Arteriosclerosis, Thrombosis, and Vascular Biology

JOURNAL OF THE AMERICAN HEART ASSOCIATION

Comparative Genome-Wide Association Studies in Mice and Humans for Trimethylamine *N*-Oxide, a Proatherogenic Metabolite of Choline and l-Carnitine

Jaana Hartiala, Brian J. Bennett, W.H. Wilson Tang, Zeneng Wang, Alexandre F.R. Stewart, Robert Roberts, Ruth McPherson, Aldons J. Lusis, Stanley L. Hazen and Hooman Allayee
CARDIoGRAM Consortium

Arterioscler Thromb Vasc Biol. 2014;34:1307-1313; originally published online March 27, 2014;
doi: 10.1161/ATVBAHA.114.303252
Arteriosclerosis, Thrombosis, and Vascular Biology is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1079-5642. Online ISSN: 1524-4636

Comparative Genome-Wide Association Studies in Mice and Humans for Trimethylamine *N*-Oxide, a Proatherogenic Metabolite of Choline and L-Carnitine

Jaana Hartiala, Brian J. Bennett, W.H. Wilson Tang, Zeneng Wang, Alexandre F.R. Stewart, Robert Roberts, Ruth McPherson, CARDIoGRAM Consortium,* Aldons J. Lusis, Stanley L. Hazen, Hooman Allayee

Objective—Elevated levels of plasma trimethylamine *N*-oxide (TMAO), the product of gut microbiome and hepatic-mediated metabolism of dietary choline and L-carnitine, have recently been identified as a novel risk factor for the development of atherosclerosis in mice and humans. The goal of this study was to identify the genetic factors associated with plasma TMAO levels.

Approach and Results—We used comparative genome-wide association study approaches to discover loci for plasma TMAO levels in mice and humans. A genome-wide association study in the hybrid mouse diversity panel identified a locus for TMAO levels on chromosome 3 (P=2.37×10^{-6}) that colocalized with a highly significant (P=1.07×10^{-20}) *cis*-expression quantitative trait locus for solute carrier family 30 member 7. This zinc transporter could thus represent 1 positional candidate gene responsible for the association signal at this locus in mice. A genome-wide association study for plasma TMAO levels in 1973 humans identified 2 loci with suggestive evidence of association (P=3.0×10^{-7}) on chromosomes 1q23.3 and 2p12. However, genotyping of the lead variants at these loci in 1892 additional subjects failed to replicate their association with plasma TMAO levels.

Conclusions—The results of these limited observational studies indicate that, at least in humans, genes play a marginal role in determining TMAO levels and that any genetic effects are relatively weak and complex. Variation in diet or the repertoire of gut microbiota may be more important determinants of plasma TMAO levels in mice and humans, which should be investigated in future studies. **(*Arterioscler Thromb Vasc Biol*. 2014;34:1307-1313.)**

Key Words: atherosclerosis ■ genetics ■ humans ■ mice ■ trimethylamine *N*-oxide

Choline is a key nutrient with various metabolic roles in lipid metabolism and cell membrane structure, and it serves as a precursor for the synthesis of the neurotransmitter acetylcholine.[1–3] Dietary choline is also an important source of methyl groups that are required for proper metabolism of certain amino acids, such as homocysteine and methionine.[3] A variety of animal studies have shown that choline deficiency adversely affects brain and cognitive development during fetal and neonatal life,[1,4–6] which has led to specific nutritional guidelines recommending adequate intake of choline for infants and pregnant or lactating women.[7,8]

One route for the initial catabolism of dietary choline (in the form of phosphotidylcholine) is mediated by intestinal microbes and leads to the formation of trimethylamine (TMA). TMA is efficiently absorbed from the gastrointestinal tract and subsequently oxidized by the liver to form TMA *N*-oxide (TMAO). This latter reaction is catalyzed by one or more of the flavin monooxygenase (FMO) enzymes, of which there are 6 gene family members in higher mammals.[9] Interestingly, mutations of *FMO3* that result in deficiency of this enzyme are the cause of trimethylaminuria, otherwise known as fish malodor syndrome.[10] This relatively rare recessive disorder is characterized by the near absence of plasma TMAO levels and highly elevated TMA levels, depending on the functional severity of the mutation in *FMO3*. The pungent odor of rotting fish that characterizes trimethylaminuria is because of the release of the volatile gas TMA through the breath, skin, and urine.

Recently, we uncovered a novel mechanism through which gut microbiota and hepatic-mediated metabolism of dietary choline promote atherosclerosis and increase the risk of

Received on: January 13, 2014; final version accepted on: March 18, 2014.

From the Department of Preventive Medicine (J.H., H.A.) and Institute for Genetic Medicine (J.H., H.A.), Keck School of Medicine of the University of Southern California, Los Angeles; Department of Genetics (B.J.B.) and Nutrition Research Institute (B.J.B.), University of North Carolina, Chapel Hill, Kannapolis; Departments of Cardiovascular Medicine (W.H.W.T., Z.W., S.L.H.) and Cellular and Molecular Medicine (W.H.W.T., Z.W., S.L.H.) and Center for Cardiovascular Diagnostics and Prevention (W.H.W.T., Z.W., S.L.H.), Cleveland Clinic, OH; John and Jennifer Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada (A.F.R.S., R.R., R.M.); and Departments of Medicine (A.J.L.), Human Genetics (A.J.L.), and Microbiology, Immunology, and Molecular Genetics (A.J.L.), David Geffen School of Medicine of UCLA.

*A full list of authors and affiliations for the CARDIoGRAM Consortium is provided in the online-only Data Supplement.

This manuscript was sent to Robert A. Hegele, Consulting Editor, for review by expert referees, editorial decision, and final disposition.

The online-only Data Supplement is available with this article at http://atvb.ahajournals.org/lookup/suppl/doi:10.1161/ATVBAHA.114.303252/-/DC1.

Correspondence to Hooman Allayee, PhD, Institute for Genetic Medicine, Keck School of Medicine of USC, 2250 Alcazar St, CSC202, Los Angeles, CA 90033. E-mail hallayee@usc.edu

Nonstandard Abbreviations and Acronyms

CAD	coronary artery disease
CARDIoGRAM	Coronary Artery Disease Genome-wide Replication And Meta-Analysis
FMO	flavin monooxygenase
GWAS	genome-wide association study
HMDP	hybrid mouse diversity panel
QTL	quantitative trait loci
SNPs	single nucleotide polymorphisms
Slc30a7	solute carrier family 30 member 7
TMA	trimethylamine
TMAO	trimethylamine *N*-oxide

coronary artery disease (CAD).[11,12] These studies demonstrated that plasma TMAO levels in humans were positively associated with the presence of multiple CAD phenotypes, including atherosclerotic plaque burden and future risk of myocardial infarction, stroke, or death in a dose-dependent fashion. A similar relationship was observed between plasma TMAO levels and aortic lesion development among various inbred mouse strains.[13] More recently, we also demonstrated that L-carnitine, a trimethylamine abundant in red meat, is also metabolized by intestinal microbiota to produce TMAO in mice and humans and that L-carnitine supplementation accelerated atherosclerosis in mice.[14] Notably, short-term administration of broad-spectrum antibiotics eliminated the production of TMAO in both mice and humans and decreased atherosclerosis in mice. Furthermore, TMAO supplementation in mice, or dietary supplementation of either choline or L-carnitine, in the presence of intact gut microbiota led to alterations in cholesterol and sterol metabolism in multiple distinct compartments, including reduction in reverse cholesterol transport, providing a mechanistic rationale for the association between TMAO levels and atherosclerotic cardiovascular phenotypes.[14] Taken together, these studies provide evidence consistent with the proatherogenic role of TMAO in mammals and support the notion that gut microbiota plays an obligatory role in the formation TMAO from dietary choline and L-carnitine.

It is reasonable to assume that variation in plasma TMAO levels could also be affected by intrinsic genetic factors of the host. However, with the exception of *FMO3*, the genes that control plasma TMAO levels are not known. Therefore, the aim of the present study was to use comparative genome-wide association study (GWAS) approaches in mice and humans to identify novel genetic determinants associated with plasma TMAO levels.

Materials and Methods

Materials and Methods are available in the online-only Supplement.

Results

Association of the *FMO* Cluster With *FMO3* Gene Expression, Plasma TMAO Levels, and CAD in Humans

In previous studies, we reported that *Fmo3* expression varied significantly among inbred strains from the hybrid mouse diversity panel (HMDP) and that a major locus regulating its expression mapped directly over *Fmo3*, suggesting *cis*-acting regulation in mice.[13] Furthermore, *Fmo3* expression was positively correlated with both plasma TMAO levels and atherosclerosis in mice. Based on these observations, we first used a targeted approach to evaluate whether genetic associations could specifically be observed with the human *FMO* locus on chromosome 1q24.3. To evaluate the association of the *FMO* cluster with hepatic *FMO3* mRNA levels, we used a previously published liver gene expression data set.[15] These analyses were performed in a subset of 151 white subjects for whom complete gene expression and genotype data were publicly available. Fifty-seven single nucleotide polymorphisms (SNPs) were available for analysis in a specified ≈451-kb region containing *FMO3*, *FMO6P*, *FMO2*, *FMO1*, and *FMO4*, including 200 kb of flanking sequence (100 kb from each end). As shown in Figure 1A, 1 SNP (rs2075988) yielded age- and sex-adjusted association with *FMO3* mRNA levels ($P=4.5\times10^{-4}$) that remained significant after correction for multiple testing (0.05/57; Bonferroni-corrected $P=8.8\times10^{-4}$). *Cis*-expression quantitative trait loci (QTL) were not observed for any other members of the *FMO* gene family at this locus (data not shown).

