Single Nucleotide Polymorphisms in Multiple Novel Thrombospondin Genes May Be Associated With Familial Premature Myocardial Infarction

Eric J. Topol, MD; Jeanette McCarthy, PhD; Stacey Gabriel, PhD; David J. Moliterno, MD; William J. Rogers, MD; L. Kristin Newby, MD; Matt Freedman, PhD; Jennifer Metivier, MS; Ruth Cannata, RN, BSN; Christopher J. O’Donnell, MD, MPH; Kandice Kottke-Marchant, MD, PhD; Gurunathan Murugesan, PhD; Edward F. Plow, PhD; Olga Stenina, PhD; George Q. Daley, MD, PhD; for the GeneQuest Investigators and Collaborators*

Background—Recent advances in high-throughput genomics technology have expanded our ability to catalogue allelic variants in large sets of candidate genes related to premature coronary artery disease.

Methods and Results—A total of 398 families were identified in 15 participating medical centers; they fulfilled the criteria of myocardial infarction, revascularization, or a significant coronary artery lesion diagnosed before 45 years in men or 50 years in women. A total of 62 vascular biology genes and 72 single-nucleotide polymorphisms were assessed. Previously undescribed variants in 3 related members of the thrombospondin protein family were prominent among a small set of single-nucleotide polymorphisms that showed a statistical association with premature coronary artery disease. A missense variant of thrombospondin 4 (A387P) showed the strongest association, with an adjusted odds ratio for myocardial infarction of 1.89 (P<0.002 adjusted for covariates) for individuals carrying the P allele. A variant in the 3’ untranslated region of thrombospondin-2 (change of thymidine to guanine) seemed to have a protective effect against myocardial infarction in individuals homozygous for the variant (adjusted odds ratio of 0.31; P=0.0018). A missense variant in thrombospondin-1 (N700S) was associated with an adjusted odds ratio for coronary artery disease of 11.90 (P=0.041) in homozygous individuals, who also had the lowest level of thrombospondin-1 by plasma assay (P=0.0019).

Conclusions—This large-scale genetic study has identified the potential of multiple novel variants in the thrombospondin gene family to be associated with familial premature myocardial infarction. Notwithstanding multiple caveats, thrombospondins specifically and high-throughput genomic technology in general deserve further study in familial ischemic heart disease. (Circulation. 2001;104:2641-2644.)

Key Words: genetics ■ coronary disease ■ myocardial infarction ■ atherosclerosis ■ genes

We began a case-control study using high-throughput genomics technology to examine the role of common genetic variants in a large number of candidate genes for premature, familial coronary artery disease (CAD) and myocardial infarction. Candidate genes were chosen for their acknowledged role in endothelial cell biology, vascular biology, lipid metabolism, and the coagulation cascade. In the present article, we describe the results of our analysis of 72 single-nucleotide polymorphisms (SNPs) drawn from 62 candidate genes in 352 CAD cases and 418 controls, an effort that entailed generating >50 000 individual genotypes; this is, to date, the largest such genetic association study.

Methods

Case and Control Population
Fifteen medical centers in the United States (see Appendix) participated in the enrollment of probands and their affected siblings. Each proband was required to have developed CAD by 45 years if male or 50 years if female, as manifest by either a myocardial infarction, surgical or percutaneous coronary revascularization, or a coronary...
angiogram with evidence of at least a 70% stenosis in a major epicardial artery. At least one sibling who also had fulfilled these criteria had to be alive to qualify for inclusion, and the proband along with affected sibling(s) answered a health questionnaire, had anthropometric measures taken, and had blood drawn for measurement of serum makers and extraction of DNA. The protocol was approved by the institutional review board at each participating institution. All patients gave informed consent to participate. For the purpose of our case-control study, a series of unrelated singleton cases was selected such that only one affected individual from each family was represented, giving preference to the sibling with the earlier age of onset. The case series was limited to white families. Controls representing a general, unselected population of white Americans were identified through random-digit phone dialing in the Atlanta, Georgia, area.

Variant Allele Discovery, Validation, and Genotyping
Cell lines derived from an ethnically diverse population were obtained and used for SNP discovery by methods previously described in detail. Genomic sequences representing the coding and partial regulatory regions of genes were amplified by polymerase chain reaction and screened using 2 independent methods: denaturing high-performance liquid chromatography or variant detector arrays (Affymetrix). An average of 114 chromosomes were screened in high-performance liquid chromatography or variant detector partial regulatory regions of genes were amplified by polymerase sequence; 17 SNPs were not identified at least once when assessed in a subset of 96 control individuals, and therefore, they were judged to be too rare to justify genotyping in the complete set of cases and controls. The final number of SNPs analyzed in the case-control study was 72.

