Association of exome sequences with plasma C-reactive protein levels in >9000 participants

Ursula M. Schick¹, Paul L. Auer¹,², Joshua C. Bis³, Honghuang Lin⁴, Peng Wei⁵, Nathan Pankratz⁷, Leslie A. Lange⁶, Jennifer Brody⁸, Nathan O. Stitziel⁹,¹⁰, Daniel S. Kim¹¹, Christopher S. Carlson¹, Myriam Fornage⁵,⁶, Jeffery Haessler¹, Li Hsu¹,¹², Rebecca D. Jackson¹⁷, Charles Kooperberg¹, Suzanne M. Leal¹⁸, Bruce M. Psaty¹³,¹⁴,¹⁵,¹⁰, Eric Boerwinkle⁵,¹⁹, Russell Tracy²¹, Diego Ardissino²², Sviati Shah²³, Cristen Willer²⁴,²⁵,²⁶, Ruth Loos²⁷,²⁸, Olle Melander²⁹, Ruth Mcpherson³⁰, Kees Hovingh³¹,³², Muredach Reilly³³, Hugh Watkins³⁴,³⁵, Domenico Girelli³⁶, Pierre Fontanillas³⁷, Daniel I. Chasman³⁸, Stacey B. Gabriel³⁷, Richard Gibbs¹⁹, Deborah A. Nickerson⁷, Alanna C. Morrison⁵, Myron D. Gross⁷, and Alex P. Reiner¹,¹⁶, *, on behalf of the Cohorts for Heart and Aging Research in Genomic Epidemiology and the National Heart, Lung, and Blood Institute GO Exome Sequencing Project‡

¹Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, ²School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA, ³Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA, ⁴Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA, ⁵Human Genetics Center, School of Public Health, ⁶Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA, ⁷Department of Lab Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA, ⁸Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA, ⁹Cardiovascular Division, Department of Medicine, ¹⁰Division of Statistical Genomics, Washington University School of Medicine, Saint Louis, MO 63110, USA, ¹¹Department of Genome Sciences, ¹²Department of Biostatistics, ¹³Department of Epidemiology, Cardiovascular Health Research Unit, ¹⁴Department of Medicine, ¹⁵Department of Health Services, ¹⁶Department of Epidemiology, University of Washington, Seattle, WA 98105, USA, ¹⁷Division of Endocrinology, Diabetes, and Metabolism, Ohio State University, Columbus, OH 43210, USA, ¹⁸Center for Statistical Genetics, Department of Molecular and Human Genetics, ¹⁹Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA, ²⁰Group Health Research Institute, Group Health Cooperative, Seattle, WA 98101, USA, ²¹Departments of Biochemistry and Pathology, University of Vermont, Burlington, VT 05401, USA, ²²Division of Cardiology, Azienda Ospedaliero-Universitaria di Parma, Parma, Italy, ²³Division of Cardiology, Department of Medicine and Center for Human Genetics, Duke University, Durham, NC, USA, ²⁴Department of Internal Medicine, Division of Cardiovascular Medicine, ²⁵Department of Computational Medicine and Bioinformatics, ²⁶Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA, ²⁷The Charles Bronfman Institute for Personalized Medicine, ²⁸The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, ²⁹Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö, Sweden, ³⁰Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, Canada, ³¹Department of Vascular Medicine, ³²Department of Experimental Vascular Medicine, Academic Medical Center, Amsterdam 1105 AZ, The Netherlands, ³³The Institute for Translational Medicine and Therapeutics and The Cardiovascular Institute, Perelman

To whom correspondence should be addressed at: 1100 Fairview Avenue N. M3-A410, Seattle, WA 98109-1024, USA. Tel: +1 206 667 2710; Fax: +1 206 667 4142; Email: apreiner@u.washington.edu
These authors contributed equally to this work.
*Authorship banner is included in Supplementary Material.

© The Author 2014. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

C-reactive protein (CRP) concentration is a heritable systemic marker of inflammation that is associated with cardiovascular disease risk. Genome-wide association studies have identified CRP-associated common variants associated in ~25 genes. Our aims were to apply exome sequencing to (1) assess whether the candidate loci contain rare coding variants associated with CRP levels and (2) perform an exome-wide search for rare variants in novel genes associated with CRP levels. We exome-sequenced 6050 European-Americans (EAs) and 3109 African-Americans (AAs) from the NHLBI-ESP and the CHARGE consortia, and performed association tests of sequence data with measured CRP levels. In single-variant tests across candidate loci, a novel rare (minor allele frequency = 0.16%) CRP-coding variant (rs77832441-A; p.Thr59Met) was associated with 53% lower mean CRP levels ($P = 2.9 \times 10^{-6}$). We replicated the association of rs77832441 in an exome array analysis of 11,414 EAs ($P = 3.0 \times 10^{-15}$). Despite a strong effect on CRP levels, rs77832441 was not associated with inflammation-related phenotypes including coronary heart disease. We also found evidence for an AA-specific association of $APOE<e2$ rs7214 with higher CRP levels. At the exome-wide significance level ($P < 5.0 \times 10^{-8}$), we confirmed associations for reported common variants of $HNF1A$, $CRP$, $IL6R$ and $TOMM40$-$APOE$. In gene-based tests, a burden of rare/lower frequency variation in $CRP$ in EAs ($P < 6.8 \times 10^{-4}$) and in retinoic acid receptor-related orphan receptor $\alpha$ ($RORA$) in AAs ($P = 1.7 \times 10^{-5}$) were associated with CRP levels at the candidate gene level ($P < 2.0 \times 10^{-5}$). This inquiry did not elucidate novel genes, but instead demonstrated that variants distributed across the allele frequency spectrum within candidate genes contribute to CRP levels.

## INTRODUCTION

C-reactive protein (CRP) is an acute-phase protein reactant produced by the liver in response to proinflammatory stimuli. CRP is a sensitive, but nonspecific heritable biomarker of systemic inflammation that is associated with a variety of inflammation-mediated diseases (1–5). Particular attention has been focused on characterizing the association between CRP and cardiovascular disease (CVD). Prospective epidemiologic studies suggest that basal CRP levels are predictive of risk of future CVD (2,6–8), though the degree of association is dependent on levels of other conventional vascular risk factors (5,9). Current consensus recommendations support the clinical use of CRP to predict CVD risk among a subset of asymptomatic adults and in the selection of statin therapy (10). Despite routine clinical use of CRP levels, data from Mendelian randomization studies suggest that CRP is unlikely to be causally related to CVD (11–17).