We next determined whether variation in the *FMO* cluster influenced plasma TMAO levels using the GWAS results from the GeneBank study, a cohort of patients undergoing elective cardiac evaluation at the Cleveland Clinic. Table 1 describes the clinical characteristics of the 3865 individuals used in the present study. As expected for a patient population undergoing coronary angiography as part of their clinical evaluation, a majority of these subjects were men, had prevalent CAD, and were taking lipid-lowering medication (Table 1). In this analysis, 471 SNPs were available, but none were significantly associated with plasma TMAO levels (Figure 1B). Finally, we evaluated whether the *FMO* locus was associated with risk of CAD in the Coronary Artery Disease Genome-wide Replication And Meta-Analysis (CARDIoGRAM) consortium, which represents a meta-analysis of GWAS data from a discovery set of ≈22 000 CAD cases and ≈65 000 controls.[16] In CARDIoGRAM, 388 SNPs were available for analyses, of which 21 yielded values of $P<0.05$ for association with CAD (Figure 1C). However, none of these associations were significant at the Bonferroni-corrected significance threshold ($P=1.3\times10^{-4}$; 0.05/388). Furthermore, the SNP that exhibited the strongest association with *FMO3* mRNA levels (rs2075988) did not demonstrate evidence for association with either plasma TMAO levels or risk of CAD (Figure 1).

GWAS for Plasma TMAO Levels in Mice

To identify novel genetic factors associated with plasma TMAO levels in mammals, we next used the HMDP to perform an unbiased GWAS in mice. This newly developed genetic platform consists of ≈100 classic inbred and recombinant inbred mouse strains that are maximally informative for association analysis and have been used to perform GWAS for other quantitative traits relevant to human diseases, including atherosclerosis, metabolites, and hepatic mRNA levels.[17–20] For the present study, we performed a GWAS for plasma TMAO levels in male mice on a chow diet and identified 1 locus on mouse chromosome 3 between 110 and

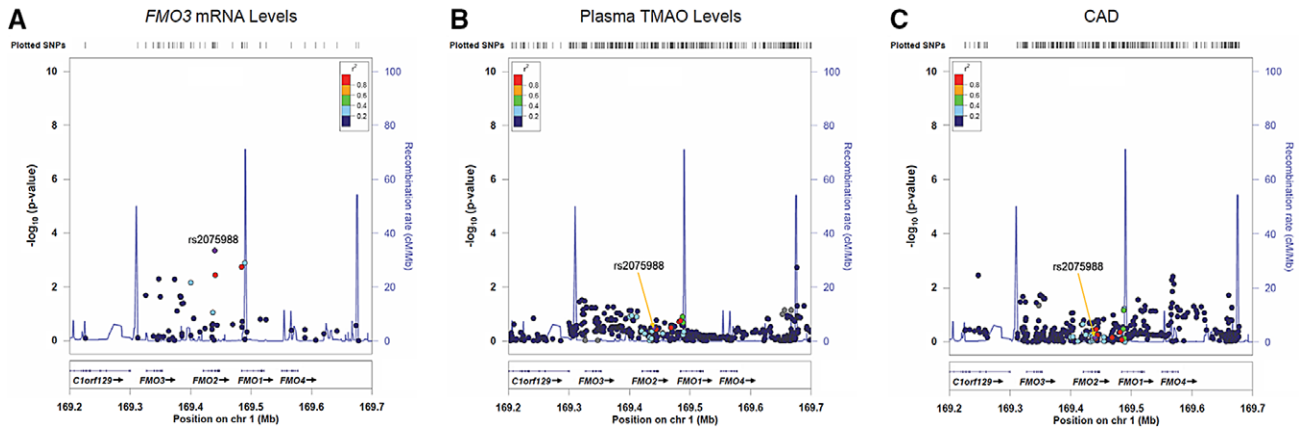


Figure 1. Association of the flavin monooxygenase (*FMO*) locus with FMO3 mRNA levels, plasma trimethylamine *N*-oxide (TMAO) levels, and risk of coronary artery disease (CAD) in humans. Using a publicly available expression quantitative trait loci liver data set, 57 single nucleotide polymorphisms (SNPs) were tested for association with hepatic *FMO3* mRNA levels, one of which (rs2075988) yielded a significant *P* value (4.5×10^{-4}) after Bonferroni correction for multiple testing (**A**). In the GeneBank cohort, none of the 471 SNPs tested in the *FMO* locus yielded significant association with plasma TMAO levels (**B**). Evaluation of the *FMO* locus with risk of CAD using 388 SNPs available from the results of the Coronary Artery Disease Genome-wide Replication And Meta-Analysis (CARDIoGRAM) consortium did not reveal any significant associations (**C**). The same genomic interval spanning ≈451 kb across the *FMO* cluster on chromosome 1q24.3 is shown for all 3 plots, and the variant most strongly associated with *FMO3* mRNA levels is given as the reference SNP (rs2075988). Chr indicates chromosome.

115 Mb that exceeded the genome-wide significance threshold for association in the HMDP ($P=2.37\times10^{-6}$; Figure 2A and 2B). The 10-Mb region centered around the lead SNP on chromosome 3 contained several genes and exhibited a highly significant *cis*-expression QTL ($P=1.07\times10^{-20}$) for the gene encoding solute carrier family 30 member 7 (*Slc30a7*; Figure 2C). The colocalization of QTLs for plasma TMAO and *Slc30a7* mRNA levels suggests that this zinc transporter could represent 1 positional candidate gene responsible for the association signal at this locus. Suggestive evidence for association of plasma TMAO levels ($P=7.62\times10^{-6}$) was also observed with a region on mouse chromosome 1 at 184 Mb (Figure 2A), although this locus did not achieve genome-wide significance. The lead SNP on chromosome 1 maps to within 40 kb of the lamin β-receptor gene but ≈20 Mb distal from the *Fmo* gene cluster (162–163 Mb).

Table 1. Clinical Characteristics of the Study Population

Trait	n=3865
Age, y	64±11
Male/female	6372/2789
Number with CAD at baseline, %	6776 (76)
CAD severity	
0 vessels, %	2766 (30)
1 or 2 vessels, %	3392 (37)
≥3 vessels, %	3003 (33)
No. of MACE, %	1285 (14)
BMI, kg/m^2	29.6±6.2
Total cholesterol, mg/dL	170±41.1
HDL cholesterol, mg/dL	40.0±13.5
LDL cholesterol, mg/dL	99.0±33.5
Triglycerides, mg/dL	151.5±110.1
TMAO, μmol/L	6.2±13.0
Taking lipid-lowering medication (%)	5751 (63)

Data are shown as mean±SD or numbers of individuals (%). BMI indicates body mass index; CAD, coronary artery disease; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MACE, major adverse cardiac events; and TMAO, trimethylamine *N*-oxide.

GWAS for Plasma TMAO Levels in Humans

To complement the mouse studies, we performed a 2-stage GWAS in GeneBank. In the first stage, ≈2.4 million genotyped and imputed autosomal SNPs were evaluated for association with plasma TMAO levels in 1973 subjects with adjustment for age and sex. The quantile-quantile plot for these analyses is shown in Figure 3A, and the observed genomic inflation factor (λ) was 1.007, indicating that the GWAS results are not confounded by underlying population stratification. As shown by the Manhattan plot in Figure 3B, 2 loci with suggestive evidence of association were identified on chromosomes 1q23.3 and 2p12. The lead SNP at the chromosome 1 locus (rs17359359; $P=2.8\times10^{-7}$) is located ≈47 kb telomeric of *NUF2*, which is a component of the kinetochore complex that is required for chromosome segregation but, to our knowledge, has no known relationship to TMAO metabolism. This locus is also located ≈8 Mb telomeric to the *FMO* gene cluster and is clearly distinct because there is no apparent long-range linkage disequilibrium between these 2 loci. By comparison, the lead SNP at the chromosome 2p12 locus (rs885187; $P=2.8\times10^{-7}$) does not map near any known gene. Based on previously observed sex differences in plasma TMAO levels, we also performed a GWAS in men and women separately. However, these analyses did not reveal sex-specific effects on chromosomes 1q23.3 and 2p12 or identify other loci (Figure I in the online-only Data Supplement).

In stage 2, we evaluated the chromosome 1 locus further by genotyping rs17359359 in 1892 additional GeneBank subjects for whom plasma TMAO levels were available. These analyses failed to replicate the association of rs17359359 with plasma TMAO levels in stage 2 ($P=0.85$), and a combined analysis with all subjects attenuated the overall association ($P=1.8\times10^{-4}$; Table 2). Based on the chromosome 3 locus identified in the

