C-reactive protein (*CRP*) promoter polymorphisms influence circulating CRP levels in a genome-wide association study of African Americans

Ayo P. Doumatey[1],∗, Guanjie Chen[1], Fasil Tekola Ayele[1], Jie Zhou[1], Michael Erdos[2], Daniel Shriner[1], Hanxia Huang[1], Jokotade Adeleye[3], Williams Balogun[3], Olufemi Fasanmade[4], Thomas Johnson[4], Johnnie Oli[5], Godfrey Okafor[5], Albert Amoah[6], Benjamin A. Eghan[7], Kofi Agyenim-Boateng[7], Joseph Acheampong[7], Clement Adebamowo[3,8], Norman P. Gerry[9], Michael F. Christman[9], Adebowale Adeyemo[1] and Charles N. Rotimi[1]

[1]Center for Research on Genomics and Global Health and [2]Genome Technology Branch, Molecular Genetics Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA, [3]University of Ibadan, Ibadan, Nigeria, [4]University of Lagos, Lagos, Nigeria, [5]University of Nigeria Teaching Hospital, Enugu, Nigeria, [6]University of Ghana Medical School, Accra, Ghana, [7]University of Science and Technology, Kumasi, Ghana, [8]Institute of Human Virology, School of Medicine, University of Maryland, Baltimore, MD, USA and [9]Coriell Institute for Medical Research, Camden, NJ, USA

C-reactive protein (CRP) is an acute phase reactant protein produced primarily by the liver. Circulating CRP levels are influenced by genetic and non-genetic factors, including infection and obesity. Genome-wide association studies (GWAS) provide an unbiased approach towards identifying loci influencing CRP levels. None of the six GWAS for CRP levels has been conducted in an African ancestry population. The present study aims to: (i) identify genetic variants that influence serum CRP in African Americans (AA) using a genome-wide association approach and replicate these findings in West Africans (WA), (ii) assess transferability of major signals for CRP reported in European ancestry populations (EA) to AA and (iii) use the weak linkage disequilibrium (LD) structure characteristic of African ancestry populations to fine-map the previously reported *CRP* locus. The discovery cohort comprised 837 unrelated AA, with the replication of significant single-nucleotide polymorphisms (SNPs) assessed in 486 WA. The association analysis was conducted with 2 366 856 genotyped and imputed SNPs under an additive genetic model with adjustment for appropriate covariates. Genome-wide and replication significances were set at $P < 5 \times 10^{-8}$ and $P < 0.05$, respectively. Ten SNPs in (*CRP pseudogene-1*) *CRPP1* and *CRP* genes were associated with serum CRP ($P = 2.4 \times 10^{-09}$ to 4.3×10^{-11}). All but one of the top-scoring SNPs associated with CRP in AA were successfully replicated in WA. *CRP* signals previously identified in EA samples were transferable to AAs, and we were able to fine-map this signal, reducing the region of interest from the 25 kb of LD around the locus in the HapMap CEU sample to only 8 kb in our AA sample.

INTRODUCTION

C-reactive protein (CRP), a protein of the pentaxin family, is produced mainly by the liver and is involved in several host defense mechanisms. Its levels are raised in the acute phase response to tissue injury, infection and other inflammatory stimuli ([1]). In cell culture, its induction is mainly under the control of interleukin-6 (IL-6) ([2]–[4]). However, *in vivo*

∗To whom correspondence should be addressed at: Center for Research on Genomics and Global Health, NHGRI/NIH, 12 South Dr. Bldg12A-4047, Bethesda, MD 20852, USA. Tel: +1 3014510195; Fax: +1 3014515426; Email: doumateya@mail.nih.gov

studies have shown that its induction is controlled by a complex network of cytokines, some of which are IL-6 independent (4,5). The role of CRP in chronic complex disease is gaining increasing recognition. For example, CRP has been shown to be elevated in obese, hypertensive and diabetic subjects (6,7). Furthermore, high basal levels of serum CRP predict future cardiovascular disease (CVD) and metabolic abnormalities (4,8,9). Consequently, it is used in clinical settings along with the lipids profile to evaluate the risk of developing CVD (4,10).

The gene encoding CRP is located on chromosome 1 (1q21–q23) and a number of studies have shown associations between polymorphisms at *CRP* locus and serum CRP levels (8,11–13). However, only a small fraction of the reported heritability, estimated at 30–50% (14), is explained by these associations. To date, six genome-wide association studies (GWAS) of CRP have been published and they are all in populations of European ancestry (EA) or Asian ancestry (8,14–18). All six studies confirmed previously reported

associations between *CRP* polymorphisms and serum CRP levels while reporting the new loci: *APOE*, *HNF1A* (*hepatocyte nuclear factor 1 homeobox A*), *IL6R* (*interleukin-6 receptor*), *LEPR* (*leptin receptor*) and *IL6*. Recently, Dehghan *et al*. (16) reported additional loci [such as *NLRP3* (*NLR family, pyrin domain containing 3*) and *IL1F10*] that may be involved in chronic inflammation. Thus far, there is no GWAS of CRP in a population of African ancestry. Notably, a number of studies have reported that CRP levels vary with ethnicity. African Americans (AA) have been shown to have the highest baseline CRP level when compared with other ethnic groups in the USA (6,19). In view of the well-known differences between ancestral populations with respect to haplotype blocks, linkage disequilibrium (LD) patterns and allele frequencies (11), conducting GWAS in multiple populations has been strongly advocated (20). In the present study, we report the first GWAS for CRP in AA enrolled from the Washington DC metropolitan area and replicated the top-scoring single-nucleotide polymorphisms (SNPs) in an independent sample of 486 West Africans (WA). Second, we attempted to replicate previously reported significant GWAS findings. Finally, we use the weak LD structures characteristic of African ancestry populations to fine-map the reported *CRP* loci in EA.

