Association study of CRP gene polymorphisms with serum CRP level and cardiovascular risk in the NHLBI Family Heart Study

Qingwei Wang,1 Steven C. Hunt,3 Qin Xu, Yuqing E. Chen, Michael A. Province, John H. Eckfeldt, James S. Pankow, and Qing Song

1Cardiovascular Research Institute and 2Clinical Research Center, Morehouse School of Medicine, Atlanta, Georgia; 3Cardiovascular Genetics Division, University of Utah, Salt Lake City, Utah; 4Division of Biostatistics, Washington University, St. Louis, Missouri; 5Department of Laboratory Medicine and Pathology and 6Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota; and 7Cardiovascular Center, University of Michigan Medical Center, Ann Arbor, Michigan

Submitted 3 November 2005; accepted in final form 25 May 2006

Recent epidemiological studies have demonstrated that plasma C-reactive protein (CRP) levels were associated with ischemic cardiovascular/cerebrovascular disease among apparently healthy individuals (3, 9, 11, 19, 27). CRP is primarily synthesized and secreted into the circulation from the liver. The level of CRP in the circulation can increase rapidly and dramatically in response to cytokines, including IL-6 and TNF-α (25). This response occurs at the translational level.

CRP levels demonstrate substantial interindividual variability. Family studies have demonstrated that CRP levels are substantially (~30–50%) determined by heritable factors (2, 10, 20, 24, 26, 40). The combination of sociodemographic, behavioral, and lifestyle factors, obesity and fat patterning, and prevalent diabetes explained 13–30% of the interindividual variability of these traits (24). Although several genetic polymorphisms in the CRP gene have been documented to be associated with CRP levels (5, 7, 12, 18, 28, 34, 35, 41), whether these local genetic polymorphisms on the CRP gene are associated with cardiovascular traits remains controversial (8, 17, 41).

The 5′- and 3′-flanking regions usually make significant contributions to the control of the quantitative amount of gene expression. A transgenic mouse model has demonstrated that 5′- and 3′-flanking regions are involved in the regulation of CRP gene expression (38). In the present study, we systematically screened the 5′- and 3′-flanking regions of the CRP gene for novel genetic polymorphisms and investigated their associations with plasma CRP levels and cardiovascular risk in the National Heart, Lung, and Blood Institute (NHLBI) Family Heart Study population.

Genetic studies can provide insight into causation, because they are less inclined to be affected by confounding and reverse causation. CRP genotypes precede physiological and pathological states, such as a low-grade chronic inflammatory condition; therefore, investigation of a direct association of CRP genotypes with cardiovascular traits may help elucidate the role of CRP elevation in development of cardiovascular disease.

METHODS

Study population. For the NHLBI Family Heart Study, probands were selected from four population-based cohorts: the Atherosclerosis Risk in Communities Study (Forsyth County, NC, and suburban centers in Minneapolis, MN), the Utah Health Family Tree Study (Salt Lake City, UT), and the Framingham Heart Study (Framingham, MA). A total of 5,381 subjects from 1,245 two- to three-generation families [657 individuals at high risk for coronary heart disease (CHD) and 588 randomly selected individuals] completed an extensive clinical examination in 1994–95. A complete description of the study can be found elsewhere (15). For the present study, analysis was limited to Caucasian subjects in whom CRP levels and other covariates were measured (n = 1,296), as described elsewhere (24).

CRP had not been measured in the African-American subjects at the time of this study. Serum CRP was measured at the Laboratory for Clinical Biochemistry Research, University of Vermont by a sensitive enzyme-linked immunosorbent assay method calibrated with World Health Organization reference material (13, 24). The intra- and interassay coefficients of variation were 3% and 6%, respectively. Alcohol consumption and smoking status were coded as current or not current.
Prevalent CHD \( (n = 174) \) was defined as medically recorded validation of reported history of myocardial infarction, coronary bypass surgery, or percutaneous transluminal coronary angioplasty. Intima-media thickness (IMT) was measured by ultrasound at the common carotid and internal carotid arteries and at the bifurcation of the carotid arteries. The mean wall thickness in the 1-cm segment of the right and left common carotid artery proximal to the dilatation of the carotid bulb was estimated. Table 1 shows the baseline characteristics of the study population. The study was approved by an institutional review committee at each site, and the subjects gave informed consent.

