Comprehensive Survey of Common Genetic Variation at the Plasminogen Activator Inhibitor-1 Locus and Relations to Circulating Plasminogen Activator Inhibitor-1 Levels

Sekar Kathiresan, MD*; Stacey B. Gabriel, PhD*; Qiong Yang, PhD; Amy L. Lochner, AB; Martin G. Larson, ScD; Daniel Levy, MD; Geoffrey H. Tofler, MD; Joel N. Hirschhorn, MD, PhD; Christopher J. O’Donnell, MD, MPH

**Background**—Using a linkage disequilibrium (LD)–based approach, we sought to comprehensively define common genetic variation at the plasminogen activator inhibitor-1 (PAI-1) locus and relate common single nucleotide polymorphisms (SNPs) and haplotypes to plasma PAI-1 levels.

**Methods and Results**—In reference pedigrees, we defined LD structure across a 50-kb genomic segment spanning the PAI-1 locus via a dense SNP map (1 SNP every 2 kb). Eighteen sequence variants that capture underlying common genetic variation were genotyped in 1328 unrelated Framingham Heart Study participants who had plasma PAI-1 antigen levels measured. Regression analyses were used to examine associations of individual SNPs and of inferred haplotypes with multivariable-adjusted PAI-1 levels. Two genetic variants, SNP rs2227631 and the 4G/5G polymorphism, were strongly associated (P<0.0001) with PAI-1 levels. SNP rs2227631 is in tight LD (D’=0.97, r²=0.78) with the 4G/5G polymorphism, which makes it difficult to distinguish which of these 2 polymorphisms is responsible for the association with PAI-1 levels. In stepwise analysis considering all polymorphisms tested, 3 SNPs, rs2227631 (or the correlated 4G/5G polymorphism), rs6465787, and rs2227674, each explained 2.5%, 1%, and 1%, respectively, of the residual variance in multivariable-adjusted PAI-1 levels (stepwise P<0.0001, P=0.04, and P=0.03, respectively). A single common haplotype, at 50% frequency among Framingham Heart Study participants, was strongly associated with higher PAI-1 levels (haplotype-specific P=0.0001). The susceptibility haplotype harbors the minor alleles of SNP rs2227631 and the 4G/5G polymorphism.

**Conclusions**—Three sequence variants at the PAI-1 locus, in sum, explain ∼5% of the residual variance in multivariable-adjusted PAI-1 levels. For quantitative cardiovascular traits such as circulating biomarkers, defining LD structure in a candidate gene followed by association analyses with both SNPs and haplotypes is an effective approach to localize common susceptibility alleles.

**Key Words:** plasminogen • epidemiology • genetics • genomics • fibrinolysis

Higher plasma levels of plasminogen activator inhibitor-1 (PAI-1), the principal circulating inhibitor of fibrinolysis, have been associated with increased risk of coronary heart disease events. In addition, increased PAI-1 is a common feature of the metabolic syndrome and predicts incident diabetes. Significant heritability of this phenotype in family and twin studies suggests that plasma PAI-1 is determined in part by genetic influences. In vitro and genetic studies have suggested that the 4G allele of the 4G/5G insertion-deletion polymorphism is associated with higher plasma PAI-1 concentrations. The 4G allele has been associated with an increased risk of myocardial infarction in a meta-analysis. It is likely that single polymorphisms at a candidate gene locus represent only part of the overall genetic variation that may contribute to phenotypic variation. It is unknown whether the 4G allele is simply correlated with other possibly causative alleles and whether additional variation at the PAI-1 locus plays a role in determining PAI-1 levels. One emerging approach is to take advantage of the complete human genome sequence and the increasingly rich collection of single nucle-
otide polymorphisms (SNPs) in public databases to study the role of the estimated 11 million SNP variants (minor allele frequency >1%) in explaining phenotypic variation.

SNP alleles at a locus are often correlated (known as linkage disequilibrium [LD]) and cohered as haplotypes. Most of the genome exists in regions of strong LD, called haplotype blocks, within which most individuals carry 1 or 2 of a few common haplotypes. Within these blocks, a relatively small number of SNPs, termed tag SNPs, can mark common haplotypes and capture most of the genetic diversity in a sample. The utilization of tag SNPs and common haplotypes formed by tag SNPs in genetic association studies has been posited as an efficient and effective method to narrow association signal and localize susceptibility variants.

