>
<td>Sex (M/F)</td>
<td>907/1,301</td>
<td>299/422</td>
<td>442/623</td>
<td>166/256</td>
<td>–</td>
</tr>
<tr>
<td>Age (years)</td>
<td>22.0 (21.0–24.0)</td>
<td>22.0 (21.0–23.0)</td>
<td>22.0 (21.0–24.0)</td>
<td>23.0 (21.0–24.0)</td>
<td>0.008</td>
</tr>
<tr>
<td>Body Mass Index (kg/m²)</td>
<td>22.6 (21.0–24.0)</td>
<td>22.5 (21.0–24.0)</td>
<td>22.7 (21.0–24.0)</td>
<td>22.5 (20.8–24.4)</td>
<td>0.454</td>
</tr>
<tr>
<td>Creatinine (µmol/L)</td>
<td>64.8 (56.4–74.6)</td>
<td>65.1 (56.7–75.4)</td>
<td>64.7 (56.4–74.4)</td>
<td>64.1 (56.3–74.1)</td>
<td>0.586</td>
</tr>
<tr>
<td>Urea (mmol/L)</td>
<td>4.30 (3.70–5.10)</td>
<td>4.40 (3.70–5.13)</td>
<td>4.30 (3.60–5.10)</td>
<td>4.20 (3.50–5.00)</td>
<td>0.030</td>
</tr>
<tr>
<td>Plasma cobalamin (pmol/L)</td>
<td>306 (231–411)</td>
<td>307 (231–421)</td>
<td>305 (232–405)</td>
<td>305 (230–411)</td>
<td>0.675</td>
</tr>
<tr>
<td>holoTC (pmol/L)</td>
<td>53.7 (39.1–72.2)</td>
<td>54.1 (39.5–72.8)</td>
<td>53.2 (38.9–71.4)</td>
<td>55.4 (39.5–73.3)</td>
<td>0.388</td>
</tr>
<tr>
<td>Homocysteine (µmol/L)</td>
<td>8.13 (7.00–9.62)</td>
<td>8.17 (6.99–9.61)</td>
<td>8.17 (7.08–9.61)</td>
<td>7.92 (6.83–9.75)</td>
<td>0.319</td>
</tr>
<tr>
<td>MMA (µmol/L)</td>
<td>0.17 (0.14–0.21)</td>
<td>0.19 (0.16–0.24)</td>
<td>0.16 (0.14–0.20)</td>
<td>0.13 (0.11–16)</td>
<td>&lt; 0.0001¹</td>
</tr>
</tbody>
</table>

*Valid HIBCH rs291466 genotypes were obtained for 2,208 of the 2,210 TSS subjects in the GWAS study. ² IQ (inter-quartile interval) represents the 25th–75th percentile. ³ The non-transformed data were compared with the Kruskal-Wallis one-way ANOVA by ranks for three or more independent groups; $p < 0.05$ was considered significant. ⁴ The association test for this SNP was also performed by simple linear regression with PLINK v.2 and under the assumption of an additive genetic model (number of minor alleles: 0, 1, and 2). p = 1.95 × 10⁻⁶ was obtained for the unadjusted log10-transformed data.
for AG and GG genotype groups, respectively), as originally observed.

Functional Studies
Because the signal associated with HIBCH rs291466 was driven by a common allele in a gene previously unknown to influence MMA and had the largest impact on MMA levels, we pursued functional characterization studies. Alterations of initiator methionine residues have been associated with loss-of-function alleles. We assessed the impact of this missense change on HIBCH (see below). To determine whether HIBCH rs291466 influences the expression of HIBCH mRNA, we analyzed RNA-seq data generated from lymphoblastoid lines.

