Common Genetic Variation in Five Thrombosis Genes
and Relations to Plasma Hemostatic Protein Level
and Cardiovascular Disease Risk

Sekar Kathiresan, Qiong Yang, Martin G. Larson, Amy L. Camargo, Geoffrey H. Tofler,
Joel N. Hirschhorn, Stacey B. Gabriel, Christopher J. O’Donnell

Objective—We undertook a linkage disequilibrium (LD)—based genetic approach to investigate the hypothesis that common sequence variants in 5 thrombosis genes influence plasma hemostatic protein levels or risk of cardiovascular disease (CVD).

Methods and Results—In a reference panel, we characterized LD structure at the fibrinogen gene cluster (fibrinogen-β [FGB], FGA, and FGG), factor VII (F7), and tissue plasminogen activator (PLAT) loci. Forty-one tag single nucleotide polymorphisms (SNPs) were genotyped in 1811 unrelated Framingham Heart Study participants. There were significant associations of 9 FGB SNPs with fibrinogen level (minimum \( P = 0.002 \)) and of 7 F7 SNPs and F7 level (minimum \( P < 0.0001 \)). SNPs at the PLAT locus were not associated with PLAT level. In stepwise analysis, a single FGB variant explained 1% of the residual variance in fibrinogen level, and 2 F7 SNPs together explained 10% of the residual variance in F7 level. Two PLAT haplotypes were associated with CVD (multivariable-adjusted global \( P = 0.0004 \)).

Conclusions—A comprehensive survey of common sequence variation demonstrates that cis-regulatory SNPs explain a modest proportion of the residual variance in circulating fibrinogen and F7 level and PLAT haplotypes increase the risk of CVD. Additional studies are warranted to confirm the association of PLAT sequence, variation, and risk of CVD. We hypothesized that cis-acting SNPs influence hemostatic protein levels or CVD risk. In Framingham Heart Study participants, a single FGB SNP explained 1% of fibrinogen level variance, and 2 F7 SNPs together explained 10% of F7 level variance. PLAT gene haplotypes were associated with CVD. (Arterioscler Thromb Vasc Biol. 2006;26:000-000.)

Key Words: coagulation ■ fibrinogen ■ genetics ■ myocardial infarction ■ plasminogen activators ■ thrombosis ■ polymorphism

Plasma levels of blood clotting and fibrinolytic proteins may represent an important intermediate phenotype for cardiovascular disease (CVD) because increased circulating levels of hemostatic proteins have been associated with CVD risk. \(^1\)–\(^3\) Both plasma levels of hemostatic proteins \(^4\) and occurrence of CVD \(^5\) have a substantial heritable component. Hence, it has been suggested that sequence variation in thrombosis genes may explain interindividual variability in plasma hemostatic protein levels and CVD risk. \(^6\)

Several sequence variants at the fibrinogen-β (FGB) and factor VII (F7; encoded by the F7 gene) loci have been associated consistently with circulating plasma levels of their respective gene products. \(^7\)–\(^9\) Previous studies have been limited by an incomplete characterization of sequence variation at these loci. The proportion of plasma protein level variation explained by cis-acting sequence variation is unclear.

Recent advances in genomics allow the comprehensive evaluation of common sequence variation at a candidate gene locus for association with phenotype. Neighboring single nucleotide polymorphism (SNP) alleles are often highly correlated (ie, linkage disequilibrium [LD]). \(^10\) Genome-wide descriptions of LD structure have been reported recently, including the International Haplotype Map (HapMap) project. \(^11\) Within regions of LD, a subset of SNPs that are nonredundant (termed “tag SNPs”) may capture the majority of common variation. \(^12\)

Thus, to test the hypothesis that common sequence variants in thrombosis genes influence plasma hemostatic protein levels and risk of CVD, we sought to: (1) define the LD pattern for common SNP variants at 5 thrombosis gene loci (FGB, FGA, FGG, F7, and tissue plasminogen activator [PLAT]); and (2) identify a set of tag SNPs that capture underlying...
Methods

Study Participants

The Framingham Heart Study (FHS) offspring cohort was recruited in 1971 with the inclusion of 5124 participants who were offspring and spouses of the original cohort established in 1948.13 Participants in the present study are from a subset of individuals (n=2933) who provided blood samples for DNA extraction during the sixth clinical examination (1995 to 1998). Participants were unrelated (ie, 1 person randomly selected from each family or biologically unrelated to any other participants), and selection was designed to include equal numbers of men and women. DNA had been collected without regard to any phenotype feature.

Using these criteria, a panel of DNA samples from 1811 unrelated participants was constructed, and these participants represent the present study sample. The institutional review board at Boston Medical Center approved the study protocols, and all participants gave written informed consent. Clinical characteristics of the study participants are presented in Table 1.

Hemostatic Factor Levels

Fasting collection of samples for hemostatic factor levels and methods of their determination during the fifth cycle examination (1991 to 1995) have been described previously.14 Plasma fibrinogen was determined by the Clauss method.15 Tissue plasminogen activator (tPA) antigen was measured by a sandwich ELISA (TintElize TPA; Biopool, AB). The intra-assay coefficients of variation were 2.6% for fibrinogen, 3.0% for F7 antigen, and 5.5% for tPA antigen.

