Epidemiology

Contribution of Clinical Correlates and 13 C-Reactive Protein Gene Polymorphisms to Interindividual Variability in Serum C-Reactive Protein Level

Sekar Kathiresan, MD; Martin G. Larson, ScD; Ramachandran S. Vasan, MD; Chao-Yu Guo, PhD;
Philimon Gona, PhD; John F. Keaney, Jr, MD; Peter W.F. Wilson, MD;
Christopher Newton-Cheh, MD, MPH; Stacy L. Musone, BA; Amy L. Camargo, BA;
Jared A. Drake, BA; Daniel Levy, MD; Christopher J. O'Donnell, MD, MPH;
Joel N. Hirschhorn, MD, PhD; Emelia J. Benjamin, MD, ScM

Background—Serum C-reactive protein (CRP) level is a heritable complex trait that predicts incident cardiovascular disease. We investigated the clinical and genetic sources of interindividual variability in serum CRP.

Methods and Results—We studied serum CRP in 3301 Framingham Heart Study (FHS) participants (mean age 61 years, 53% women). Twelve clinical covariates explained 26% of the variability in CRP level, with body mass index alone explaining 15% ($P < 0.0001$) of the variance. To investigate the influence of genetic variation at the *CRP* gene on CRP levels, we first constructed a dense linkage disequilibrium map for common single-nucleotide polymorphisms (SNPs) spanning the *CRP* locus (1 SNP every 850 bases, 26 kilobase [kb] genomic region). Thirteen *CRP* SNPs were genotyped in 1640 unrelated FHS participants with measured CRP levels. After adjustment for clinical covariates, 9 of 13 SNPs were associated with CRP level ($P < 0.05$). To account for correlation among SNPs, we conducted forward stepwise selection among all 13 SNPs; a triallelic SNP (rs3091244) remained associated with CRP level (stepwise $P < 0.0001$). The triallelic SNP (C→T→A; allele frequencies 62%, 31%, and 7%), located in the promoter sequence, explained 1.4% of total serum CRP variation; haplotypes harboring the minor T and A alleles of this SNP were associated with higher CRP level (haplotype $P = 0.0002$ and 0.004).

Conclusions—In our community-based sample, clinical variables explained 26% of the interindividual variation in CRP, whereas a common triallelic *CRP* SNP contributed modestly. Studies of larger samples are warranted to assess the association of genetic variation in *CRP* and risk of cardiovascular disease. (***Circulation***. 2006;113:1415-1423.)

Key Words: epidemiology ■ inflammation ■ genetics ■ C-reactive protein ■ risk factors

C-reactive protein (CRP) is a marker of systemic inflammation, and circulating CRP levels predict incident cardiovascular disease (CVD).[1] Current evidence suggests that serum CRP level is a complex trait, influenced by both clinical and genetic factors.[2] Clinical and environmental correlates of CRP reported in the literature include age, sex, smoking, body mass index, hormone replacement therapy, and statin use, among others.[3] In addition, heritability estimates for CRP levels have ranged from 27% to 40%,[4,5] which suggests a role for DNA sequence variation in determining serum protein level. However, the extent to which clinical and genetic variables together account for the total variation in serum CRP level is unclear.

Clinical Perspective p 1423

Recent advances in genomics allow for the comprehensive evaluation of candidate gene regions for roles in determining phenotypic variation. There are an estimated 11 million single-nucleotide polymorphisms (SNPs) with a minor allele frequency >1%, and it recently has been recognized that groups of neighboring sequence variants are correlated, a pattern termed "linkage disequilibrium" (LD).[6] A gene-

mapping approach that takes advantage of LD offers promise in mapping genes for complex traits.[7] Specifically, such an approach involves characterization of sequence variation in a reference sample, selection of tag SNPs ("tag" refers to a subset of SNPs that capture the underlying genetic variation in a reference panel and are taken forward to genotype in an association study), and testing of tag SNPs for association with phenotype.[7]

CRP sequence variants that explain interindividual variability in CRP level may aid in the prediction of clinical CVD, may provide targets for drugs to alter gene expression, and represent strong candidate alleles to test for association with CVD. We hypothesized that clinical variables are related to serum CRP levels; that a limited number of tag polymorphisms capture the majority of common genetic variation present at the *CRP* locus; that after accounting for clinical variables, tag polymorphisms in the *CRP* gene are associated with serum CRP level; and that *CRP* polymorphisms related to serum CRP level will be associated with risk of CVD. Accordingly, we investigated LD structure and tag SNP selection at the *CRP* locus in a reference DNA panel and the clinical and genetic correlates of serum CRP level in Framingham Heart Study (FHS) participants.

Methods

Study Participants
The Framingham Offspring cohort (n=5124 participants) was recruited in 1971; the participants have been examined approximately every 4 to 8 years. Of the 3539 Offspring participants who attended the seventh examination (1998–2001),[8] 3301 individuals were available for the analyses of clinical correlates of CRP, after the exclusion of participants with missing CRP level (n=238). DNA was available in 1809 unrelated individuals who provided blood samples for DNA extraction during the sixth examination cycle (1995–1998). The genotype-phenotype association sample consisted of the overlap of participants with available CRP level and available DNA (n=1640). The Institutional Review Board at Boston Medical Center approved the study, and all participants gave written informed consent.