Table 2. Subject Characteristics of the TUDA Replication Cohort, with Participants Grouped by HIBCH rs291466 Genotype

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All* Median (IQI)</th>
<th>CC (T/? Median (IQI))</th>
<th>CT (T/Met Median (IQI))</th>
<th>TT (Met/Met Median (IQI))</th>
<th>Kruskal-Wallis Test p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Individuals</td>
<td>1,481</td>
<td>460</td>
<td>723</td>
<td>298</td>
<td>–</td>
</tr>
<tr>
<td>Sex (M/F)</td>
<td>493/991</td>
<td>168/293</td>
<td>231/494</td>
<td>94/204</td>
<td>–</td>
</tr>
<tr>
<td>Age (years)</td>
<td>79.0 (72.6–83.9)</td>
<td>79.5 (72.8–84.7)</td>
<td>78.6 (72.0–83.2)</td>
<td>78.9 (73.2–84.1)</td>
<td>0.285</td>
</tr>
<tr>
<td>Body Mass Index (kg/m²)</td>
<td>27.1 (23.6–30.8)</td>
<td>27.3 (23.6–31.0)</td>
<td>27.0 (23.8–30.6)</td>
<td>26.8 (23.6–30.6)</td>
<td>0.742</td>
</tr>
<tr>
<td>Creatinine (µmol/L)</td>
<td>82.0 (69–101)</td>
<td>82.0 (69.0–101)</td>
<td>82.0 (69.5–101)</td>
<td>81.0 (71.0–103)</td>
<td>0.878</td>
</tr>
<tr>
<td>Hemoglobin (g/dL)</td>
<td>12.8 (11.7–13.8)</td>
<td>12.8 (11.7–14.0)</td>
<td>12.8 (11.7–13.8)</td>
<td>12.7 (11.7–13.8)</td>
<td>0.630</td>
</tr>
<tr>
<td>Urea (mmol/L)</td>
<td>6.7 (5.4–8.7)</td>
<td>6.6 (5.4–8.8)</td>
<td>6.8 (5.4–8.7)</td>
<td>6.8 (5.5–8.7)</td>
<td>0.726</td>
</tr>
<tr>
<td>Serum cobalamin (µmol/L)</td>
<td>185 (131–287)</td>
<td>177 (128–274)</td>
<td>187 (135–286)</td>
<td>194 (128–307)</td>
<td>0.270</td>
</tr>
<tr>
<td>holoTC (µmol/L)</td>
<td>39.0 (25.8–57.0)</td>
<td>38.4 (25.6–55.0)</td>
<td>38.8 (25.9–57.5)</td>
<td>39.9 (25.8–60.9)</td>
<td>0.550</td>
</tr>
<tr>
<td>Homocysteine (µmol/L)</td>
<td>16.2 (13.0–20.5)</td>
<td>15.9 (12.5–19.7)</td>
<td>16.3 (13.2–20.9)</td>
<td>16.2 (13.3–20.3)</td>
<td>0.097</td>
</tr>
<tr>
<td>MMA (µmol/L)</td>
<td>0.34 (0.24–0.51)</td>
<td>0.42 (0.29–0.62)</td>
<td>0.33 (0.24–0.50)</td>
<td>0.26 (0.19–0.37)</td>
<td>&lt; 0.0001c</td>
</tr>
</tbody>
</table>

M, male; F, female.
*IQI (inter-quartile interval) represents the 25th–75th percentile.
*The non-transformed data were compared with the Kruskal-Wallis one-way ANOVA by ranks for three or more independent groups; p < 0.05 was considered significant.
*The association test for this SNP was also performed by simple linear regression with PLINK v.2 and under the assumption of an additive genetic model (number of minor alleles: 0, 1, and 2). p = 4.01 × 10⁻³⁸ was obtained for the unadjusted log₁₀-transformed data.

MMA, methionine; holoTC, holotranscobalamin.