RESULTS

The sample included 837 unrelated AA subjects (340 men and 497 women); 15.4% of the subjects had type 2 diabetes (T2D). The anthropometric and clinical characteristics of the study subjects are given in Table 1. Men and women were similar in age with a mean age of 47 years. Women were heavier (31.5 versus 29.2 kg/m^2, $P < 0.0001$) and had higher circulating high-sensitivity C-reactive protein (hsCRP) levels (0.63 versus 0.50 mg/dl, $P = 0.15$), but this was not statistically significant.

P-values for the associations of each SNP with serum hsCRP levels according to chromosome number and position are shown on the Manhattan plot (Supplementary Material, Fig. S1). The top 50 SNPs associated with serum hsCRP adjusting for age, sex, body mass index (BMI), T2D and the

Table 1. Anthropometric and clinical characteristics of study subjects

Parameters	Male	Female	P-value
n	340	497	
Age (years)	47.7 ± 13	48.6 ± 14.6	0.37
BMI (kg/m^2)	29.2 ± 7.7	31.5 ± 9.1	<0.0001
Type 2 diabetic (%)	12.9	17.1	0.1
Smoking (%)	58.7	41.7	<0.0001
Hypertensive (%)	49.7	47.5	0.56
SBP (mmHg)	131 ± 21.7	128.5 ± 22.4	0.12
DBP (mmHg)	82.8 ± 14.1	79 ± 12.1	<0.0001
Cholesterol (mg/dl)	191.2 ± 42	199 ± 42.8	0.012
HDL cholesterol (mg/dl)	50.1 ± 16.1	53.8 ± 16.8	0.0014
LDL cholesterol (mg/dl)	115.7 ± 38.3	120.6 ± 39	0.073
Triglycerides (mg/dl)	107.4 ± 59.6	105.7 ± 85.2	0.74
HsCRP			
Mean (mg/dl)	0.50 ± 1.47	0.63 ± 1.0	0.15
Geometric mean (mg/dl)	0.19	0.29	

Values are mean $\pm$ SD except for hypertension, type 2 diabetes and smoking which are frequencies. Mean values were compared by Student's *t*-test and the frequencies by the χ^2 test.

Table 2. SNPs associated with ln(hsCRP) at a genome-wide significance level in African Americans

| SNP	Chromosome	Position[a]	Type	CA[b] (CAF)[c]	Closest gene	Distance to gene (bp)	Beta (SE)	P-value
rs16827466[d]	1	159649700	INTERGENIC	T (0.19)	*CRPP1*	25 073	0.47 (0.07)	4.3×10^{-11}
rs726640	1	159655518	INTERGENIC	A (0.17)	*CRPP1*	19 255	0.50 (0.08)	5.0×10^{-11}
rs7531832	1	159654216	INTERGENIC	G (0.17)	*CRPP1*	20 557	0.50 (0.08)	5.9×10^{-11}
rs16842525	1	159664624	INTERGENIC	C (0.19)	*CRPP1*	10 149	0.45 (0.07)	6.8×10^{-10}
rs1341666	1	159667059	INTERGENIC	G (0.19)	*CRPP1*	7714	0.44 (0.07)	6.8×10^{-10}
rs3093058	1	159685315	UPSTREAM	A (0.21)	*CRP*	-936	0.43 (0.07)	7.6×10^{-10}
rs6667499	1	159677654	UPSTREAM	A (0.21)	*CRPP1*	-2257	0.42 (0.07)	9.4×10^{-10}
rs9628671	1	159687972	UPSTREAM	T (0.21)	*CRP*	-3593	0.42 (0.07)	1.3×10^{-09}
rs16842520	1	159657850	INTERGENIC	G (0.18)	*CRPP1*	16 923	0.45 (0.07)	1.8×10^{-09}
rs12239267	1	159670928	DOWNSTREAM	T (0.18)	*CRPP1*	3845	0.44 (0.07)	2.4×10^{-09}