Assessment of Outcomes

Using prespecified criteria,16 a panel of 3 physicians blinded to genotype status confirmed the occurrence of CVD end points after review of all available hospital records, clinic notes, and pathology reports. For suspected stroke or transient ischemic attacks, a separate review panel including an FHS neurologist adjudicated events.

CVD end points occurring through the end of examination cycle 5 (1991 to 1995) were considered. The composite CVD end point (n=155) was defined by the presence of coronary heart disease (recognized or unrecognized myocardial infarction [MI], coronary insufficiency, and angina pectoris), stroke or transient ischemic attack, intermittent claudication, or congestive heart failure. More restrictive end points included coronary heart disease (n=114) and MI (n=52).

SNP Selection and Genotyping in Reference Panel

From the public National Center for Biotechnology Information dbSNP database and the SeattleSNPs variation discovery resource, we selected markers to encompass the coding region: 10 kb upstream of the transcription initiation site and 5 kb downstream of each of the 5 genes. An SNP map was constructed at each locus meeting the following criteria: (1) 1 common SNP (minor allele frequency ≥5%) approximately every 5 kb, and (2) minimum of 6 common SNPs per LD block (see below for LD block definition). Our effort to characterize LD structure was independent of the HapMap Consortium project conducted at the Broad Institute.17

SNPs were genotyped using the Sequenom MassARRAY platform (Sequenom)17 in a reference panel of 12 multigenerational white family pedigrees from Utah/Centre d’Etude du Polymorphisme Humain (CEPH) panel (Coriell Institute for Medical Research, Camden, NJ).15 The reference panel included 93 individuals of European ancestry. Supplemental Table I (available online at http://atvb.ahajournals.org) summarizes the genotyping performed in CEPH reference panel.

TABLE 1. Characteristics of FHS Sample

<table>
<thead>
<tr>
<th></th>
<th>Men (n=905)</th>
<th>Women (n=906)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>56±10</td>
<td>55±10</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>28±4</td>
<td>27±5</td>
</tr>
<tr>
<td>Current cigarette smoking, %</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>129±18</td>
<td>125±20</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>76±10</td>
<td>73±10</td>
</tr>
<tr>
<td>Hypertension treatment, %</td>
<td>21.9</td>
<td>17.7</td>
</tr>
<tr>
<td>Total cholesterol, mg/dL</td>
<td>200±36</td>
<td>210±38</td>
</tr>
<tr>
<td>HDL-C, mg/dL</td>
<td>43±11</td>
<td>56±16</td>
</tr>
<tr>
<td>Triglycerides, mg/dL</td>
<td>161±122</td>
<td>141±131</td>
</tr>
<tr>
<td>Diabetes, %</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Alcohol consumption, ounces per week</td>
<td>4±4</td>
<td>1.8±2.5</td>
</tr>
<tr>
<td>Postmenopausal, %</td>
<td>NA</td>
<td>66</td>
</tr>
<tr>
<td>Estrogen replacement therapy, %</td>
<td>NA</td>
<td>18</td>
</tr>
<tr>
<td>Lipid-lowering therapy, %</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>MI, %</td>
<td>5.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Coronary heart disease, %</td>
<td>9.4</td>
<td>3.5</td>
</tr>
<tr>
<td>CVD, %</td>
<td>11.5</td>
<td>6.0</td>
</tr>
<tr>
<td>Fibrinogen, mg/dL</td>
<td>307±59</td>
<td>314±56</td>
</tr>
<tr>
<td>F7 antigen, % control</td>
<td>98±16</td>
<td>104±17</td>
</tr>
<tr>
<td>Plasminogen activator inhibitor-1, mg/mL</td>
<td>24±15</td>
<td>21±16</td>
</tr>
<tr>
<td>Tissue plasminogen activator, ng/mL</td>
<td>10±4</td>
<td>8.6±4</td>
</tr>
</tbody>
</table>

Continuous measures, mean±SD; categorical variables, percentage. HDL-C indicates high-density lipoprotein cholesterol. Baseline data from the fifth examination cycle (1995 to 1998) are presented. Biopool, AB. The intra-assay coefficients of variation were 2.6% for fibrinogen, 3.0% for F7 antigen, and 5.5% for tPA antigen.
threshold for nominal significance at $P<0.01$. For CVD end point analyses, logistic regression analysis was performed with each of the SNPs to test the null hypothesis that CVD status did not differ by SNP genotype (3 genotype classes per SNP). Haplotype-based association analyses for the hemostatic factor level and CVD phenotypes were conducted using a weighted-regression approach.\textsuperscript{21} For covariates and a fuller description of haplotype analyses, please see the supplemental Methods, available online at http://atvb.ahajournals.org).

**Results**

**LD Structure and Haplotypes at Candidate Gene Loci in the Reference Panel**

In the Figure, we present the genomic organization and LD structure defined in the reference panel for the fibrinogen gene cluster ($FGB$, $FGA$, and $FGG$ genes, respectively) on chromosome 4. Forty-four SNPs span a genomic region of 61.8 kb (average spacing $\approx$ 1 SNP every 1.4 kb) with 3 blocks of strong LD (extending 12, 13, and 17 kb, respectively). In each block, a small number (4 to 6) of common haplotypes represented $>95\%$ chromosomes. Sixteen tag SNPs predicted all haplotypes $\geq 5\%$ frequency at the fibrinogen gene cluster.