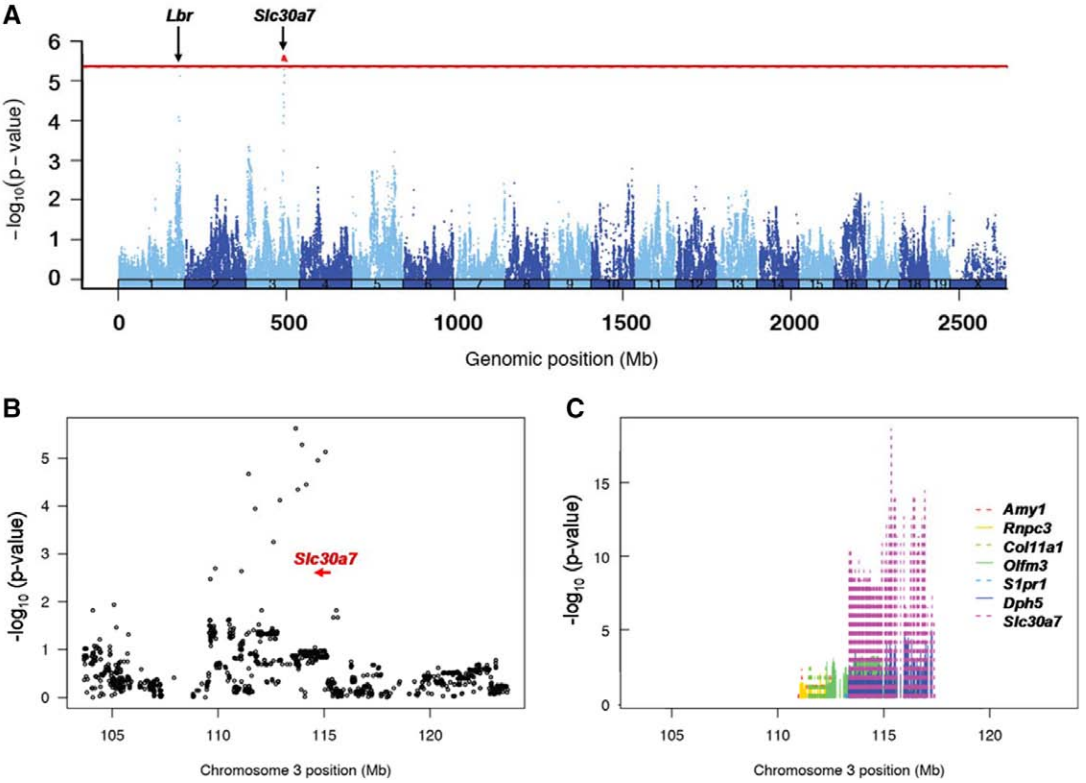


Figure 2. Manhattan plot for genome-wide association study (GWAS) of plasma trimethylamine *N*-oxide (TMAO) levels in mice. A GWAS for plasma TMAO levels in the hybrid mouse diversity panel (HMDP) identifies a significant locus over the solute carrier family 30 member 7 (*Slc30a7*) gene (red dot) at 110 to 115 Mb on chromosome 3 and a suggestive locus on chromosome 1 ≈40 kb away from the *Lbr* gene (**A**). A regional plot for chromosome 3 shows the location and transcriptional orientation of *Slc30a7* (indicated by red arrow) in relation to the peak single nucleotide polymorphisms (SNPs) in this region (**B**). Of the genes in this locus, a highly significant (P=1.07×10^{-20}) *cis*-acting expression quantitative trait loci is observed for *Slc30a7* (**C**). The red line indicates the genome-wide threshold for significance in the HMDP (P=4.1×10^{-6}). Plasma TMAO and hepatic mRNA levels were quantified in male mice from ≈100 HMDP strains (n=3–8 mice per strain) and analyzed for association with ≈107 000 SNPs, after correcting for population structure using the EMMA algorithm.

HMDP (Figure 2A and 2B), we used synteny mapping and evaluated the association of plasma TMAO levels with SNPs in the 1-Mb genomic region centered on the human *SLC30A7* ortholog located on chromosome 1p21.2. In the GWAS data set (n=1973), 1 SNP located ≈225 kb telomeric to *SLC30A7* (rs12402441) demonstrated nominal association (P=0.006) with plasma TMAO levels (Table 2). However, the association of rs12402441 with plasma TMAO levels did not replicate in the stage 2 samples, and a combined analysis with all subjects was not significant (Table 2). In the combined data set, there was also no evidence for an interaction between sex and either SNP on plasma TMAO levels (rs17359359, P_{int}=0.33; rs12402441, P_{int}=0.11). The sex-specific effects of rs17359359 and rs12402441 when men and women were analyzed separately are shown in Table I in the online-only Data Supplement.

Discussion

Using a combined mouse–human GWAS approach, we sought to identify the genetic determinants of plasma TMAO levels in mammals. Several factors served as the motivation for these studies, including recent studies demonstrating that TMAO can be generated from gut microbiota–mediated metabolism of either dietary choline or L-carnitine and that elevated plasma levels are strongly proatherogenic in both mice and humans.[11,12,14] Subsequent reports further showed that plasma TMAO levels in mice are regulated by both sex hormones, which could account, in part, for the observed dimorphism between male and female mice and increased *Fmo3* gene expression via bile acid–mediated activation of the farnesoid X receptor.[13] Of note, in humans, no differences in plasma TMAO levels were observed between men and women.[12] The collective results of these comprehensive, albeit limited, observational studies indicate that genes play a marginal role in determining TMAO levels and that any genetic effects are either complex or relatively weak. This is particularly true in humans and raises the possibility that variation in dietary composition or the repertoire of gut microbiota may be more important determinants of plasma TMAO levels.

Using the HMDP, we identified 1 locus on chromosome 3, which contains *Slc30A7* and was associated with plasma TMAO levels in male mice at the genome-wide significance threshold. This locus also exhibited evidence for *cis* gene regulation of *Slc30a7*, which is a subfamily member of the cation diffusion facilitator family of transporters and has essential functions in dietary zinc absorption.[21] Although a biological mechanism for how *Slc30a7* would regulate plasma TMAO levels is not directly evident, it has been reported that the zinc finger protein, YY1, regulates the expression of both rabbit and human *FMO1*.[22] Interestingly, the activity of certain bacterial monoooxygenases has also been shown to use zinc

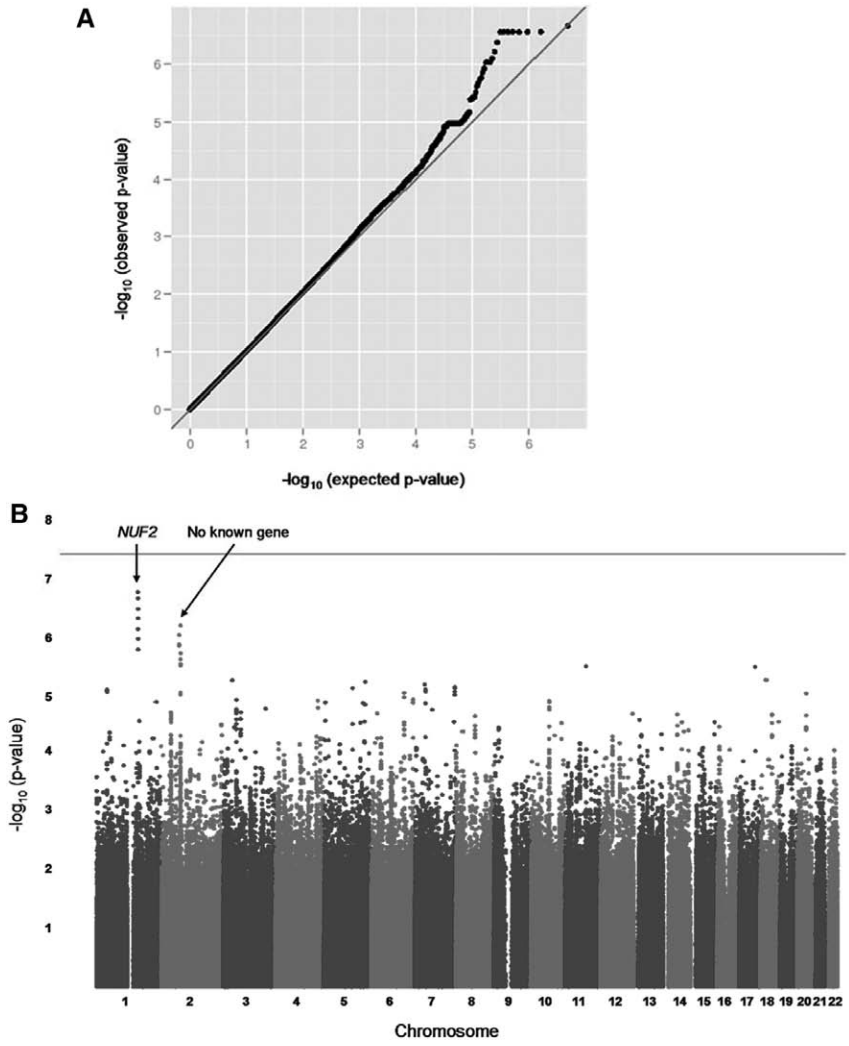


Figure 3. Results of a genome-wide association study (GWAS) for plasma trimethylamine *N*-oxide (TMAO) levels in humans. The quantile-quantile plot of the GWAS results for plasma TMAO levels in humans (n=1973) shows slight deviation of the observed *P* values from the expected distribution under the null hypothesis of no association (**A**). The observed genomic control factor in these analyses was 1.007, indicating that the results are not confounded by underlying population stratification. A GWAS analysis in humans identifies 2 loci on chromosomes 1 and 2 exhibiting suggestive evidence of association with plasma TMAO levels but no locus that exceeds the genome-wide threshold for significance (indicated by the horizontal red line; **B**).

as a cofactor.[23] However, more detailed functional studies will be required to determine whether *Slc30a7* could affect TMAO levels by influencing zinc-mediated activity of ≥1 of the FMOs in mice. We also note that although this locus on mouse chromosome 3 also yielded a highly significant *cis*-expression QTL for *Slc30a7*, we cannot exclude the possibility that another gene in this interval harboring functional coding variation is the causal genetic factor for plasma TMAO levels. Because our GWAS with the HMDP was only with male mice, it is also possible that inclusion of females would provide additional support for the association of the *Slc30a7* locus, as well as identify other genomic regions controlling plasma TMAO levels that are potentially specific to females. For example, we previously reported that plasma TMAO levels are several-fold higher in female mice compared with males, a portion of which is attributable to differences in sex hormones.[13]

As a comparative analysis to our studies with the HMDP, we also performed a GWAS for plasma TMAO levels in the GeneBank cohort. This analysis identified 2 suggestive loci in the discovery phase, but our attempt to replicate the *NUF2* locus on chromosome 1 was unsuccessful. Based on the *Slc30a7* locus identified in the mouse GWAS, we also evaluated the syntenic region on human chromosome 1p21.2 for association with plasma TMAO levels. Although 1 SNP in this region yielded nominal association with plasma TMAO levels in humans, this signal also did not replicate in the stage 2 samples. Given the high concordance rate (>98.8%) for genotypes of the same DNA samples used in stages 1 and 2, we do not think technical variability to have been a factor for the lack of replication in stage 2 and conclude that these loci likely represent false-positive signals. However, despite the lack of genetic variation around the human *SLC30A7* ortholog being associated with plasma TMAO levels, it is possible that this transporter still plays a biological role in regulating TMAO levels in both species. Furthermore, we did not obtain any evidence for sex-specific effects at these loci or identify any others when the GWAS was performed in men and women separately. Taken together, these results suggest that variation in plasma TMAO levels in humans may be because of weak genetic effects and that larger sample sizes will be required to identify the underlying regulatory factors.

