Common Genetic Variation in the Prothrombin Gene, Hormone Therapy, and Incident Nonfatal Myocardial Infarction in Postmenopausal Women

Lucia A. Hindorff1,2, Bruce M. Psaty1,2,3,4, Christopher S. Carlson5, Susan R. Heckbert1,2, Thomas Lumley2,6, Nicholas L. Smith1,2, Rozenn N. Lemaitre2, Mark J. Rieder5, Deborah A. Nickerson5, and Alexander P. Reiner1,2

1 Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA.
2 Cardiovascular Health Research Unit, School of Medicine, University of Washington, Seattle, WA.
3 Department of Medicine, School of Medicine, University of Washington, Seattle, WA.
4 Department of Health Services, School of Public Health, University of Washington, Seattle, WA.
5 Department of Genome Sciences, School of Medicine, University of Washington, Seattle, WA.
6 Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA.

Received for publication September 1, 2005; accepted for publication October 19, 2005.

Genetic variants in coagulation factors are associated with myocardial infarction and may modify the association between hormone therapy and cardiovascular disease risk. This study assessed whether common variation in the prothrombin gene was associated with incident nonfatal myocardial infarction in postmenopausal women and whether this association differed according to current estrogen use. Eight variants representing 98% of common prothrombin variants were selected using publicly available genomic variation data. These variants and the functional G20210A variant were genotyped and used to infer haplotypes in a population-based Washington State case-control study of postmenopausal Caucasian women (1995–1999; 273 cases and 788 controls). Women carrying a nonsynonymous polymorphism in exon 6 (C5467T) had an increased risk of myocardial infarction (for each additional copy, relative to women with one fewer copy, odds ratio = 1.4, 95% confidence interval: 1.0, 1.8). Prothrombin haplotypes were also associated with myocardial infarction (with minimal adjustment, global \( p = 0.056 \); with full adjustment, \( p = 0.034 \)). Associations between haplotypes and myocardial infarction were similar among users of hormone therapy and nonusers (global \( p = 0.61 \)), though statistical power was limited. These preliminary results suggest that common genetic variants in the prothrombin gene or other variants in linkage disequilibrium are associated with myocardial infarction in postmenopausal women.

case-control studies; estrogen replacement therapy; female; genetics; myocardial infarction; polymorphism, genetic; postmenopause; prothrombin

Abbreviations: CI, confidence interval; GHC, Group Health Cooperative; OR, odds ratio; SNP, single nucleotide polymorphism.

The precursor prothrombin (factor II), when activated to thrombin, plays the penultimate role in forming a fibrin clot by cleaving fibrinogen to fibrin, after which fibrin is cross-linked by factor XIII. In addition to its well-understood role in thrombosis, knowledge about prothrombin has expanded to include roles in platelet aggregation, inflammation, and vascular development (1, 2). Prothrombin levels are highly heritable, but previously characterized prothrombin gene sequence variants explain only a small fraction of the heritability (3, 4). The most well-characterized genetic variant, the 20210 G/A transition in the 3\(^\prime\)-untranslated region of the gene, has been associated with arterial thrombosis in some
studies, but not all (5). The G/A nucleotide substitution appears to increase prothrombin levels by altering mRNA processing efficiency (6, 7). Other single variants associated with increased prothrombin activity have been identified but have not yet been characterized with regard to risk of arterial thrombotic disease (8).

The hypothesis that prothrombotic genetic variants may modify the association between hormone therapy and myocardial infarction has received considerable attention following the results of clinical trials (9, 10). Estrogens are known to have a favorable effect on lipid metabolism but increase expression of coagulation and inflammation genes, including prothrombin (11). We previously reported that among postmenopausal women with hypertension, the risk of myocardial infarction associated with hormone therapy was increased in women with prothrombin G20210A but not increased in women without the prothrombin variant (12). In the same population, the risk of myocardial infarction associated with hormone therapy was significantly elevated in women who carried two or more variant alleles of coagulation factor XIII as compared with women with one or no variant factor XIII alleles (13).