Genome-wide association studies (GWASs) have sought to characterize genetic determinates of CRP levels. This approach has been successful at identifying ~25 loci associated with CRP levels among individuals of European (17–22), Asian (23–25) and African (26,27) descent (Supplementary Material, Table S1). To further interrogate known loci and to search for novel loci associated with CRP levels, we applied exome sequencing, which captures sequence variation in the protein-coding portion of the genome. Exome sequencing has proved useful for identifying rare causal variants for several Mendelian disorders (28–32). Furthermore, recent studies illustrate the application of exome sequencing to identify variation underlying complex traits (33–38).

In this study, we apply exome sequencing to a large sample of European-American (EA) and African-American (AA) ascertained through seven population-based cohorts [Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk Development in Young Adults (CARDIA), Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), Jackson Heart Study (JHS), Multi-Ethnic Study of Atherosclerosis (MESA) and the Women’s Health Initiative (WHI)] that compose the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. Our specific aims were to (1) assess whether known CRP loci harboring common variants also contain rare coding variants associated with CRP levels and (2) perform an exome-wide search for rare variants in novel genes associated with CRP levels. Follow-up HumanExome BeadChip genotyping data...
from an independent sample derived from the WHI and JHS
cohorts were used as replication for discovery findings.

RESULTS

Participant characteristics

Race-stratified characteristics of discovery and validation
cohorts are summarized in Supplementary Material, Table S2.
Overall, compared with EA, AA had a greater proportion of
women, higher prevalence of hypertension and type 2 diabetes,
higher body mass index (BMI) and higher median CRP levels.

Single-variant test results

Summary information for uniquely annotated variants included
in the ESP–CHARGE meta-analyses is detailed in Supplementary
Material, Table S3. The meta-analyzed exome-wide
single-variant association results for CRP levels in EA, AA
and combined races are summarized in Figure 1 and Supplementary
Material, Figure S1A. Overall results for significant association
signals were consistent with no significant between-study
heterogeneity (Tables 1 and 2).

Confirmation of common coding variants associated
with CRP levels

Nine coding variants from four distinct chromosomal regions
reached exome-wide significance in the combined sample
(Table 1). All nine of these significant variants are common
[minor allele frequency (MAF) >10% in both EA and AA]
and have been previously reported or are in linkage disequilibrium
(LD) with known CRP-associated single-nucleotide polymorphisms
(SNPs) (2,19,25,39–41). The genes and their variants are hepatocyte nuclear factor 1 homeobox A [HNF1A (MIM
142410); rs2464196 (p.Ser487Asn), rs2259820 (p.Leu459=),
rs1169288 (p.Ile27Leu) and rs1169289 (p.Leu17=), 12q24.2],
interleukin 6 receptor [IL6R (MIM 147880); rs2228145 (p.Asp
358Ala), 1q21.31], leptin receptor [LEPR (MIM 601007);
rs1805096 (p.Pro1019=), 1p31.3], translocase of outer mitochondrial
membrane 40 [TOMM40 (MIM 608061); rs157581
(p.Phe113=), rs11556505 (p.Phe131=), 19q13.32] and apolipoprotein E [APOE (MIM 107741); rs429358 (p.Cys130Arg),
19q13.32]. These significant hits in HNF1A, IL6R, LEPR,
TOMM40 and APOE correspond to four independent association
signals. Consistent with the results of prior GWAS meta-
analyses, common intronic variants in CRP, LEPR, HNF1A
and TOMM40 were also associated with CRP levels at an exome-
wide significance level (Supplementary Material, Table S4).
Supplementary Material, Table S5 summarizes effect estimates
for all variants reaching significance by the analytic group (ESP,
CHARGE-ARIC, CHARGE-CHS or CHARGE-FHS).

Lower frequency variation associated with CRP levels

Two lower frequency synonymous variants (TOMM40 rs112
849259, p.Asp209=, MAF = 3.1% and CRP rs1800947,
rs.Leu184=, MAF = 4.4%) reached exome-wide significance

Figure 1. Manhattan plots of −log_{10}(P-values) from single-variant analyses. (A) Combined race, (B) EA and (C) AA. Variants in the 25 candidate loci identified through CRP GWAS are highlighted in red, candidate loci with variant reaching exome-wide significance are labeled in italics and rare variants rs77832441 and rs112849259 are also labeled.
rs112849259 is not present on the Exome Array, we were previously reported as associated with CRP levels. Since in the combined sample of EA and AA participants (Table 2).

Table 1. Exome-wide significant single-variant associations with CRP ($P < 5.00 \times 10^{-8}$) for common coding variants