To date, *FMO3* is the only genetic factor known to affect plasma TMAO levels in humans. *FMO3* is composed of 10 exons spanning 26.9 kb on chromosome 1q24.3 and encodes a 532-residue enzyme. At the amino acid level, FMO3 shares ≥79% homology with the mouse Fmo3 protein and other members of the human FMO family. Interestingly, we previously demonstrated that FMO1, FMO2, and FMO3 were able

Table 2. Effect of Single Nucleotide Polymorphisms Identified Through GWAS in Humans and Mice on Plasma TMAO Levels in the GeneBank Cohort

	rs17359359				rs12402441			
Stage	GG	AG	AA	*P* Value*	AA	AG	GG	*P* Value*
GWAS	5.3±8.0 (n=1727)	8.2±20.3 (n=238)	9.9±8.1 (n=8)	2.8×10^{-7}	5.8±10.8 (n=1773)	4.5±4.7 (n=186)	4.5±3.7 (n=14)	0.006
Replication	6.6±14.5 (n=1495)	8.3±23.2 (n=186)	4.0±2.9 (n=9)	0.71	6.6±13.3 (n=1598)	8.8±31.2 (n=158)	4.5±1.5 (n=10)	0.68
Combined	5.9±11.5 (n=3222)	8.3±21.6 (n=424)	6.8±6.5 (n=17)	1.1×10^{-4}	6.2±12.1 (n=3371)	6.5±21.5 (n=344)	4.5±2.9 (n=24)	0.14

Mean (±SD) plasma TMAO levels (μmol/L) are shown as a function of genotype. GWAS indicates genome-wide association study; and TMAO, trimethylamine *N*-oxide.

**P* values obtained using linear regression with natural log-transformed values after adjustment for age and sex.

to generate TMAO from TMA but that FMO3 was by far the most active family member.[13] Because the Mendelian disease trimethylaminuria is caused by rare mutations that lead to *FMO3* deficiency, we leveraged our own data in GeneBank and those from public sources to evaluate whether common variants at the *FMO* locus were associated with *FMO3* gene expression, plasma TMAO levels, and risk of CAD. However, these analyses in humans did not reveal any strong associations with SNPs surrounding *FMO3*. It is possible that the imputed genotypes from the GWAS data we used did not provide sufficient coverage of the variation around *FMO3* (or the entire *FMO* locus). Based on data for subjects of European ancestry from the 1000 Genomes Project, 59 tagging SNPs with minor allele frequencies ≥1% would cover *FMO3* at an $r^2\geq0.8$. However, only 15 tagging SNPs across *FMO3* were present in our analyses of plasma TMAO levels in GeneBank. In addition, rare variants in *FMO3* that could influence gene expression, TMAO production, and risk of CAD would also not necessarily be represented by our imputed GWAS data. By comparison, our previous studies in mice revealed a relatively strong *cis*-expression QTL for *Fmo3* expression over the *Fmo* locus. However, the present analyses for plasma TMAO levels in the HMDP did not yield association with the *Fmo* locus at the genome-wide level (data not shown). These observations suggest that the relationship between *FMO3* gene expression and plasma TMAO levels in both mice and humans is complex and that other regulatory mechanisms, including post-transcriptional and post-translational modifications, may exist.

The discordance between rare mutations in *FMO3* that dramatically reduce plasma TMAO levels and the lack of common genetic determinants associated with this metabolite implies that variation in TMAO levels in humans and mice may be influenced by other factors, such as gut microbial and dietary composition. For example, we previously defined the relative abundances of bacteria at each taxonomic level in relation to the production of TMAO through pyrosequencing of 16S rRNA genes in both mice and humans. One notable difference in these analyses was the source of gut bacteria because the contents of the cecum were used for mice, whereas stool samples were used for the human analyses. This may explain, at least in part, why a direct comparison of bacterial taxa associated with plasma TMAO concentrations did not identify any genre common to both species. This observation is consistent with previous reports indicating that microbes identified from the distal gut of the mouse do not necessarily represent those typically detected in humans.[24,25] Thus, although sharing many taxa, the microbial composition observed in mice is architecturally and globally different than in humans. Despite these differences, we were still able to demonstrate associations between dietary patterns (eg, vegan/vegetarian versus omnivore or normal chow versus choline/carnitine supplemented) and both plasma TMAO levels and proportions of specific taxa of fecal microbes in humans and cecal microbes in mice.[14] These observations suggest that high dietary intake of L-carnitine or choline would lead to increased plasma TMAO levels, particularly if specific bacterial taxa that metabolize these nutrients to TMA are present in the gut. It is possible that the effects of host genetic factors would also manifest under such dietary conditions. However, compared with mice housed under standardized environmental conditions, the diet in free-ranging humans is far more heterogeneous, which would add further complexity and diversity to any potential interactions with the gut microbiome.

Despite our comprehensive efforts to identify loci associated with plasma TMAO levels, we also note several potential limitations of our study. First, we used GWAS approaches in mice and humans that mostly test association with common genetic variation, which would not necessarily detect the effects of rare variants on plasma TMAO levels. Second, our human GWAS was performed in subjects of European ancestry, and it is possible that genetic variants that are either specific to or present at higher frequency in other ethnicities could influence TMAO levels. Third, although including ≈100 inbred strains, it is still possible that the HMDP does provide sufficient genetic variation to capture all of the effects on plasma TMAO levels in mice compared with the substantially greater genetic diversity present in outbred human populations. In addition, the pathways leading to variation in TMAO levels in mice and humans may not be entirely similar. Finally, as discussed above, variability in dietary composition, particularly in humans, and the gut microbiome clearly factor into plasma TMAO levels and are thus likely to be strong confounding variables that our study did not take into consideration.

Conclusions

Our results indicate that *Slc30a7* may represent a novel gene for TMAO levels in mice but that the contribution of genetic factors in humans is more complex. These observations

suggest that the inter-relationships between dietary choline and L-carnitine levels with the composition of gut microbes are perhaps more likely determinants of variation in plasma TMAO levels. Exploring such interactions as part of future studies may help to identify the intrinsic genetic factors that influence plasma TMAO levels and their influence on the development of atherosclerosis.

Sources of Funding

This study was supported, in part, by National Institutes of Health (NIH) grants K99HL102223, P01HL30568, P01HL28481, R01HL103866, P20HL113452, R01ES021801, a pilot project award from the Southern California Clinical and Translational Science Institute through NIH grant UL1TR000130, and American Heart Association Scientist Development grant 12SDG12050473. GeneBank was supported in part by NIH grants P01HL098055, P01HL076491, and R01HL103931. R. Roberts has received research funding from Canadian Institutes of Health Research MOT82810 and Canada Foundation for Innovation 11966. S.L. Hazen is also partially supported by a gift from the Leonard Krieger Fund. Mass spectrometry instrumentation used was housed within the Cleveland Clinic Mass Spectrometry Facility with partial support through a Center of Innovation by AB SCIEX.

Disclosures

S.L. Hazen is named as coinventor on pending and issued patents held by the Cleveland Clinic relating to cardiovascular diagnostics. He reports having been paid as a consultant or speaker for the following companies: Cleveland Heart Laboratory, Inc, Esperion, Liposciences Inc, Merck & Co, Inc, and Pfizer Inc. He reports he has received research funds from Abbott, Cleveland Heart Laboratory, Esperion, and Liposciences, Inc. He has the right to receive royalty payments for inventions or discoveries related to cardiovascular diagnostics from Abbott Laboratories, Cleveland Heart Laboratory, Inc, Frantz Biomarkers, LLC, and Siemens. The other authors report no conflicts.

