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 ≥ 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) = σg2K and Var(e) = σe2I, where K represents an identity-by-descent kinship matrix across all genotypes. A restricted maximum likelihood estimate of σg2 and σe2 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 $\beta\neq0$.

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.