CRPP1, CRP pseudogene1.
[a]Position of the SNPs was derived from dbSNP genome build 37.
[b]CA, coded allele.
[c]CAF, coded allele frequency.
[d]rs16827466 has been merged into rs10494326.
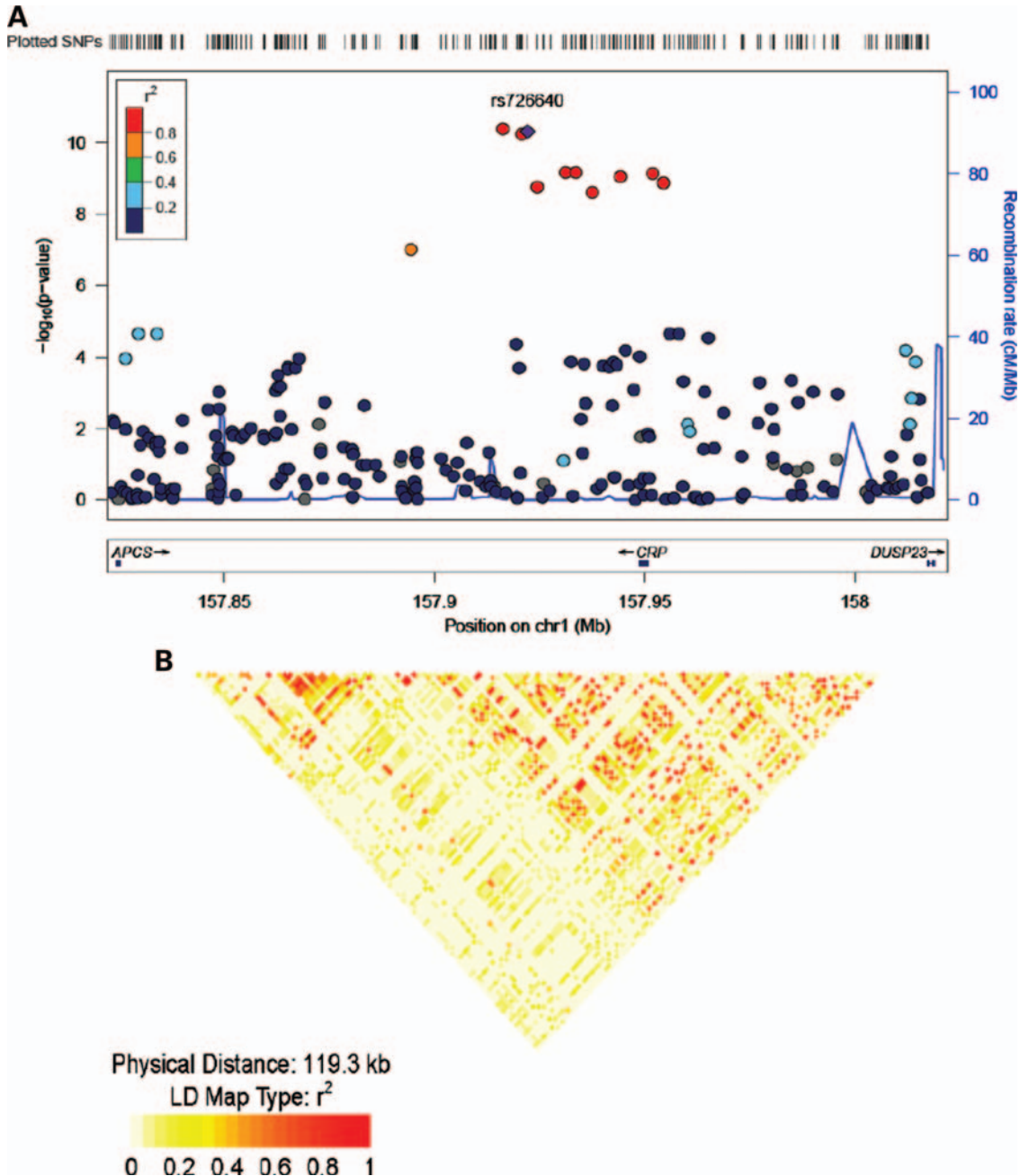
Figure 1. Regional plots of top-scoring SNPs identified in the GWAS and linkage disequilibrium (LD) plot in African Americans. (**A**) SNPs are plotted by chromosomal position (in megabases, according to NCBI build 36, dbSNP build 126) on the *x*-axis against association ($-10\log_{10}$ *P*-value) with hsCRP on the *y*-axis. Light blue lines show recombination rates taken from HapMap Phase II data to reflect local LD structure around the associated SNP. The lower panel shows gene annotations taken from the UCSC genome browser. The purple diamond represents one of the SNP that shows the strongest association in the plotted region. The LD of other SNPs and the best ranking SNP in the plotted region is shown in a scale from minimum (blue) to maximum (red). (**B**) The LD plot at the *CRP/CRPP1* loci in African Americans.

first two principal components (PCs) of genotypes are presented in Supplementary Material, Table S1. Additional adjustment for hypertension and smoking status (Supplementary Material, Table S2) did not change the association results.