Our investigation sought to comprehensively examine the role of common SNPs and haplotypes at the PAI-1 locus in determining PAI-1 levels measured in the Framingham Heart Study (FHS), a well-characterized, community-based sample. Thus, our objectives were (1) to define the LD patterns for common genetic variants at the PAI-1 locus, (2) to determine associations of PAI-1 single genetic variants and multimarker haplotypes with plasma PAI-1 levels, and (3) to determine the relation between the 4G/5G polymorphism and other common variation at the PAI-1 locus.

Methods

Study Participants

The FHS offspring cohort began in 1971 with enrollment of 5124 men and women. Approximately every 4 years, participants have undergone a routine medical history and physical examination, as well as laboratory assessment of cardiovascular disease risk factors. The Institutional Review Board at Boston Medical Center approved the study, and all participants gave written informed consent.

Of the 3799 participants who attended the fifth offspring examination (1991–1995), blood testing for PAI-1 antigen was completed on 2837 eligible participants. DNA was available in a panel of 1811 unrelated individuals representing 96 independent chromosomes of European ancestry. Assays were considered successful if they met the following criteria: (1) at least 75% success for genotyping calls, (2) Hardy-Weinberg equilibrium (HWE) P > 0.01, and (3) mendelian transmission errors ≤ 1. In addition, we imposed a minor allele frequency threshold and defined “common” for the present study as a minor allele frequency > 2%. Overall, for 24 SNPs (22 noncoding, 2 nonsynonymous coding: rs6090 and rs6092) and the 4G/5G insertion-deletion polymorphism, we developed successful assays and observed a minor allele frequency > 2%.

LD Structure in Reference Pedigrees:

Identification of Haplotype Blocks and Tag SNPs

Haplotype blocks were defined and tag SNPs were selected with the “spine of LD” setting in the publicly available Haplovie software package version 2.03 (Jeffrey C. Barrett and Mark J. Daly, http://www.broad.mit.edu/personal/jbarrett/haplovie). Briefly, for each pair of markers, the absolute value of D’ (an estimate of the strength of LD) and a logarithm of the odds score (LOD; an estimate of the significance of LD) were calculated. On the basis of these 2 measures, each pairwise marker comparison was categorized into 1 of 3 groups: (1) no or minimal evidence of historical recombination (D’ = 1/LOD > 0.2 or 0.2 ≤ D’ < 1/LOD ≤ 0.5), (2) strong evidence of historical recombination (D’ < 1/LOD < 2 or D’ < 0.5/any LOD), and (3) uninformative (D’ = 1.0/LOD ≤ < 2).

In the spine of LD setting, haplotype blocks are assigned based on each end marker of a block having a D’ > 0.8 with all intervening pairwise marker comparisons excepting 1 comparison. Tag SNPs were selected by ranking SNPs within a block on the basis of successful genotyping percentage and then selecting SNPs one at a time from this ranked list until all haplotypes > 2% frequency within the block were uniquely tagged. With this procedure, 11 tag SNPs were required to mark common haplotypes. For the 20 SNPs that fell into 2 blocks, the 11 tag SNPs captured the 9 unmeasured SNPs with a mean pairwise $r^2$ of 0.78. Eight of the 9 unmeasured SNPs were captured with a pairwise $r^2$ > 0.70. Thus, in reference pedigrees, our tag SNPs captured well the unmeasured SNPs.

Genotyping in the FHS

In the FHS sample, we genotyped 11 tag SNPs, the 4G/5G polymorphism, and 7 SNPs that were redundant in CEPH pedigrees. Redundant SNPs were typed to help assess LD block structure similarity between CEPH and FHS samples (data not shown). Thus, 19 SNPs were genotyped in FHS. The $\chi^2$ test was used to compare observed genotype frequencies with their estimates under HWE. One tag SNP, rs7242, was out of HWE ($P < 0.01$) in the FHS sample and thus was excluded from further analysis. The remaining 18 sequence variants were in HWE ($P > 0.05$).