The recent availability of detailed genomic sequence variation data in the prothrombin gene enabled us to extend our previous work to other common prothrombin variants. Using a linkage disequilibrium-based single nucleotide polymorphism (SNP) selection algorithm, LDSelect (available at http://pga.gs.washington.edu/), and SNP discovery data from the SeattleSNPs Program for Genomic Applications (University of Washington, Seattle, Washington), we selected a minimal set of SNPs representing 98 percent of the SNPs with minor allele frequency greater than or equal to 5 percent. By adopting this approach, our analyses were not limited to SNPs previously characterized in the literature. Additionally, data from multiple SNPs can be used to infer gene-based haplotypes, which in some cases may be more predictive of disease than individual SNPs (14). The aims of this study were to determine whether common genetic variation in the prothrombin gene, at the level of either SNPs or haplotypes, was associated with incident nonfatal myocardial infarction in postmenopausal women and whether these associations differed according to current use of hormone therapy.

**MATERIALS AND METHODS**

**Study setting and eligibility**

Participants in this study were part of an ongoing case-control study of myocardial infarction at the Group Health Cooperative (GHC), a large health maintenance organization based in Seattle, Washington (12). Cases were postmenopausal female GHC members who had an incident nonfatal myocardial infarction during 1995–1999. Myocardial infarction cases were identified through computerized discharge records for GHC hospitals and GHC claims databases for services provided by non-GHC facilities. Controls were a random sample of postmenopausal female GHC members without a history of myocardial infarction, frequency-matched to cases on decade of age, year of identification, and presence or absence of pharmacologically treated hypertension. The few female GHC study participants who were African-American (3 percent) were excluded from the current study. All women included in the study identified themselves as Caucasian, were less than 80 years old, were postmenopausal, and had at least four GHC physician visits. A total of 279 cases and 819 controls met these criteria. We excluded women who used progestin, estrogen creams, or estrogen patches without estrogen pills ($n = 33$) or who were missing data on all genotype assays ($n = 4$). Following these exclusions, data on 273 cases and 788 controls were available. The GHC institutional review board reviewed and approved the study, and subjects gave written informed consent.

**Data collection and definitions**

Data on each participant’s characteristics prior to the index date were collected from the GHC outpatient record. For cases, the index date was the date of the myocardial infarction; for controls, the index date was a computer-generated random date within the calendar year for which they had been selected. A woman was classified as postmenopausal if her medical record noted a cessation of menses, if she had symptoms of menopause (among women who had undergone a hysterectomy), or, in the absence of information on symptoms and menses, if she was aged 55 years or older at the index date. History of cardiovascular disease was defined as a record of angina, stroke, claudication, or vascular procedures, including coronary artery bypass grafting, angioplasty, carotid endarterectomy, or peripheral vascular bypass. Missing values for demographic and clinical characteristics were uncommon and were imputed for no more than 4 percent of subjects for any variable by using the method of best-subset regression (implemented in Intercooled Stata 8.0; Stata Corporation, College Station, Texas).

Data on medication use were obtained from the GHC computerized pharmacy database, which includes a record of all prescriptions dispensed to GHC enrollees since 1977. These data contain information on the drug name, the date the prescription was filled, the quantity dispensed, the dose, and dosing instructions. A woman was classified as a current user of hormone therapy if, for the most recent hormone therapy prescription, enough medication had been dispensed prior to the index date to last until the index date, assuming 80 percent compliance (12). Duration of hormone therapy use was calculated for the hormone currently being used at the index date and was defined as the total number of days of intended use for consecutive prescriptions, assuming 80 percent compliance and permitting gaps of no more than 90 days between the run-out date and the subsequent refill (15). During the time period from which subjects were recruited, esterified estrogens were the standard oral estrogens prescribed at GHC (15).