In this sample of AA, 10 SNPs were associated with hsCRP at a genome-wide level of significance ($P < 5 \times 10^{-8}$) (Table 2, Fig. 1). All 10 SNPs were clustered in and near the *CRP* locus. Two of these SNPs (rs3093058, $P = 7.6 \times$

10^{-10} and rs9628671, $P = 1.3 \times 10^{-10}$) are upstream of the *CRP* locus in the promoter region; the other eight SNPs, including the strongest association signal in this study, were located downstream from the *CRP* locus in the *CRP pseudogene-1* (*CRPP1*) (Fig. 1). Three other SNPs (rs2027471, rs1341665, rs7553007) at the *CRP* locus resulted in *P*-values $<10^{-4}$ (Supplementary Material, Table S3). To investigate whether the association signals can be solely

Table 3. Replication of selected SNPs in a cohort of West Africans and meta-analysis results in the discovery and replication data sets

Chromosome	SNP	Genes	Africa America Diabetes Mellitus study (n = 486) CA[a] (CAF)[b]	P-value	Beta (SE)	Howard University Family Study (n = 837) CA[a] (CAF)[b]	P-value	Beta	Meta-analysis (n = 1323) P-value	Beta (SE)	Dir.	P(Het)*	I^2	P-value**
1	rs16827466[c]	CRPP1	T (0.48)	0.46	0.05 (0.07)	T (0.19)	4.3×10^{-11}	0.47 (0.07)	9.6×10^{-08}	0.27 (0.05)	++	4.3×10^{-05}	94	1.0×10^{-5}
1	rs7531832	CRPP1	G (0.22)	0.004	0.27 (0.09)	G (0.17)	5.9×10^{-11}	0.50 (0.08)	2.8×10^{-12}	−0.40 (0.06)	−	5.9×10^{-02}	72.1	0.37
1	rs726640	CRPP1	A (0.21)	0.0008	0.34 (0.10)	A (0.17)	5.0×10^{-11}	0.50 (0.08)	1.8×10^{-13}	0.44 (0.06)	++	2.1×10^{-01}	36.3	0.47
1	rs16842525	CRPP1	C (0.20)	0.002	0.31 (0.10)	C (0.19)	6.8×10^{-10}	0.44 (0.07)	5.9×10^{-12}	−0.40 (0.06)		2.9×10^{-01}	11.7	0.86
1	rs1341666	CRPP1	G (0.21)	0.002	0.31 (0.10)	G (0.19)	6.8×10^{-10}	0.44 (0.07)	7.6×10^{-12}	−0.40 (0.06)	−	2.6×10^{-01}	20.9	0.72
1	rs12239267	CRPP1	T (0.21)	0.005	0.28 (0.10)	T (0.18)	2.4×10^{-09}	0.44 (0.07)	6.0×10^{-11}	0.39 (0.06)	++	2.2×10^{-01}	34.9	0.59
1	rs6667499	CRPP1	A (0.21)	0.002	0.31 (0.10)	A (0.21)	9.4×10^{-10}	0.42 (0.07)	7.0×10^{-12}	0.38 (0.06)	++	3.5×10^{-01}	0	1.0
1	rs3093058	CRP	A (0.21)	0.003	0.29 (0.10)	A (0.21)	7.6×10^{-10}	0.43 (0.07)	1.0×10^{-11}	0.39 (0.06)	++	2.7×10^{-01}	19.2	1.0
1	rs9628671	CRP	T (0.23)	0.006	0.26 (0.10)	T (0.21)	1.3×10^{-09}	0.42 (0.07)	4.5×10^{-11}	0.38 (0.06)	++	1.9×10^{-01}	42.5	0.73
1	rs16827462	CRPP1	A (0.22)	0.001	0.33 (0.10)	A (0.19)	9.8×10^{-08}	0.40 (0.07)	4.2×10^{-10}	0.37 (0.06)	++	5.8×10^{-01}	0	0.60

[a]CA, coded allele.
[b]CAF, coded allele frequency.
[c]rs16827466 has been merged into rs10494326.
*P-value for heterogeneity test.
**P-value for the test of allele frequency difference between WA and AA.

explained by the leading SNP (rs3093058) in the *CRP* locus, we conditioned on this SNP and repeated the association analyses. There was a significant drop in the strength of association of the other *CRPP1/CRP* SNPs, suggesting that the leading SNP is driving most of the association. For example, after adjusting for the lead SNP, the most significant *P*-value in the *CRPP1/CRP* region was 0.02 (Supplementary Material, Table S4). We noted a number of SNPs with suggestive association (*P*-value $\sim 10^{-6}$) with hsCRP levels in this sample (Supplementary Material, Table S3). These SNPs are clustered in three distinct chromosomal regions: seven on chromosome 2 in *low-density lipoprotein receptor-related protein 1B* (*LRP1B*; Gene ID 53353); six on chromosome 5 near *mannosidase-alpha class 2 A member 1* (*MAN2A1*; Gene ID 4124) and five on chromosome 16 near (*excision repair cross-complementing rodent repair deficiency, complementation group-4*) *ERCC4* (Gene ID 2072).

The characteristics of 486 unrelated subjects in the WA replication sample are summarized in Supplementary Material, Table S5. Overall, the mean age in this cohort was 49 years and, like the discovery sample, women were heavier (by ~ 3.6 BMI units, $P < 0.0001$) and had slightly higher circulating CRP than men but this was not statistically significant ($P = 0.45$). Considering the 10 SNPs significantly and 1 SNP marginally associated with CRP in the discovery sample, 10 passed our quality control filters and were included in this analysis. The results of the replication conducted under an additive model with adjustment for age, gender, BMI and T2D are shown in Table 3. Of the 10 SNPs that were successfully genotyped in this cohort, nine were significantly associated with CRP at a *P*-value of <0.05 among WA. The combined analysis (meta-analysis) showed that all nine SNPs were significantly associated with CRP and with the same direction of effect in both samples. The test of heterogeneity showed some heterogeneity between the individual data sets, especially marked for rs16827466 that showed no association in the replication data set.