**Experimental approach of novel single-nucleotide polymorphism discovery.** This project was focused on the 5′- and 3′-flanking regions, which are complementary to the target regions of the Seattle Single-Nucleotide Polymorphism (SNP) Project with a \( \sim 1,000 \text{-bp overlap} \) (Fig. 1). A systematic search for novel genetic variants in the CRP gene was performed by denaturing high-performance liquid chromatography with the WAVE system (Transgenomic) on DNA samples from Coriell Repositories (African-American AA-100 and European-American Cau-100 panels, \( n = 46 \) unrelated subjects for each). PCR primers for overlapping amplicons were designed to cover the entire targeted regions. Genomic DNA (10 ng) was amplified by PCR using AmpliTaq Gold (Applied Biosystems), denatured at 94°C for 5 min, and slowly reannealed at room temperature for heteroduplex formation. Samples were then analyzed on the WAVE system, with buffer gradient and oven temperature controlled by WAVEmaker software. Heterozygous amplicons were identified by elution profiles. For each type of elution profile, three representative heterozygous samples were further analyzed on an automatic sequencer (model 3100, Applied Biosystems) in forward and reverse directions. Sequence trace files were analyzed using Mutation Surveyor software (SoftGenetics) to locate nucleotide variants and verified by manual inspections. The SNPs were designated by their positions relative to the transcription initiation site for SNPs in the 5′-flanking region or by their positions relative to the last nucleotide of the last CRP exon (exon 2) for SNPs in the 3′-flanking region.

**VISTA analysis.** VISTA genome sequence global alignment tools (21) were used to locate the evolutionarily conserved regions of the CRP gene. The genomic DNA sequences of human [University of California Santa Cruz (UCSC) Human Genome Assembly, May 2004; chromosome (Chr) 1: 156,559,891..156,366,619], mouse (UCSC Mouse Genome Assembly, May 2004; Chr 1: 172,771,763..172,843,629), and rat (UCSC Rat Genome Assembly, June 2003; Chr 13: 88,841,127..88,912,037) were aligned with a threshold of 75% identity in a 100-bp window.

**Genotype determination.** Genotyping was carried out on a pyrosequencer (model PSQ96MA, Pyrosequencing, Uppsala, Sweden) following the manufacturer’s protocol. Genotyping quality control was achieved by 1) PedCheck software (23) to identify misinheritances among the family pedigrees, 2) regenotyping of 12% total assays, 3) complete linkage disequilibrium (LD) among neighboring SNPs \( \sim 4944 \) and \( \sim 224, \sim 757 \) and \( \sim 3679 \), and 4) Hardy-Weinberg equilibrium (HWE).
Tag SNP selection and haplotype estimation. In combination with the SNPs in current databases (Seattle SNP, dbSNP, and Celera), there were 84 polymorphisms in the CRP gene, 21 of which were common [minimum allele frequency (MAF) > 0.05] in Caucasians. To select the tag SNPs for further association analysis, we first genotyped these 21 SNPs in a subset of 96 individuals. We then used EMLD software to analyze the pairwise correlations among these SNPs. Tag SNPs were selected by the criterion of complete LD ($r^2 = 1$). Haplotypes were estimated from the six SNPs using PHASE (version 2.0) software (32, 33). There were five common haplotypes with frequencies $\geq 5\%$, resulting in 10 diplotypes with frequencies $\geq 4\%$.