Statistical Analysis

Because of a skewed distribution, serum PAI-1 antigen levels were logarithmically transformed (natural log). Sex-specific standardized
residuals from multivariable-adjusted PAI-1 levels were calculated with SAS25 and served as the phenotype. On the basis of reported determinants of PAI-1 levels in the literature, covariates included in the multivariable models were age, body mass index, current cigarette smoking, systolic blood pressure, diastolic blood pressure, hypertension treatment, alcohol consumption, total cholesterol, HDL cholesterol, triglycerides, diabetes, prevalent cardiovascular disease, and menopause status and estrogen replacement therapy for women.

Regression analysis was performed with each of the SNPs to test the null hypothesis that the phenotype means did not differ by marker genotype. We assumed a general model of inheritance and used a 2–degrees-of-freedom (df) test for each SNP. Analyses were performed on standardized PAI-1 residuals. To further identify a subset of SNPs that significantly explained variance in PAI-1 levels when adjusted for the effects of other SNPs, we conducted a stepwise selection of the SNPs in multivariable linear regression models.25,26

Association analyses of haplotypes from a single block were conducted with a weighted-regression approach as implemented in the haplo.score program.26,27 All compatible haplotype configurations of a multimarker genotype were used in the regression, with weights being the corresponding posterior likelihood of such a configuration estimated with the expectation-maximization (EM) algorithm.28 An n-1 df score statistic, with n being the total number of haplotypes, tested all haplotypes simultaneously to detect any departure from the null hypothesis of no association. A 1-df haplotype-specific score statistic tested whether trait differences exist between a single haplotype and all other haplotypes combined. Haplotype-specific effect was also estimated to measure the mean difference in phenotype between carriers of 1 or 2 copies of a haplotype compared with those without the haplotype.

In joint analyses of haplotypes from 2 haplotype blocks, a weighted regression model with 2 predictors that represented the haplotypes from each block was performed with SAS.25 The weights in the regression were the posterior likelihood of the joint haplotype configurations of the 2 blocks. The haplotype configurations and the posterior likelihood of each configuration were estimated with a similar EM algorithm implemented in SNPHAP (http://www-gene.cimr.cam.ac.uk/clayton/software/snphap.txt). Multivariable logistic regression analysis was performed to test the null hypothesis that coronary heart disease status did not differ by marker genotype. Covariates in the analyses included age, sex, and traditional cardiovascular risk factors including hypertension (defined by systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or hypertension treatment), diabetes (defined by fasting blood glucose >125 mg/dL or current medication use for diabetes), total cholesterol, HDL cholesterol, lipid-lowering therapy, cigarette smoking (coded as yes if current smoker or quit within 1 year), body mass index, and triglycerides. For all analyses, a nominal P<0.05 was considered significant.

Results

LD Structure at the PAI-1 Locus in Reference Pedigrees

The LD structure at the PAI-1 locus in reference CEPH pedigrees is displayed in the Figure. Two extended segments of LD were evident in the PAI-1 gene region, labeled block 1 and block 2. Within these segments, a small number of common, ancestral haplotypes were observed. Eight haplotypes in block 1 (16-kb genomic segment) accounted for 95% of the chromosomes in reference pedigrees, and likewise, 6 haplotypes in block 2 (13-kb genomic segment) represented 95% of the chromosomes. Six tag SNPs predict common haplotypes in block 1, and an additional 5 tag SNPs predict common haplotypes in block 2.

LD structure at the PAI-1 gene locus in reference pedigrees. A, PAI-1 locus position on chromosome 7 (chr 7) in the human genome July 2003 assembly (hg16). Genes encoded in this genomic region include SERPINE1 and AP1S1 on the plus strand and VGF on the minus strand. Rs identification numbers for SNPs 1 to 25 are listed in sequence order underneath the black bar, which lists the genomic coordinates across the gene. B, Haplotype block structure panel shows that 25 SNPs fell into 2 haplotype blocks. SNPs 1 to 9 fell into block 1, and SNPs 10 to 20 fell into block 2. Each haplotype is shown within a rectangle containing the alleles, a frequency percentage, and a bar proportional to haplotype frequency. For the 4G/5G insertion/deletion polymorphism (ss3172207), “A” refers to the 4G allele and “C” refers to the 5G allele. Within block 1, 8 common haplotypes were present, and within block 2, there were 6 common haplotypes. This limited haplotype diversity within each block illustrates the strength of LD across this region. C, 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. Figure prepared with LocusView 2.0 (T. Petryshen, et al, Broad Institute; available at http://www.broad.mit.edu/mpg/locusview/).