<table>
<thead>
<tr>
<th>Gene (GRCh37 coordinate)</th>
<th>rs# (function; coded/ non-coded allele)</th>
<th>Previous report (GWAS variant if different than reported, LD measures)</th>
<th>Sample (N)</th>
<th>MAF</th>
<th>$\beta$ (SE)</th>
<th>P-value</th>
<th>Het. $I^2$ ($P$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEPR (1:66102257)</td>
<td>rs1805096 (Syn; G/A)</td>
<td>Reiner et al. (26)</td>
<td>EA (6050)</td>
<td>0.37</td>
<td>-0.11 (0.019)</td>
<td>3.56 $\times 10^{-8}$</td>
<td>70.4 (0.012)</td>
</tr>
<tr>
<td>IL6R (1:154426970)</td>
<td>rs2228145 (nSyn; C/A)</td>
<td>Curocichin et al. (39)</td>
<td>EA (6050)</td>
<td>0.40</td>
<td>-0.11 (0.016)</td>
<td>8.34 $\times 10^{-12}$</td>
<td>56.7 (0.044)</td>
</tr>
<tr>
<td>HNF1A (12:12143542)</td>
<td>rs2259820 (Syn; T/C)</td>
<td>Reiner et al. (19)</td>
<td>EA (6050)</td>
<td>0.31</td>
<td>-0.12 (0.020)</td>
<td>1.83 $\times 10^{-9}$</td>
<td>0.669 (0.761)</td>
</tr>
<tr>
<td>HNF1A (12:121435427)</td>
<td>rs2464196 (nSyn; A/G)</td>
<td>Reiner et al. (19)</td>
<td>EA (6050)</td>
<td>0.32</td>
<td>-0.12 (0.020)</td>
<td>9.29 $\times 10^{-9}$</td>
<td>0.71 (0.071)</td>
</tr>
<tr>
<td>HNF1A (12:12146622)</td>
<td>rs1169289 (Syn; G/C)</td>
<td>Kong and Lee (25)</td>
<td>EA (6050)</td>
<td>0.46</td>
<td>-0.12 (0.019)</td>
<td>8.98 $\times 10^{-11}$</td>
<td>8.5 (0.34)</td>
</tr>
<tr>
<td>HNF1A (12:12146650)</td>
<td>rs1169288 (nSyn; C/A)</td>
<td>Curocichin et al. (39)</td>
<td>EA (6050)</td>
<td>0.34</td>
<td>-0.06 (0.028)</td>
<td>2.16 $\times 10^{-2}$</td>
<td>0 (1)</td>
</tr>
<tr>
<td>TOMM40 (19:45395714)</td>
<td>rs157581 (Syn; C/T)</td>
<td>Middleburg et al. (40)</td>
<td>EA (6050)</td>
<td>0.21</td>
<td>-0.16 (0.023)</td>
<td>2.42 $\times 10^{-12}$</td>
<td>0.68 (0.32)</td>
</tr>
<tr>
<td>TOMM40 (19:45396144)</td>
<td>rs1556605 (Syn; T/C)</td>
<td>Middleburg et al. (40)</td>
<td>EA (6050)</td>
<td>0.47</td>
<td>-0.09 (0.027)</td>
<td>5.07 $\times 10^{-9}$</td>
<td>0.88 (0.37)</td>
</tr>
<tr>
<td>APOE (19:45411941)</td>
<td>rs429538 (nSyn; C/T)</td>
<td>Chasman et al. (41)</td>
<td>EA (6050)</td>
<td>0.26</td>
<td>-0.11 (0.018)</td>
<td>1.36 $\times 10^{-9}$</td>
<td>0 (0.90)</td>
</tr>
</tbody>
</table>

nSyn, nonsynonymous; syn, synonymous; MAF, minor allele frequency; Het. $I^2$, heterogeneity; EA, European-American; AA, African-American; SE, standard error; (P), P-value; $\beta$, beta; SNP, single-nucleotide polymorphism; GRCh37, Genome Reference Consortium Human Build 37; rs#, reference SNP ID number.

Table 2. Single-variant associations of rare and low frequency coding variants with CRP levels

<table>
<thead>
<tr>
<th>Gene (GRCh37 coordinate)</th>
<th>rs# (function; coded/ non-coded allele)</th>
<th>Previously report</th>
<th>Sample (N)</th>
<th>MAF</th>
<th>$\beta$ (SE)</th>
<th>P-value</th>
<th>Het. $I^2$ ($P$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP (1:159683438)</td>
<td>rs1800947 (Syn; G/C)</td>
<td>Ridker et al. (2)</td>
<td>EA (6050)</td>
<td>0.0607</td>
<td>-0.27 (0.039)</td>
<td>4.82 $\times 10^{-12}$</td>
<td>0 (0.56)</td>
</tr>
<tr>
<td>CRP (1:159683814)</td>
<td>rs77832441 (nSyn; A/G)</td>
<td>Not reported</td>
<td>EA (6050)</td>
<td>0.0022</td>
<td>-0.75 (0.197)</td>
<td>1.39 $\times 10^{-4}$</td>
<td>0 (0.44)</td>
</tr>
<tr>
<td>TOMM40 (19:45397307)</td>
<td>rs112849259 (Syn; C/T)</td>
<td>Not reported</td>
<td>EA (6050)</td>
<td>0.0005</td>
<td>-0.26 (0.065)</td>
<td>6.65 $\times 10^{-4}$</td>
<td>0 (0.82)</td>
</tr>
</tbody>
</table>

nSyn, nonsynonymous; syn, synonymous; MAF, minor allele frequency; Het. $I^2$, heterogeneity; EA, European-American; AA, African-American; SE, standard error; (P), P-value; $\beta$, beta; SNP, single-nucleotide polymorphism; GRCh37, Genome Reference Consortium Human Build 37; rs#, reference SNP ID number.

in the combined sample of EA and AA participants (Table 2). The CRP rs1800947 synonymous variant has previously been associated with lower CRP levels in EA, independently of the more common CRP-lowering haplotype at the chromosome 1q23 CRP locus (12.20.42). TOMM40 rs112849259 has not previously been reported as associated with CRP levels. Since rs112849259 is not present on the Exome Array, we were unable to assess its association with CRP in our genotype-based validation sample.

In addition to the CRP rs1800947 synonymous variant, we identified a novel association of a rare nonsynonymous CRP variant rs77832441 (Thr59Met; MAF = 0.16%) with lower CRP levels in the combined discovery sample (A allele, beta (\(\beta\)) standard error (SE) = -0.88 (0.19); \(P = 2.90 \times 10^{-6}\))...
that was significant at the candidate gene level (significance threshold \( P < 1.87 \times 10^{-4} \)). We subsequently validated this finding among 11,414 independent EA Exome Array samples, where the A allele (MAF = 0.31%) was strongly associated with CRP levels (\( \beta (SE) = -0.90 (0.11), P = 3.00 \times 10^{-15} \)). The combined discovery and replication \( P \)-value was 3.86 \( \times 10^{-16} \).