**Statistical analysis.** Phenotypic data were analyzed using generalized estimating equations with an exchangeable correlation matrix in a general linear model (GENMOD, SAS) (31). Family number was considered a random variable in the model; therefore, the intraclass correlations could be used to adjust the standard error estimates for the familial dependencies among the subjects. Score statistics from GENMOD were tested by $\chi^2$ analysis to test for the significance of the diplotypes or genotypes. Covariates included in each model were age, gender, field center, body mass index, current smoker, and current alcohol intake. These covariates were selected because they were significantly associated with CRP in the majority of the regression analyses. The diplotypes or individual SNP genotypes were the independent variables, and CRP and IMT were the dependent variables. CHD prevalence was tested for association with each SNP or the haplotype pairs by the same model but with logit transformation of the dependent variable. The diplotypes were tested as genotypes for association before individual SNP testing. Only 10 of the 15 diplotypes with frequencies $\geq 4\%$ were analyzed; the rarer diplotypes were set to missing. For dependent variables that showed significance in global diplotype tests ($P < 0.05$), individual SNPs were tested using an $F$ test with two degrees of freedom on the three genotype means to indicate which SNP(s) may be responsible for the association. Before the association of SNP $-757$ with the dependent variables was tested, subjects with the C/C genotype were combined with subjects with the C/T phenotype, because only seven subjects had the C/C phenotype. Because of skewness, CRP was logarithmically transformed before the analysis.

**RESULTS**

**Novel SNP identification, tag SNP selection, and haplotype configuration.** We identified 56 SNPs and 4 insertion/deletion polymorphisms in our experimental approach. Among these 60 polymorphisms, 20 SNPs have been reported in public and private databases (dbSNP, Celera, NHLBI Program for Genomic Applications, and Seattle SNP) or publications and 40 polymorphisms were novel. The combination of our new polymorphisms with the current database SNPs resulted in 84 polymorphisms in the CRP gene; among these, 37 polymorphisms had MAF $> 0.05$, including 1 Caucasian-unique SNP, 15 African-American-unique SNPs, and 21 SNPs common to both populations.

Pairwise correlations between SNPs were analyzed on genotype data from a subset of 96 unrelated individuals on 21 SNPs. Except for one SNP (+7598), which is located at the 3’ end of this region and was not in LD with any other SNPs, the other 20 SNPs belong to 4 complete-LD groups ($r^2 = 1$; Table 2). On the basis of this information, we selected one SNP from each cluster [−7180 (rs1341665), −757 (rs3093059), −717 (rs2794521), +224 (rs11130864), and +7598 (rs876583)] for tag SNPs. A triallelic SNP, −286 (rs3091244), is in complete LD with SNP −757 (rs3093059), and alleles C and T of SNP −286 are derived from allele T of SNP −757. This triallelic SNP was also included as a tag SNP for genotyping on the full set of Family Study subjects. Two redundant SNPs [−4944 (rs3122012) and +3679 (rs3093077), in complete LD with +224 and −757, respectively] were also included as genotyping quality controls. In total, eight SNPs (−7180, −4944, −757, −717, −286, +224, +3679, and +7598) were genotyped in the entire sample set for association analysis. This tag SNP set ensures that the entire set of common SNPs on the CRP gene locus is represented in this study without loss of genetic information.

The allele and estimated haplotype frequencies are shown in Table 3. Three haplotypes (0.354, 0.311, and 0.185) are very common, and two additional haplotypes (0.074 and 0.062) have estimated frequency $> 0.05$.

**Genotyping.** Among the eight SNPs that we genotyped in the entire sample set, SNP −4944 and +224 (in complete LD with −4944) were not in HWE ($P = 0.024$). All other SNPs were in HWE. Deviation from HWE is usually used as an accepted test for genotyping error. However, we independently genotyped two SNPs (−4944 and +224) that are in complete LD ($r^2 = 1$). In addition, the genotyping error was also monitored by misinheritances among the family pedigrees and regenotyping (0.17% error rate among 12% total regenotyping samples). It is unlikely that genotyping error is the cause of the deviation of these two SNPs from HWE.

**Association analyses.** $P$ values for the test of association of the diplotypes with each of the study variables are shown in Table 4. A trend toward significant association ($P = 0.045$) of CRP diplotype with serum CRP, which explained 1.4% of the CRP phenotypic variance, prompted the single-locus analysis that followed. The highest CRP mean was 50% higher than the lowest diplotype mean (Table 5). Approximately 18% of the sample had diplotypes with mean CRP $> 3.0$ mg/L. CRP diplotype was not associated with IMT at any site or with CHD. Because of the CRP diplotype results, individual SNPs were tested to determine whether one or more SNPs explained the majority of the association signal (Table 6). SNP −757 (rs3093059) was highly associated with serum CRP. The additional separation of the −757 T allele into two alleles at the SNP −286 (rs3091244) locus seemed to further increase the association significance, despite the increase in degrees of freedom of the test. These associations remained significant at $P < 0.05$ after Bonferroni’s adjustment for multiple testing of the six SNPs in the diplotype. SNP −7180 (rs1341665) showed a borderline association ($P = 0.06$) with serum CRP, but the association was not borderline after multiple-comparison adjustment.