Clinical Examination and CRP Testing
Risk factor information was collected from routine medical history, physical examination, and laboratory assessment as described previously (please see online-only Data Supplement for risk factor definitions). CVD was defined by the presence of coronary heart disease (recognized or unrecognized myocardial infarction, coronary insufficiency, angina pectoris, or coronary death), stroke, transient ischemic attack, intermittent claudication, or heart failure. CVD diagnoses were determined by a panel of 3 investigators who reviewed all available medical and hospital records, as described previously.[9] CRP level was measured once for each participant with a Dade Behring BN100 nephelometer (Deerfield, Ill) on fasting morning serum samples that had been stored at −70°C and thawed at room temperature. The mean intra-assay coefficient of variation was 3.2%.

SNP Selection and Genotyping in a Reference Panel
To construct a dense SNP map at the *CRP* locus, we searched the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP) as of June 2004 and selected every SNP within a 26-kilobase (kb) region spanning the *CRP* gene (2-kb coding region; 15 kb upstream and 9 kb downstream) on chromosome 1 (gene symbol *CRP*, accession number NM_000567). Ninety-nine SNPs were genotyped in a panel of 30 white trios (90 individuals total, with each trio consisting of 2 parents and an offspring) from Utah/Centre d'Etude du Polymorphisme Humain (CEPH) panel (Coriell Institute for Medical Research, Camden, NJ). Genotyping was performed with matrix-assisted laser desorption ionization–time of flight mass spectrometry in the Sequenom MassARRAY platform (San Diego, Calif) as described previously.[10,11]

Assays were considered successful if they met the following criteria: (1) at least 75% success for genotyping calls; (2) Hardy-Weinberg equilibrium $P>0.01$; 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 $\geq 5\%$. Overall, 30 SNPs passed quality control and had a minor allele frequency $\geq 5\%$ (Data Supplement Table I).

LD Structure in Reference Pedigrees
LD structure was assessed by calculating D'[12] and a logarithm of the odds score[12] for each pair of SNPs as implemented in the publicly available Haploview version 2.03 software.[13] Haplotype blocks were defined by the criteria of Gabriel et al.[10]

Tag SNP Selection
We used 3 approaches to select tag SNPs: (1) a pairwise correlation method[14]; (2) R_h^2, a method of selecting haplotype-tag SNPs to predict common haplotypes[15]; and (3) R_s^2, a method of selecting haplotype-tag SNPs to predict unmeasured SNPs (see Data Supplement for details).[15]

In each of the 3 methods, 7 tag SNPs were required to capture the underlying common genetic variation. These 3 sets of 7 SNPs partially overlapped and yielded a total of 10 tag SNPs (Data Supplement Table I). In addition, we genotyped 3 previously studied SNPs: rs1417938,[16] rs1130864,[17] and rs3091244.[18] Overall, 13 *CRP* SNPs were genotyped in the FHS sample. All SNPs genotyped were in Hardy-Weinberg equilibrium (χ^2 test $P>0.05$).

Statistical Analysis

Clinical Correlates
Serum CRP levels had a skewed distribution and were natural log transformed. We constructed stepwise multivariable linear regression models (SAS REG procedure[19]) to identify clinical covariates associated with log CRP, with age and sex forced in as covariates in all models. Additional candidate covariates are detailed in Table 2. $P<0.05$ was the significance criterion for covariates to be retained in the forward stepwise-selection models.

Genotype–CRP Level Association
Log CRP values were adjusted for 15 covariates to create a multivariable-adjusted CRP residual level for each participant. A participant's multivariable-adjusted CRP level served as the sole phenotype in genotype-phenotype and haplotype-phenotype association analyses.

For genotype-phenotype association analyses, we assumed a general model of inheritance (3 genotype classes per biallelic SNP and 6 genotype classes for triallelic SNP rs3091244). We conducted multivariable linear regression analyses (SAS Proc GLM,[19] SAS Institute, Cary, NC) to test the null hypothesis that log CRP residuals did not differ by SNP genotype. To identify a subset of SNPs that significantly explained variance in CRP levels when adjusted for the effects of other SNPs, we conducted forward stepwise selection of the SNPs using a $P<0.05$ criterion. In exploratory analyses, we evaluated specific interaction terms in the multivariable-adjusted model, including an SNP×covariate interaction term for each clinical covariate significantly related to serum CRP level. Haplotype-based association analyses were conducted with weighted regression as implemented in the haplo.stats program.[20]

Genotype–Cardiovascular Disease Association
Because we identified a single CRP polymorphism (triallelic SNP rs30921244) that strongly related to CRP level, in a post hoc analysis, we hypothesized that this sequence variant was related to either prevalent or incident CVD among participants. Multivariable logistic

TABLE 1. Framingham Study Sample Characteristics

	Clinical Correlates Sample		Genetic Association Sample	
	Men (n=1536)	Women (n=1765)	Men (n=823)	Women (n=817)
Age, y	61±10	61±9	62±9	62±9
Systolic blood pressure, mm Hg	128±17	126±20	129±18	127±20
Diastolic blood pressure, mm Hg	76±10	73±10	75±10	72±10
Body mass index, kg/m²	28.8±4.6	27.6±5.9	28.8±4.6	27.4±5.6
Total/HDL cholesterol ratio	4.5±1.4	3.7±1.2	4.5±1.4	3.7±1.2
Cigarette smoking, %	13	13	12	13
Fasting glucose, mg/dL	109±29	101±25	110±33	102±26
Triglycerides, mg/dL	143±100	132±77	142±100	131±74
Diabetes, %	17	11	18	13
Hypertension treatment, %	37	31	38	33
Lipid-lowering medication, %	25	17	24	18
Hormone replacement therapy, %	. . .	34	. . .	36
Aspirin treatment, %	43	26	46	28
Prevalent CVD, %	18	9	19	9
CRP, mg/L, median (Q1, Q3)	1.9 (0.9, 4.4)	2.6 (1.1, 5.9)	1.9 (1.0, 4.9)	2.7 (1.2, 5.8)

Q1 indicates quartile 1; Q3, quartile 3.
Continuous measures, mean±SD, except CRP, for which median (1st and 3rd quartile) values are presented.

regression analysis was performed to test the null hypothesis that prevalent CVD status (n=210 at examination 7, 1998 to 2001) did not differ by marker genotype, and multivariable Cox proportional hazards modeling was performed to test the null hypothesis that incident CVD (n=52 at postexamination 7 to December 2004) did not differ by marker genotype (see Data Supplement for details of genetic statistical analyses). For all analyses, a nominal $P<0.05$ was considered significant.