A blood specimen was collected from each participant at an in-person visit following the index date. Blood was drawn from each participant’s antecubital vein and kept at 4°C until processed. DNA was extracted from white blood cells using standard salting-out procedures (16) and was stored at –70°C.
Selection of tagSNPs, genotyping, and haplotype inference

The SeattleSNPs Program for Genomic Applications (http://pga.gs.washington.edu/) resequenced coding, non-coding, and flanking regions of the prothrombin gene (22.1 kilobases in total) as part of its variation discovery efforts in genes related to the inflammatory response. In the 23 persons of European descent, 57 SNPs were identified; 46 of these were common, defined here as a minor allele frequency greater than or equal to 5 percent. The LDSelect algorithm was used to select the maximally informative set of common SNPs, or tagSNPs. LDSelect classified the 46 common SNPs into eight bins such that, within each bin, at least one tagSNP would be in linkage disequilibrium with all other SNPs at a linkage disequilibrium threshold of $r^2 = 0.64$ (17). These eight SNPs were either direct or indirect markers of 98 percent (45 out of 46) of the common SNPs.

TagSNPs were genotyped using TaqMan Assays-by-Design under standard conditions (TaqMan SNP Genotyping Assays protocol, revision B, part #4332856b; Applied Biosystems, Foster City, California). Six percent of samples were duplicates; among these, discrepancies in genotyping were noted in less than 0.5 percent. Genotypes were missing for an average of 4.6 percent (per-SNP range, 2.4–9.6 percent) of subjects. The 20210 G/A substitution was genotyped using a polymerase chain reaction/restriction fragment length polymorphism approach (creating a HindIII restriction site) (18). Haplotypes were inferred probabilistically from multilocus SNP data on all nine SNPs using the Stephens-Smith-Donnelly algorithm, implemented in PHASE 2.0 (19).

Statistical analysis

Statistical analyses were conducted using Intercooled Stata 8.0. All $p$ values were two-sided. Comparisons between cases and controls were made using linear regression for continuous variables or $\chi^2$ tests for discrete variables. Allele frequencies were estimated using gene counting. All SNPs were in Hardy-Weinberg equilibrium. For G10998A, only heterozygotes were observed. The single homozygote for the 20210A allele was grouped with heterozygotes. Odds ratios and 95 percent confidence intervals, approximating relative risks of myocardial infarction, were estimated using logistic regression. Heterozygotes and homozygotes for each minor allele were initially modeled as indicator variables. If a likelihood ratio test comparing this model to a model where genotype was modeled as 0, 1, or 2 copies of the minor allele suggested that two separate terms were not needed, genotypes were modeled assuming a log-additive model. This model assumes that the odds ratio is constant for each additional copy of the minor allele, relative to women with one fewer copy. For instance, the odds ratio for C5467T estimates the relative risk of myocardial infarction for the comparison of CT with CC or the comparison of TT with CT. The comparison of TT with CC is the square of the odds ratio. All regressions were adjusted for the minimal set of case-control matching variables (age, index year, and presence or absence of treated hypertension). For some analyses, estimates were fully adjusted for case-control matching variables plus variables jointly associated with hormone therapy use and myocardial infarction risk—glucose level, diabetes, cholesterol level, systolic blood pressure, history of cardiovascular disease, and smoking.

Haplotype associations were modeled using weighted logistic regression and a robust variance estimator to account for uncertainty in haplotype estimation. The weights in this model were the posterior probabilities, estimated by PHASE, for each combination of haplotypes. Two types of haplotype models were estimated. The first model estimated odds ratios and 95 percent confidence intervals for each haplotype, relative to all other haplotypes combined. In this model, the odds ratio for each haplotype estimates the relative risk of myocardial infarction associated with each additional copy of the haplotype, relative to women with one fewer copy of the haplotype. The second haplotype model included all haplotype terms except for the reference haplotype; this model permitted a global test of the association between all haplotypes and myocardial infarction. In this multiple-haplotype model, the odds ratio for each haplotype estimates the risk of myocardial infarction associated with each additional copy of the haplotype, relative to the reference haplotype. Odds ratios and 95 percent confidence intervals for combinations of haplotypes were linear combinations of the estimates from the multiple haplotype model. Frequency counts for haplotype combinations accounted for uncertainty in haplotype estimation and were obtained by summing the posterior probabilities obtained by PHASE.