FHS Participants

Characteristics of the study sample (mean age 55 years; 54% women) are shown in Table 1. Unadjusted mean PAI-1 levels were higher in men than women.

Single Allelic Variants and PAI-1 Levels in Framingham Participants

Results of association analyses of each of 18 genetic variants with multivariable-adjusted plasma PAI-1 levels in FHS
participants are shown in Table 2. SNP rs2227631 was significantly associated with PAI-1 levels ($P<0.0001$). The 4G/5G polymorphism, located 172 bases downstream of rs2227631, was associated as strongly as rs2227631 ($P=0.0001$). SNP rs2227631 is in tight LD with the 4G/5G polymorphism ($D^2=0.97$, $r^2=0.78$). Additionally, 9 other variants were also associated with PAI-1 levels (each $P<0.05$). Similar results were seen with more basic regression models, including a model that used only age-adjusted, sex-specific plasma PAI-1 levels as the phenotype.

**Overall Contribution of Individual Genetic Variants in Stepwise Models**

To distinguish between genetic variants that are strongly correlated with each other and those that independently contribute to the variation in PAI-1 levels, we conducted stepwise analysis that included all 18 genetic variants. In stepwise selection, the 4G/5G polymorphism explained 2.5% of the residual variance in multivariable-adjusted PAI-1 levels (stepwise $P<0.0001$), with the 4G allele associated with higher PAI-1 levels. An additional 1% was explained by rs6465787 (stepwise $P=0.02$) and a further 1% by rs2227674 (stepwise $P=0.03$). When we repeated stepwise selection forcing in rs2227631, the variant in tight LD with the 4G/5G polymorphism, 3 variants were significantly associated with PAI-1 levels—rs2227631, rs6465787, and rs2227674—and the 4G/5G polymorphism was no longer significant.

Mean unadjusted PAI-1 levels by genotype for 4 variants associated in stepwise models are presented in Table 3. Association analyses were conducted with multivariable-adjusted log-transformed PAI-1 level, as noted in Methods.

**Associations of Haplotypes With PAI-1 Levels**

We inferred haplotype frequencies in the FHS sample via an EM algorithm using 7 SNPs (6 tag SNPs and the 4G/5G polymorphism) in block 1 and 4 tag SNPs in block 2 (the fifth tag SNP in block 2—rs7424—was eliminated owing to failure to achieve HWE). The resulting haplotypes in the larger FHS sample were very similar to those initially predicted in reference CEPH pedigrees (Tables 4 and 5; Figure).

There were significant overall differences among the 8 common haplotypes in block 1 with respect to mean PAI-1 levels (global $P=0.0001$). When we compared individual haplotypes with all the other haplotypes combined, Hap 1A and Hap II were positively associated with PAI-1 levels (haplotype-specific $P=0.00001$ and $P=0.05$, respectively; Table 4). Hap 1A harbors the 4G allele, and Hap II harbors...
the minor T allele of rs6465787; thus, the single SNP findings from stepwise selection (Table 3) and the haplotype findings represent the same result.

Association analyses for haplotypes in block 2 are displayed in Table 5. The 5 common haplotypes differed with respect to mean PAI-1 (global \( P = 0.04 \)). A single common haplotype in block 2, Hap 2A, was positively associated with plasma PAI-1 (haplotype-specific \( P = 0.006 \)). Hap 2E was associated with lower PAI-1 levels (Hap \( P = 0.05 \)). Hap 2E harbors the minor allele of rs2227674; thus, the single SNP rs2227674 finding from stepwise selection (Table 3) and the Hap 2E finding represent the same result.