**Additional characterization of CRP p.Thr59Met rare variant association signal**

To demonstrate that the CRP missense variant rs77832441 p.Thr59Met association signal is independent of other common and lower frequency CRP locus SNPs, we performed conditional regression analyses using either the ESP discovery exome-sequenced samples or the genotype-based Exome Array validation samples. As shown in Supplementary Material, Table S6, the association of rs77832441 p.Thr59Met with CRP levels was independent of CRP variants known to be associated with CRP through prior GWAS (rs1417938, rs3093059 and rs1801278) (43). The p.Thr59Met missense variant is located at residue 41 of the \( \beta \)-sheet (residues 32–41) proximate to a 3/10 alpha helical domain (residues 43–45). The 40–42 region of the CRP protein is involved in interprotomer interactions with residues 115–123 on the adjoining monomer (Supplementary Material, Fig. S2). Therefore, the non-conservative threonine-to-methionine amino acid change [which is predicted to be functional by \( \text{in silico} \) protein conservation algorithms (44,45)] has the potential to reduce CRP pentamerization or reduce native pentameric CRP stability.

To follow up on this association, we sought to test whether other inflammation-related phenotypes were associated with reduced native pentameric CRP stability.

**Additional analysis of common coding variants at the chromosome 19q13**

Two common synonymous variants in TOMM40 (rs157581-C allele, MAF = 30%; and rs11556505-T allele, MAF = 13%) were associated with a \( \sim 12\% \) lower mean CRP level (\( P \leq 8.62 \times 10^{-9} \)) in the combined EA and AA single-variant meta-analysis. On the basis of previous studies (39,46), we hypothesized that these TOMM40 common variants and/or the low frequency (rs112849259) variant may be in LD with the functional \( \text{APOE} \) \( \epsilon4 \)-allele-defining variant rs429358 p.Cys130Arg by performing conditional analysis using the ESP exome sequencing data. Conditional analysis in EA and AA demonstrated that all three TOMM40 synonymous CRP-lowering variants (rs157581, rs11556505 and rs112849259) are conditionally dependent on the \( \text{APOE} \) \( \epsilon4 \)-allele-defining SNP rs429358 (Table 3).

The \( \text{APOE} \) \( \epsilon2 \) allele-defining variant rs7412 p.Arg176Cys failed quality control due to poor sequencing coverage in both ESP and CHARGE discovery samples; however, rs7412 genotype was available in 13,794 individuals from our validation sample genotyped on the Exome Array. Association analyses using the Exome Array samples suggested that the rs7412 minor allele (T) was associated with higher CRP levels in

<table>
<thead>
<tr>
<th>rs# (function)</th>
<th>MAF (EA/AA/combined)</th>
<th>Sample (N)</th>
<th>1000 Genomes pairwise ( r^2 )</th>
<th>Without adjustment for rs429358 ( \beta (SE) )</th>
<th>With adjustment for rs429358 ( \beta (SE) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs112849259</td>
<td>0.026/0.039/</td>
<td>EA (1832)</td>
<td>CEU (1, 0.07, 0.005)</td>
<td>–0.40 (0.11) P &lt; 0.01</td>
<td>–0.17 (0.12) P = 0.13</td>
</tr>
<tr>
<td>(Synonymous)</td>
<td>0.031</td>
<td>AA (1528)</td>
<td>YRI (rs112849259 NA in YRI)</td>
<td>–0.30 (0.10) P &lt; 0.01</td>
<td>–0.15 (0.11) P = 0.15</td>
</tr>
<tr>
<td>rs157581</td>
<td>0.21/0.47/0.30</td>
<td>EA (1832)</td>
<td>CEU (0.07, 1, 0.58)</td>
<td>–0.16 (0.044) P &lt; 0.01</td>
<td>–0.030 (0.058) P = 0.60</td>
</tr>
<tr>
<td>(Synonymous)</td>
<td>0.20/0.47/0.30</td>
<td>AA (1528)</td>
<td>YRI (NA, 1, 0.11)</td>
<td>–0.089 (0.038) P &lt; 0.01</td>
<td>–0.025 (0.043) P = 0.56</td>
</tr>
<tr>
<td>rs11556505</td>
<td>0.14/0.12/0.13</td>
<td>EA (1832)</td>
<td>CEU (0.005, 0.583, 1)</td>
<td>–0.18 (0.052) P &lt; 0.01</td>
<td>–0.014 (0.068) P = 0.84</td>
</tr>
<tr>
<td>(Synonymous)</td>
<td>0.14/0.12/0.13</td>
<td>AA (1528)</td>
<td>YRI (NA, 0.11, 1)</td>
<td>–0.051 (0.056) P &lt; 0.01</td>
<td>–0.032 (0.057) P = 0.57</td>
</tr>
</tbody>
</table>

MAF, minor allele frequency; Het, \( r^2 \), heterogeneity; EA, European-American; AA, African-American; SE, standard error; \( P \), \( P \)-value; \( \beta \), beta; rs#, reference SNP ID number.
Gene-based test results

Results for collapsing approach (T1) and sequence kernel association test (SKAT) gene-based tests are summarized in Supplementary Material, Figures S1B and S1C, respectively. None of the genes reached exome-wide significance (significance threshold $P < 2.5 \times 10^{-8}$) in EA or AA alone. In the race-combined sample, only rare variation in CRP was associated with CRP levels (T1 $P = 2.36 \times 10^{-8}$) at an exome-wide significance level.

Among the 25 candidate genes from GWAS (significance threshold $P < 2.00 \times 10^{-3}$), the most significant finding was that rare variation in the CRP locus was associated with CRP levels in EA with both the T1 test ($P = 6.80 \times 10^{-4}$) and SKAT ($P = 1.71 \times 10^{-8}$; Supplementary Material, Tables S8 and S9). Results for both tests of EA on the Exome Array robustly replicated the CRP gene-based association ($P_{SKAT} = 3.21 \times 10^{-15}$; $P_{T1} = 2.54 \times 10^{-14}$). Removal of the rare CRP variant rs77832441 (Thr59Met) from the EA T1 and SKAT tests eliminated significant signal for the gene-based test ($P_{SKAT} = 0.96$; $P_{T1} = 0.77$), suggesting that the CRP gene-based association signal is driven solely by rs77832441. Gene-based results for CRP in AA were less significant ($P_{SKAT} = 0.06$; $P_{T1} = 1.05 \times 10^{-3}$), only reaching candidate gene level of significance in the T1 test.