References

1. Zeisel SH, Blusztajn JK. Choline and human nutrition. *Annu Rev Nutr*. 1994;14:269–296.
2. Buchman AL, Ament ME, Sohel M, Dubin M, Jenden DJ, Roch M, Pownall H, Farley W, Awal M, Ahn C. Choline deficiency causes reversible hepatic abnormalities in patients receiving parenteral nutrition: proof of a human choline requirement: a placebo-controlled trial. *JPEN J Parenter Enteral Nutr*. 2001;25:260–268.
3. Hollenbeck CB. The importance of being choline. *J Am Diet Assoc*. 2010;110:1162–1165.
4. Meck WH, Williams CL. Metabolic imprinting of choline by its availability during gestation: implications for memory and attentional processing across the lifespan. *Neurosci Biobehav Rev*. 2003;27:385–399.
5. Cermak JM, Holler T, Jackson DA, Blusztajn JK. Prenatal availability of choline modifies development of the hippocampal cholinergic system. *FASEB J*. 1998;12:349–357.
6. Craciunescu CN, Albright CD, Mar MH, Song J, Zeisel SH. Choline availability during embryonic development alters progenitor cell mitosis in developing mouse hippocampus. *J Nutr.* 2003;133:3614–3618.
7. Yates AA, Schlicker SA, Suitor CW. Dietary Reference Intakes: the new basis for recommendations for calcium and related nutrients, B vitamins, and choline. *J Am Diet Assoc*. 1998;98:699–706.
8. Zeisel SH. Choline: an essential nutrient for humans. *Nutrition*. 2000;16:669–671.
9. Krueger SK, Williams DE. Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol Ther*. 2005;106:357–387.
10. Phillips IR, Shephard EA. Flavin-containing monooxygenases: mutations, disease and drug response. *Trends Pharmacol Sci*. 2008;29:294–301.
11. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011;472:57–63.
12. Tang WH, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, Wu Y, Hazen SL. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med*. 2013;368:1575–1584.
13. Bennett BJ, de Aguiar Vallim TQ, Wang Z, Shih DM, Meng Y, Gregory J, Allayee H, Lee R, Graham M, Crooke R, Edwards PA, Hazen SL, Lusis AJ. Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell Metab*. 2013;17:49–60.
14. Koeth RA, Wang Z, Levison BS, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med.* 2013;19:576–585.
15. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol*. 2008;6:e107.
16. Schunkert H, König IR, Kathiresan S, et al.; Cardiogenics; CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011;43:333–338.
17. Bennett BJ, Farber CR, Orozco L, et al. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res*. 2010;20:281–290.
18. Farber CR, Bennett BJ, Orozco L, et al. Mouse genome-wide association and systems genetics identify Asxl2 as a regulator of bone mineral density and osteoclastogenesis. *PLoS Genet*. 2011;7:e1002038.
19. Ghazalpour A, Bennett B, Petyuk VA, et al. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet*. 2011;7:e1001393.
20. Park CC, Gale GD, de Jong S, Ghazalpour A, Bennett BJ, Farber CR, Langfelder P, Lin A, Khan AH, Eskin E, Horvath S, Lusis AJ, Ophoff RA, Smith DJ. Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Syst Biol*. 2011;5:43.
21. Huang L, Yu YY, Kirschke CP, Gertz ER, Lloyd KK. Znt7 (Slc30a7)-deficient mice display reduced body zinc status and body fat accumulation. *J Biol Chem*. 2007;282:37053–37063.
22. Luo Z, Hines RN. Regulation of flavin-containing monooxygenase 1 expression by ying yang 1 and hepatic nuclear factors 1 and 4. *Mol Pharmacol.* 2001;60:1421–1430.
23. Ensign SA, Allen JR. Aliphatic epoxide carboxylation. *Annu Rev Biochem*. 2003;72:55–76.
24. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*. 2005;102:11070–11075.
25. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. Evolution of mammals and their gut microbes. *Science*. 2008;320:1647–1651.

Significance

Elevated plasma levels of trimethylamine *N*-oxide (TMAO), a metabolite generated from dietary choline and carnitine by intestinal bacteria, have recently been identified as a novel risk factor for coronary artery disease. Notably, elimination of bacteria in the gut through administration of antibiotics reduced TMAO levels in mice and humans and decreased atherosclerosis in mice. However, the genes that control plasma TMAO levels are not well defined. The present study uses a comparative genome-wide association study approach in mice and humans to identify the genetic determinants of plasma TMAO levels. In mice, genetic variants near solute carrier family 30 member 7 were significantly associated with plasma TMAO levels, whereas no locus was identified in humans. Our findings suggest that, at least in humans, plasma TMAO levels are under complex genetic regulation, that the effects of any underlying genes are relatively weak, and that variation in gut bacteria may be more important in determining TMAO levels.

Materials and Methods

Study Population. The Cleveland Clinic GeneBank study is a single site sample repository generated from consecutive patients undergoing elective diagnostic coronary angiography or elective cardiac computed tomographic angiography with extensive clinical and laboratory characterization and longitudinal observation. Subject recruitment occurred between 2001 and 2006. Ethnicity was self-reported and information regarding demographics, medical history, and medication use was obtained by patient interviews and confirmed by chart reviews. All clinical outcome data were verified by source documentation. CAD was defined as adjudicated diagnoses of stable or unstable angina, MI (adjudicated definition based on defined electrocardiographic changes or elevated cardiac enzymes), angiographic evidence of $\geq$ 50% stenosis of one or more major epicardial vessel, and/or a history of known CAD (documented MI, CAD, or history of revascularization). Prospective cardiovascular risk was assessed by the incidence of major adverse cardiac events (MACE) during three years of follow-up from the time of enrollment, which included nonfatal MI, nonfatal stroke, and all-cause mortality. Nonfatal events were defined as MI or stroke in patients who survived at least 48 hours following the onset of symptoms. Adjudicated outcomes ascertained over the ensuing 3 years for all subjects following enrollment were confirmed using source documentation. The GeneBank Study has been used previously for discovery and replication of novel genes and risk factors for atherosclerotic disease [1-6].

Animal Husbandry. Male mice were purchased from The Jackson Laboratory and housed in vivaria accredited by the Association for Assessment and Accreditation of Laboratory Animal Care. Animals were on a 12hr light–dark cycle and maintained on a chow diet with 6% fat by

weight (Ralston-Purina Co.) until sacrifice at 16 weeks of age. Following a 16hr fast, mice were bled retro-orbitally under isoflurane anesthesia and plasma TMAO levels were determined as described below. All procedures were approved by the UCLA Institutional Animal Care and Use Committee.

Measurement of Plasma TMAO Levels. TMAO levels in human and mouse plasma was quantitated using stable isotope dilution high performance liquid chromatography (HPLC) with on line electrospray ionization tandem mass spectrometry on an API 5000 triple quadrupole mass spectrometer (AB SCIEX, Foster City, CA) interfaced with a Cohesive HPLC (Franklin, MA) equipped with a phenyl column (4.6 × 2505mm, 5μm RexChrom Phenyl; Regis, Morton Grove, IL). Separation was performed using a gradient starting from 10mM ammonium formate over 0.5 min, then to 5mM ammonium formate, 25 % methanol and 0.1 % formic acid over 3min, held for 8min, followed by 100% methanol and water washing for 3min. TMAO was monitored in multiple reaction monitoring (MRM) mode using characteristic parent-daughter ion transitions at m/z 76→58. The internal standards TMAO-trimethyl-d9 and choline-trimethyl-d9, were added to plasma samples prior to protein precipitation and similarly monitored in MRM mode at m/z 85→68 and m/z 113→69. Various concentrations of TMAO standards and a fixed amount of internal standards were spiked into control plasma to prepare the calibration curves for quantification of plasma analytes.

Genome-wide Association Mapping and Significance Threshold in Mice. A GWAS for plasma TMAO levels in mice was carried out using the Hybrid Mouse Diversity Panel (HMDP). The HMDP is comprised of ~100 inbred and recombinant inbred (RI) strains that have

previously been assembled for high-resolution association mapping and subjected to extensive metabolic, transcriptional, and proteomic profiling [7]. Genotypes of single nucleotide polymorphisms (SNPs) for the inbred strains in the HMDP were obtained from the Broad Institute (www.broadinstitute.org/mouse/hapmap) and combined with the genotypes from the Wellcome Trust Center for Human Genetics (WTCHG). Genotypes of RI strains at the Broad SNPs were inferred from the WTCHG genotypes by imputing alleles at polymorphic SNPs among parental strains, with ambiguous genotypes labeled as "missing." Of the 140,000 SNPs available, 107,145 were informative with an allele frequency greater than 5% and used in the present GWAS analysis.

We applied the following linear mixed model to account for the population structure and genetic relatedness among strains: $y=\mu+x\beta+u+e$ where μ represents mean TMAO levels, x represents the SNP effect, u represents random effects due to genetic relatedness with $Var(u) = \sigma g2K$ and $Var(e) = \sigma e2I$, where K represents an identity-by-descent kinship matrix across all genotypes. A restricted maximum likelihood estimate of $\sigma g2$ and $\sigma e2$ were computed using an efficient mixed model association algorithm (EMMA) [8], and the association mapping was performed based on the estimated variance component with a standard F test to test $\beta \neq 0$.

Genome-wide significance threshold in the HMDP was determined by the family-wise error rate as the probability of observing one or more false positives across all SNPs per phenotype. We ran 100 different sets of permutation tests and parametric bootstrapping of size 1,000 and observed that the mean and standard error of the genome-wide significance threshold at FWER of 0.05 were $3.9 \times 10^{-6} \pm 0.3 \times 10^{-6}$ and $4.0 \times 10^{-6} \pm 0.3 \times 10^{-6}$, respectively. This is approximately an order of magnitude larger than the threshold obtained by Bonferroni correction

(4.6×10^{-7}), which would be an overly conservative estimate of significance since nearby SNPs among inbred mouse strains are highly correlated with each other.

Genotyping. Genome-wide genotyping of SNPs in humans was performed on the Affymetrix Genome-Wide Human Array 6.0 chip. Using these data and those from 120 phased chromosomes from the HapMap CEU samples (HapMap r22 release, NCBI build 36), genotypes were imputed for untyped autosomal SNPs across the genome using MACH 1.0 software. All imputations were done on the forward (+) strand using 562,554 genotyped SNPS that had passed quality control (QC) filters. QC filters for the imputed dataset excluded SNPs with HWE p-values < 0.0001, call rate less than 97% or minor allele frequencies < 1%, and individuals with less than 90% call rates. This resulted in 2,421,770 autosomal SNPs that were available for a GWAS analysis in 1973 GeneBank subjects. Genotyping of the two SNPs selected for replication in stage 2 was performed using the TaqMan Allelic Discrimination system from Applied Biosystems, Inc. (Foster City, CA). In control samples from the GWAS dataset that were also genotyped by Taqman, the concordance rate with genotypes obtained from the Affymetrix chip was >98.8% for the two replication SNPs.