Hap 1A was associated with higher PAI-1 levels, as was Hap 2A. The haplotypes in block 1 and block 2 were correlated (multiallelic \( D^* = 0.74 \); most individuals with Hap 2A also carry Hap 1A. To address whether the association signal could be narrowed to Hap 1A, we considered both Hap 1A and Hap 2A jointly in a model. In this analysis, Hap 1A remained significantly associated with plasma PAI-1 level (\( P < 0.0001 \)), whereas Hap 2A was no longer significant (\( P = 0.32 \)).

**Joint Consideration of SNPs and Hap 1A**

We considered the effects of SNPs and Hap 1A jointly by constructing models with the dependent variable being multivariable-adjusted plasma PAI-1 level and the 2 independent variables being Hap 1A and an individual SNP. Eighteen separate models were constructed, one for each of the 18 SNPs tested at the locus. When both Hap 1A and SNPs were considered, Hap 1A was significantly related to plasma PAI-1 in all of the models except when combined with SNP rs2227631 or 4G/5G (data not shown). In a model with 4G/5G and Hap 1A, 4G/5G was significantly associated with PAI-1 level, whereas Hap 1A was not. Similarly, in a model with rs2227631 and Hap1A, rs2227631 was associated with PAI-1 level, whereas Hap 1A was not. These results imply that these 2 SNP effects are stronger than the Hap 1A effect.

**Genetic Variants and Coronary Heart Disease**

We evaluated the relations between prevalent coronary heart disease (at the time of the FHS Offspring Study examination cycle 5) and the 2 variants most significantly associated with PAI-1 level. SNP rs2227631 was not associated with prevalent coronary heart disease (\( n = 114 \) for coronary heart disease; \( P = 0.41 \) in an age- and sex-adjusted model, and \( P = 0.59 \) in a multivariable-adjusted model). The 4G/5G polymorphism also was not associated with prevalent coronary heart disease (\( n = 111 \) for coronary heart disease; \( P = 0.71 \) in an age- and sex-adjusted model, and \( P = 0.88 \) in a multivariable-adjusted model). Compared with the 5G/5G genotype as the referent, the 4G/4G genotype OR was 1.16 (95% CI 0.61 to 2.21).

**Discussion**

**Principal Findings**

In reference pedigrees, we characterized LD structure across a 50-kb genomic segment spanning the PAI-1 locus using a

---

**TABLE 4.** PAI-1 Haplotypes in Block 1 and Relations to Multivariable-Adjusted Plasma PAI-1 Levels in FHS

<table>
<thead>
<tr>
<th>Hap</th>
<th>rs757722</th>
<th>rs6959121</th>
<th>rs4729662</th>
<th>rs6465787</th>
<th>rs2227631</th>
<th>4G/5G</th>
<th>Hap Frequency</th>
<th>Hap Effect*</th>
<th>Hap P</th>
<th>Global P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hap 1A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>4G</td>
<td>0.50</td>
<td>0.21</td>
<td>0.00001</td>
</tr>
<tr>
<td>Hap 1B</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>5G</td>
<td>G</td>
<td>0.17</td>
<td>-0.21</td>
<td>0.0008</td>
</tr>
<tr>
<td>Hap 1C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>5G</td>
<td>0.10</td>
<td>-0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>Hap 1D</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>5G</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Hap 1E</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>5G</td>
<td>G</td>
<td>0.05</td>
<td>-0.07</td>
<td>0.51</td>
</tr>
<tr>
<td>Hap 1F</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>5G</td>
<td>A</td>
<td>0.03</td>
<td>-0.09</td>
<td>0.53</td>
</tr>
<tr>
<td>Hap 1G</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>4G</td>
<td>G</td>
<td>0.02</td>
<td>0.15</td>
<td>0.34</td>
</tr>
<tr>
<td>Hap 1H</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>5G</td>
<td>G</td>
<td>0.02</td>
<td>-0.43</td>
<td>0.02</td>
</tr>
<tr>
<td>Hap 1I</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>5G</td>
<td>G</td>
<td>0.01</td>
<td>0.43</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Hap indicates haplotype.
Haplotypes of \( \geq 1\% \) in frequency are displayed.
Because of missing genotype data, overall \( n = 877 \).
*Hap Effect refers to the percentage of 1-SD change in multivariable-adjusted PAI-1 residual per copy of haplotype.