Discussion
Three aspects of this GWAS of plasma MMA concentration are particularly noteworthy. First, we report the strong association of HIBCH rs291466 with MMA levels in a young, healthy, ethnically homogeneous cohort from Ireland, as well as its replication in two ethnically matched Irish cohorts representing significantly different stages of life. Second, the value of GWASs is emphasized given that this study revealed that the largest genetic influence on MMA levels is not related to cobalamin metabolism and is instead related to a little-studied pathway involved in valine catabolism. Last, we note the potential complexity of determining the mode of function of any phenotype-associated variant, as shown by the unexpected apparent mechanism through which the HIBCH rs291466 (p.Met1?) association signal achieves its effect on MMA levels. In this case the variant associated with the strongest GWAS signal would be expected to produce a null or hypomorphic allele, i.e., one that produces a protein that is not properly targeted to the mitochondrion. Our experiments reveal the opposite; the missense allele associated with higher levels of MMA produces more mature enzyme and acts at the level of mRNA. It is unclear whether this allele directly influences transcription or whether this activity resides with another linked SNP.
The clinical utility of MMA concentration as a confirmatory biomarker of cobalamin deficiency makes this metabolite an important candidate for GWASs. Recent GWAS results for cobalamin have reported a strong association with fucosyltransferase 2 (FUT2 [MIM: 182100]), and in several of these studies, MUT was also identified as a significant determinant of cobalamin at a genome-wide level. We observed no such associations in our GWAS of MMA concentration, and the top MMA GWAS result is in fact unrelated to cobalamin metabolism. This is consistent with two mechanisms for elevated levels of circulating MMA. First, a reduction in cobalamin reduces activity of the MUT enzyme, resulting in accumulation of MMA. Second, a cobalamin-independent mechanism of increased flux through the valine catabolism pathway (e.g., via increased levels of HIBCH due to the C [?] allele of rs291466) can result in increased production of MMA. The effect of the latter was not anticipated, but is very strong.

Although HIBCH rs291466 is the more common variant (frequency of allele associated with elevated MMA: 0.57 versus 0.05) with greater impact on plasma MMA concentration (contribution to variance: 9.1% versus 2.2%), we also genotyped the top ACSF3 polymorphism (rs1054747) in our replication cohorts. ACSF3 rs1054747 was significantly associated with MMA levels in the sample of ~1,400 older Irish adults, but not in our longitudinal pregnancy cohort (Table 4). The combination of the low MAF and the substantially smaller sample size (~180 pregnant women) most likely diminished our power to detect the effect of this SNP on MMA levels. In support of this, we note that their newborns did show a borderline association (Table 4).

Although HIBCH rs291466 is the more common variant (frequency of allele associated with elevated MMA: 0.57 versus 0.05) with greater impact on plasma MMA concentration (contribution to variance: 9.1% versus 2.2%), we also genotyped the top ACSF3 polymorphism (rs1054747) in our replication cohorts. ACSF3 rs1054747 was significantly associated with MMA levels in the sample of ~1,400 older Irish adults, but not in our longitudinal pregnancy cohort (Table 4). The combination of the low MAF and the substantially smaller sample size (~180 pregnant women) most likely diminished our power to detect the effect of this SNP on MMA levels. In support of this, we note that their newborns did show a borderline association (Table 4).

![Figure 2. Replication of the HIBCH rs291466 Association with Plasma MMA Concentration in the TUDA Cohort](image-url)
connection between ACSF3 activity and steady-state levels of MMA in plasma.

We also confirmed the association of MMA concentration and HIBCH rs291466. This SNP was not only the top genome-wide association signal in a young, healthy Irish cohort, it was also significant in an older Irish cross-sectional cohort and in a longitudinal cohort of pregnant Irish women and their newborns. The persistence of this genetic effect is remarkable considering the known potential for altered MMA levels at each of these stages of life. Maternal circulating cobalamin decreases over the course of pregnancy as a result of hemodilution and other pregnancy-related factors.41 This is accompanied by changes of pregnancy related factors.41 This is accompanied by changes in MMA levels as a result of factors such as impaired cobalamin absorption and decreased renal function. The ability to detect the GWAS signal of MMA in plasma.