Gene-based testing with SKAT additionally showed that rare putatively deleterious variation in the retinoic acid receptor-related orphan receptor $\alpha$ [RORA (MIM 600825)] locus was associated with CRP levels in AA ($P = 1.73 \times 10^{-3}$), but not in EA ($P = 0.83$). Using HumanExome BeadChip genotype data from an independent African-American validation sample [$P = 0.06$], the gene-based test did not reach a statistical significance. However, only one RORA rare variant was shared between exome sequencing discovery and HumanExome BeadChip validation platform, which may explain our inability to replicate the finding.

**DISCUSSION**

By combining exome sequence data from two large consortia, we have identified and validated a novel association between lower CRP levels and the rs7732441 CRP missense variant. Though strongly associated with CRP levels, rs7732441 was not associated with other inflammation-mediated phenotypes including CHD, stroke, systolic blood pressure, waist-to-hip ratio and BMI. In addition to the previously described CRP-lowering association signal attributable to the canonical APOE e4 allele, we demonstrate a second, independent association signal from the APOE e2 allele that is specific to AA. Finally, a targeted gene-based analysis of known CRP-associated genes suggested possible associations between rare coding variants in RORA (in AA) on CRP levels.

In this study, we did not identify any novel genes through aggregated gene-based tests. Meta-analysis of cohort-specific results in this study was carried out in skatMeta, which limited available gene-based methodologies to SKAT and burden tests. Methods development for rare variant association analyses is an active area of research (47). Future efforts may extend available options for meta-analysis, which may be capable of identifying additional rare variant association signals in our dataset.

Previous studies suggest that tens of thousands of samples may be necessary to be sufficiently powered to detect associations between rare variants and complex traits (48,49). Though we uncover associations between CRP levels and rare-to-low frequency variants of larger effect, we recognize that larger sample sizes will be required to detect associations of rare coding variants of more moderate effect sizes. Power calculations based on our discovery sample size of ~6000 EAs suggest we had 80% power to detect a variant with ~2 mg/l effect on CRP at 3% MAF but considerably less power to detect the same effect size at MAF ≤ 1% (Supplementary Material, Fig. S3). Beyond limited power, factors such as heterogeneity of study sampling design (inclusive of both extreme sampling for other phenotypes and randomly selected individuals) or exome sequencing platform, and minor differences in variant calling and quality control procedures between ESP and CHARGE consortia, may have limited our ability to detect novel rare variants associated with CRP levels and/or limit generalizability of our findings.

**CRP**

Previous studies have identified and replicated associations between CRP levels and variants in the CRP locus in populations of European (17,18,20), African (26,27), Asian (17,23,24,39) and Hispanic ancestry (26). These GWAS have identified three independently associated common alleles (rs33116653, rs12093699 and rs1205) and a single low frequency synonymous variant (rs1800947) in the CRP gene associated with lower CRP levels in EA (Supplementary Material, Table S10). GWAS and candidate gene studies (12,42) have identified an additional AA-specific allele (rs16827466) associated with higher CRP levels.

Extending these previous findings, we identify a novel rare CRP variant (rs7732441) that is both strongly associated with CRP level in single-variant tests and drives the gene-based results for CRP. On the basis of conditional analyses (Supplementary Material, Table S6) and LD estimates (Supplementary Material, Table S7), this variant appears to represent an independent signal from known GWAS CRP SNPs. Despite the magnitude of its effect, rs7732441 only contributed 0.5% of the overall natural log(CRP) phenotypic variance. Our results
extend the existence of allelic heterogeneity at the CRP locus and suggest that rare coding variation in the CRP gene contributes to CRP phenotypic variation.

Functional prediction algorithms taking into account conservation and protein structure, SIFT (45) and Polyphen-2 (44) predict the p.Thr59Met variant to be deleterious [SIFT = 0.03 (damaging) and Polyphen-2 = 1.0 (probably damaging)]. Similarly, crystal structures of CRP indicate that the variant could disrupt the site involved in monomer subunit interactions (43), which may reduce CRP pentamer formation or stability. In vitro, the monomeric and native pentameric isoforms of CRP exhibit distinct physicochemical and inflammatory properties (50–52), which have potential implications for the role of CRP in atherothrombotic disorders.

It remains possible that, since CRP levels were measured by immunoassay, the p.Thr59Met missense variant may alter an epitope recognized by the monoclonal antibody used for CRP capture or detection, leading to artificially low CRP values. We investigated the possibility that the p.Thr59Met amino acid substitution interferes with CRP detection by examining CRP measurements among a subset of 3442 CARDIA participants for whom both polyclonal and monoclonal antibody assays were performed at study Year 15. We hypothesized that the polyclonal assay would be less susceptible to artifactual bias due to the amino acid substitution at residue 59. Comparison of the results failed to demonstrate a large difference in CRP effect size between monoclonal and polyclonal assays in a subset of the discovery sample (Supplementary Material, Fig. S4). CRP levels detected by monoclonal antibody and polyclonal antibody assays were highly correlated in this sample for p.Thr59Met carriers \(r = 0.97\) and non-carriers \(r = 0.94\). These results suggest that p.Thr59Met may confer a true biological reduction in circulating CRP levels, though further studies are warranted to fully characterize the functional significance of this variant.

Despite the large, likely direct effect of rs77832441 on mean CRP levels, investigation of the relationship of the identified rare variant to other inflammation-mediated phenotypes failed to demonstrate an association. Similar to previous studies (11–17,53–55), these findings provide additional evidence that CRP levels may not elicit direct effects on CHD, stroke, systolic blood pressure, hip-to-waist ratio, type 2 diabetes and BMI. While CRP is a robust marker of CHD risk (2,6–8), Mendelian randomization studies have consistently failed to demonstrate that CRP is causally related to CHD (11–17). Our finding of a lack of association between rare, putatively deleterious rs77832441 and CHD provides additional support that CRP is not in the causal pathway for CHD.