There is a single block of LD at the $F7$ locus and a single block at the $PLAT$ locus (supplemental Figures I and II, respectively, available online at http://atvb.ahajournals.org). In sum for all 5 gene regions, a total of 25 tag SNPs were identified and genotyped in the FHS cohort along with an additional 16 SNPs (supplemental Table I).

The LD structure defined in the present study agreed well with that seen in the HapMap database as of March 1, 2006 (data not shown).

**Haplotype Correlation Between Framingham Participants and Reference Panel**

We inferred common haplotypes (frequency $\geq 5\%$) within each block in the FHS sample ($\approx$3600 chromosomes), and we assessed the correlation with haplotypes initially defined in the reference panel (96 chromosomes). The haplotype frequencies in the FHS sample were highly correlated with frequencies in the reference sample (Pearson $r=0.92$; $P<0.0001$; supplemental Figure III). Of the 22 haplotypes $\geq 5\%$ frequency in FHS, all were observed in the CEPH reference panel.

LD structure at the fibrinogen gene cluster in reference panel. Chr 4 panel depicts the $FGB$, $FGA$, and $FGG$ positions on chromosome 4 in the human genome July 2003 assembly (hg16). Haplotype block structure panel shows that 44 SNPs fell into 3 haplotype blocks. Six tag SNPs in block 1 are labeled with a blue rectangle. Six common haplotypes from block 1 are labeled A through F. Alleles shown may be that from the positive or negative strand of the genome. LD structure panel displays the LD relationships between pairs of markers in the region, with each square representing the pairwise strength and significance of LD. Red coloring indicates no or minimal evidence of historic recombination, white indicates weak LD, and blue uninformative LD. Figure prepared using LocusView 2.0, T. Petryshen et al, Broad Institute.
HindIII circulating fibrinogen (also known in previous literature as polymorphism reported previously to be associated with levels. Included among these SNPs was FGB_001437, a alleles of all 9 SNPs were associated with higher fibrinogen SNPs were associated with fibrinogen level (Table 2), 9 of 14

In association analyses of each of 14 FGB SNPs, Haplotypes, and Plasma Fibrinogen Level

Fibrinogen Gene Cluster SNPs, Haplotypes, and Plasma Fibrinogen Level

In association analyses of each of 14 FGB SNPs with plasma fibrinogen levels in FHS participants (Table 2), 9 of 14 FGB SNPs were associated with fibrinogen level (P<0.05). Minor alleles of all 9 SNPs were associated with higher fibrinogen levels. Included among these SNPs was FGB_001437, a polymorphism reported previously to be associated with circulating fibrinogen (also known in previous literature as β-455G/A). FGB_001437 is in strong LD (D'=1) with HindIII β-148C/T (FGB_001744 in present study), another polymorphism related previously to fibrinogen level. To resolve whether the 14 SNP associations reflected distinct SNP effects or multiple associations attributable to LD, we conducted stepwise analysis. In stepwise analysis, FGB_001437 was the only significant predictor of fibrinogen level (stepwise P=0.026). FGB_001437 explained 1% of the residual variance in multivariable-adjusted fibrinogen level. Higher fibrinogen levels were observed in individuals harboring either 2 copies or 1 copy of the minor A allele of FGB_001437 (314±5.4 mg/dL, 316±5.6 mg/dL, and 308±5.9 mg/dL for the AA, AG, and GG genotypes, respectively). In a global test for associations of mean plasma fibrinogen level with FGB haplotypes within the single 12-kb block of LD (Figure, block 1), mean plasma fibrinogen level differed across haplotypes (global P=0.03; supplemental Table II). Differences were also noted for individual haplotypes compared with all other haplotypes combined. Hap C and Hap G were each associated with higher fibrinogen level (individual haplotype P=0.02 and P=0.04, respectively; supplemental Table II), whereas Hap A was associated with lower fibrinogen level (individual haplotype P=0.04). Hap C contains the minor alleles of all 9 SNPs associated with fibrinogen level, (Figure). Thus, the haplotype findings are consistent with those for individual SNPs. Although there were haplotype-specific associations with fibrinogen level in block 1, there were no significant associations by the global test in fibrinogen gene cluster blocks 2 and 3.

F7 SNPs, Haplotypes, and Plasma F7 Level

In association analyses of plasma levels of F7 with each of 8 F7 SNPs, 7 SNPs were associated with F7 level (Table 3). Minor alleles of 5 SNPs were associated with lower F7 levels, whereas the minor alleles of 2 SNPs were associated with higher F7 level. In stepwise selection including all 8 genetic variants, a promoter sequence variant, SNP rs1755685, explained 9% of the residual variance in multivariable-adjusted F7 level (stepwise P<0.0001). SNP rs964617 explained an additional 1% of the residual variance in F7 level (stepwise P=0.003). Individuals harboring either 1 or 2 copies of the minor A allele of rs1755685 had lower F7 level (80%±14%, 95%±16%, and 104%±16% for the AA, AC, and CC genotypes, respectively). Individuals harboring either 1 or 2 copies of the minor T allele of rs964617 had higher F7 level (105%±17%, 104%±15%, and 99%±17% for the TT, CT, and CC genotypes, respectively). Notably, the previously studied Arg/Gln variant (SNP rs6046) was not retained in the stepwise model, although it was in LD with SNP rs1755685 (D'=0.96; r²=0.79), and it explained a similar proportion of the residual variance in individual SNP analysis.