Table 3. Subject Characteristics of the Longitudinal Pregnancy Replication Cohort at First Antenatal Clinic Visit, with Participants Grouped by HIBCH rs291466 Genotype

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All</th>
<th>CC (/?/)</th>
<th>CT (/?/Met)</th>
<th>TT (Met/Met)</th>
<th>Kruskal-Wallis Testb p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Individuals</td>
<td>185</td>
<td>70</td>
<td>80</td>
<td>35</td>
<td>–</td>
</tr>
<tr>
<td>Age (y)</td>
<td>30.0 (25.0–33.0)</td>
<td>25.9 (22.5–28.2)</td>
<td>24.3 (21.0–27.6)</td>
<td>24.8 (21.5–28.5)</td>
<td>0.044</td>
</tr>
<tr>
<td>Body Mass Index (kg/m²)</td>
<td>24.8 (21.9–28.2)</td>
<td>24.3 (21.0–28.2)</td>
<td>24.8 (21.5–28.5)</td>
<td>0.292</td>
<td></td>
</tr>
<tr>
<td>Gestation (weeks)</td>
<td>14.2 (12.9–15.4)</td>
<td>14.0 (12.4–15.6)</td>
<td>13.9 (13.3–14.7)</td>
<td>0.914</td>
<td></td>
</tr>
<tr>
<td>Creatinine (μmol/L)</td>
<td>46.6 (42.4–51.3)</td>
<td>46.7 (42.4–51.9)</td>
<td>48.1 (42.9–51.3)</td>
<td>0.700</td>
<td></td>
</tr>
<tr>
<td>Serum cobalamin (pmol/L)</td>
<td>171 (120–215)</td>
<td>180 (133–225)</td>
<td>184 (140–217)</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>holoTC (pmol/L)</td>
<td>33.6 (24.7–43.7)</td>
<td>36.4 (27.5–46.4)</td>
<td>36.7 (25.2–43.7)</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td>Homocysteine (μmol/L)</td>
<td>6.29 (5.35–7.34)</td>
<td>6.06 (5.06–7.27)</td>
<td>6.50 (5.51–7.63)</td>
<td>0.097</td>
<td></td>
</tr>
</tbody>
</table>

*IQI (inter-quartile interval) represents the 25th–75th percentile.

Table 4. Plasma MMA Concentrations in the Longitudinal Pregnancy Replication Cohort and in Cord Blood with Participants Grouped by HIBCH rs291466 Genotype

<table>
<thead>
<tr>
<th>Gestation Time Point</th>
<th>All</th>
<th>CC (/?/)</th>
<th>CT (/?/Met)</th>
<th>TT (Met/Met)</th>
<th>Replication testb p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 weeks (μmol/L)</td>
<td>0.15 (0.11–0.21), n = 184</td>
<td>0.18 (0.13–0.25), n = 69</td>
<td>0.15 (0.12–0.21), n = 80</td>
<td>0.10 (0.08–0.14), n = 35</td>
<td>1.95 × 10⁻⁸</td>
</tr>
<tr>
<td>24 weeks (μmol/L)</td>
<td>0.18 (0.14–0.25), n = 132</td>
<td>0.23 (0.17–0.29), n = 49</td>
<td>0.18 (0.15–0.25), n = 53</td>
<td>0.14 (0.12–0.17), n = 30</td>
<td>1.05 × 10⁻⁵</td>
</tr>
<tr>
<td>34 weeks (μmol/L)</td>
<td>0.21 (0.16–0.31), n = 129</td>
<td>0.25 (0.19–0.36), n = 48</td>
<td>0.21 (0.17–0.31), n = 57</td>
<td>0.15 (0.14–0.20), n = 24</td>
<td>7.39 × 10⁻⁵</td>
</tr>
<tr>
<td>Cord blood, delivery (μmol/L)</td>
<td>0.40 (0.30–0.58), n = 110</td>
<td>0.47 (0.30–0.67), n = 43</td>
<td>0.40 (0.34–0.58), n = 49</td>
<td>0.26 (0.23–0.3), n = 18</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

*IQI (inter-quartile interval) represents the 25th–75th percentile.