**TOMM40-APOE**

19q13 is a gene-dense region including TOMM40, APOE and apolipoprotein C-I (APOC1; MIM 107710). Variants in all three of these genes have been previously associated with CRP levels through GWAS or candidate gene analyses (17,26,39–41,46). In our exome sequencing analysis, the APOE e4-tagging missense variant (rs429358) and three synonymous TOMM40 variants were associated with CRP levels (rs157581, rs11556505 and rs112849259). Evidence from previous study and 1000 Genomes sequencing suggest that rs157581 and rs11556505 are in moderate-to-strong LD with rs429358, the variant that defines the APOE e4 allele (56,57). Similarly, we show in our dataset that the CRP-associated TOMM40 variants (rs157581, rs11556505 and rs112849259) do not represent an independent signal from the APOE rs429358 variant. Furthermore, we identify and replicate an additional AA-specific rs429358-independent novel association between APOE e2-tagging variant rs7412 and higher CRP level.

APOE is a major constituent of lipoproteins, and the APOE e2 and e4 alleles are important genetic determinants of CVD and Alzheimer’s disease (58,59). The p.Arg158Cys missense variant encoded by e2 exhibits reduced affinity for the low-density lipoprotein receptor (LDLR) and reduced clearance of apoE-containing triglyceride-rich lipoprotein particles (such as very low-density lipoprotein), and is associated with lower low-density lipoprotein (LDL) cholesterol levels (41) and type III hyperlipoproteinemia (59). In contrast, the p.Cys112Arg variant encoded by e4 exhibits more rapid clearance of apoE-containing triglyceride-rich lipoproteins due to increased LDLR affinity and is associated with increased LDL-cholesterol (LDL-c) levels and increased cardiovascular risk (41,58,60,61). In addition to its role in cholesterol transport, APOE has antiatherogenic and anti-inflammatory effects in experimental systems (59,62). However, the mechanisms for these pleiotropic effects as well as the paradoxical association of e2 with lower LDL-c/higher CRP levels and e4 with higher LDL-c/lower CRP levels are not well understood.

**RORA**

RORA encodes a nuclear receptor with strong homology to the retinoic acid receptor. Data from knockout RORA mice models suggest the importance of RORA in regulating immune and inflammatory responses, atherosclerosis susceptibility and ischemia-induced angiogenesis (63,64). Previous GWAS indicate that common variation in RORA is associated with CRP (18) and liver enzyme levels (65,66) at a genome-wide significance level in EA. The previous GWAS meta-analysis of CRP identified a common variant rs340029-T allele as associated with increased CRP levels (18); however, GWAS in AA and Hispanic Americans found no association between rs340029 and CRP levels (26). Further characterization of rare variation of this locus through sequencing efforts is warranted to follow up on our discovery finding of a burden of rare variants at the locus.

**Summary**

Overall, our results suggest that variants distributed across the allele frequency spectrum within biological candidate genes identified by GWAS contribute to CRP levels. As suggested by other studies, robustly associating rare coding variants with modest effects on complex traits will require extremely large sample sizes (26,48). Collaborative efforts involving meta-analysis of exome sequence or exome BeadChip genotype data will be necessary to amass the large sample sizes required to identify additional rare coding variants contributing to the phenotypic variance of CRP levels.
MATERIALS AND METHODS

Study subjects and CRP measurements

Our discovery sample consisted of exome sequence data from 3360 individuals from the NHLBI-ESP and 5799 individuals from the CHARGE project with valid CRP measures. In total, these 9159 participants included 6050 EA and 3109 AA sampled from seven population-based cohorts: ARIC (N = 4827), CARDIA (N = 190), CHS (N = 946), FHS (N = 1144), JHS (N = 346), MESA (N = 399) and the WHI (N = 1307) as part of the NHLBI-ESP and an independent sample from three of the same population-based cohorts (ARIC, FHS and CHS) as part of the CHARGE consortium (67). CRP levels were measured by high-sensitivity immunoassay in all seven cohorts. Detailed descriptions of each of the seven cohorts and the techniques used to measure circulating CRP levels are provided in previous publications (68–73) and summarized in Supplementary Material, Table S11. Clinical information was collected through self-report and in-person examination. All participants provided written informed consent as approved by local human subjects committees.

Sampling design

The CHARGE cohorts’ study participants included here were selected as part of a large random cohort sample or for extreme values for at least one of the following phenotypes: age at menopause, electrocardiogram QT interval, fasting blood glucose, fibrinogen level, renal function, Stamlar—Kannel-like extremes of risk factors selected by principle components (PCs) and waist-to-hip ratio. The sampling design of ESP included extremes, we adjusted for sampling design to minimize bias. The CHARGE cohorts’ study participants included here were selected as part of a large random cohort sample or for extreme

Exome sequencing and variant calling

In ESP, the processes of library construction, exome capture, sequencing and mapping were performed as previously described (26,37,74). Sequencing was performed at the University of Washington (UW) and the Broad Institute of MIT/Harvard (Broad). Briefly, exome capture was performed using Roche Nimblegen SeqCap EZ or Agilent SureSelect Human All Exon 50 Mb. Paired-end sequencing (2 × 76 bp) was performed on Illumina GAII and HiSeq instruments. Single-nucleotide variants (SNVs) were called using a maximum-likelihood approach (75) implemented in the UMAKExe pipeline at the University of Michigan, which allowed all samples to be analyzed simultaneously, both for variant calling and filtering. Binary Alignment/Map (BAM) (76) files summarizing Burrows–Wheeler Alignment (BWA) (75) alignments generated at the UW and the Broad were used as an input. These BAM files summarized alignments mapped to the Genome Reference Consortium Human Build 37 (GRCh37), refined by duplicate removal, re-alignment and indel re-alignment using the Genome Analysis ToolKit (GATK) (77). We excluded all reads that were not confidently mapped (Phred-scored mapping quality <20) from further analysis. Mean depth was 127 × in targeted regions. We then computed genotype likelihoods for exome targeted regions and 50 flanking bases, accounting for per base alignment quality using SAMtools (76). Variable sites and their allele frequencies were identified using a maximum-likelihood model, implemented in gmlnMingle (78). These analyses assumed a uniform prior probability of polymorphism at each site. The final call-set was performed on 6823 samples (referred to as the ESP6800 call-set).