Statistical Analyses. A GWAS for plasma TMAO levels in humans was carried out with adjustment for age and sex. Linear regression analyses were performed using natural log transformed values under an additive genetic model. A publicly available liver expression dataset [9] was used to evaluate the association of SNPs in the region containing the *FMO* gene cluster for cis expression quantitative trait loci (eQTL). Seventy nine SNPs were available in this dataset for the *FMO* region, of which 22 were excluded due to missing genotypes. Linear

regression was carried out with the remaining 57 SNPs under an additive genetic model with adjustment for age and sex to identify eQTLs for hepatic *FMO3* mRNA levels. The results of the Coronary Artery Disease Genome-wide Replication And Meta-Analysis (CARDIoGRAM) Consortium were used to determine whether variants at the *FMO* locus were associated with CAD. CARDIoGRAM represents a GWAS meta-analysis of CAD comprising a discovery set of ~22,000 cases and ~65,000 controls, in which logistic regression was first used in each cohort to test for association with CAD using a log-additive model with adjustment for age and gender and taking into account the uncertainty of possibly imputed genotypes. Subsequently, a meta-analysis was performed separately for every SNP from each study that passed the quality control criteria using a fixed effect model with inverse variance weighting [10]. The results of this default meta-analysis were used to determine whether SNPs spanning the *FMO* cluster on chromosome 1 were associated with CAD. All genetic analyses were performed using PLINK 1.07 [11] or SAS 9.3 (SAS Institute Inc, Cary, NC).

References

1. Bhattacharyya T, Nicholls SJ, Topol EJ, Zhang R, Yang X, Schmitt D, Fu X, Shao M, Brennan DM, Ellis SG, Brennan ML, Allayee H, Lusis AJ, Hazen SL. Relationship of paraoxonase 1 (PON1) gene polymorphisms and functional activity with systemic oxidative stress and cardiovascular risk. *JAMA*. 2008;299:1265-76.
2. Hartiala J, Li D, Conti DV, Vikman S, Patel Y, Tang WH, Brennan ML, Newman JW, Stephensen CB, Armstrong P, Hazen SL, Allayee H. Genetic contribution of the leukotriene pathway to coronary artery disease. *Hum Genet*. 2011;129:617-27.
3. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011;472:57-63.
4. Tang WH, Wu Y, Hartiala J, Fan Y, Stewart AF, Roberts R, McPherson R, Fox PL, Allayee H, Hazen SL. Clinical and genetic association of serum ceruloplasmin with cardiovascular risk. *Arterioscler Thromb Vasc Biol*. 2012;32:516-22.
5. Tang WH, Hartiala J, Fan Y, Wu Y, Stewart AF, Erdmann J, Kathiresan S, Roberts R, McPherson R, Allayee H, Hazen SL. Clinical and genetic association of serum

paraoxonase and arylesterase activities with cardiovascular risk. *Arterioscler Thromb Vasc Biol*. 2012;32:2803-2812.
6. Reiner AP, Hartiala J, Zeller T, et al. Genome-wide and gene-centric analyses of circulating myeloperoxidase levels in the charge and care consortia. *Hum Mol Genet*. 2013;22:3381-93.
7. Ghazalpour A, Rau CD, Farber CR, et al. Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mamm Genome*. 2012;23:680-92.
8. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient control of population structure in model organism association mapping. *Genetics*. 2008;178:1709-23.
9. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol*. 2008;6:e107.
10. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333-8.
11. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-75.

Supplemental Material

Comparative Genome-Wide Association Studies in Mice and Humans for Trimethylamine *N*-oxide, a Pro-Atherogenic Metabolite of Choline and L-Carnitine

Hartiala et al.

The CARDIoGRAM Consortium

Executive Committee: Sekar Kathiresan[1,2,3], Muredach P. Reilly[4], Nilesh J. Samani[5,6], Heribert Schunkert[7,79]

Executive Secretary: Jeanette Erdmann[7,79]

Steering Committee: Themistocles L. Assimes[8], Eric Boerwinkle[9], Jeanette Erdmann[7,79] Alistair Hall[10], Christian Hengstenberg[11], Sekar Kathiresan[1,2,3], Inke R. König[12], Reijo Laaksonen[13], Ruth McPherson[14], Muredach P. Reilly[4], Nilesh J. Samani[5,6], Heribert Schunkert[7,79], John R. Thompson[15], Unnur Thorsteinsdottir[16,17], Andreas Ziegler[12]

Statisticians: Inke R. König[12] (chair), John R. Thompson[15] (chair), Devin Absher[18], Li Chen[19], L. Adrienne Cupples[20,21], Eran Halperin[22], Mingyao Li[23], Kiran Musunuru[1,2,3], Michael Preuss[12,7], Arne Schillert[12], Gudmar Thorleifsson[16], Benjamin F. Voight[2,3,24], George A. Wells[25]

Writing group: Themistocles L. Assimes[8], Panos Deloukas[26], Jeanette Erdmann[7,79], Hilma Holm[16], Sekar Kathiresan[1,2,3], Inke R. König[12], Ruth McPherson[14], Muredach P. Reilly[4], Robert Roberts[14], Nilesh J. Samani[5,6], Heribert Schunkert[7,79], Alexandre F. R. Stewart[14]

ADVANCE: Devin Absher[18], Themistocles L. Assimes[8], Stephen Fortmann[8], Alan Go[27], Mark Hlatky[8], Carlos Iribarren[27], Joshua Knowles[8], Richard Myers[18], Thomas Quertermous[8], Steven Sidney[27], Neil Risch[28], Hua Tang[29]

CADomics: Stefan Blankenberg[30]*, Tanja Zeller[30]*, Arne Schillert[12], Philipp Wild[30], Andreas Ziegler[12], Renate Schnabel[30]*, Christoph Sinning[30]*, Karl Lackner[31], Laurence Tiret[32], Viviane Nicaud[32], Francois Cambien[32], Christoph Bickel[30], Hans J. Rupprecht[30], Claire Perret[32], Carole Proust[32], Thomas Münzel[30]

CHARGE: Maja Barbalic[33], Joshua Bis[34], Eric Boerwinkle[9], Ida Yii-Der Chen[35], L. Adrienne Cupples[20,21], Abbas Dehghan[36], Serkalem Demissie-Banjaw[37,21], Aaron Folsom[38], Nicole

Glazer[39], Vilmundur Gudnason[40,41], Tamara Harris[42], Susan Heckbert[43], Daniel Levy[21], Thomas Lumley[44], Kristin Marciante[45], Alanna Morrison[46], Christopher J. O´Donnell[47], Bruce M. Psaty[48], Kenneth Rice[49], Jerome I. Rotter[35], David S. Siscovick[50], Nicholas Smith[43], Albert Smith[40,41], Kent D. Taylor[35], Cornelia van Duijn[36], Kelly Volcik[46], Jaqueline Whitteman[36], Vasan Ramachandran[51], Albert Hofman[36], Andre Uitterlinden[52,36]

deCODE: Solveig Gretarsdottir[16], Jeffrey R. Gulcher[16], Hilma Holm[16], Augustine Kong[16], Kari Stefansson[16,17], Gudmundur Thorgeirsson[53,17], Karl Andersen[53,17], Gudmar Thorleifsson[16], Unnur Thorsteinsdottir[16,17]

GERMIFS I and II: Jeanette Erdmann[7,79], Marcus Fischer[11], Anika Grosshennig[12,7], Christian Hengstenberg[11], Inke R. König[12], Wolfgang Lieb[54], Patrick Linsel-Nitschke[7], Michael Preuss[12,7], Klaus Stark[11], Stefan Schreiber[55], H.-Erich Wichmann[56,58,59], Andreas Ziegler[12], Heribert Schunkert[7,79]

GERMIFS III (KORA): Zouhair Aherrahrou[7,79], Petra Bruse[7,79], Angela Doering[56], Jeanette Erdmann[7,79], Christian Hengstenberg[11], Thomas Illig[56], Norman Klopp[56], Inke R. König[12], Patrick Diemert[7], Christina Loley[12,7], Anja Medack[7,79], Christina Meisinger[56], Thomas Meitinger[57,60], Janja Nahrstedt[12,7], Annette Peters[56], Michael Preuss[12,7], Klaus Stark[11], Arnika K. Wagner[7], H.-Erich Wichmann[56,58,59], Christina Willenborg[7,79], Andreas Ziegler[12], Heribert Schunkert[7,79]

LURIC/AtheroRemo: Bernhard O. Böhm[61], Harald Dobnig[62], Tanja B. Grammer[63], Eran Halperin[22], Michael M. Hoffmann[64], Marcus Kleber[65], Reijo Laaksonen[13], Winfried März[63,66,67], Andreas Meinitzer[66], Bernhard R. Winkelmann[68], Stefan Pilz[62], Wilfried Renner[66], Hubert Scharnagl[66], Tatjana Stojakovic[66], Andreas Tomaschitz[62], Karl Winkler[64]

MIGen: Benjamin F. Voight[2,3,24], Kiran Musunuru[1,2,3], Candace Guiducci[3], Noel Burtt[3], Stacey B. Gabriel[3], David S. Siscovick[50], Christopher J. O'Donnell[47], Roberto Elosua[69], Leena Peltonen[49], Veikko Salomaa[70], Stephen M. Schwartz[50], Olle Melander[26], David Altshuler[71,3], Sekar Kathiresan[1,2,3]

OHGS: Alexandre F. R. Stewart[14], Li Chen[19], Sonny Dandona[14], George A. Wells[25], Olga Jarinova[14], Ruth McPherson[14], Robert Roberts[14]

PennCATH/MedStar: Muredach P. Reilly[4], Mingyao Li[23], Liming Qu[23], Robert Wilensky[4], William Matthai[4], Hakon H. Hakonarson[72], Joe Devaney[73], Mary Susan Burnett[73], Augusto D. Pichard[73], Kenneth M. Kent[73], Lowell Satler[73], Joseph M. Lindsay[73], Ron Waksman[73], Christopher W. Knouff[74], Dawn M. Waterworth[74], Max C. Walker[74], Vincent Mooser[74], Stephen E. Epstein[73], Daniel J. Rader[75,4]