<table>
<thead>
<tr>
<th>Allelic Variant</th>
<th>Variant Type</th>
<th>Major Allele→Minor Allele</th>
<th>Minor Allele Frequency</th>
<th>P</th>
<th>Effect Direction</th>
<th>Minor Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1025154</td>
<td>Intrinsic</td>
<td>A→G</td>
<td>0.19</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGB_000472 (rs1800789)</td>
<td>5’ flanking</td>
<td>G→A</td>
<td>0.20</td>
<td>0.002</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>FGB_001038 (rs1800791)</td>
<td>5’ flanking</td>
<td>G→A</td>
<td>0.15</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGB_001437† (rs1800790)</td>
<td>5’ flanking</td>
<td>G→A</td>
<td>0.20</td>
<td>0.005</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>FGB_001643 (rs1800788)</td>
<td>5’ flanking</td>
<td>C→T</td>
<td>0.20</td>
<td>0.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGB_003582 (rs2227399)</td>
<td>Intrinsic</td>
<td>T→G</td>
<td>0.19</td>
<td>0.002</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>FGB_004110 (rs2227401)</td>
<td>Intrinsic</td>
<td>C→T</td>
<td>0.21</td>
<td>0.003</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>FGB_004950 (rs2042642)</td>
<td>Intrinsic</td>
<td>G→C</td>
<td>0.20</td>
<td>0.003</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>FGB_006550 (rs6056)</td>
<td>Coding, synonymous</td>
<td>C→T</td>
<td>0.19</td>
<td>0.03</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>FGB_009487 (rs4220)</td>
<td>Coding, nonsynonymous</td>
<td>G→A</td>
<td>0.17</td>
<td>0.03</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>FGB_009952 (rs2227421)</td>
<td>Coding, synonymous</td>
<td>C→G</td>
<td>0.21</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGB_009977 (rs2227423)</td>
<td>3’ untranslated</td>
<td>A→C</td>
<td>0.32</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGB_010447 (rs2227244)</td>
<td>3’ flanking</td>
<td>T→C</td>
<td>0.18</td>
<td>0.007</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>FGB_011079 (rs1044291)</td>
<td>3’ flanking</td>
<td>C→T</td>
<td>0.32</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Phenotype was standardized residual of plasma fibrinogen level adjusted for the following covariates: age, sex, body mass index, current cigarette smoking, systolic blood pressure, diastolic blood pressure, hypertension treatment, total cholesterol, high-density lipoprotein cholesterol, triglycerides, diabetes, alcohol consumption, prevalent CVD, and, in women, menopause status and estrogen replacement therapy.

*In a stepwise selection model including all 14 FGB variants tested, only SNP FGB_001437 was significantly related to plasma fibrinogen level (P=0.03). †partial r² for SNP FGB_001437 was 0.01; †alleles shown may be that from the positive or negative strand of the genome; §+ refers to an association with higher fibrinogen level.
Mean plasma F7 levels differed across F7 gene haplotypes (global $P<10^{-6}$; supplemental Table III). Compared with all haplotypes combined, Hap D was associated with lower levels of F7 (Hap $P<10^{-5}$), whereas Hap B and Hap E were each associated with higher levels of F7 (Hap $P<10^{-5}$ and Hap $P=0.008$, respectively). Hap D harbors the minor allele of rs1755685 and haplotypes B and E, the minor allele of rs964617 (supplemental Figure I).

There were no significant associations of the SNPs or haplotypes at the PLAT locus with plasma tPA level (all $P>0.05$).

### Associations of Thrombosis Gene SNPs and Haplotypes with CVD Risk

In analysis of individual variants, a single SNP at the PLAT locus, rs2020918, was associated with increased risk of CVD (Table 4). In age- and sex-adjusted analyses, the CT and TT genotypes were associated with odds ratios (ORs) of 1.5 and 2.6 for CT; general model $P=0.004$) and MI (OR, 4.0; 95% CI, 1.6 to 10.0; general model $P=0.006$). In the multivariable-adjusted model, the magnitude of the association was modestly attenuated but remained statistically significant (Table 4).

**PLAT** gene haplotypes were significantly associated with risk of CVD (global $P=0.0004$; supplemental Table IV). Hap B (at 28% frequency) was associated with increased risk of CVD (Hap $P=0.02$) and Hap D (at 7% frequency) was similarly associated with an increased risk of CVD (Hap $P=0.0006$). Conversely, Hap C and Hap F were associated with decreased risk of CVD (Hap $P=0.02$ for each). The T allele of SNP rs2020918 distinguishes the 2 haplotypes associated with increased risk of CVD (haplotypes B and D) from all others.

There were no associations with CVD for SNPs at FGA, FGB, FGG, and F7 or of other SNPs at the PLAT locus (all $P>0.01$). Similarly, haplotypes at the fibrinogen gene cluster and F7 were not associated with CVD (global $P$ for each block $>0.05$).