In CHARGE, DNA samples were constructed into Illumina paired-end pre-capture libraries according to the manufacturer’s protocol. The complete protocol and oligonucleotide sequences are accessible from the Human Genome Sequencing Center (HGSC) website. Either four or six pre-capture libraries were pooled together and then hybridized to Nimblegen exome capture array (HGSC VCRome 2.1 design (79); (42 Mb, NimbleGen]) and sequenced in paired-end mode in a single lane on the Illumina HiSeq 2000 platform. Illumina sequence analysis was performed using the HGSC Mercury analysis pipeline. Pooled samples were de-multiplexed using the Consensus Assessment of Sequence and Variation (CASAVA) software. Reads were then mapped to the GRCh37 human reference sequence using BWA producing SAMtools BAM files. Aligned reads were then recalibrated using GATK (77) along with BAM sorting, duplicate read marking and realignment near indels. The Atlas2 (80) suite was used to call variants and produce high-quality variant call files (VCF) (81). The VCF includes a filter indicating variants with apparent strand-bias, low allele fraction, low coverage or low quality to produce a high-quality variant list. Specifically, the poor quality fields included variants with a posterior probability of less than 0.95, <3 variant reads, variant read ratio <0.1, >99% variant reads in a single-strand direction, total coverage <6 and homozygous reference alleles with less than 6 × coverage.

Quality control

ESP used a support vector machine classifier to separate likely true-positive and false-positive variant sites, applying a series of variant-level filtering steps. Variant-level quality metrics included allelic balance (the proportional representation of each allele in likely heterozygotes), base quality distribution for sites supporting the reference and alternate alleles, and the distribution of supporting evidence between strands and sequencing cycle, among others. These steps were followed by quality control on individual samples within each study. We used as the positive training set variants identified by dbSNP (82) or 1000 Genomes (57), and we used variants that failed multiple filters as the negative training set. We found this method to be effective at removing sequencing artifacts while preserving good quality data, as indicated by the transition–transversion (ti–tv) ratio for previously known and newly identified variant sites, the proportion of high-frequency variants overlapping with dbSNP, and the ratio of synonymous to nonsynonymous variants, as well as attempts at the validation of a subset of sites. We excluded variants with read depth >500, variants with >2 observed alleles in CHARGE and any genotype containing a copy of the less frequent alternate allele in ESP, or missing rates ≥10% for ESP and >20% for CHARGE, and HWE P-value of <5 × 10−8 and <5 × 10−6 for ESP and CHARGE, respectively.
In ESP, samples with discrepant self-reported race and ancestry derived from principal component analysis (PCA) performed on exome sequencing data in PLINK (83) as well as ancestry outliers by PCA were removed. Samples having very low concordance (<90%) with previously obtained SNP array data were considered likely sample swaps and were also dropped from further analysis. In CHARGE, within each cohort, a sample was excluded if it fell beyond 6 standard deviations of any of four selected measures that were calculated by the cohort and ancestry group: number of singletons, heterozygote to homozygote ratio, mean depth or ti–tv ratio.

**Variant annotation**

To facilitate meta-analysis between CHARGE and ESP, we created a combined variant annotation file including all quality-controlled variant sites observed in either study: 2 706 877 variants in CHARGE and 1 907 911 in ESP6800. We first annotated variants in the two studies separately using an in-house pipeline built on ANNOVAR (84) and dbNSFP v2.0 (85) according to the reference genome GRCh37 and National Center for Biotechnology Information RefSeq. The majority of the exonic variants were annotated to a unique gene and functional category. A variant, however, can be annotated to multiple genes by ANNOVAR and can have more than one of the following functional categories: stop-gain, stop-loss, splicing, nonsynonymous, noncoding RNA splicing, synonymous, exonic, 5′ untranslated region (UTR5), 3′ untranslated region (UTR3), noncoding RNA exonic, upstream, intronic, noncoding RNA intronic, downstream and intergenic, where the first five categories are considered ‘functional’ variants to be included in the rare variant burden tests. We then merged the CHARGE and ESP annotated variant lists to ensure that a variant that was present in both studies has the same reference allele and functional annotation. The combined CHARGE-ESP SNP info file that was used in the skatMeta package included a total of 3 494 971 unique autosomal sites present in either or both ESP and CHARGE.

**Association analyses**

Samples with very high CRP values (>100 mg/l) were excluded from analyses and measured CRP values below the lower limit of detection were replaced with the assay lower limit value, leaving 3109 AA and 6050 EA available for association testing. CRP values were naturally log (ln) transformed to normalize the distribution of CRP levels. We performed two types of tests, single variant (common and lower frequency variants) and gene burden (rare and lower frequency variants only), as detailed further below and summarized in Supplementary Material, Table S12. Cohort-level analyses were carried out using the R (86) skatMeta package. In the CHARGE cohorts (ARIC, FHS and CHS), the ‘skatCohort’ or ‘skatFamcohort’ functions were used to create datasets for meta-analysis. In ESP, the six cohorts were pooled into a single dataset and were analyzed using the ‘skatCohort’ function.

Meta-analyses of the ‘skatCohort’ and ‘skatFamcohort’ results were conducted at two independent sites to ensure concordance in findings. Both race-stratified meta-analysis considering a single ethnicity, and combined meta-analyses including both ethnicities were carried out because most CRP-associated loci identified to date have shown consistent patterns of association between EA and AA (26). All meta-analyses were conducted in the skatMeta package using the ‘singlesnpMeta’ function for single-variant meta-analyses, and the ‘burdenMeta’ and ‘skatMeta’ functions for gene-level meta-analyses. We considered only variants on autosomal chromosomes in all analyses. Confirmatory analyses in METAL (87) evaluate between-study heterogeneity of significant results using the heterogeneity $I^2$ metric. Based on adjustment for 23 tests of significant CRP-associated SNV in the discovery sample, heterogeneity was considered to be statistically significant (a $P$-value of $<2.17 \times 10^{-3}$).

**Single-variant tests**

Using the skatMeta, we ran race-stratified study-specific objects (ESP, CHARGE-ARIC, CHARGE-FHS and CHARGE-CHS) for downstream meta-analyses. Within each meta-analysis group (EA, AA and combined) and for each variant site with five or more minor alleles detected, we tested for association with CRP levels via linear regression with an additive genetic model. We included as covariates, race-specific PCs as needed, age, sex and BMI. In the ESP samples, a dummy variable correcting for the sampling procedure, cohort and capture target was included in the model.