The authors had full access to the data and take responsibility for its integrity. All authors have read and agree to the manuscript as written.

Results

Framingham Heart Study Sample Characteristics
Clinical characteristics of the FHS sample (mean age 61 years, 53% women) are shown in Table 1. The genotype-phenotype association study sample (n=1640) was similar to the sample with CRP levels.

Clinical Correlates of Serum CRP Levels
In stepwise multivariable linear regression models that included age and sex, 10 clinical factors were significantly related to serum CRP levels (Table 2). Clinical correlates explained 26% of the interindividual variability in serum CRP level, with body mass index alone explaining 15% of the total variance.

LD Pattern at the *CRP* Locus in a Reference Panel
In 90 CEPH individuals (30 trios) of European ancestry, we constructed a dense SNP map (1 SNP every 830 bases) at the *CRP* locus (genomic distance 26 kb, 15 kb upstream, 2-kb gene length, 9 kb downstream) using 30 common SNPs (minor allele frequency ≥5%; Data Supplement Table I). We observed 1 extended haplotype block of 16 kb in this region (Figure). There was limited haplotypic diversity, with 7

haplotypes with ≥5% frequency accounting for 94% of chromosomes.

Performance of Tag SNPs
We evaluated the performance of the tag SNPs genotyped in the present study by comparing the correlation between our tag SNPs and *CRP* sequence variation described in the SeattleSNPs Program for Genomics Applications (http://pga.gs.washington.edu/). In the SeattleSNPs project, 23 individuals of European ancestry were resequenced for 6.8 kb spanning the *CRP* gene. Ten biallelic SNPs were observed with a minor allele frequency ≥5%. This SeattleSNPs collection should represent complete ascertainment for common SNPs in 6.8 kb of genomic distance spanning the *CRP* gene. Six of the biallelic tag SNPs typed in the present study perfectly captured the 10 SeattleSNPs variants (mean $r^2=1.0$).

Associations of Single SNPs With Multivariable-Adjusted Serum CRP Level
We examined the relations between each of 13 SNPs at the *CRP* locus and multivariable-adjusted serum CRP residual; 9 SNPs were nominally associated (each $P<0.05$, minimum $P<0.0001$; Table 3). Several SNPs were highly correlated with each other. For example, SNP rs3116653 and rs1417938 were near-perfect proxies for each other ($r^2=0.98$).

To account for correlation among SNPs, we conducted stepwise selection among all 13 SNPs and found that triallelic SNP rs3091244 remained associated with CRP level (stepwise $P<0.0001$). This triallelic SNP was highly correlated with haplotypes composed of 2 other SNPs (rs1205 and rs2808630). In stepwise models that excluded the triallelic SNP, these 2 SNPs, rs1205 and rs2808630, were also separately associated with CRP level ($P<0.0001$ and 0.003,

TABLE 2.   Clinical Correlates of Log-CRP Level: Stepwise Linear Regression Model R^2=0.26

	Fold Change in CRP (95% CI)	Partial R^2†	P
Age, per 10 y	1.16 (1.12–1.21)	0.02*	<0.0001
Sex, female vs male	1.17 (1.08–1.26)	0.01*	<0.0001
Body mass index, per 5 kg/m²	1.46 (1.42–1.51)	0.15	<0.0001
Hormone replacement therapy, yes/no	1.79 (1.62–1.97)	0.03	<0.0001
Cigarette smoking, yes/no	1.45 (1.32–1.61)	0.02	<0.0001
Total/HDL cholesterol ratio	1.06 (1.02–1.09)	0.01	<0.0001
Hypertension treatment, yes/no	1.19 (1.10–1.29)	0.01	<0.0001
Lipid-lowering therapy, yes/no	0.77 (0.71–0.85)	0.004	<0.0001
Prevalent CVD, yes/no	1.28 (1.15–1.42)	0.01	<0.0001
Triglycerides, per 100 mg/dL	1.08 (1.03–1.14)	0.003	0.0009
Systolic blood pressure, per 20 mm Hg	1.07 (1.03–1.12)	0.002	0.01
Diastolic blood pressure, per 10 mm Hg	0.95 (0.91–0.99)	0.001	0.02

CRP indicates C-reactive protein.

*Age and sex were forced into the multivariable model.

†Partial R^2 indicates the increment to R^2 as each variable was added in the order listed.

Candidate covariates in stepwise selection included body mass index, hormone replacement therapy, cigarette smoking, total/HDL cholesterol ratio, hypertension treatment, lipid-lowering therapy, prevalent CVD, triglycerides, systolic blood pressure, diastolic blood pressure, diabetes, fasting glucose, and aspirin treatment.

respectively). The triallelic SNP (C→T→A, allele frequencies 62%, 31%, and 7%), located in the promoter sequence, explained 1.8% of the residual variation in a model that included all clinical covariates (Table 3). Because clinical covariates explained 26% of serum CRP variation, this 1.8% of variation in CRP residual represents 1.4% of the total variation in CRP level (ie, [1−0.26]×1.8%).