WTCCC: Nilesh J. Samani[5,6], John R. Thompson[15], Peter S. Braund[5], Christopher P. Nelson[5], Benjamin J. Wright[76], Anthony J. Balmforth[77], Stephen G. Ball[78], Alistair S. Hall[10], Wellcome Trust Case Control Consortium

CARDIoGRAM Affiliations

1 Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, MA, USA; 2 Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 3 Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, MA, USA; 4 The Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA, USA; 5 Department of Cardiovascular

Sciences, University of Leicester, Glenfield Hospital, Leicester, UK; 6 Leicester National Institute for Health Research Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester, LE3 9QP, UK; 7 Medizinische Klinik II, Universität zu Lübeck, Lübeck, Germany; 8 Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; 9 University of Texas Health Science Center, Human Genetics Center and Institute of Molecular Medicine, Houston, TX, USA; 10 Division of Cardiovascular and Neuronal Remodelling, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, UK; 11 Klinik und Poliklinik für Innere Medizin II, Universität Regensburg, Regensburg, Germany; 12 Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany; 13 Science Center, Tampere University Hospital, Tampere, Finland; 14 The John & Jennifer Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Canada; 15 Department of Health Sciences, University of Leicester, Leicester, UK; 16 deCODE Genetics, 101 Reykjavik, Iceland; 17 University of Iceland, Faculty of Medicine, 101 Reykjavik, Iceland; 18 Hudson Alpha Institute, Huntsville, Alabama, USA; 19 Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, 40 Ruskin Street, Ottawa, Ontario, Canada, K1Y 4W7; 20 Department of Biostatistics, Boston University School of Public Health, Boston, MA USA; 21 National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA; 22 The Blavatnik School of Computer Science and the Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Tel-Aviv, Israel, and the International Computer Science Institute, Berkeley, CA, USA; 23 Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA; 24 Department of Medicine, Harvard Medical School, Boston, MA, USA; 25 Research Methods, Univ Ottawa Heart Inst; 26 Department of Clinical Sciences,

Hypertension and Cardiovascular Diseases, Scania University Hospital, Lund University, Malmö, Sweden; 27 Division of Research, Kaiser Permanente, Oakland, CA, USA; 28 Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA; 29 Dept Cardiovascular Medicine, Cleveland Clinic; 30 Medizinische Klinik und Poliklinik, Johannes-Gutenberg Universität Mainz, Universitätsmedizin, Mainz, Germany; 31 Institut für Klinische Chemie und Laboratoriumsmediizin, Johannes-Gutenberg Universität Mainz, Universitätsmedizin, Mainz, Germany; 32 INSERM UMRS 937, Pierre and Marie Curie University (UPMC, Paris 6) and Medical School, Paris, France; 33 University of Texas Health Science Center, Human Genetics Center, Houston, TX, USA; 34 Cardiovascular Health Resarch Unit and Department of Medicine, University of Washington, Seattle, WA USA; 35 Cedars-Sinai Medical Center, Medical Genetics Institute, Los Angeles, CA, USA; 36 Erasmus Medical Center, Department of Epidemiology, Rotterdam, The Netherlands; 37 Boston University, School of Public Health, Boston, MA, USA; 38 University of Minnesota School of Public Health, Division of Epidemiology and Community Health, School of Public Health (A.R.F.), Minneapolis, MN, USA; 39 University of Washington, Cardiovascular Health Research Unit and Department of Medicine, Seattle, WA, USA; 40 Icelandic Heart Association, Kopavogur Iceland; 41 University of Iceland, Reykjavik, Iceland; 42 Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda MD, USA; 43 University of Washington, Department of Epidemiology, Seattle, WA, USA; 44 University of Washington, Department of Biostatistics, Seattle, WA, USA; 45 University of Washington, Department of Internal Medicine, Seattle, WA, USA; 46 University of Texas, School of Public Health, Houston, TX, USA; 47 National Heart, Lung and Blood Institute, Framingham Heart Study, Framingham, MA and Cardiology Division,

Massachusetts General Hospital, Boston, MA, USA; 48 Center for Health Studies, Group Health, Departments of Medicine, Epidemiology, and Health Services, Seattle, WA, USA; 49 The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; 50 Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle; 51 Boston University Medical Center, Boston, MA, USA; 52 Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; 53 Department of Medicine, Landspitali University Hospital, 101 Reykjavik, Iceland; 54 Boston University School of Medicine, Framingham Heart Study, Framingham, MA, USA; 55 Institut für Klinische Molekularbiologie, Christian-Albrechts Universität, Kiel, Germany; 56 Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany; 57 Institut für Humangenetik, Helmholtz Zentrum München, Deutsches Forschungszentrum für Umwelt und Gesundheit, Neuherberg, Germany; 58 Institute of Medical Information Science, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Germany; 59 Klinikum Grosshadern, Munich, Germany; 60 Institut für Humangenetik, Technische Universität München, Germany; 61 Division of Endocrinology and Diabetes, Graduate School of Molecular Endocrinology and Diabetes, University of Ulm, Ulm, Germany; 62 Division of Endocrinology, Department of Medicine, Medical University of Graz, Austria; 63 Synlab Center of Laboratory Diagnostics Heidelberg, Heidelberg, Germany; 64 Division of Clinical Chemistry, Department of Medicine, Albert Ludwigs University, Freiburg, Germany; 65 LURIC non profit LLC, Freiburg, Germany; 66 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University Graz, Austria; 67 Institute of Public Health, Social and Preventive Medicine, Medical Faculty Manneim, University of Heidelberg, Germany; 68 Cardiology Group Frankfurt-Sachsenhausen, Frankfurt, Germany; 69

Cardiovascular Epidemiology and Genetics Group, Institut Municipal d'Investigació Mèdica, Barcelona; Ciber Epidemiología y Salud Pública (CIBERSP), Spain; 70 Chronic Disease Epidemiology and Prevention Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland; 71 Department of Molecular Biology and Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, USA; 72 The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 73 Cardiovascular Research Institute, Medstar Health Research Institute, Washington Hospital Center, Washington, DC 20010, USA; 74 Genetics Division and Drug Discovery, GlaxoSmithKline, King of Prussia, Pennsylvania 19406, USA; 75 The Institute for Translational Medicine and Therapeutics, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 76 Department of Cardiovascular Surgery, University of Leicester, Leicester, UK; 77 Division of Cardiovascular and Diabetes Research, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, LS2 9JT, UK; 78 LIGHT Research Institute, Faculty of Medicine and Health, University of Leeds, Leeds, UK; 79 Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK), Universität zu Lübeck, Lübeck, Germany

*Present Address: Department of General and Interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany.

CARDIoGRAM Sources of Funding

The **ADVANCE** study was supported by a grant from the Reynold's Foundation and NHLBI grant HL087647.

Genetic analyses of **CADomics** were supported by a research grant from Boehringer Ingelheim. Recruitment and analysis of the CADomics cohort was supported by grants from Boehringer Ingelheim and PHILIPS medical Systems, by the Government of Rheinland-Pfalz in the context of the "Stiftung Rheinland-Pfalz für Innovation", the research program "Wissen schafft Zukunft" and by the Johannes-Gutenberg University of Mainz in the context of the "Schwerpunkt Vaskuläre Prävention" and the "MAIFOR grant", by grants from the Fondation de France, the French Ministry of Research, and the Institut National de la Santé et de la Recherche Médicale.

The **deCODE** CAD/MI Study was sponsored by NIH grant, National Heart, Lung and Blood Institute R01HL089650-02.

The German MI Family Studies (**GerMIFS I-III (KORA)**) were supported by the Deutsche Forschungsgemeinschaft and the German Federal Ministry of Education and Research (BMBF) in the context of the German National Genome Research Network (NGFN-2 and NGFN-plus), the EU funded integrated project Cardiogenics (LSHM-CT-2006-037593) and ENGAGE, and the bi-national BMBF/ANR funded project CARDomics (01KU0908A).

LURIC has received funding from the EU framework 6 funded Integrated Project "Bloodomics" (LSHM-CT-2004-503485), the EU framework 7 funded Integrated Project AtheroRemo (HEALTH-F2-2008-201668) and from Sanofi/Aventis, Roche, Dade Behring/Siemens, and AstraZeneca.
The **MIGen** study was funded by the US National Institutes of Health (NIH) and National Heart, Lung, and Blood Institute's STAMPEED genomics research program through R01 HL087676. Ron Do from the MIGen study is supported by a Canada Graduate Doctoral Scholarship from the Canadian Institutes of Health Research.

Recruitment of **PennCATH** was supported by the Cardiovascular Institute of the University of Pennsylvania. Recruitment of the **MedStar** sample was supported in part by the MedStar Research Institute and the Washington Hospital Center and a research grant from GlaxoSmithKline. Genotyping of PennCATH and Medstar was performed at the Center for Applied Genomics at the Children's Hospital of Philadelphia and supported by GlaxoSmithKline through an Alternate Drug Discovery Initiative research alliance award (M. P. R. and D. J. R.) with the University of Pennsylvania School of Medicine.

The **Ottawa Heart Genomic Study** was supported by CIHR #MOP--82810 (R. R.), CFI #11966 (R. R.), HSFO #NA6001 (R. McP.), CIHR #MOP172605 (R. McP.), CIHR #MOP77682 (A. F. R. S.).

The **WTCCC** Study was funded by the Wellcome Trust. Recruitment of cases for the WTCCC Study was carried out by the British Heart Foundation (BHF) Family Heart Study Research Group and supported by the BHF and the UK Medical Research Council. N. J. S. and S. G. B. hold chairs funded by the British Heart Foundation. N. J. S. and

A.H.G are also supported by the Leicester NIHR Biomedical Research Unit in Cardiovascular Disease and the work described in this paper is part of the research portfolio of the Leicester NIHR Biomedical Research Unit.