**Gene burden tests**

Using the skatMeta package, we ran two different types of gene-level tests. The first was adopted from the T1 collapsing approach (15). For the T1 tests, we only considered polymorphic variants having a within race MAF of $\leq 0.01$ that was calculated from the entire ESP–CHARGE call-set. The second gene-level test was adopted from the SKAT approach (47). For the SKAT tests, we only considered variants having an MAF of $\geq 0.05$ that was calculated from the entire ESP–CHARGE call-set. Both T1 and SKAT tests considered only SNVs annotated as nonsynonymous, stop-gain, stop-loss, noncoding RNA splice or splice variants in the shared ESP/CHARGE annotation file. Burden tests considered only considered variants that passed quality control in either or both ESP and CHARGE. All gene-level tests were adjusted for age, sex, BMI and principal components (as needed). In addition, the ESP analyses included a dummy variable correcting for the sampling procedure, cohort and capture target.

**Significance thresholds**

For the analysis of rare variants (MAF $<0.05$) in the 25 known CRP-associated genes, we evaluated 267 coding variants and thus, we used a Bonferroni-corrected threshold of $P < 1.87 \times 10^{-4}$ to declare significance at the single-variant level. In the burden test of candidate genes, we corrected for the number of candidate genes surveyed ($N = 25$) to assign the threshold of significance at $P < 2.0 \times 10^{-3}$. In the exome-wide exploratory analyses, an association was deemed to be statistically significant at $P \leq 5.0 \times 10^{-8}$ for single variants (the standard GWAS common variant criteria for assessing significance based on a million test), which is a stringent threshold considering that we only tested $\sim 640 000$ variant sites. Significance was
assessed at $P < 2.5 \times 10^{-6}$ for the gene-based rare variant association tests based on a correction for 19 230 genes. Supplementary Material, Table S12 summarizes the significance thresholds for all the reported tests.

Validation of CRP association signals in additional samples

Significant association findings from the discovery sample were followed up in an independent sample of 13 794 participants of the WHI with CRP measurement available [11 414 EA participants and a sub-sample of 2380 AA participants from the WHI-SHARE project (26)] and 2201 AA from JHS. Replication samples were independent from samples used in the ESP and CHARGE exome sequencing discovery sample. The JHS and WHI validation samples were genotyped using the Illumina HumanExome v1.0 BeadChip at Broad Institute or at the Translational Genomics Research Institute (Phoenix, AZ), respectively. Genotype calls were assigned using GenomeStudio v2010.3. We removed samples with call-rates <98%, SNPs with call-rates <95% or HWE $P$-values less than 5 \times 10^{-6}. We checked concordance of genotype calls across hundreds of duplicated samples and SNPs with concordance rates <99% were excluded from analysis. High-sensitivity CRP was measured on with the use of a latex-particle enhanced immunoturbidimetric assay. Consistent with the discovery analysis, only samples with CRP values <100 mg/l were included in the replication analyses. Single-variant and gene burden tests for association tests were performed as described above.

Investigation of the association between rs77832441 and inflammation-related phenotypes

Samples from the WHI EA replication with available Illumina Human Exome Array data were used to investigate the association between rs77832441 and inflammation-mediated phenotypes stroke, systolic blood pressure, waist-to-hip ratio and BMI. To investigate the association with CHD, we used EA samples from the Myocardial Infarction Genetics Exome Array Consortium with available Illumina HumanExome BeadChip data (NCHD cases = 14 727, N controls = 30 232). Availability of EA samples from 6474 type 2 diabetes cases and 6370 controls from the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in Ethnic Samples Consortium with available exome sequence data permitted us to evaluate whether rs77832441-A allele was associated with case status. Linear or logistic regression was used to assess association between the predictor (rs77832441—per A-allele) and the continuous or binary inflammation-mediated phenotype with adjustment for covariates as needed. Statistical significance was assessed at $P < 0.05$.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of the National Heart, Lung, and Blood Institute (NHLBI) and the contributions of the research institutions, study investigators, field staff and study participants in creating this resource for biomedical research. Funding for GO ESP was provided by NHLBI grants RC2 HL-103010 (HeartGO), RC2 HL-102923 (LungGO) and RC2 HL-102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL-102925 (BroadGO) and RC2 HL-102926 (SeattleGO). Funding support for ‘Building on GWAS for NHLBI-diseases: the US CHARGE consortium’ was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL 102419). Data for ‘Building on GWAS for NHLBI-diseases: the US CHARGE consortium’ were provided by Eric Boerwinkle on behalf of the Atherosclerosis Risk in Communities (ARIC) Study, L. Adrienne Cupples, principal investigator for the Framingham Heart Study, and Bruce Psaty, principal investigator for the Cardiovascular Health Study. Sequencing was carried out at the Baylor Genome Center (U54 HG003273). The ARIC Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268 201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C and HHSN 268201100012C), R01HL087641, R01HL59367 and R01HL 086694. The authors thank the staff and participants of the ARIC study for their important contributions. The Framingham Heart Study is conducted and supported by the NHLBI in collaboration with Boston University (Contract No. N01-HC-25195), and its contract with Affymetrix, Inc., for genome-wide genotyping services (contract no. N02-HL-6-4278), for quality control by Framingham Heart Study investigators using genotypes in the SNP Health Association Resource (SHArE) project. A portion of this research was conducted using the LinGA Cluster for Genetic Analysis (LinGA) computing resources at Boston University Medical Campus. This CHS research was supported by contracts HHSN268201200036C, HHSN268 200800007C, N01 HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086 and grants HL080295, HL087652, HL105756 from the National Heart, Lung, and Blood Institute (NHLBI) with additional contribution from National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the National Institutes on Aging (NIA). A full list of CHS principal investigators and institutions can be found at CHS-NHLBI.org. Supported in part by grant R25CA094880 from the National Cancer Institute and by R01HL071862 from NHLBI. The Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) project was supported by NIH grant U01DK085526.

Conflict of Interest statement. None declared.

REFERENCES