CRP level by genotype at the triallelic SNP rs3091244 is presented in Table 4. Unadjusted values are presented for ease of interpretation. Major allele homozygotes (CC genotype) had the lowest CRP level. Major allele heterozygotes (CT or CA genotypes) had an intermediate level of CRP, whereas individuals with only the minor T or A alleles had the highest CRP level. In exploratory analyses, there were nominally significant SNP×covariate interaction terms for the following: hypertension treatment×rs3116654, triglycerides×rs1417938, triglycerides×rs1130864, and body mass index×rs1205 (each interaction term P=0.04; Data Supplement Table II).

## Associations of Haplotypes With Multivariable-Adjusted Serum CRP Level

Mean serum multivariable-adjusted CRP level differed across 8 common haplotypes observed in FHS (global P<0.0001; Table 5). Compared with all other haplotypes combined, haplotypes A and B were associated with a higher CRP level (haplotype P=0.0002 and 0.004, respectively). Haplotype A harbored the minor T allele, and haplotype B harbored the minor A allele of the triallelic SNP. A 2-SNP combination (rs2808630 and rs1205) served as a proxy for the triallelic SNP and also distinguished haplotypes A and B from all others (Table 5).

Triallelic SNP and Proxies in the International Haplotype Map

Given that the triallelic SNP was our best association result at the *CRP* locus, we studied whether genetic variants highly correlated to the triallelic SNP existed beyond the 26-kb genomic segment surveyed.[21] Using the HapMap catalog of >3.5 million SNPs in the CEPH sample of European ancestry,[22] we assessed a 2-megabase genomic segment encompassing the *CRP* locus and found that there were no perfect proxies (r^2=1.0) for the triallelic SNP (as represented by the A-G haplotype at rs2808630 and rs1205). Seven variants in this 2-megabase region were correlated with the A-G haplotype to a lesser degree (each r^2=0.74).

Triallelic SNP Genotype and CVD

Because triallelic SNP rs3091244 was strongly related to CRP level, and CRP level has been consistently related to CVD, we performed a post hoc analysis to evaluate the relations between SNP rs3091244 genotype and prevalent (at the time of FHS Offspring Study examination cycle 7) or incident (after examination 7 to December 2004) CVD. SNP rs3091044 was not associated with prevalent CVD (n=210; P=0.62 in multivariable-adjusted model). Compared with the major allele homozygotes (CC) as referent genotype, the OR for the group that included major allele heterozygotes (CT and CA) was 0.96 (95% CI 0.67 to 1.37), and for the group with both minor alleles (TA, TT, and AA), it was 0.79 (95% CI 0.48 to 1.29). SNP rs3091044 was not associated with incident CVD (n=52; P=0.18 in multivariable-adjusted model). Compared with the major allele homozygotes (CC) as referent genotype, the hazard ratio for the group that included major allele heterozygotes (CT and CA) was 0.82
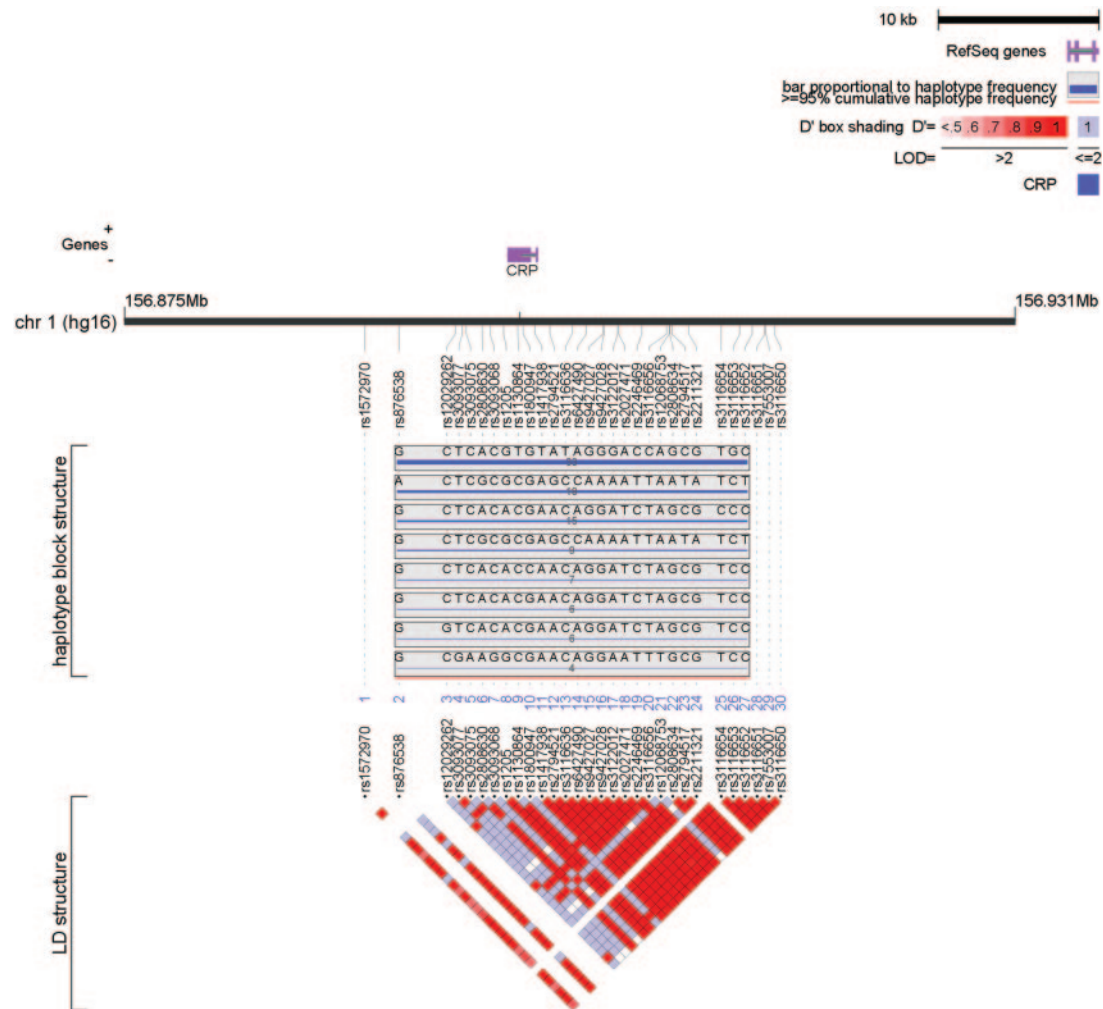
LD structure at the *CRP* gene locus in a reference panel. Chr 1 panel depicts the *CRP* position on chromosome 1 in the human genome July 2003 assembly (hg16). Haplotype block structure panel shows that 26 SNPs fell into a single haplotype block. LD structure panel displays the LD relations between pairs of markers in the region, with each square representing the pairwise strength and significance of LD. Red indicates no or minimal evidence of historical recombination, white indicates weak LD, and blue indicates uninformative LD. Note that the triallelic SNP rs3091244 was not genotyped in the reference panel and thus is not displayed here. The chromosomal position of SNP rs3091244 falls between SNP 11 and SNP 12. Figure prepared with LocusView 2.0 (T. Petryshen et al, Broad Institute; available at: http://www.broad.mit.edu/mpg/locusview/).