**TABLE 5.** PAI-1 Haplotypes in Block 2 and Relations to Multivariable-Adjusted Plasma PAI-1 Levels in FHS

<table>
<thead>
<tr>
<th>Hap</th>
<th>rs2227660</th>
<th>rs2227674</th>
<th>rs2227692</th>
<th>rs2070683</th>
<th>Hap Frequency</th>
<th>Hap Effect*</th>
<th>Hap P</th>
<th>Global P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hap 2A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>0.43</td>
<td>0.11</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Hap 2B</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>0.22</td>
<td>-0.09</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Hap 2C</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>0.14</td>
<td>0.01</td>
<td>0.88</td>
<td>0.04</td>
</tr>
<tr>
<td>Hap 2D</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>0.13</td>
<td>0.00</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>Hap 2E</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>0.08</td>
<td>-0.15</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

Hap indicates haplotype.
Haplotypes of \( \geq 1\% \) in frequency are displayed.
Because of missing genotype data, overall \( n = 1149 \).
*Hap Effect refers to the percentage of 1-SD change in multivariable-adjusted PAI-1 residual per copy of haplotype.
dense SNP map and found 2 blocks of sequence variants in strong LD. In association analyses of single genetic variants with PAI-1 levels, we identified 2 genetic variants, rs2227631 and the 4G/5G polymorphism, which are in tight LD and each strongly associated with plasma PAI-1 levels. After we accounted for either rs2227631 or the 4G/5G polymorphism, 2 additional variants further explained the variation in PAI-1 levels. After we accounted for clinical covariates, 3 sequence variants, in sum, explained ~5% of the residual variance in circulating PAI-1 levels. In haplotype-phenotype association, a single common haplotype, Hap 1A (50% frequency), was strongly associated with increased PAI-1 level. Hap 1A harbors the minor alleles of both rs2227631 and the 4G/5G polymorphism.

These findings confirm prior reports that the 4G allele is associated with increased plasma PAI-1 levels.29 The percent of residual variance in the multivariable-adjusted PAI-1 level explained by the 4G/5G polymorphism, which was 2.5% in the present study, is consistent with prior reports, in which it ranged from 0.63%30 to 1.6%.31

With the genotyping of 25 SNPs in reference pedigrees and 18 PAI-1 SNPs in our association study, the present findings extend prior work by comprehensively studying the association of common genetic variants with PAI-1 levels in a large community-based sample. Our work directly addresses 2 questions: (1) Can the contribution of the 4G variant to PAI-1 levels be distinguished from LD with other possibly causative alleles? and (2) Are there additional allelic contributors at the PAI-1 locus to variance in PAI-1 levels?

Common Genetic Variation at the PAI-1 Locus and PAI-1 Level
The LD structure defined in the present study with publicly available common SNPs is in good agreement with the haplotypes defined by resequencing at the PAI-1 locus.32 In the Seattle SNPs Program for Genomics Applications project, 23 individuals of European ancestry were resequenced for ~13 kb spanning the PAI-1 gene.32 This resequencing effort and the FHS sample yielded similar Hap 1A frequencies, 46% and 50%, respectively. In addition, the percent of individuals discordant for the minor alleles of 4G/5G and rs2227631 was similar at 2% in the Seattle SNPs European-descent sample and 5% in the FHS cohort.

Using an LD-based approach, we narrowed the association signal at the PAI-1 locus to a specific region of limited haploptic diversity in the 5’ end of the gene. The 2 strongest known causal candidates present on Hap 1A are rs2227631 and the 4G/5G polymorphism. We are unable to distinguish the possible effect of rs2227631 and the 4G/5G polymorphism. In general, functional studies may help differentiate the effects of genetic variants. However, whereas previous functional studies have evaluated the 4G/5G polymorphism,8 an evaluation of the function of rs2227631 is lacking.