Interactions were estimated by introducing a multiplicative term into multivariate models, and significance was assessed using a likelihood ratio test of nested models (for SNP models and unweighted regressions) or a Wald test statistic (for weighted regressions). When multiple haplotype-hormone therapy interactions were modeled, a global test of all interaction terms was conducted using a multivariate Wald test statistic.

To determine how robust the primary analyses were to assumptions about haplotype uncertainty and hormone use, several sensitivity analyses were conducted. First, haplotype analyses were restricted to the most likely haplotype for each individual. Next, the definition of current hormone therapy was relaxed to include current users of estrogen creams or patches. In addition, women who used estrogen only were modeled separately from those who used estrogen plus progestin. Finally, current users of hormone therapy were also subdivided into women who had used hormone therapy for up to 1 year and women who had used hormone therapy for 1 year or more.

RESULTS

In this population of 1,061 postmenopausal women, cardiovascular disease risk factors were distributed in the expected ways among myocardial infarction cases and controls (table 1). Cases were significantly more likely to be smokers, to be diabetic or hyperlipidemic, and to have a history of cardiovascular disease, higher systolic blood pressure,

<table>
<thead>
<tr>
<th></th>
<th>Cases (n = 273)</th>
<th>Controls (n = 788)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (years)†</td>
<td>67.5 (8.8)‡</td>
<td>68.0 (8.7)</td>
</tr>
<tr>
<td>Mean body mass index§</td>
<td>29.3 (6.9)</td>
<td>28.4 (6.6)</td>
</tr>
<tr>
<td>Mean systolic blood pressure (mmHg)</td>
<td>143 (20)*</td>
<td>138 (20)</td>
</tr>
<tr>
<td>Mean diastolic blood pressure (mmHg)</td>
<td>80 (10)</td>
<td>79 (10)</td>
</tr>
<tr>
<td>Mean cholesterol level (mg/dl)</td>
<td>244 (45)*</td>
<td>232 (41)</td>
</tr>
<tr>
<td>Mean glucose level (mg/dl)</td>
<td>126 (63)*</td>
<td>107 (35)</td>
</tr>
<tr>
<td>Treated hypertension (%)†</td>
<td>46.9</td>
<td>51.4</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>21.3*</td>
<td>7.5</td>
</tr>
<tr>
<td>Hyperlipidemia (%)</td>
<td>11.7*</td>
<td>7.4</td>
</tr>
<tr>
<td>Smoking (%)</td>
<td>26.4*</td>
<td>8.6</td>
</tr>
<tr>
<td>History of cardiovascular disease (%)¶</td>
<td>25.3*</td>
<td>11.7</td>
</tr>
<tr>
<td>No. of physician visits in past year (%)</td>
<td>6.4*</td>
<td>5.5</td>
</tr>
<tr>
<td>Current estrogen use (%)</td>
<td>37.7</td>
<td>40.2</td>
</tr>
</tbody>
</table>

* p < 0.05 (cases vs. controls).
† Matching factor.
‡ Numbers in parentheses, standard deviation.
§ Weight (kg)/height (m)².
¶ Includes history of any of the following: angina, stroke, claudication, coronary artery bypass grafting, angioplasty, carotid endarterectomy, and peripheral vascular bypass.