(95% CI 0.46 to 1.45), whereas it was 0.31 (95% CI, 0.09 to 1.07) for the group with both minor alleles (TA, TT, and AA).

Statistical Power
Because our sample included a modest number of CVD events (210 prevalent, 52 incident), we assessed power to detect an association between *CRP* SNP genotype and CVD using separate analyses of prevalence and incidence but combining results into a single, combined test (4 degrees of freedom, ie, 2 degrees of freedom each for the prevalence and incidence analyses). At the 5% significance level, we had power of 0.55 for an SNP genotype with an OR of 1.3 per risk allele (that is, OR 1.00, 1.30, and 1.69 for carriers of 0, 1 and 2 risk alleles), power of 0.80 for an OR of 1.40, and power of 0.94 for an OR of 1.5 per risk allele, assuming a risk allele frequency of 30% (the T allele of triallelic SNP rs3091244 has an allele frequency of 31%).

Discussion

Principal Findings
In a community-based sample, we sought to define the contribution of clinical correlates and 13 *CRP* gene polymorphisms to interindividual variation in serum CRP level. The present study yielded several findings: (1) in a reference sample of European ancestry, a single block of LD stretched across the *CRP* locus, and a limited number of tag polymorphisms captured the majority of common sequence variation; (2) the proportion of CRP variation explained by clinical variables was ≈26%, and for *CRP* SNPs, it was ≈1.4%; (3) of 13 CRP common sequence variants studied, a triallelic SNP in the promoter sequence was the strongest cis-acting variant associated with CRP level; (4) there were no perfect proxies for the triallelic SNP across a 2-megabase region that spanned the *CRP* locus[21]; and (5) the *CRP* triallelic SNP was not associated with CVD in analyses with

TABLE 3. Association of Individual *CRP* Gene Variants and Multivariable-Adjusted Serum CRP Level* in FHS

Allelic Variant	Variant Type	Major Allele→Minor Allele	Minor Allele Frequency	Partial R^2†	P
rs3116653	5′ Upstream	C→G	0.31	0.009	0.0008
rs3116654	5′ Upstream	T→C	0.12	0.004	0.04
rs2794517	5′ Upstream	C→T	0.30	‡	0.99
rs3122012	5′ Upstream	A→G	0.31	0.008	0.001
rs3091244	5′ Upstream	C→T	T: 0.31	0.018	<0.0001
		C→A	A: 0.07		
rs1417938	Intronic	A→T	0.31	0.008	0.001
rs1800947	Synonymous, coding	G→C	0.06	‡	0.09
rs1130864	3′ Untranslated	C→T	0.31	0.009	0.0006
rs1205	3′ Flanking	G→A	0.33	0.016	<0.0001
rs2808630	3′ Flanking	A→G	0.28	‡	0.43
rs3093077	3′ Flanking	T→G	0.07	0.008	0.002
rs12029262	3′ Flanking	C→G	0.09	0.007	0.002
rs876538	3′ Flanking	G→A	0.20	‡	0.76

*Phenotype was log-transformed serum CRP level adjusted for age, sex, and 13 candidate covariates listed in legend to Table 2.

†Partial R^2 refers to the proportion of multivariable-adjusted CRP level explained by the specific SNP.

‡Partial R^2 is only provided for the 9 CRP sequence variants significantly associated with multivariable-adjusted serum CRP level.

modest power to detect a genetic effect. The triallelic SNP association with CRP level strongly supports recent work described below and represents a reproducible association between genetic variation and this biomarker of high interest.