*Replication was performed by simple linear regression of the unadjusted log₁₀-transformed data via PLINK v.2 and under the assumption of an additive genetic; p < 0.05 was considered significant.
possible deficiency (holoTC < 30 pmol/L). Thus, 28% of all CC (/??) individuals might be falsely classified as cobalamin deficient when using MMA values alone. In contrast, among TT (Met/Met) individuals, far fewer (25%) had MMA concentrations above 0.37 μmol/L, and the majority of those with high MMA levels also had low cobalamin status (73%) (holoTC < 30 pmol/L), consistent with a true cobalamin deficiency. Choosing other thresholds for total cobalamin (110 pmol/L) or holoTC (35 pmol/L) changed the proportion of individuals within deficient or sufficient categories but did not remove the strong genotype effect. Thus, this common HIBCH polymorphism is a cobalamin-independent determinant of plasma MMA concentration that influences its effectiveness as a cobalamin status indicator in older adults when applying current clinical cut-offs.

The unanticipated discovery of a variant in HIBCH as a strongly significant contributor to the variance of plasma MMA concentration highlights this little-studied enzyme of branched-chain amino acid metabolism that for several decades has been cited as an unusual feature within the valine oxidation pathway. β-oxidation of odd-chain fatty acids, catabolism of branched-chain amino acids, and metabolism of cholesterol side-chain moieties all follow pathways that generate methylmalonyl-CoA, with coenzyme A (CoA) being added early in the catabolic process. Unlike leucine and isoleucine, valine degradation contains a unique intermediate step, carried out by mitochondrial HIBCH, whereby CoA is removed from 3-hydroxyisobutyryl-CoA, producing 3-hydroxyisobutyric acid, an easily diffusible free acid. Two steps later, CoA is reintroduced when methylmalonic semialdehyde is converted to propionyl-CoA (Figure 4). HIBCH is reported to be abundant in human liver and might be important to remove the preceding toxic intermediate, methacrylyl-CoA. Animal studies also suggest that release of the HIBCH product, 3-hydroxybutyric acid, might be a mechanism to conserve the gluconeogenic potential of valine. Severe mutations in HIBCH are extremely rare and all reported affected children have died early in life.

What are the consequences of this common but previously unstudied polymorphism? It is possible that the reduced amount of HIBCH enzyme produced by the methionine-encoding allele causes a decreased flux of 3-hydroxyisobutyryl-CoA through HIBCH, perhaps leading to higher mitochondrial concentrations of upstream intermediates and moderately altered C4-hydroxy carnitines in the blood of TT (Met/Met) individuals. This model is consistent with what is observed in subjects with rare severe mutations in this enzyme. Methacrylyl-CoA, an intermediate compound in valine catabolism, is highly reactive and forms toxic conjugates with cysteine. The concentration of this compound might be increased when HIBCH activity is low (such as T [Met] homozygotes compared to C [?] homozygotes). Higher concentrations of methacrylyl-CoA within mitochondria could inhibit enzymes involved in oxidative phosphorylation. It is also possible that the methionine-encoding allele results in a less efficient gluconeogenic system. We did not measure these metabolites, and we did not detect any consistent changes in liver function tests by genotype in our discovery or our replication cohort of older adults (Tables 1 and 2). Additional studies of the phenotypic and metabolic consequences of this polymorphism are warranted. No other phenotypes have been reported to be associated with HIBCH or ACSF3 in the GWAS catalog. Recently, Raffler and colleagues reported an association between variation in HIBCH and an unidentified compound in urine. We anticipate that this compound will be a cleared derivative of valine catabolism. The absence of any known associations between these two loci and reported phenotypes might mean that elevated MMA levels in general do not directly contribute to disease, but rather report on its presence or its potential to develop.
Beyond the impact of this SNP, the mechanism of action of the HIBCH rs291466 signal requires elucidation. Although T (Met) is the minor allele for HIBCH rs291466 in this Irish population, phylogenetic analysis suggests that it is likely to be the ancestral allele. T (Met) is also the major allele in the publicly genotyped HapMap CHB, JPT, and YRI samples. The derived C (?) allele results in higher MMA concentrations. Given that MMA is a downstream product of valine catabolism, our metabolic results suggest that the resulting enzyme is more active or more abundant. Because this polymorphism is located at the start of the mitochondrial leader sequence, it might alter the efficiency at which the protein is imported into mitochondria. We tested this model in cell culture studies. Both isoforms of the enzyme were equally efficiently translocated into the mitochondrial matrix (Figure S4).