TABLE 2. Characteristics of prothrombin (factor II) single nucleotide polymorphisms (SNPs) and their association with myocardial infarction in postmenopausal women, Seattle, Washington, 1995–1999

<table>
<thead>
<tr>
<th>SNP*</th>
<th>SNP Database identifier</th>
<th>Location</th>
<th>Frequency (%)</th>
<th>Odds ratio‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cases (2N = 546)</td>
<td>Controls (2N = 1,576)</td>
</tr>
<tr>
<td>T2713C</td>
<td>rs¶ 3136434</td>
<td>Intron 3</td>
<td>6.9</td>
<td>8.5</td>
</tr>
<tr>
<td>G2890A</td>
<td>rs 3136435</td>
<td>Intron 4</td>
<td>6.2</td>
<td>6.6</td>
</tr>
<tr>
<td>T4992C</td>
<td>rs 2070851</td>
<td>Intron 4</td>
<td>23.2</td>
<td>21.8</td>
</tr>
<tr>
<td>C5389G</td>
<td>rs 2070852</td>
<td>Intron 5</td>
<td>30.7</td>
<td>29.4</td>
</tr>
<tr>
<td>C5467T</td>
<td>rs 5896</td>
<td>Exon 6</td>
<td>15.8</td>
<td>12.3</td>
</tr>
<tr>
<td>G10998A**</td>
<td>rs 3136524</td>
<td>Intron 11</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>T20603C</td>
<td>rs 3136512</td>
<td>Intron 12</td>
<td>42.9</td>
<td>47.5</td>
</tr>
<tr>
<td>A21239G</td>
<td>rs 3136516</td>
<td>Intron 13</td>
<td>42.8</td>
<td>45.7</td>
</tr>
<tr>
<td>G20210A**</td>
<td>rs 1799963</td>
<td>3′-untranslated region</td>
<td>2.8</td>
<td>1.6</td>
</tr>
</tbody>
</table>

* SNPs are numbered according to the Program for Genomic Applications reference sequence, except for SNP 20210A, which is numbered according to Poort et al. (18). SNP 20210 corresponds to nucleotide 21538 in the Program for Genomic Applications reference sequence.
‡ Odds ratios are for each additional copy of the allele, relative to women with one fewer copy.
§ Adjusted for age, index year, and treated hypertension.
¶ OR, odds ratio; CI, confidence interval; rs, reference SNP.
# Adjusted for the minimal adjustment variables plus glucose level, diabetes, cholesterol level, systolic blood pressure, history of cardiovascular disease, and smoking.
** For SNP G10998A, only heterozygotes were observed. Homozygotes for the 20210A allele (n = 1) were grouped with the heterozygotes.
haplotype was inferred with greater than 90 percent probability in 93 percent of women, and for 80 percent of women, only one likely pair of haplotypes was inferred. Haplotype inference from the nine SNPs resolved five haplotypes whose frequency exceeded 5 percent in the control population (table 3). The most common haplotype, with a control frequency of 47.9 percent, comprised major alleles for all of these SNPs. Several of the SNPs (2713, 2890, 5467, and 20210) tagged only one common haplotype; others (4992, 5389, 20603, and 21239) were found on more than one common haplotype. The 20210A allele was observed primarily on the background of the haplotype with all common alleles; this haplotype was observed with a frequency of 1.5 percent in controls. Altogether, these six haplotypes comprised 92.3 percent of the inferred haplotypes observed in the control population; other haplotypes, which were grouped together for analysis, accounted for the remaining 7.7 percent.