### Discussion

Using an LD-based approach, we comprehensively surveyed common genetic variation at 5 thrombosis gene loci and associated with cardiovascular events in 1754 men and women in FHS.

**TABLE 4. Association of SNP rs2020918 at the PLAT Locus With Cardiovascular Events in 1754 Men and Women in FHS**

<table>
<thead>
<tr>
<th>OR (95% CI)</th>
<th>Unadjusted</th>
<th>Age/sex adjusted</th>
<th>Fully adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total CVD</strong></td>
<td>1.5 (1.0–2.2)</td>
<td>1.5 (1.0–2.3)</td>
<td>1.5 (0.9–2.2)</td>
</tr>
<tr>
<td><strong>MI</strong></td>
<td>2.9 (1.4–5.9)</td>
<td>3.0 (1.4–6.3)</td>
<td>2.3 (1.0–5.0)</td>
</tr>
</tbody>
</table>

*For each genotype, the No. of men and women were as follows: CC, n=676; CT, n=862; TT, n=216; †includes MI (both recognized and unrecognized), angina pectoris, coronary insufficiency, intermittent claudication, coronary heart disease death, transient ischemic attack, congestive heart failure, and atherothrombotic stroke; §analyses were adjusted for the following covariates: age, sex, hypertension, diabetes, total and high-density lipoprotein cholesterol, lipid-lowering therapy, smoking, body mass index, and triglycerides; ¶includes MI (both recognized and unrecognized), angina pectoris, coronary insufficiency, and coronary heart disease death; ‡includes recognized and unrecognized MI.

Analyses performed in the genotype model (3 genotype classes per SNP) with the CC genotype as the referent.
sessed for associations with both circulating levels of hemostatic factors and risk of CVD outcomes in 1811 men and women. Several SNPs located in cis-regulatory sequence are associated with plasma protein levels or risk of CVD. After accounting for clinical covariates and correlation among SNPs, a single polymorphic variant at the FGB gene explains 1% of the residual variance in circulating fibrinogen level and 2 F7 gene variants explain 10% of the residual variance in F7 level. There was no evidence for association of SNPs in FGB or F7 with CVD, but a common SNP in the PLAT gene region was associated with 2.5-fold increased risk of CVD.

Our results provide the following: (1) extend current knowledge about the role of cis-regulatory genetic variation at the fibrinogen gene cluster, FVII, and PLAT loci in relation to plasma hemostatic protein levels and CVD outcomes; (2) provide strong support for the ability to use the association approach to identify variants that affect complex phenotypes; and (3) show that this approach may be particularly useful in the search for functional alleles that affect plasma protein levels.

**Fibrinogen**

Since an initial report in 1987, several individual FGB polymorphisms have been related to plasma fibrinogen level.\(^8\)

We extend previous work by comprehensively and jointly assessing the role of common variation at the FGB locus. We identified ≥9 individual FGB variants that are associated with fibrinogen level. These individual SNP results confirm multiple previous reports that the minor A allele of FGB_001437 (β-455G/A\(^25\) in previous literature) and other SNPs in LD\(^8\) with this variant (including HindIII β-148C/T) are associated with higher fibrinogen level.

When all individually associated variants are considered jointly, FGB_001437 was most strongly associated with fibrinogen level, and this polymorphism explained a modest proportion (≈1%) of the variance in multivariable-adjusted fibrinogen level. This suggests that a single polymorphic site at the FGB locus influences fibrinogen level. Several in vitro studies have shown that the minor allele of FGB_001437 binds less well to a putative repressor protein complex and thus is associated with increased FGB chain transcription.\(^23\)

Additional functional studies investigating each of the 9 associated variants may help clarify whether FGB_001437 is the true causal site.\(^24\)

**Factor VII**

In previous reports, 5 individual F7 SNPs were reported to be related with plasma F7 levels: 4 in the 5' sequence and 1 nonsynonymous coding variant.\(^25\)–\(^29\) The minor alleles of the coding variant R353Q (also termed Arg/Gln in the literature) and 3 others have been related to lower F7 level, whereas the minor allele of the fifth variant has been associated with increased F7 level. We used stepwise selection to identify the number of independent SNP effects at the F7 locus and found that there were 2 distinct SNP effects at the F7 locus: 1 variant strongly relating to lower F7 levels and 1 variant relating to higher F7 levels.

**Tissue Plasminogen Activator**

We found no association of either PLAT SNPs or PLAT haplotypes with circulating levels of tPA antigen. A recent study reported a similar absence of association with steady-state plasma tPA level.\(^30\) In contrast, when examining a dynamic tPA phenotype, vascular release of tPA after stress, Ladenvall et al reported that carriage of the minor T allele at -7351 C/T (or rs2020918) was associated with a 50% decreased vascular release of tPA after stress.\(^31\) Therefore, this finding suggests that decreased local release of tPA, but not altered systemic levels, may increase risk of thrombotic CVD.\(^31\)