In the *in silico* replication study of the published CRP GWAS findings in populations of EA and Indian Asian ancestry, we successfully replicated the *CRP* locus (rs7553007, $P = 2.9 \times 10^{-5}$ and rs11265260, $P = 0.03$) in AA (Table 4). In parallel, we used an LD-based replication strategy to seek additional signals that were not directly replicated in our study. By searching all the SNPs in LD ($r^2 \geq 0.3$) in the HapMap CEU with each of the reported SNP in a 500 kb window around the SNP of interest, we found 11 additional SNPs significantly associated with hsCRP at the *CRP* locus (Table 5) and one SNP at the *LEPR* locus. Because some of the previously reported SNPs were discovered using a cohort comprising Europeans and Asian Indians, a search using HapMap (Gujarati Indians in Houston) GIH LD data found six *CRP* SNPs significantly associated with hsCRP in our study as well as one SNP in *LEPR* (rs3828033) (Table 5).

Finally, we attempted to refine the *CRP* signal taking advantage of the weak LD structure in our AA sample (Fig. 2). We generated LD plots centered on one of the SNPs (rs7553005) associated with the CRP level across populations and examined the plots in an EA population (HapMap CEU) as well as in our study. LD around rs7553005 extends an interval of 25 kb in HapMap CEU and contains all the SNPs previously

Table 4. Direct replication of SNPs previously reported to be associated with CRP in African Americans

Reference	Populations	SNPs	CA	Locus	Gene name	Reported P-value	Observed P-value	Reported beta	Observed beta	CA (CAF) in HUFS	CA (CAF) in CEU HapMap
(8)	Europeans and Indian Asians	rs7553007	A	1q23.2	*CRP*	8×10^{-44}	2.9×10^{-05a}	-0.21	-1.3^b	A (0.22)	A (0.33)
(14)	Self-reported European ancestry	rs11265260	?	1q23.2	*CRP*	7×10^{-6}	0.03^a	NR	1.04^b	G (0.08)	G (0.05)
(15)	Women of European ancestry	rs3091244[c]	?	1q23.2	*CRP*	6×10^{-28}	0.03^a	0.21	-0.05	T (0.26)	T (0.33)[d]
(8)	Europeans and Indian Asians	rs6700896	T	1p31.3	*LEPR*	3×10^{-14}	0.80^e	-0.15	-0.04^b	T (0.45)	T (0.35)
(15)	Women of European ancestry	rs1892534	A	1p31.3	*LEPR*	7×10^{-21}	0.88	-0.17	-0.004	A (0.45)	A (0.34)
(8)	Europeans and Indian Asians	rs4537545	T	1q21.3	*IL-6R*	2×10^{-14}	0.25	-0.12	-0.30^b	C (0.40)	T (0.37)
(15)	Women of European ancestry	rs8192284[f]	?	1q23.3	*IL-6R*	2×10^{-8}	0.41^e	-0.10	0.03	C (0.12)	C (0.32)[d]
(15)	Women of European ancestry	rs780094	A	2p23.3	*GCKR*	7×10^{-15}	0.35	0.14	0.03	A(0.16)	A (0.39)
(14)	Self-reported European ancestry	rs1169310	A	12q24.31	*HNF1A*	2×10^{-8}	0.89^e	NR	0.04	A (0.14)	A (0.38)
(15)	Women of European ancestry	rs7310409	A	12q24.31	*HNF1A*	7×10^{-17}	0.52^e	-0.15	0.02	A (0.31)	A (0.42)
(8)	Europeans and Indian Asians	rs1183910	T	12q24.31	*HNF1A*	1×10^{-30}	0.71^e	-0.14	0.13^b	T (0.12)	T (0.29)
(15)	Women of European ancestry	rs10778213	G	12q23.2	Unknown	1×10^{-10}	0.70	-0.12	0.01	A (0.27)	G (0.51)
(14)	Self-reported European ancestry	rs2075650	?	19q13.32	*APOE*	1×10^{-7}	NA	NR	NA	—	—
(15)	Women of European ancestry	rs769449	?	19q13.32	*APOE*	9×10^{-21}	NA	-0.26	NA	—	—
(8)	Europeans and Indian Asians	rs4420638	G	19q13.32	*APO complex*	5×10^{-27}	0.46	-0.22	-0.21	G (0.21)	G (0.18)

NA, SNP not available in HUFS; NR, not reported; ?, risk allele was not reported.
[a]SNP replicated in HUFS.
[b]Observed beta was converted to adhere to uniformity in units and log transformation across studies. Our beta was multiplied by a factor of 4.35 ($\sim$10/2.3) to convert it from mg/dl to mg/l and from natural log to log10.
[c]SNP was not available in Affymetrix data set but was genotyped in HUFS on Sequenom platform.
[d]Frequency from 1000 Genome.
[e]Imputed SNP.
[f]rs8192284 has been merged into rs2228145.

associated with CRP in EA but was restricted to 8 kb in our AA and contained two SNPs (rs7553007, rs11265260) of the three major SNPs associated with CRP in EA.