There was a suggestion of an association of IMT of the common carotid artery ($P = 0.016$) for SNP −7180. However, after multiple-comparison adjustment for these secondary hypothesis tests, the associations did not remain significant. None

---

**Table 2. Complete LD clusters of CRP SNPs in Caucasians**

<table>
<thead>
<tr>
<th>CRP Cluster</th>
<th>SNP Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>−7180 −5009 −2609 +1082 +4499 +7304</td>
</tr>
<tr>
<td>II</td>
<td>−4944 −2114 −2104 +224 intron 62633</td>
</tr>
<tr>
<td>III</td>
<td>−757 +1951 +3402 +3679 +7095 +7144 −286</td>
</tr>
<tr>
<td>IV</td>
<td>−717 +2447</td>
</tr>
<tr>
<td>V</td>
<td>+7598</td>
</tr>
</tbody>
</table>

LD, linkage disequilibrium ($r^2 = 1$); SNP, single-nucleotide polymorphism.
of CRP SNPs were associated with IMT of the internal carotid artery or IMT at the bifurcation of the carotid arteries (data not shown). The association of SNP −7180 was not associated with CHD prevalence \((P = 0.12)\).

**Evolutionary conservation by VISTA analysis.** VISTA analysis revealed four highly conserved regions within our SNP discovery target area (excluding the CRP coding regions): 1) from the transcription starting point to −300 bp, 2) at −7 kb in the 5′-flanking region, 3) at 2 kb downstream from the last CRP exon in the 3′-flanking region, and 4) 7–9 kb downstream from the last CRP exon and immediately upstream from the CRP pseudogene (Fig. 1).

**DISCUSSION**

In this study, we used a complete SNP map on the CRP gene (7,180 bp in the 5′-flanking region, exons, and intron and 7,598 bp in the 3′-flanking region) to systematically investigate the associations between CRP genotypes, serum CRP level, and cardiovascular phenotypes. We found that CRP diplotypes and genotypes were associated with serum CRP levels in the NHLBI Family Heart Study cohort. However, CRP genotypes were not associated with IMT or prevalence of CHD.

Several genetic polymorphisms on the CRP gene have been documented to be associated with CRP levels (5, 7, 17, 18, 28, 34, 35, 41). These CRP polymorphisms included the triallelic SNP −286 (rs3091244, also called 1444 in the literature) in exon 2, and SNP +224 (rs1130864, also called +1444 in the literature) and SNP +1082 (rs1205) in the 3′-flanking region. Conversely, a lack of association with serum CRP levels was also documented for SNP −286 (SNP 1059G>C (1, 5, 18), SNP +224 (28), SNP +1082 (8), and SNP −717 (5, 8, 18). In our study, SNP −757 and SNP −286 were significantly associated with serum CRP levels, and SNP −7180 had a borderline association with serum CRP levels. Because of the complete LD relation \((r^2 = 1)\) among SNPs in our cohort, our results confirmed the associations of SNP −286 (rs3091244) and possibly −7180 (rs1341665, in LD with SNP +1082 (rs1205)) with serum CRP levels. In addition, SNP −757 (rs3093059), which represents a set of biallelic SNPs that are in complete LD with triallelic SNP 286, was added to the list of SNPs that are associated with serum CRP levels. We also confirmed the lack of association of SNP −717 by 10.220.33.6 on March 31, 2017 http://ajpheart.physiology.org/ Downloaded from
(rs2794521) with serum CRP levels. The intronic SNP –62633 (rs1417938) and SNP +224 (rs1130864) were in complete LD in our cohort and were not significantly associated with serum CRP levels. These results demonstrated that serum CRP levels could be influenced by local genetic polymorphisms in the CRP gene.