The **Age, Gene/Environment Susceptibility Reykjavik Study** has been funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament).

The **Cleveland Clinic GeneBank** study was supported by NHLBI grants P01HL098055, P01HL076491, R01HL103866, P20HL113452, and R01HL103931.

The collection of clinical and sociodemographic data in the **Dortmund Health Study** was supported by the German Migraine- & Headache Society (DMKG) and by unrestricted grants of equal share from Astra Zeneca, Berlin Chemie, Boots Healthcare, Glaxo-Smith-Kline, McNeil Pharma (former Woelm Pharma), MSD Sharp & Dohme and Pfizer to the University of Muenster. Blood collection was done through funds from the Institute of Epidemiology and Social Medicine, University of Muenster.

The **EPIC-Norfolk study** is supported by the Medical Research Council UK and Cancer Research UK.

The **EpiDREAM study** is supported by the Canadian Institutes of Health Research, Heart and Stroke Foundation of Ontario, Sanofi-Aventis, GlaxoSmithKline and King Pharmaceuticals.
Funding for Andrew Lotery from the **LEEDS** study was provided by the T.F.C. Frost charity and the Macular Disease Society.

The **Rotterdam Study** is supported by the Erasmus Medical Center and Erasmus University Rotterdam; the Netherlands Organization for Scientific Research; the Netherlands Organization for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly; The Netherlands Heart Foundation; the Ministry of Education, Culture and Science; the Ministry of Health Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. Support for genotyping was provided by the Netherlands Organization for Scientific Research (NWO) (175.010.2005.011, 911.03.012), the Netherlands Genomics Initiative (NGI)/ NWO project nr. 050-060-810 and Research Institute for Diseases in the Elderly (RIDE). Abbas Dehghan is supported by a grant from NWO (Vici, 918-76-619).

The **SAS** study was funded by the British Heart Foundation.

The Swedish Research Council, the Swedish Heart & Lung Foundation and the Stockholm County Council (ALF) supported the **SHEEP** study.

SMILE was funded by the Netherlands Heart foundation (NHS 92345). Dr Rosendaal is a recipient of the Spinoza Award of the Netherlands Organisation for Scientific Research (NWO) which was used for part of this work.

The **Verona Heart Study** was funded by grants from the Italian Ministry of University and Research, the Veneto Region, and the Cariverona Foundation, Verona.

The **Atherosclerosis Risk in Communities Study** is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022. The authors thank the staff and participants of the ARIC study for their important contributions.

The **KORA** (Kooperative Gesundheitsforschung in der Region Augsburg) research platform was initiated and financed by the Helmholtz Zentrum München - National Research Center for Environmental Health, which is funded by the German Federal Ministry of Education, Science, Research and Technology and by the State of Bavaria. Part of this work was financed by the German National Genome Research Network (NGFN-2 and NGFNPlus) and within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ.

Work described in this paper is part of the research portfolio supported by the Leicester NIHR Biomedical Research Unit in Cardiovascular Disease.

This work forms part of the research themes contributing to the translational research portfolio of Barts and the London Cardiovascular Biomedical Research Unit which is supported and funded by the National Institute of Health Research.

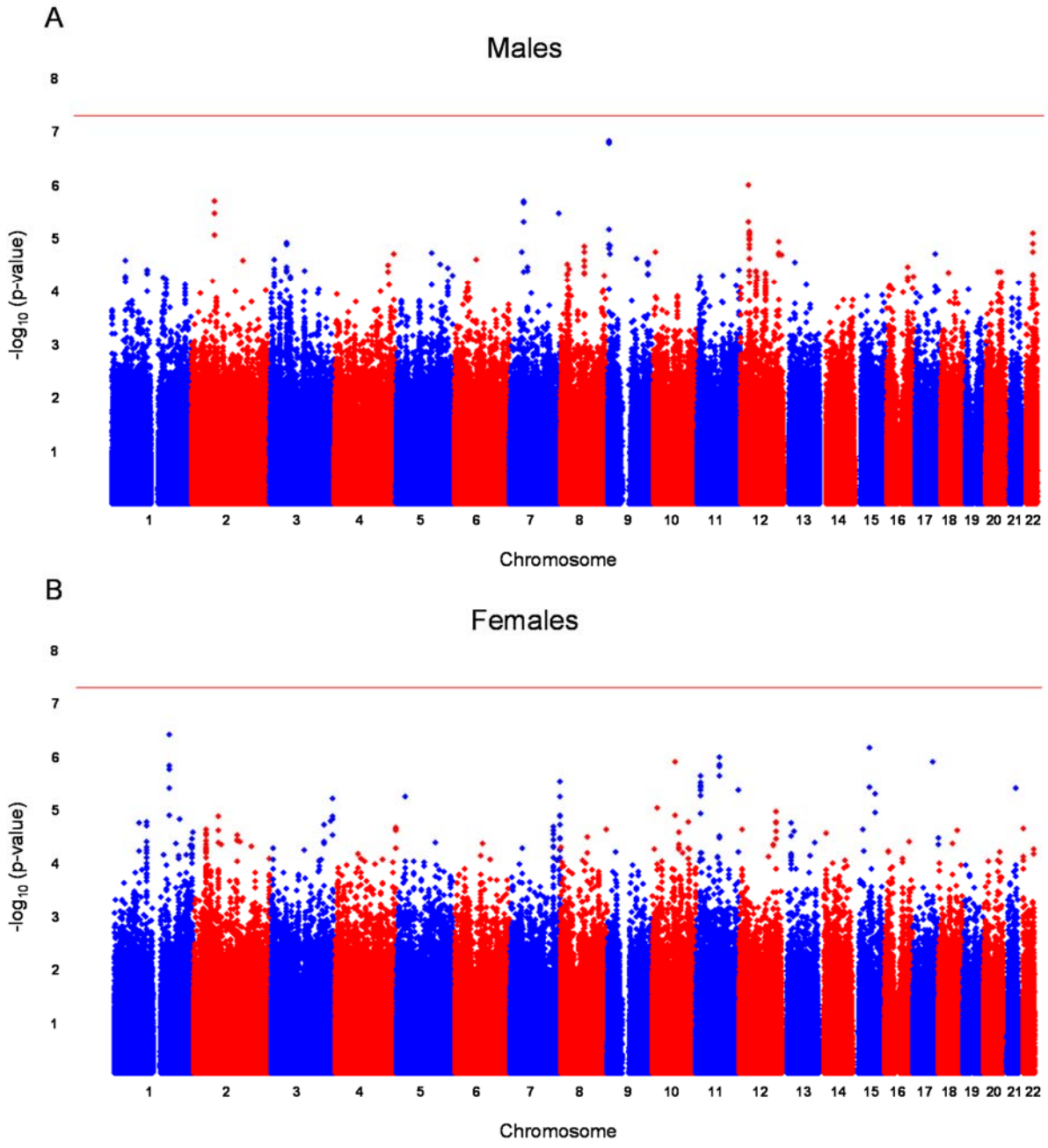
Supplemental Table I. Sex-stratified Association of SNPs Identified in Human and Mouse GWAS for Plasma TMAO Levels.

Males	rs17359359				rs12402441			
Stage	**GG**	**AG**	**AA**	**[a]p-value**	**AA**	**AG**	**GG**	**[a]p-value**
GWAS	5.0 ± 6.5 (n=1231)	8.9 ± 24.9 (n=154)	7.7 ± 5.8 (n=5)	0.0002	5.6 ± 10.9 (n=1238)	4.2 ± 4.7 (n=147)	4.4 ± 2.6 (n=5)	0.008
Replication	6.3 ± 11.8 (n=836)	9.9 ± 30.0 (n=108)	3.1 ± 2.2 (n=7)	0.73	7.0 ± 15.8 (n=950)	5.5 ± 7.7 (n=93)	4.9 ± 1.0 (n=7)	0.65
Combined	5.6 ± 9.1 (n=2094)	9.3 ± 27.0 (n=262)	5.0 ± 4.5 (n=12)	0.02	6.2 ± 13.2 (n=2188)	4.7 ± 6.1 (n=240)	4.7 ± 1.7 (n=12)	0.02

Females	rs17359359				rs12402441			
Stage	**GG**	**AG**	**AA**	**[a]p-value**	**AA**	**AG**	**GG**	**[a]p-value**
GWAS	6.1 ± 10.8 (n=496)	6.8 ± 5.6 (n=84)	13.6 ± 11.5 (n=3)	0.0003	6.4 ± 10.7 (n=531)	5.0 ± 3.7 (n=50)	1.5 ± 0.7 (n=2)	0.31
Replication	6.9 ± 17.7 (n=612)	6.3 ± 5.8 (n=71)	5.0 ± 1.1 (n=3)	0.47	6.1 ± 8.4 (n=648)	13.6 ± 47.6 (n=65)	3.6 ± 2.8 (n=3)	0.22
Combined	6.6 ± 15.0 (n=1108)	6.6 ± 5.6 (n=155)	9.3 ± 8.7 (n=6)	0.003	6.2 ± 9.5 (n=1179)	9.8 ± 36.0 (n=115)	2.7 ± 2.0 (n=5)	0.74

Mean (± SD) plasma TMAO levels (μM) are shown as a function of genotype.
[a]P-values obtained using linear regression with natural log transformed values after adjustment for age.

Supplemental Figure I



Supplemental Figure I. Sex-stratified GWAS results for plasma TMAO levels. Manhattan plots derived from GWAS analyses performed in males (A) and females (B) separately did not reveal sex-specific effects on chromosomes 1q23.3 and 2p12 or identify other loci that exceed the genome-wide threshold for significance (indicated by the horizontal red line).