DISCUSSION

CRP is a well-known acute phase reactant in the inflammatory process whose levels are elevated not just in infection, but also in chronic non-communicable diseases such as CVD (4). Most studies of the genetic variants influencing circulating CRP levels have used a candidate gene approach with only a handful using a genome-wide strategy. In the present study, we conducted the first genome-wide association analysis of CRP levels in AA. In common with the previous GWAS in populations of EA and Asian ancestry (8,14–18), we found that the strongest associations were with SNPs located within and around the *CRP* locus. Similar to findings in Europeans, the SNPs with the lowest P-values in this study of AA (rs3093058, rs9628671) are located in the *CRP* promoter region; however, the risk alleles for these two SNPs are very rare in populations of EA as demonstrated in the HapMap CEU reference and NHANES data (21). Two previous candidate gene studies conducted in AA—Carlson *et al.* (22) and

Table 5. SNPs associated with hsCRP levels in African Americans using the LD-based replication method

SNPs[a]	Locus	Gene name	Reported P-value	SNPs in LD	Beta$_{AA}$/beta $_{EA}$	Uncorrected P-value	Corrected P-value	df
Using HapMap CEU LD								
rs7553007[b]	1q23.2	*CRP*	8×10^{-44}	rs2592887	$-0.25/-0.16$	4.5×10^{-05}	7.1×10^{-05}	1.58
				rs1470515	-0.24/NA	1.9×10^{-04}	3.2×10^{-04}	1.58
				rs2808624	-0.28/NA	1.4×10^{-04}	2.1×10^{-04}	1.58
				rs11265257	-0.28/NA	1.5×10^{-04}	2.4×10^{-04}	1.58
				rs876537	-0.28/NA	1.9×10^{-04}	3.0×10^{-04}	1.58
				rs2808628	-0.27/NA	1.4×10^{-04}	2.2×10^{-04}	1.58
				rs2808629	-0.28/NA	1.7×10^{-04}	2.6×10^{-04}	1.58
				rs2794520	-0.28/NA	6.4×10^{-05}	1.0×10^{-04}	1.58
				rs1205	$-0.29/-0.21$	9.5×10^{-05}	1.5×10^{-04}	1.58
				rs2027471[c]	$-0.30/-0.20$	2.3×10^{-05}	3.7×10^{-05}	1.58
				rs1341665[c]	-0.30/NA	2.3×10^{-05}	3.7×10^{-05}	1.58
rs1892534	1p31.3	*LEPR*	7×10^{-21}	rs4291477	-0.20/NA	4.1×10^{-03}	7.6×10^{-03}	1.87
rs6700896[b]	1p31.3	*LEPR*	3×10^{-14}	rs4291477	-0.20/NA	4.1×10^{-03}	7.7×10^{-03}	1.89
Using HapMap GIH LD								
rs7553007[b]	1q23.2	*CRP*	8×10^{-44}	rs2592887	$-0.25/-0.16$	4.5×10^{-05}	8.5×10^{-05}	1.87
				rs1572970	-0.23/NA	1.9×10^{-04}	3.5×10^{-04}	1.87
				rs876537	-0.27/NA	1.9×10^{-04}	3.6×10^{-04}	1.87
				rs2808629	-0.28/NA	1.7×10^{-04}	3.1×10^{-04}	1.87
				rs2794520	-0.28/NA	6.4×10^{-05}	1.2×10^{-04}	1.87
				rs1205	$-0.29/-0.21$	9.5×10^{-05}	1.78×10^{-04}	1.87
rs6700896[b]	1p31.3	*LEPR*	3×10^{-14}	rs3828033	-0.14/NA	0.01	0.03	2.16

Beta $_{EA,}$ beta in EA are taken from reference (15); beta$_{AA}$, beta in AA; NA; not available; df, degree of freedom.
[a]Only SNPs that showed association with CRP levels are reported.
[b]SNPs were among the top 100 SNPs associated with hsCRP in initial analysis.
[c]SNPs reported in Europeans and Asian Indians.

Crawford *et al.* (21)—confirmed rs3093058 as being one of three *CRP* SNPs contributing to ethnic differences in CRP levels. The index SNP (rs3093058) associated with CRP in our study is different from those previously reported in EA (8,14–16) but tagging the same signal as shown by the LD plots of EA and AA (Fig. 2). The difference seen across populations in the reported SNPs associated with a phenotype is not uncommon, especially if one of the populations has an African ancestry background that is characterized by shorter LD blocks (23). These findings illustrate an important issue often observed in genetic studies of admixed populations such as AA. The reported genetic variants associated with a phenotype may be population-specific due to differences in the population's ancestral background, haplotype structure and allele frequencies as demonstrated by our study and by re-sequencing of the *CRP* locus in several populations (24).