Two types of haplotype analyses were implemented (table 3). First, each haplotype was compared with all other haplotypes combined. The haplotype comprising minor alleles 20603C and 21239G (haplotype B) was significantly associated with a decreased risk of myocardial infarction (for each additional copy, relative to women with one fewer copy, minimally adjusted OR = 0.68, 95 percent CI: 0.51, 0.91; p = 0.010). Haplotype F, carrying the 20210A allele, was associated with a 1.9-fold increased risk of myocardial infarction for each additional copy; this association approached statistical significance (95 percent CI: 0.94, 3.7; p = 0.075). The remaining haplotypes were not significantly associated with myocardial infarction. The second haplotype analysis estimated the relative risk associated with an additional copy of each haplotype, relative to haplotype A, the most common haplotype and the haplotype comprising all major alleles. The global p value for the test of all haplotypes in relation to myocardial infarction approached statistical significance (95 percent CI: 0.94, 3.7; p = 0.056), suggesting a general association between prothrombin haplotypes and myocardial infarction. The strongest associations were for haplotype B (for each additional copy, relative to haplotype A, OR = 0.67, 95 percent CI: 0.49, 0.91) and haplotype F (OR = 1.7, 95 percent CI: 0.86, 3.5). When estimates were fully adjusted for matching factors plus several additional traditional cardiovascular disease risk factors, the overall association between prothrombin haplotypes and myocardial infarction was stronger (global p = 0.034). Relative to haplotype A, the estimates for each additional copy of haplotype B (OR = 0.63, 95 percent CI: 0.44, 0.88) and haplotype F (OR = 2.1, 95 percent CI: 0.95, 4.6) were slightly more pronounced. When the analysis was restricted to the most likely haplotype for each woman, estimates were similar, and the overall p value for the multiple-haplotype model was 0.058.

Odds ratios and 95 percent confidence intervals for each haplotype combination, relative to the most common combination, haplotype A/haplotype A, are displayed in supplemental table 1 on the Journal’s website (www.aje.oxfordjournals.org), along with the observed counts of each combination by case-control status. Haplotype combinations in which both of the haplotypes had frequencies of

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Single nucleotide polymorphism</th>
<th>Frequency (%)</th>
<th>Odds ratio*</th>
<th>95% CI</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>A T G T C C G T A G</td>
<td>51.2</td>
<td>1.0</td>
<td>Referent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B T G T C C G C G G</td>
<td>11.1</td>
<td>0.68</td>
<td>0.51, 0.91</td>
<td>0.67</td>
<td>0.49, 0.91</td>
</tr>
<tr>
<td>C T G C G T G C G G</td>
<td>14.8</td>
<td>1.2</td>
<td>0.89, 1.6</td>
<td>1.1</td>
<td>0.83, 1.5</td>
</tr>
<tr>
<td>D C G C G C G C G G</td>
<td>7.1</td>
<td>0.86</td>
<td>0.59, 1.3</td>
<td>0.80</td>
<td>0.54, 1.2</td>
</tr>
<tr>
<td>E T A T G C G C G G</td>
<td>6.6</td>
<td>0.98</td>
<td>0.67, 1.4</td>
<td>0.94</td>
<td>0.64, 1.2</td>
</tr>
<tr>
<td>F T G T C C G T A A</td>
<td>2.8</td>
<td>1.9</td>
<td>0.94, 3.7</td>
<td>1.7</td>
<td>0.86, 3.5</td>
</tr>
<tr>
<td>Other</td>
<td>6.4</td>
<td>7.7</td>
<td>0.94</td>
<td>0.58, 1.2</td>
<td></td>
</tr>
</tbody>
</table>

* Estimates were adjusted for age, index year, and treated hypertension.

** Estimates were adjusted for age, index year, and treated hypertension.

† OR, odds ratio; CI, confidence interval.
less than 10 percent generally had small (<5) numbers of cases and controls. The haplotype associations estimated in this study were based on a relatively small number of observed haplotype combinations. For example, the five most common haplotype combinations (AA, AB, AC, AD, and AO) accounted for 65.6 percent of the observed combinations. Several potentially informative haplotype combinations (for example, haplotype B in combination with haplotype F) were not observed with sufficient frequency to permit precise estimates.