Statistical Analysis
All analyses were done using the SAS statistical package (Version 8.0, SAS Institute Inc.). Differences between cases and controls were assessed with a chi2 statistic for categorical covariates and the Wilcoxon statistic for continuous covariates. Significance was determined using a continuity-adjusted chi2 or Fisher’s exact test for each genotype compared with the homozygous wild-type for that locus. Odds ratios were calculated and presented with 95% confidence intervals. Multivariate logistic regression was used to adjust for sex, presence of hypertension, diabetes, and body mass index using the LOGISTIC procedure in SAS.

After identifying thrombospondin-1 as a gene implicated in premature CAD, the plasma samples from 240 cases that were previously collected in citrate anticoagulant and stored at −70°C were evaluated with an ELISA assay developed from a published procedure. All samples were analyzed in duplicate. Repeat measures of plasma thrombospondin were strongly correlated over all ranges of thrombospondin (r2=0.96).

Results
The demographic characteristics of the 352 cases and 418 controls are presented in the Table. Cases were more likely than controls to be male, older, diabetic, hypertensive, and have a higher body mass index. The most common event that led to the inclusion of a case into the study was myocardial infarction (54%). Cases were enrolled in the study, on average, 9 years after their qualifying event, suggesting a survivor bias.

Genotype distributions for cases and controls are shown in the Online Table (can be found at http://www.circulationaha.org) for all loci examined. Eleven SNPs in 9 genes showed statistically significant differences (P<0.05) between cases and controls for CAD, myocardial infarction, or both. A variant in only one of these genes, MTHFR (C677T), has been the subject of conflicting reports of association with CAD. Besides MTHFR, the other variants represent novel associations. Our study did not confirm statistical association for several other variants that have been previously linked with a risk of thrombotic cardiovascular disease or CAD, including the PLA2 allele (L33P) of the platelet glycoprotein IIb/IIIa receptor, factor V Leiden (R506Q), prothrombin G20210A, hemochromatosis (C282Y), or P-selectin (T715P) variants.

The THBS4 variant A387P in the third repeat type II unit may affect the secondary structure of the protein and disrupt the Ca2+ binding site. The THBS1 N700S variant occurs in the first type III unit. Patients who were homozygous for the THBS1 variant (SS) had the highest odds ratio for myocardial infarction (8.16) and a significantly lower plasma level of THBS1 than other genotypes (median levels, 88 ng/mL for SS, 235 ng/mL for NS, and 189 ng/mL for NN genotypes; P=0.0019). Correcting for 100 independent hypotheses (~50 genes tested for two outcomes) resulted in none of the associations reaching P<0.05. Only THBS4 reached a level of significance of P<0.10 for the association with myocardial infarction.

Test for Population Stratification
Underlying population stratification, caused by unequal proportions of ethnicities in the case and control populations, or ethnic admixture, can lead to spurious associations in case-control studies. Because we used a general population control
group with self-reported ethnicity, we tested for the presence of population stratification in our sample using the method described by Reich and Goldstein.\textsuperscript{12} We selected 96 evenly spaced, unlinked SNPs from the SNP consortium database and successfully genotyped 72 of the markers in 100 randomly selected cases and 100 of the controls. Given the extremely low probability that any of these markers are linked to a casual variant, we expect that in the absence of population stratification, the $\chi^2$ values would be distributed as a $\chi^2$ distribution with 1 degree of freedom. The value of this distribution is 1. The $\chi^2$ distribution generated by our markers has a mean value of 1.2, which was not significantly different from the null hypothesis ($P>0.05$). Thus, the results of our analysis show that the general population controls used in this study are well matched to the cases. According to the population genetic simulations performed by Pritchard and Rosenberg\textsuperscript{13} for similar study designs, the probability that spurious associations will be obtained at the candidate locus (at the 0.05 level) is $\approx 5\%$.