The confirmation or replication of reported signals is often more successful in populations with a similar ancestry as the original discovery population (25). Therefore, we chose to replicate our findings in WA and successfully replicated all but one SNP (rs16827466) with directionally consistent effect size estimates. The lack of replication of rs16827466 in WA may be due to heterogeneity in both allele frequencies and effect size (as shown in Table 3). Although heterogeneity between individual studies can affect replication efforts, the robustness of the *CRP* association is quite remarkable, seeing that our study successfully replicated three previously reported CRP-associated SNPs (rs3091244, rs11265260 and rs7553007) despite the sharp differences in allele frequency (especially rs3091244 and rs7553007) between EA populations and our study sample. Differences in allele frequency between populations can easily obscure associations and it

has been estimated that a change in minor allele frequency (MAF) <0.1 can result in a drop in power to replicate from 80% to as low as 20% (26). In this study, we declared replication of a locus based on an adjusted P-value (<0.05) and consistent effect size direction. It should be noted that the weak LD structure seen in our AA samples is not always an advantage when doing 'exact replication' because the large LD block containing all the signals in EA was segmented into smaller LD blocks in AA. On the other hand, this weak LD in African ancestry populations is often an advantage in following up an association signal because it could result in considerable refinement of the original locus, as has been successfully demonstrated with loci associated with height, uric acid and bilirubin (23,27,28). In the present study, we were able to refine the localization of the *CRP* signal by successfully fine-mapping the 25 kb interval of LD around the *CRP* locus in the HapMap CEU sample by more than 2-fold to 8 kb in our AA sample (assuming a single causal variant in the region common to both populations). However, it should be noted that fine mapping may be complicated by multiple causal variants in the same region (some of which may be population-specific). Future studies using sequence-based approaches would probably provide examples of such situations.

The present study did not find novel loci, which is not surprising given the limited power of our small discovery panel for detecting weak effect sizes. On the other hand, using an LD-based replication strategy, we successfully replicated additional SNPs in the *LEPR* gene previously reported to be associated with CRP levels by two independent studies in Europeans (8,15,16). In addition to *CRP* and *LEPR*, other genes have been associated with CRP levels in Europeans
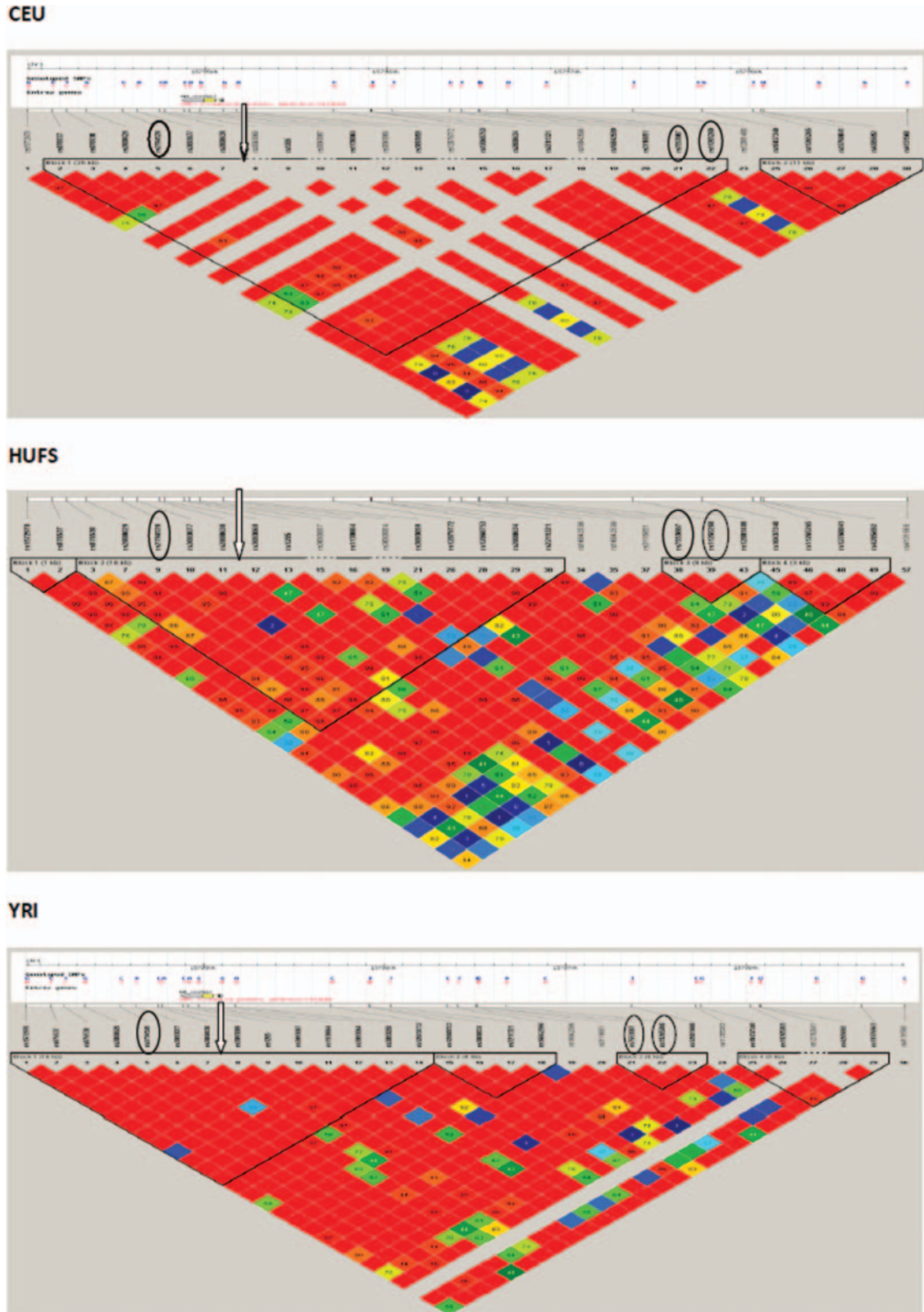
Figure 2. Comparison of linkage disequilibrium in the HapMap Phase II CEU, YRI data and our HUFS sample at *CRP* locus. Circled SNPs are the top-ranking SNPs associated with CRP levels in populations of European ancestry. Arrow indicates the index SNP (rs7553007) used to build the LD plot in the region. Arrow indicates the position of the index SNP in our data (rs3093058).