The above-mentioned SNPs were inconsistently associated with serum CRP levels in different studies. Besides spurious associations, one reason for this discrepancy may be the distinct LD patterns among different populations. It is postulated that the LD pattern is a function of the ethnohistory of populations or subpopulations. It is likely that the tag SNPs used in association studies were not the functional SNPs; instead, they might be in LD, to a certain degree, with unidentified functional polymorphisms in their vicinity. Because the variation in CRP levels explained by the CRP gene is very small, the signal could be easily missed, especially in small samples of subjects; this could be one reason for the inconsistent findings among studies.

Our study did not provide direct evidence about which SNPs are functional in regulation of CRP expression. However, some suggestions were provided. The VISTA tool was built on a widely accepted notion that the highly evolutionary conserved genomic regions among species are likely to be functionally important, and, vice versa, functional constraints may have left footprints on nucleotide sequence by making some genomic regions evolve more slowly than average after species divergence. Therefore, the SNPs appearing on a highly conserved region are more likely to have a functional role. For example, SNP –7180 had borderline associations with serum CRP levels and IMT at the common carotid artery. Among the SNPs that are in complete LD with –7180, only –7180 and +7304 reside in highly conserved regions (Fig. 1). An early transgenic study showed that the 3′-flanking sequence of the CRP gene plays an important role in CRP gene expression (38). SNP –286 is also in the highly conserved peak, as revealed by VISTA analysis. Recent functional analysis in cell culture suggested that this triallelic SNP is a functional one (7, 36). It binds to the upstream stimulatory factor-1 in vitro, and different alleles may have different binding affinities (7, 36).

IMT of carotid arteries is a predictor of coronary artery disease. It has been previously reported that increased CRP levels are associated with the angiographically documented coronary atherosclerotic disease in hypo-α-lipoproteinemia patients (30) and the carotid IMT in elderly individuals (39). In our study, CRP diplotype was not related to IMT of the common carotid artery. In a single SNP analysis, SNP –7180 showed an association with IMT at the common carotid artery, which became nonsignificant after multiple-testing adjustment. Previous reports showed that plasma CRP levels were associated with IMT at the common carotid artery (4, 6, 14, 16, 22, 29, 37) but were not associated after adjustment for age, body mass index, visceral fat, and insulin (4, 13, 29, 37). Given the small percentage of CRP variation explained by CRP genetic polymorphisms, one would not expect these polymorphisms to explain a significant proportion of IMT or CHD.

In conclusion, in this large cross-sectional family-based study using the NHLBI Family Heart Study cohort, we found a significant impact of local SNPs of the CRP gene on plasma CRP levels, but we did not find direct evidence of genetically controlled CRP elevations by local CRP SNPs contributing to cardiovascular phenotypes. Individuals’ plasma CRP elevation may be a combined result of local genetic variants, polymorphisms on other genes, and environmental responses such as inflammation. Because CRP is significantly heritable, genes other than the structural CRP gene must help regulate CRP levels. Therefore, further genetic characterization of CRP levels may be required to further investigate the role of CRP in cardiovascular disease.

ACKNOWLEDGMENTS

We thank the investigators, staff, and participants in the NHLBI Family Heart Study.

GRANTS

This work was partially supported by a starting grant from Morehouse Cardiovascular Research Institute [National Institutes of Health (NIH) Enhancement of Cardiovascular and Related Research Areas Grant 5U11 HL-03676], Morehouse Clinical Research Center (NIH Center of Clinical Research Excellence Grant S5U4 RR-014758-05), Morehouse School of Medicine (NIH Enhancement of the Capacity of Biomedical Research-Research Centers of Minority Institutions Grant 5G12 RR-003054-18), and Morehouse School of Medicine (NIH Stroke Prevention/Intervention Research Program Grant 5U54NS046798-02) and a Beginning Grant-in-Aid from the American Heart Association (Q. Song). Support was also provided by NHLBI Cooperative Agreement Grants U01 HL-56563, U01 HL-56564, U01 HL-56565, U01 HL-56566, U01 HL-56567, U01 HL-56568, and U01 HL-56569 to the University of North Carolina, Wake Forest University, the University of Minnesota, Boston University/Framingham, the University of Utah, and Washington University (St. Louis, MO).

REFERENCES