**Discussion**

The work described here exploits high-throughput genomic technologies to perform a large-scale case-control genetic association study in patients with familial premature CAD. In total, some 72 SNPs were analyzed in $\approx 770$ individuals, thus representing $>50,000$ genotypes. This study represents an attempt to use contemporary genomic technology to associate a large set of SNPs in a group of candidate genes implicated in arterial thrombosis and vascular biology with familial, premature coronary disease. Three novel SNPs from 3 distinct thrombospondin genes emerged as among the most highly associated variants. Each variant allele formed the basis of “at risk” genotypes that were each significantly associated with familial, premature myocardial infarction. The thrombospondin family of 5 extracellular matrix glycoproteins are known to play a pivotal role in cell adhesion: modulating vascular injury, coagulation, and angiogenesis and serving as a key ligand for CD36, an oxidized LDL receptor, and for integrins, including $\alpha_\text{v}\beta_\text{3}$\textsuperscript{14–16}. Thrombospondins have been demonstrated in atherosclerotic plaque,\textsuperscript{17,18} and thrombospondin deficiency has been associated with increased levels of matrix metalloproteinase-2, a protein linked to the vulnerability of atherosclerotic plaque.\textsuperscript{19}

The common SNP variant in the type 3 repeat of thrombospondin-4 (A387P) is predicted to affect folding and secretion of the protein and disruption of the calcium binding site.\textsuperscript{20,21} Indeed, although not previously implicated, the thrombospondin protein family members play critical roles in vascular integrity and thrombosis and may be particularly susceptible to playing a role, if altered, in premature atherosclerosis and myocardial infarction.

It is clear that considerable additional work is needed to extend our findings. Replication of our novel statistical associations in independent populations of patients with familial, premature myocardial infarction, as well as nonfamilial, older onset disease will be critical to determining which of our observations may be generalizable. The identification of additional SNPs in and around these genes will be necessary to determine which individual variants or haplotype are the true underlying cause of the observed associations. Defining the molecular mechanism linking the thrombospondin SNPs to adverse clinical outcomes will be key to understanding the pathophysiology of this pathway. Finally, multivariate models that incorporate the validated SNP associations and assess their interactive role with each other and additional covariates will provide the basis of comprehensive risk assessment models for ultimate use in clinical practice.

Given the large number of tests performed in our study, we anticipate that some of the apparent statistical associations we report will, in fact, have occurred by chance and that such associations will be indistinguishable from those that reflect some true underlying biological predisposition. Our initial attempts to replicate the association of the THBS-4 variant in 2 smaller series of patients with early-onset premature CAD have failed to confirm the generality of our observation. Such failures to replicate genetic associations in independent populations are a vexing aspect of genetic association studies\textsuperscript{22} and reflect the difficulty of defining uniform clinical end points and the effects of confounding environmental influence and population-specific genetic modifiers. Inclusion in our patient cohort required that 2 members of a family were both affected by premature CAD. However, neither of the independent populations we tested with early-onset CAD was defined using such a stringent inclusion criterion, and the subset of myocardial infarction cases was not rigorously evaluated; these differences may account for our failure to replicate. The coincidence of finding association with 3 distinct SNPs in thrombospondin family members and the functional correlation of low plasma levels with the highest risk genotype for thrombospondin-1 strengthens the hypothesis of a potential biological link between thrombospondin variation and early-onset CAD and warrants further study to validate the association and elucidate the biological mechanism.

Coronary atherosclerosis is still the most important cause of death, and one would expect that a continuum in genetic liability exists between premature and typical CAD and myocardial infarction. The thrombospondin variants identified are interesting, but the interpretation of their actual importance relies on considerable further study, requiring independent replication and proof of a cause-and-effect relationship for the variants directly influencing the disease. Further work with genome-wide scanning of our sibships may prove helpful, as has been the case of identifying potential genes in a Finnish population of premature coronary disease\textsuperscript{23} or in sibships associated with longevity.\textsuperscript{24} Our study highlights some of the limitations in high-throughput candidate gene investigation and emphasizes that such work should be considered as exploratory and hypothesis-generating.

**Appendix**

**GeneQuest Investigators and Collaborators**

Cleveland Clinic Foundation, Cleveland, Ohio: Eric J. Topol (Study Chairperson), David J. Moliterno, Gurunathan Murugesan, Olga Stenia, Kandace Kotke-Marchant, Edward F. Plow, Ruth Cannata, Patricia Welsh, and Monique Rosenthal; Emory University Hospital, Atlanta, Ga: Spencer B. King III, William Anderson, Joe Jean Borowski, and Kris Anderberg; Mayo Clinic, Rochester, Minn: David R. Holmes, Jr, Charanjit Rihal, and Sharon McIntire-
References


Single Nucleotide Polymorphisms in Multiple Novel Thrombospondin Genes May Be Associated With Familial Premature Myocardial Infarction


for the GeneQuest Investigators and Collaborators

_Circulation_. 2001;104:2641-2644; originally published online October 29, 2001;
doi: 10.1161/hc4701.100910

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2001 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/104/22/2641

Data Supplement (unedited) at:
http://circ.ahajournals.org/content/suppl/2001/11/05/hc4701.100910.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/