Consistent with this hypothesis, we found a 2.5-fold increased risk of CVD and a 3.1-fold increased risk of MI associated with the T allele of rs2020918. rs2020918 is located 7.2 kb upstream of the transcription initiation site, within an enhancer region. Several in vitro studies have demonstrated this region to contain a binding site for the Sp1 transcription factor.\(^32\)–\(^34\) The C to T substitution at rs2020918 has been shown reduce the affinity of Sp1 to this site, suggesting a mechanism by which rs2020918 is associated with lower tPA transcription.\(^31\)\(^,\)\(^34\)

Our PLAT association results for the combined CVD end-point replicate 2 previous studies in small samples of an association of rs2020918 with MI or stroke. In a population-based nested case-control study (61 MI cases and 120 controls), individuals bearing the minor T allele had a significant 2.7-fold increased odds of MI compared with the CC genotype, a risk estimate similar to our finding.\(^35\) Separately, in a sample of 182 cases of ischemic stroke and 301 controls, the TT genotype has been associated with increased risk of ischemic stroke (OR, 1.9; 95% CI, 1.01 to 3.6).\(^36\)

**Study Limitations**

Our study has several limitations. First, our association study sample is entirely white. Our findings may not generalize to other ethnicities/races, in particular those groups such as blacks with much different LD patterns. However, a recent study suggests that the effects of gene variants may be largely consistent across different “racial” or “ethnic” groups.\(^37\) Second, population stratification may lead to spurious associations attributable to underlying population substructure with differences in allele frequencies.\(^38\) However, this possibility is lessened in our sample given the ethnic homogeneity.

Third, our study examined prevalent CVD at the fifth clinic examination. The apparent association between the PLAT variant and CVD may represent a survival bias if individuals with the risk T allele had a lower case fatality rate than other genotypes. Fourth, because of a limited number of individual event types, our clinical end-point investigation used a composite outcome. However, our composite outcome included both MI and stroke, and these phenotypes have been associated previously with PLAT SNP rs2020918.

Fifth, the small number of CVD events may have reduced our statistical power to detect SNP–CVD associations. Sixth, we may have failed to detect SNP-biomarker associations because of insufficient statistical power. We estimate that with a sample size of 1500 unrelated individuals and a significance level 0.001, and under a conservative assumption that nongenetic factors explain 10% of trait variance, we had
### Implications and Conclusions

Our implementation of an LD-based approach at 5 medically relevant genes comes on the heels of the recent completion of the HapMap project and offers several possible insights. First, studying LD in a reference panel is efficient and generates haplotype frequency estimates similar to those seen in the larger population, at least within a population of European descent.

Second, the strength of associations between genetic variants and circulating blood protein levels supports the general use of circulating biomarker “intermediate phenotypes” to test hypotheses regarding cis-regulatory effects. Sequence variants causally related to circulating protein levels may in turn serve as strong candidate alleles to test for association with disease.

In conclusion, several cis-regulatory sequence variants at the FGB and F7 loci are associated with circulating levels of their respective protein products, and a PLAT cis-regulatory variant increases risk of CVD. Further studies to replicate the PLAT SNP association in diverse populations are warranted.

### Acknowledgments

This work was supported in part through the National Heart, Lung and Blood Institute Framingham Heart Study (contract No. N01-HC-25195) and the Cardiogenomics Programs for Genomic Applications (HL66582). S.K. was supported by the ACC Foundation/Merck (HL66582). S.K. was supported by the ACC Foundation/Merck (HL66582). S.K. was supported by the ACC Foundation/Merck (HL66582).

### References


Common Genetic Variation in Five Thrombosis Genes and Relations to Plasma Hemostatic Protein Level and Cardiovascular Disease Risk
Sekar Kathiresan, Qiong Yang, Martin G. Larson, Amy L. Camargo, Geoffrey H. Tofler, Joel N. Hirschhorn, Stacey B. Gabriel and Christopher J. O'Donnell

Arterioscler Thromb Vasc Biol. published online April 13, 2006;
Arteriosclerosis, Thrombosis, and Vascular Biology is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2006 American Heart Association, Inc. All rights reserved.
Print ISSN: 1079-5642. Online ISSN: 1524-4636

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://atvb.ahajournals.org/content/early/2006/04/13/01.ATV.0000222011.13026.25.citation

Data Supplement (unedited) at:
http://atvb.ahajournals.org/content/suppl/2006/05/04/01.ATV.0000222011.13026.25.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Arteriosclerosis, Thrombosis, and Vascular Biology 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 Arteriosclerosis, Thrombosis, and Vascular Biology is online at:
http://atvb.ahajournals.org//subscriptions/
SUPPLEMENTARY METHODS

Statistical Analysis

Multivariable regression analyses were conducted using SAS. For each hemostatic factor, levels were logarithmically transformed because of skewed distributions. Sex-specific standardized residuals derived from multivariable-adjusted hemostatic factor levels served as the phenotype. Linear regression analysis was performed to test the null hypothesis that the mean level of the standardized residual did not differ by SNP genotype (3 genotype classes per SNP). Covariates (continuous measures unless otherwise specified) were measured during the fifth examination and included in the multivariable models were age, body mass index (kg/m²), current cigarette smoking (yes if current smoker or quit within 1 year, no), systolic blood pressure (mmHg), diastolic blood pressure (mmHg), hypertension treatment (yes, no), total cholesterol (mg/dL), high density lipoprotein cholesterol (mg/dL), triglycerides (mg/dL), diabetes (yes defined by fasting blood glucose >125 or current medication use for diabetes, no), alcohol consumption (ounces/week), prevalent CVD (yes, no), and in women, menopause status (yes, no) and estrogen replacement therapy (yes, no). To identify SNPs with the strongest contribution to overall variance, we further conducted forward stepwise selection that included all SNPs genotyped at each locus. In the individual SNP analysis, we set a threshold for nominal significance at P<0.01.