Besides rs2227631 and the 4G/5G polymorphism, 2 additional variants may contribute to the variance in PAI-1 level. SNP rs6465787 is located 3 kb upstream of the 4G/5G polymorphism, and rs2227674 is located in intron 4. Both rs6465787 (minor allele frequency 0.02) and rs2227674 (minor allele frequency 0.21) are less common than the 4G/5G polymorphism. To the best of our knowledge, the present report is the first to describe the contribution of either rs6465787 or rs2227674 to variation in PAI-1 levels, and these findings will need to be replicated.

The variants identified in sum explain ~5% of the residual variability in multivariable-adjusted plasma PAI-1 level. The discovery of genetic variants that explain interindividual variation in biomarker phenotypes is important for several reasons. First, PAI-1 biomarker levels may predict incident clinical disease, and thus, genetic variants related to biomarker levels are strong candidate alleles to test for association with disease. Second, alleles that have been shown to be causally related to PAI-1 levels may be appropriate targets for drugs to alter gene expression. Third, a constellation of “risk” alleles may aid in predicting incident clinical disease. With complex traits such as biomarker phenotypes, the expected effect of any specific allele is expected to be modest, and our data are consistent with this expectation.

Implications of a Comprehensive Approach for Candidate Gene Association Studies
The limitations of single SNP association studies have recently been highlighted, because there is substantial difficulty in interpreting both a negative and positive result.14 Thus, it has been proposed that candidate gene association studies move to a staged approach that involves 2 steps: (1) to comprehensively define common patterns of SNP variation at a locus through a study of local LD and (2) to screen for association signal with a subset of nonredundant tag SNPs at the locus.17 With the present study, we have demonstrated that such an approach can effectively localize common susceptibility alleles. In fact, using an unbiased approach with publicly available SNPs, the implicated susceptibility haplotype would have been defined without knowledge of the 4G/5G polymorphism.

Study Limitations and Strengths
Our study has several potential limitations. First, in our definition of LD at the PAI-1 locus and the subsequent selection of tag SNPs, we restricted our focus mainly to common genetic variants. Multiple rare variants may influence a trait, and these variants may be identified by resequencing.33 However, the frequency spectrum of susceptibility variants for complex phenotypes is likely to include common variants, and our approach is most appropriate for the discovery of such variants.17

Second, nonexonic conserved regions, which show conservation comparable to exons, have been postulated to be important in the regulation of gene expression.34 Variation in these nonexonic conserved regions around the PAI-1 locus may influence PAI-1 level. We did not explicitly examine such regions in the present study, although common variation in these regions may have been captured by LD.

Third, in genomic regions of high LD, the optimal methodology to select a subset of nonredundant markers has yet to be defined. We are currently comparing various methods of tag SNP selection, and alternative methods may prove to be more efficient than the one used in the present report.35–37
Fourth, as noted previously, the extent of LD in block 1 may extend upstream of the first rs757722. The International HapMap project has been designed to comprehensively catalog patterns of LD across the human genome, and we reviewed this SNP catalog for the 100-kb region encompassing the PAI-1 locus. We found that LD breaks down ~700 bases further upstream of rs757722, although the present HapMap SNP density in this genomic region is insufficient to be certain about this conclusion. Thus, there remains the possibility that yet unidentified variants upstream of rs757722 are the true causal variants on Hap 1A.

Fifth, our sample was predominantly white, which limits the generalizability of our results to other ethnic groups. Sixth, given that a large systematic review found that the 4G/4G genotype was associated with a modest 1.2-fold increased risk of myocardial infarction, we were underpowered to detect a small effect of this magnitude.10 Finally, type I error may occur when testing for association between a phenotype and multiple genetic variants. We have not accounted for multiple testing in the present analyses. In genetic association studies, the Bonferroni correction may be overly conservative owing to a high degree of correlation among the tests performed.39 However, in the present study, the association between Hap 1A, 4G/5G polymorphism, and PAI-1 levels would have survived even a Bonferroni correction.

Strengths of the present investigation include the comprehensive assessment of common genetic variation at the PAI-1 locus, the use of single allelic variants and haplotypes in association analyses, the large sample size, and the use of multivariable analyses.