These results suggest that a downstream methionine residue is used to initiate translation and the shortened leader efficiently directs mitochondrial import. Lymphoblastoid RNA-seq data and expression quantitative trait loci (eQTL) analysis demonstrated that, compared to the T (Met) allele, the C (?) allele is associated with a significant increase of HIBCH mRNA levels, presumably resulting in the increased enzyme concentrations that we observed in cultured lymphoblastoid cells. This effect is observed in eQTL analysis ($p < 5.5 \times 10^{-06}$) of more than 30 tissue types, including liver (GTEx Portal, v.6). On the basis of the genetic code alone, it is predicted that the loss of the initiating methionine would alter HIBCH function via translation or at the level of the produced protein. However, our data support the conclusion that it is likely that this SNP or another variant linked to it produces the metabolic phenotype by influencing mRNA transcription or stability. There are 33 SNPs in strong LD with HIBCH rs291466 ($r^2 > 0.95$ in the EUR cohort), all of which are intronic or 5′ of the transcriptional start site and reside in an area containing chromatin marks consistent with active transcription. Further study is required to determine the identity and mechanism of the functional variant underlying the HIBCH rs291466 association with MMA levels.

In conclusion, we have identified two genetic loci that significantly affect the plasma concentration of MMA. These loci provide important data underpinning the current understanding of the pathways leading to MMA production. Our data suggest that genetic variation in HIBCH might be especially relevant in situations where protein catabolism is elevated, i.e., states of acute disease and in the elderly.

**Accession Numbers**

The accession number for the TSS genotype and phenotype data reported in this paper is dbGaP: phs000789.v1.p1.

**Supplemental Data**

Supplemental Data include five figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.03.005.
Acknowledgments

The authors acknowledge the contributions made by the participants in the Trinity Student Study, the Trinity Ulster Department of Agriculture study, and the longitudinal pregnancy study. The TSS GWAS work was supported in part by the Intramural Research Programs of the National Human Genome Research Institute (NHGRI), the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the NIH, and the Health Research Board, Dublin. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (https://bpc.nih.gov/systems). C.D.C. is the recipient of an NHGRI Health Disparities Research Fellowship. The TUDA study was funded by the Irish Department of Agriculture, Food and the Marine’s Food Institutional Research Measure initiative and the Northern Ireland Department for Employment and Learning under the Cross-Border Research and Development Programme “Strengthening the all-Island Research Base.” Holotranscobalamin and methylmalonic acid measurements in the TUDA cohort were sponsored by Axis-Shield Diagnostics.

Received: August 12, 2015
Accepted: March 8, 2016
Published: April 28, 2016

Web Resources

GWAS Catalog v.1.0.1, accessed February 2, 2016, http://www.ebi.ac.uk/gwas/
International HapMap GTeX Portal, http://www.gtexportal.org/home/

References