Clinical Correlates of Serum CRP

We confirmed prior reports on individual clinical correlates of CRP including age, sex, cardiovascular risk factors[23] (body mass index, cigarette smoking, total/HDL cholesterol ratio, triglycerides, systolic blood pressure, and diastolic blood pressure), medications (lipid-lowering therapy,[24] hormone replacement therapy[25]), and prevalent CVD.[23] We estimated the relative contribution of each of these factors to CRP level variation in the community and demonstrated that body mass index was by far the strongest correlate, explaining 15% of the overall variation in CRP. A possible mechanism behind the adiposity and CRP association has been suggested; adipocytes synthesize a substantial proportion of circulating interleukin-6,[26] the principal regulator of CRP synthesis in the liver.

Comparison of SNP Associations With Prior Studies

Several prior studies have examined the relations between *CRP* polymorphisms and CRP levels and reported individual associations between rs1800947 (also referred to as 1059 G/C),[27] rs1130864 (+1444C→T),[17] rs1205,[28] rs1417938,[16] rs3091244,[2,18,29,30] and CRP level. With regard to *CRP* SNPs

and CVD end points, Zee and Ridker[27] reported that rs1800947 was not associated with a composite end point of nonfatal myocardial infarction, nonfatal stroke, or cardiovascular death, whereas Chen et al[31] found that among Han Chinese, there was an association between a *CRP* SNP (rs2794521) and coronary heart disease.

Until recently, it was not clear whether the reported SNP-CRP level associations represented several distinct independent SNP associations or multiple associations due to LD. Carlson et al[2] addressed this issue by evaluating 7 tag SNPs in the 6.8 kb surrounding the *CRP* gene and demonstrated that in individuals of European ancestry, 2 specific haplotypes were associated with higher CRP levels.

We have assayed 13 polymorphisms in a 26-kb region surrounding the *CRP* locus and found the same 2 haplotypes as in the study by Carlson et al[2] to be associated with higher CRP levels. The minor A and T alleles of triallelic SNP rs3091244 (SNP 1440 in the study by Carlson et al[2]) are exclusively present on the 2 risk haplotypes (haplotypes A and B). In addition, we extend prior work by demonstrating that the triallelic SNP rs3091244 explains the greatest proportion of overall CRP variance. In stepwise selection among all 13 SNPs, only the triallelic SNP (rs3091244) or a 2-SNP proxy for it (rs1205 and rs2808630) remained associated with CRP level.

TABLE 4. CRP Level by Genotype for 6 Genotypes at Triallelic SNP rs3091244

	CC (n=623)	CT (n=633)	TA (n=66)	CA (n=131)	TT (n=148)	AA (n=8)
Median CRP, mg/dL (Q1, Q3)	1.8 (0.9,4.5)	2.5 (1.1,5.0)	3.0 (1.6,5.3)	3.0 (1.2,7.0)	3.2 (1.4,6.6)	3.4 (1.5,8.2)

Q1 indicates quartile 1; Q3, quartile 3.

Unadjusted values are presented for reference. Note that association analyses were conducted with multivariable-adjusted log-transformed CRP level as the phenotype.

TABLE 5.   Association of *CRP* Locus Haplotypes With Multivariable-Adjusted Serum CRP Level* in FHS

Haplotype	rs876538	rs12029262	rs3093077	rs2808630‡	rs1205‡	rs1130864	rs1800947	rs1417938	rs3091244‡	rs3122012	rs2794517	rs3116654	rs3116653	Haplotype Frequency	Haplotype Score†	Haplotype P	Global P
A	G	C	T	A	G	T	G	T	T	G	C	T	G	0.29	3.74	0.0002	<0.0001
B	G	C	G	A	G	C	G	A	A	A	C	T	C	0.05	2.86	0.004	
C	A	C	T	G	G	C	G	A	C	A	T	T	C	0.19	−0.28	0.78	
D	G	C	T	G	G	C	G	A	C	A	T	T	C	0.08	−0.91	0.36	
E	G	C	T	A	A	C	C	A	C	A	C	T	C	0.05	−1.98	0.048	
F	G	G	T	A	A	C	G	A	C	A	C	T	C	0.09	−2.21	0.03	
G	G	C	T	A	A	C	G	A	C	A	C	C	C	0.12	−2.34	0.02	
H	G	C	T	A	A	C	G	A	C	A	C	T	C	0.05	−2.78	0.005	

*Phenotype was log-transformed serum CRP level adjusted for age, sex, and 13 candidate covariates listed in legend Table 2.

†Haplotype score statistic tests whether trait differences exist between a single haplotype and all other haplotypes combined.

‡Note that in stepwise selection including all 13 CRP SNP variants, triallelic SNP rs3091244 or a 2-SNP proxy (rs2808630 and rs1205) was significantly associated with multivariable-adjusted CRP level. The 2 minor alleles of the triallelic SNP (T and A) or the A-G haplotype at rs2808630-rs1205 distinguishes haplotype A and haplotype B from all others.

Triallelic SNP Function

Several lines of evidence strongly suggest that the triallelic SNP rs3091244 is functional. First, prior in vitro studies have directly demonstrated a functional role for SNP rs3091244.[2,29] Szalai et al[29] reported that SNP rs3091244 resides in a consensus E-box element (ie, a DNA sequence that binds basic helix-loop-helix transcription factors such as upstream stimulatory factor-1) and that transcription factor binding occurred when the variant T allele was present. They went on to demonstrate that haplotypes harboring the variant T or A alleles were associated with higher promoter activity, as assessed by luciferase reporter assays, consistent with our findings in the present study. Second, SNP rs3091244 is located in promoter sequence, 301 bases upstream of the *CRP* gene transcription start site. Third, Cha-Molstad and colleagues[32,33] have shown that an interplay of transcription factors binding to the proximal promoter segment regulates CRP gene expression. Lastly, we observed that the region surrounding this SNP represents a nonexonic evolutionarily conserved region, which suggests functional importance (http://ecrbrowser.dcode.org/).[34] This evolutionarily conserved region represents a 75% sequence identity compared with murine sequence (www.ecrbrowser.com), extends for 287 bases, and encompasses SNP rs3091244.[34]