For CVD endpoint analyses, logistic regression analysis was performed with each of the SNPs to test the null hypothesis that CVD status did not differ by SNP genotype (3 genotype classes per SNP). Covariates in the multivariable-adjusted model were measured during the fifth examination and included age, sex, body mass index, cigarette
smoking, hypertension (defined by systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or hypertension treatment), total cholesterol, high density lipoprotein cholesterol, triglycerides, lipid lowering therapy, and diabetes.

Haplotype-based association analyses for the hemostatic factor level and CVD phenotypes were conducted using a weighted-regression approach (haplo.stats). All compatible common haplotype configurations of a multi-marker genotype were used in the regression weighted by the corresponding posterior probability of each configuration. The posterior probability was estimated using an expectation-maximization algorithm. In order to conduct haplotype analyses for all genotyped subjects, including the small number of subjects with incomplete genotype data at some of the SNPs, missing genotypes at FGB and PLAT were imputed using an algorithm implemented in haplo.stats.

A n-1 degree of freedom global score statistic tested all n haplotypes simultaneously to detect any departure from the null hypothesis of no association. A one degree of freedom haplotype-specific score statistic tested whether trait differences exist between a single haplotype and all other haplotypes combined. Haplotypes above a threshold of 0.5% were considered in analyses. Haplotype analyses were adjusted for separate sets of covariates (listed above) for the hemostatic factor level and CVD phenotypes. In haplotype analyses, we set a threshold for nominal significance for each haplotype block at a global P <0.05.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Symbol</th>
<th>Locus Link ID*</th>
<th>Gene Size (kb)</th>
<th># SNPs attempted in CEPH</th>
<th># Common, Working SNPs in CEPH †</th>
<th>Genomic Span of Common, Working SNPs in CEPH</th>
<th>Average inter-SNP distance (kb) in CEPH</th>
<th>LD Blocks in CEPH</th>
<th># Haplotypes ≥5% frequency in CEPH</th>
<th># Tag SNPs genotyped in FHS</th>
<th>Total # of SNPs genotyped in FHS (Tag + Additional)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibrinogen Beta</td>
<td>FGB</td>
<td>2244</td>
<td>8.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrinogen Alpha</td>
<td>FGA</td>
<td>2243</td>
<td>5.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrinogen Gamma</td>
<td>FGG</td>
<td>2266</td>
<td>8.5</td>
<td>93</td>
<td>44</td>
<td>61.8</td>
<td>1.4</td>
<td>3</td>
<td>5 in Block 2</td>
<td>16</td>
<td>27</td>
</tr>
<tr>
<td>Factor VII</td>
<td>F7</td>
<td>2155</td>
<td>14.2</td>
<td>38</td>
<td>20</td>
<td>64.5</td>
<td>3.2</td>
<td>1</td>
<td>4 in Block 1</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Tissue Plasminogen Activator</td>
<td>PLAT</td>
<td>5327</td>
<td>32.4</td>
<td>39</td>
<td>7</td>
<td>45.5</td>
<td>6.5</td>
<td>1</td>
<td>4 in Block 1</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Abbreviations: CEPH denotes Centre d'Étude du Polymorphisme Humain; LD, linkage disequilibrium; kb, kilobase; SNP, single nucleotide polymorphism; and FHS, Framingham Heart Study.

* LocusLink refers to curated database of sequence information by the National Center for Biotechnology Information.

† SNPs were deemed "Common, Working" if they met the following criteria: 1) ≥75% success for genotyping calls; 2) Hardy-Weinberg equilibrium P>0.01; 3) Mendelian transmission errors ≤2; and 4) minor allele frequency ≥5%.
**TABLE II. Association of Fibrinogen-β Gene Haplotypes with Multivariable-Adjusted Plasma Fibrinogen Level in FHS**

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>FGB_001038</th>
<th>FGB_001643</th>
<th>FGB_003582</th>
<th>FGB_006550</th>
<th>FGB_009952</th>
<th>FGB_011079</th>
<th>Haplotype Frequency</th>
<th>Haplotype P</th>
<th>Effect Direction</th>
<th>Global P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype A</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>0.33</td>
<td>0.04</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Haplotype B</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>0.19</td>
<td>0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>0.18</td>
<td>0.02</td>
<td>+</td>
<td>0.03</td>
</tr>
<tr>
<td>Haplotype D</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>0.16</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype E</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>0.12</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype F</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>0.02</td>
<td>0.04</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: FHS denotes Framingham Heart Study; and FGB, fibrinogen-β.