The association between prothrombin SNPs or haplotypes and myocardial infarction was also examined separately in women who were current users of hormone therapy and those who were not using hormone therapy on the index date. The main effect of estrogen therapy in this population of postmenopausal women was a slightly increased risk of myocardial infarction associated with current estrogen use (fully adjusted OR = 1.2, 95 percent CI: 0.90, 1.7). Upon stratification by current use of estrogen, the associations of individual SNPs with myocardial infarction, or of each haplotype (relative to all others) with myocardial infarction, were generally similar between current users and non-users of estrogen therapy (data not shown; p > 0.1 for each SNP- or haplotype-estrogen interaction). The results of the multiple-haplotype model, stratified by use of hormone therapy, are shown in table 4. The p value for the global test of all haplotype-estrogen interaction terms was 0.61, suggesting little statistical evidence of an interaction between prothrombin haplotypes and estrogen use. When the definition of current hormone therapy was relaxed to include current users of estrogen creams or patches, odds ratios were similar and the overall p value for all interaction terms was 0.60. However, small numbers, particularly among cases who had less common haplotypes and used estrogen, limited statistical power. Small numbers further limited power for analyses that stratified current use by type of hormone therapy (estrogen vs. estrogen plus progestin) or duration of use (<1 year vs. ≥1 year) (all p’s > 0.5). Compared with previous results (12), this analysis included 41 additional myocardial infarction cases and 65 additional controls recruited over 1 extra year of follow-up. Haplotype analysis confirmed the previous interaction between haplotype F (tagged by the G20210A variant) and hormone therapy in hypertensive women. The odds ratio associated with one or two copies of haplotype F, relative to zero copies, was 5.9 (95 percent CI: 1.1, 32) in women who were current users of hormone therapy and 2.9 (95 percent CI: 0.61, 14) in women who were not current users.

**DISCUSSION**

In this population-based case-control study of postmenopausal women enrolled in a large health maintenance organization, common genetic variation in the prothrombin (factor II) gene at the levels of both single SNPs and haplotypes was associated with incident nonfatal myocardial infarction. A common nonsynonymous SNP in exon 6 with 12 percent frequency among controls (C5467T) was associated with an increased risk of myocardial infarction (for
each additional copy, relative to women with one fewer copy, minimally adjusted OR = 1.4, 95 percent CI: 1.0, 1.8). However, haplotype analyses, which incorporated data from women missing genotypes for C5467T, showed a weaker association between the haplotype uniquely tagged by the 5467T allele and myocardial infarction (for each additional copy of haplotype C vs. all other haplotypes, minimally adjusted OR = 1.2, 95 percent 0.89, 1.6). The minor alleles 20603C and 21239G were associated with a slightly decreased risk of myocardial infarction, but only when major alleles for the other SNPs were present (for each additional copy of haplotype B, relative to women with one fewer copy, minimally adjusted OR = 0.68, 95 percent CI: 0.51, 0.91). This common haplotype was observed with 15 percent frequency among controls. Finally, our results suggest that hormone therapy did not modify the association of prothrombin haplotypes (global p = 0.61) with myocardial infarction.

The linkage disequilibrium-based approach taken here suggests that if the associations of the 5467T allele or haplotype B with myocardial infarction are due to functional genetic variants, the associations could be direct (implicating these variants) or indirect (implicating variants in linkage disequilibrium with these variants). In the SNP discovery population in which SNP selection was performed, the C5467T polymorphism was in linkage disequilibrium with 10 other SNPs at a pairwise linkage disequilibrium threshold of 0.64 (SeattleSNPs website (http://pga.gs.washington.edu/)). The location of the C5467T polymorphism in an exon and the other nine SNPs in introns suggests that any functional effects would most likely be attributed to the 5467T SNP. The 5467 C/T polymorphism results in a nonsynonymous change of methionine for threonine at amino acid 165. A homology-based algorithm that predicts whether coding-sequence changes affect protein function (SIFT (20)) suggests that in the context of the surrounding amino acid sequence, this nonsynonymous change may have functional consequences for the protein structure or activity (SIFT output available on SeattleSNPs website (http://pga.gs.washington.edu/)). With the exception of this study, this SNP remains uncharacterized with respect to myocardial infarction in the literature.