Association and Linkage for Complex Quantitative Traits

In a genome-wide linkage scan that we recently conducted for the quantitative trait of CRP level, we did not observe evidence for linkage at the *CRP* locus on chromosome 1.[5] The apparent discrepancy highlights recent concepts raised by researchers comparing association studies with linkage analysis. Association approaches offer greater statistical power to detect common alleles with modest genetic effects.[35] Indeed, for a range of blood protein level phenotypes including plasminogen activator inhibitor-1[11] and monocyte chemoattractant protein-1,[36] we and others have shown that SNPs in the gene encoding the respective blood protein explain, on average, 1% to 2.5% of phenotypic variation. Association studies have greater power than linkage studies in equivalently sized cohorts to detect such modest effect sizes.

Study Limitations and Strengths

The large community-based sample, comprehensive characterization of common sequence variation surrounding the *CRP* gene, unbiased ascertainment of measured covariates, multivariable analyses, and the combination of SNP-phenotype and haplotype-phenotype analyses are strengths of the present investigation. Despite these assets, our study has some limitations. First, we addressed the role of common *CRP* genetic variation in relation to CRP level. Multiple rare variants have been shown to influence quantitative cardiovascular traits such as HDL cholesterol.[37] Resequencing of the *CRP* gene locus in individuals with extreme phenotypes will be required to address the possibility that aggregation of multiple rare variants contributes to interindividual variation in CRP level in the general population. Second, regulatory variants acting in *trans* (ie, sequence variants outside of the *CRP* locus) likely affect serum CRP level. The present study did not address this hypothesis. Third, because our genotype×phenotype interaction analyses were exploratory and only nominally significant, they need to be confirmed in another cohort.

Finally, our sample consisted of a modest number of CVD events, and thus, the lack of association between the *CRP* triallelic SNP genotype and CVD may be due to insufficient statistical power. Our estimates suggest that we had >90% power to detect an OR of 1.5 per risk allele if we assumed a risk allele frequency of 30%. Studies with larger number of individuals affected with CVD are warranted to assess for more modest effect sizes, and such efforts are currently in progress.

Implications

Our findings may have several implications. First, body mass index is by far the strongest correlate of CRP levels, and this finding suggests that maintaining a healthy weight may be the most effective measure to keep serum CRP levels low. Second, genetic association studies that use blood protein levels as phenotypes may effectively identify functional

noncoding alleles in the genome. Finally, genetic association studies using the finite set of *CRP* tag SNPs described in the present study could be used to clarify whether CRP is causally related to clinical disease. Whereas serum CRP level has been shown to predict a range of clinical phenotypes, including myocardial infarction, stroke, diabetes, and hypertension, it is debated whether serum CRP level marks an underlying inflammatory process or causally contributes to that clinical disease. A bona fide association between a *CRP* gene variant and clinical disease would suggest a causal role for CRP.

References

1. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N Engl J Med*. 1997;336:973–979.
2. Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, Liu K, Williams OD, Iribarren C, Lewis EC, Fornage M, Boerwinkle E, Gross M, Jaquish C, Nickerson DA, Myers RM, Siscovick DS, Reiner AP. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *Am J Hum Genet*. 2005;77:64–77.
3. Ledue TB, Rifai N. Preanalytic and analytic sources of variations in C-reactive protein measurement: implications for cardiovascular disease risk assessment. *Clin Chem*. 2003;49:1258–71.
4. Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH, Tracy RP. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis*. 2001;154:681–689.
5. Dupuis J, Larson MG, Vasan RS, Massaro JM, Wilson PW, Lipinska I, Corey D, Vita JA, Keaney JF Jr, Benjamin EJ. Genome scan of systemic biomarkers of vascular inflammation in the Framingham Heart Study: evidence for susceptibility loci on 1q. *Atherosclerosis*. 2005;182:307–314.
6. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature*. 2001;411:199–204.
7. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature*. 2005;437:1299–1320.
8. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham offspring study. *Am J Epidemiol*. 1979;110:281–290.
9. Shearman AM, Cupples LA, Demissie S, Peter I, Schmid CH, Karas RH, Mendelsohn ME, Housman DE, Levy D. Association between estrogen receptor alpha gene variation and cardiovascular disease. *JAMA*. 2003;290:2263–2270.
10. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science*. 2002;296:2225–2229.
11. Kathiresan S, Gabriel SB, Yang Q, Lochner AL, Larson MG, Levy D, Tofler GH, Hirschhorn JN, O'Donnell CJ. 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.

12. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet*. 2001;69:1–14.
13. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263–265.
14. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet*. 2004;74:106–120.
15. Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Pike MC. Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. *Hum Hered*. 2003;55:27–36.
16. Suk HJ, Ridker PM, Cook NR, Zee RY. Relation of polymorphism within the C-reactive protein gene and plasma CRP levels. *Atherosclerosis*. 2005;178:139–145.
17. Brull DJ, Serrano N, Zito F, Jones L, Montgomery HE, Rumley A, Sharma P, Lowe GD, World MJ, Humphries SE, Hingorani AD. Human CRP gene polymorphism influences CRP levels: implications for the prediction and pathogenesis of coronary heart disease. *Arterioscler Thromb Vasc Biol*. 2003;23:2063–2069.
18. Kovacs A, Green F, Hansson LO, Lundman P, Samnegard A, Boquist S, Ericsson CG, Watkins H, Hamsten A, Tornvall P. A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. *Atherosclerosis*. 2005;178:193–198.
19. *SAS/STAT User's Guide* [computer program]. Version 8.1. Cary, NC: SAS Institute; 2000.
20. 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;70:425–434.
21. Lawrence R, Evans DM, Morris AP, Ke X, Hunt S, Paolucci M, Ragoussis J, Deloukas P, Bentley D, Cardon LR. Genetically indistinguishable SNPs and their influence on inferring the location of disease-associated variants. *Genome Res*. 2005;15:1503–1510.
22. International Haplotype Map Project. Available at: http://www.hapmap.org/. Accessed December 8, 2005.
23. Mendall MA, Patel P, Ballam L, Strachan D, Northfield TC. C reactive protein and its relation to cardiovascular risk factors: a population based cross sectional study. *BMJ*. 1996;312:1061–1065.
24. Ridker PM, Rifai N, Pfeffer MA, Sacks F, Braunwald E; the Cholesterol and Recurrent Events (CARE) Investigators. Long-term effects of pravastatin on plasma concentration of C-reactive protein. *Circulation*. 1999;100:230–235.
25. Cushman M, Meilahn EN, Psaty BM, Kuller LH, Dobs AS, Tracy RP. Hormone replacement therapy, inflammation, and hemostasis in elderly women. *Arterioscler Thromb Vasc Biol*. 1999;19:893–899.
26. Mohamed-Ali V, Goodrick S, Rawesh A, Katz DR, Miles JM, Yudkin JS, Klein S, Coppack SW. Subcutaneous adipose tissue releases interleukin-6, but not tumor necrosis factor-alpha, in vivo. *J Clin Endocrinol Metab*. 1997;82:4196–4200.
27. Zee RY, Ridker PM. Polymorphism in the human C-reactive protein (CRP) gene, plasma concentrations of CRP, and the risk of future arterial thrombosis. *Atherosclerosis*. 2002;162:217–219.
28. Russell AI, Cunninghame Graham DS, Shepherd C, Roberton CA, Whittaker J, Meeks J, Powell RJ, Isenberg DA, Walport MJ, Vyse TJ. Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. *Hum Mol Genet*. 2004;13:137–147.
29. Szalai AJ, Wu J, Lange EM, McCrory MA, Langefeld CD, Williams A, Zakharkin SO, George V, Allison DB, Cooper GS, Xie F, Fan Z, Edberg JC, Kimberly RP. Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *J Mol Med*. 2005;83:440–447.
30. Miller DT, Zee RY, Suk DJ, Kozlowski P, Chasman DI, Lazarus R, Cook NR, Ridker PM, Kwiatkowski DJ. Association of common CRP gene variants with CRP levels and cardiovascular events. *Ann Hum Genet*. 2005;69(pt 6):623–638.
31. Chen J, Zhao J, Huang J, Su S, Qiang B, Gu D. -717A>G polymorphism of human C-reactive protein gene associated with coronary heart disease in ethnic Han Chinese: the Beijing atherosclerosis study. *J Mol Med*. 2005;83:72–78.

32. Cha-Molstad H, Agrawal A, Zhang D, Samols D, Kushner I. The Rel family member P50 mediates cytokine-induced C-reactive protein expression by a novel mechanism. *J Immunol*. 2000;165:4592–4597.

33. Agrawal A, Cha-Molstad H, Samols D, Kushner I. Transactivation of C-reactive protein by IL-6 requires synergistic interaction of CCAAT/enhancer binding protein beta (C/EBP beta) and Rel p50. *J Immunol*. 2001;166:2378–2384.

34. Ovcharenko I, Nobrega MA, Loots GG, Stubbs L. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res*. 2004;32(Web Server issue): W280–W286.

35. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996;273:1516–1517.

36. McDermott DH, Yang Q, Kathiresan S, Cupples LA, Massaro JM, Keaney JF Jr, Larson MG, Vasan RS, Hirschhorn JN, O'Donnell CJ, Murphy PM, Benjamin EJ. CCL2 polymorphisms are associated with serum monocyte chemoattractant protein-1 levels and myocardial infarction in the Framingham Heart Study. *Circulation*. 2005;112: 1113–1120.

37. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*. 2004;305:869–872.

CLINICAL PERSPECTIVE

C-reactive protein (CRP) is a marker of systemic inflammation, and circulating CRP levels predict incident cardiovascular disease. Circulating CRP levels vary considerably between individuals, and at least part of this variability has been demonstrated to be due to genetic factors. In a community-based sample of 3301 individuals, we evaluated the extent to which clinical and genetic factors explained serum CRP level variability. Twelve clinical factors were significantly related to serum CRP level and combined explained ≈26% of CRP level variation. Of these factors, body mass index alone was the strongest explanatory factor, contributing to 15% of the variance. To isolate genetic factors related to CRP level, we studied 13 common polymorphisms that represented the genetic variation at the *CRP* gene region and found that a triallelic polymorphism (most single-nucleotide polymorphisms have 2 alleles, whereas a few, such as that in the present study, have 3 alleles) was strongly related to CRP level. The triallelic *CRP* polymorphism explained ≈1.4% of the overall variability in CRP level. Using a comprehensive approach that went beyond single polymorphisms, we were able to fully characterize the association of *CRP* genetic variation on circulating CRP level. The specific triallelic *CRP* SNP identified in the present study is of interest for future studies to assess whether this polymorphism potentially may aid in clinical disease risk prediction and in targeting therapeutic interventions. Our findings imply that maintaining a healthy weight may be the most effective measure to keep serum CRP level low.