Conclusions

In summary, using a powerful approach that is now possible from knowledge of the human genome sequence, we have comprehensively defined the role of common genetic variation at the PAI-1 locus in determining plasma PAI-1 levels. Our results suggest that approaches to define LD structure in candidate gene regions followed by conduct of association analyses with SNPs and multimarker haplotypes will be effective for localizing common susceptibility alleles.

Acknowledgments

This work was supported through National Institutes of Health/National Heart, Lung, and Blood Institute contract N01-HC-25195 and the Cardiogenomics Programs for Genomic Applications (HL66582). Dr Kathiresan is supported by the American College of Cardiology Foundation/Merck Adult Cardiology Research Fellowship Award and a Young Investigator Award from the GlaxoSmithKline Research & Education Foundation for Cardiovascular Disease. The authors thank Dr Christopher Newton-Cheh for his critical review of the manuscript.

References

Circulating blood fibrinolytic factors such as the plasminogen activator inhibitor 1 (PAI-1) have been linked to risk for coronary heart disease. Previous studies identified the genetic determinants of circulating PAI-1 by examining limited numbers of polymorphisms, such as the PAI-1 4G/5G polymorphism. Completion of the human genome sequence now allows comprehensive study of the role of common genetic variation across gene regions in determining the variability in circulating proteins. We characterized patterns of common DNA sequence variation in the PAI-1 gene region and their associations with circulating PAI-1 levels. We discovered several single nucleotide polymorphisms (SNPs) in the region of the gene encoding PAI-1 that are highly significantly associated with the level of circulating PAI-1. Several polymorphisms are highly correlated (are in “linkage disequilibrium”), including the strongly associated promoter region SNP rs2227631 and 4G/5G, and are inherited together in segments of DNA, or haplotypes, which makes it difficult to distinguish which of these 2 polymorphisms is responsible for the association with PAI-1 levels. SNP rs2227631, as well as 2 other SNPs, account for \( \approx 5\% \) of the residual variance of circulating PAI-1. These highly statistically significant findings suggest that comprehensive approaches that go beyond single polymorphisms will be needed to fully characterize the impact of genetic variation on circulating blood proteins and clinical cardiovascular disease. The specific PAI-1 SNPs identified in our study are of interest for future disease association studies to assess whether these polymorphisms alone or in combination may aid in prediction of clinical disease risk and efficacy of therapeutic intervention.

**CLINICAL PERSPECTIVE**

Circulating blood fibrinolytic factors such as the plasminogen activator inhibitor 1 (PAI-1) have been linked to risk for coronary heart disease. Previous studies identified the genetic determinants of circulating PAI-1 by examining limited numbers of polymorphisms, such as the PAI-1 4G/5G polymorphism. Completion of the human genome sequence now allows comprehensive study of the role of common genetic variation across gene regions in determining the variability in circulating proteins. We characterized patterns of common DNA sequence variation in the PAI-1 gene region and their associations with circulating PAI-1 levels. We discovered several single nucleotide polymorphisms (SNPs) in the region of the gene encoding PAI-1 that are highly significantly associated with the level of circulating PAI-1. Several polymorphisms are highly correlated (are in “linkage disequilibrium”), including the strongly associated promoter region SNP rs2227631 and 4G/5G, and are inherited together in segments of DNA, or haplotypes, which makes it difficult to distinguish which of these 2 polymorphisms is responsible for the association with PAI-1 levels. SNP rs2227631, as well as 2 other SNPs, account for \( \approx 5\% \) of the residual variance of circulating PAI-1. These highly statistically significant findings suggest that comprehensive approaches that go beyond single polymorphisms will be needed to fully characterize the impact of genetic variation on circulating blood proteins and clinical cardiovascular disease. The specific PAI-1 SNPs identified in our study are of interest for future disease association studies to assess whether these polymorphisms alone or in combination may aid in prediction of clinical disease risk and efficacy of therapeutic intervention.
Comprehensive Survey of Common Genetic Variation at the Plasminogen Activator Inhibitor-1 Locus and Relations to Circulating Plasminogen Activator Inhibitor-1 Levels

_Circulation_. 2005;112:1728-1735
doi: 10.1161/CIRCULATIONAHA.105.547836
_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 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/112/12/1728

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/