The two minor alleles found on haplotype B (20603C and 21239G), which was associated with a decreased risk of myocardial infarction, were not in linkage disequilibrium with any other SNPs at the prespecified linkage disequilibrium threshold. The 21239G allele has been described by Ceelie et al. (8), who reported a dose-dependent increase in prothrombin activity associated with the G allele in healthy controls. Greater prothrombin activity might suggest an increased thrombotic tendency, and therefore an increased risk of myocardial infarction might be predicted. However, the observation of a decreased relative risk of myocardial infarction associated with a haplotype carrying the 21239G allele is not necessarily in conflict. A haplotype effect is one possibility. In the presence of minor alleles at positions 2713, 2890, 4992, 5389 or 5467 (haplotypes C, D, and E), the decreased relative risk of myocardial infarction associated with the combination of 20603C and 21239G was not observed. Alternately, if the results are not due to chance, a true association between this haplotype and myocardial infarction might be attributed to either the functional effects of these SNPs or at least one other (possibly rarer) SNP in linkage disequilibrium that was not identified in the SNP discovery sample. This unidentified SNP would probably have an even stronger protective association with myocardial infarction than that of haplotype B, despite its lower frequency.

Strengths of this study include the availability of thorough data on common genetic variants in the prothrombin gene, the population-based design, objective determination of hormone therapy use, and the dual SNP and haplotype approaches. In the context of multiple functional SNPs or causal alleles arising on a specific ancestral haplotype, haplotype approaches may be particularly relevant (14). However, several factors may limit our findings. First, the haplotype approach taken in table 3 minimizes the association with the 5467T allele, which in the SNP analysis was associated with a greater and significant risk of myocardial infarction. The less marked odds ratios in table 3 are due to the larger percentage of missing data for this SNP, which, when filled in by haplotype inference, reduced the haplotype frequency among cases. Second, except for haplotype A, with a frequency nearing 50 percent, homozygotes for the other haplotypes comprised a relatively small number of subjects (see supplemental table 1 (www.aje.oxfordjournals.org)). For example, haplotype B homozygotes were virtually nonexistent in cases and comprised 1.6 percent of the control probability. Thus, odds ratios for homozygotes at either the SNP level or the haplotype level must be interpreted with caution. In addition, our power to examine interactions with hormone therapy was limited. With the exception of the previously characterized interaction of hormone therapy with 20210A in this population (12), the associations between each haplotype and myocardial infarction were similar in current users and nonusers, and all confidence intervals overlapped 1. Next, if they are not due to chance, our results might reflect either genetic associations with myocardial infarction or genetic associations with survival from myocardial infarction, since the study only included nonfatal myocardial infarction cases. Finally, we note that the application of SNP discovery data to population-based data on the basis of allele frequency and linkage disequilibrium and not on the basis of biologic function might naturally place our study in a hypothesis-generating role rather than a hypothesis-confirming role. In the context of multiple hypothesis-testing, our results should also be regarded as hypothesis-generating.

If confirmed, our findings support the hypothesis that common genetic variants in the prothrombin gene may contribute to the risk of myocardial infarction. Furthermore, single variants and combinations of variants may both be relevant. Although the contributions of any single variant or haplotype are likely to be small (i.e., small relative risk), common genetic variants, perhaps in combination with other genetic or environmental factors, may be associated with a moderate absolute risk of disease. Our preliminary results require replication in additional epidemiologic populations to establish validity and generalizability to other
populations and support from mechanistic studies to link genotype to biologic function.

ACKNOWLEDGMENTS

This study was funded by grants HL68639, HL43201, and HL74745 from the National Heart, Lung, and Blood Institute. L. A. H. was a Howard Hughes Medical Institute Predoctoral Fellow and was supported by grant T32-HL007902 from the National Heart, Lung, and Blood Institute.

Conflict of interest: none declared.

REFERENCES