Phenotype was standardized residual of plasma fibrinogen level adjusted for the following covariates: age, sex, body mass index, current cigarette smoking, systolic blood pressure, diastolic blood pressure, hypertension treatment, total cholesterol, high density lipoprotein cholesterol, triglycerides, diabetes, alcohol consumption, prevalent cardiovascular disease and in women, menopause status and estrogen replacement therapy.

Due to missing covariates and/or fibrinogen level, total n=1516 for haplotype association analyses.

*Alleles shown may be that from the positive or negative strand of the genome.

*(−)* refers to an association with lower fibrinogen level and *(+)* refers to an association with higher fibrinogen level compared with all other haplotypes combined.
TABLE III. Association of *F7* Gene Haplotypes with Multivariable-Adjusted Plasma Factor VII Level in FHS

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>rs10665</th>
<th>rs9464617</th>
<th>rs6039</th>
<th>rs1475931</th>
<th>Haplotype Frequency</th>
<th>Haplotype P</th>
<th>Effect Direction</th>
<th>Global P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype A</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>0.43</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype B</td>
<td>A</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>0.22</td>
<td>&lt;0.0001</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Haplotype C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>0.19</td>
<td>0.32</td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Haplotype D</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>0.12</td>
<td>&lt;0.0001</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Haplotype E</td>
<td>A</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>0.03</td>
<td>0.008</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: *F7* denotes Factor VII; and FHS, Framingham Heart Study.
Phenotype was standardized residual of plasma factor VII level adjusted for covariates listed in Table II. Due to missing covariates, factor VII level, and/or genotypes, total n=1253 for haplotype association analyses.

(+) refers to as an association with lower factor VII level and (+) refers to an association with higher factor VII level compared with all other haplotypes combined.
# TABLE IV. Association of PLAT Haplotypes with Cardiovascular Events*†‡ in FHS

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>rs2020918</th>
<th>rs2299609</th>
<th>rs879293</th>
<th>rs73612</th>
<th>rs2285169</th>
<th>rs1050275</th>
<th>Haplotype Frequency</th>
<th>Haplotype P</th>
<th>Effect Direction§</th>
<th>Global P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C C A A T T</td>
<td>0.37</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>T G G C C C</td>
<td>0.28</td>
<td>0.02</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>C C G A T T</td>
<td>0.16</td>
<td>0.02</td>
<td>-</td>
<td>0.0004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>T G G C C T</td>
<td>0.07</td>
<td>0.0006</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>C G G C C T</td>
<td>0.05</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>C C G C C T</td>
<td>0.03</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: FHS denotes Framingham Heart Study; and CI, confidence interval.

*Includes myocardial infarction (both recognized and unrecognized), angina pectoris, coronary insufficiency, intermittent claudication, coronary heart disease death, transient ischemic attack, congestive heart failure, and atherothrombotic stroke.

†Analyses were adjusted for the following covariates: age, sex, hypertension, diabetes, total and high-density lipoprotein cholesterol, lipid lowering therapy, smoking, BMI, and triglycerides.

‡Alleles shown may be that from the positive or negative strand of the genome.

§(-) refers to as an association with lower risk of CVD and (+) refers to an association with increased risk of CVD compared with all other haplotypes combined.

n=1791 for fully adjusted analyses.
SUPPLEMENTARY Figure I.

Linkage disequilibrium (LD) structure at the factor VII (F7) locus in reference panel. Chr 13 panel depicts the F7 locus position on chromosome 13 in the human genome July 2003 assembly (hg16). Haplotype block structure panel shows that 9 of the 20 SNPs fell into a single haplotype block. Alleles shown may be that from the positive or negative strand of the genome. LD structure panel displays the LD relations between pairs of markers in the region with each square representing the pairwise strength and significance of LD. Red coloring indicates no or minimal evidence of historical recombination, white indicates weak LD, and blue uninformative LD. Figure prepared using LocusView 2.0, T. Petryshen et al, Broad Institute (http://www.broad.mit.edu/mpg/locusview/).
SUPPLEMENTARY Figure II.

Linkage disequilibrium (LD) structure at the tissue plasminogen activator (PLAT) locus in reference panel. Chr 8 panel depicts the PLAT locus position on chromosome 8 in the human genome July 2003 assembly (hg16). Haplotype block structure panel shows that 7 SNPs fell into a single haplotype block. Alleles shown may be that from the positive or negative strand of the genome. LD structure panel displays the LD relations between pairs of markers in the region with each square representing the pairwise strength and significance of LD. Red coloring indicates no or minimal evidence of historical recombination, white indicates weak LD, and blue uninformative LD. Figure prepared using LocusView 2.0, T. Petryshen et al, Broad Institute (http://www.broad.mit.edu/mpg/locusview/).
SUPPLEMENTARY Figure III.

Correlation in Haplotype Frequency in Reference Panel and the Framingham Heart Study Sample. Haplotypes composed of tag SNPs within each block were constructed in the reference panel (96 independent chromosomes) and Framingham Heart Study sample (up to 3622 chromosomes). The haplotype frequencies in the reference panel were highly correlated with frequencies in the Framingham Heart Study sample (Pearson r=0.92, P<0.0001).
SUPPLEMENTARY Figure I
SUPPLEMENTARY Figure III

![Haplotype Frequency in Reference Panel vs. Fibrinogen Gene Cluster](chart.png)
References


(2) Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70(2):425-34.