including *HNF1A* (gene encoding a hepatic transcriptional regulator), *IL6R*, *APO* complex and *GCKR* (gene encoding a hepatic and pancreatic glucokinase regulator protein). However, none of these was replicated in the present study. It should be noted that the associated SNPs vary considerably in frequency between populations. For example, rs1183910 (T) in *HNF1A* has an MAF of 0.292 in HapMap CEU in contrast to 0.092 and 0.048 in the HapMap (Yoruba in Ibadan) YRI and AA from Perlegen samples, respectively. This, apart from sample size issues, is a source of heterogeneity between studies.

Besides the commonly reported loci, we found suggestive associations between serum CRP levels and *LRP1B*, *MAN2A1*, *ERCC4*. Although the precise biological mechanisms by which these genes could influence CRP levels are currently unclear, there is indirect evidence that they play a role in CRP regulation through processes that are mediated by obesity. For example, *ERCC4*, a gene involved in nucleotide excision repair, has been associated with abdominal obesity in the GWAS of metabolic syndrome in Indian Asian men (29) and *LRP1B*, which encodes an LRP1B, is associated with BMI in a meta-analysis (conducted by GWAS) of 249 796 individuals of EA (30).

In conclusion, we have conducted a genome-wide association analysis of CRP levels in AA and confirmed that the genetic variants with the strongest effects are all within and around the *CRP* locus. We also fine-mapped the *CRP* signal to an 8 kb interval. Although the power to detect weak effect sizes was limited by the sample size, we were able to successfully replicate a secondary locus, the *LEPR* gene. As additional studies of AA and other ethnicities are conducted, meta-analysis would facilitate the discovery of more loci regulating circulating CRP levels.

MATERIALS AND METHODS

The participants used in this study were drawn from a larger population-based study of AA from the Washington, DC metropolitan area. This study is described in detail elsewhere (31). Briefly, the goal of the study was to enroll and examine families as well as unrelated AA to study the genetics of common complex traits. Participants were not ascertained based on any phenotype. During clinic visits, participants provided written informed consent and completed a set of questionnaires. Anthropometric and blood pressure measurements were taken and blood samples were drawn for measurements of several biochemical parameters. All individuals included in this GWAS for CRP were unrelated. The study protocol was approved by Howard University Internal Review Board (IRB).

HsCRP was measured using a clinical auto analyzer (COBAS Integra 400 plus). This is a particle-enhanced turbidimetric assay. Human CRP is made to react with latex particles coated with monoclonal anti-CRP antibodies and the precipitate is determined turbidimetrically at 552 nm. The measuring range for this assay is 0.01–2 mg/dl.

Genome-wide genotyping was performed using the Affymetrix® Genome-Wide Human SNP Array 6.0. DNA samples were prepared and hybridized following the manufacturer's instructions. After processing, chips were scanned and

genotype calls were made using the Birdseed 2 Algorithm. All samples used in the analysis achieved a chip-wide call rate of ≥95%. Individual SNPs were excluded if they had a call rate of <95% across all individuals, an MAF of ≤0.01 or had a Hardy–Weinberg equilibrium test *P*-value of $<1 \times 10^{-3}$. The average call rate for this set of SNPs in these individuals was 99.55%. The concordance of blind duplicates was 99.74%. A check for population stratification or structure was sought by conducting non-parametric clustering of genotypes as previously described (32). The genomic inflation factor (λ) of 0.9961 and the quantile–quantile plot (Supplementary Material, Fig. S1) showed minimal evidence of population stratification. To adjust for any potential residual population stratification and admixture, all association analyses were adjusted for the first two PCs of the genotypes computed using EIGENSOFT (33).To increase genomic coverage, imputation was performed using the MACH algorithm and with the HapMap reference populations as previously described (23). After quality control filters were applied, ~1.6 million successfully imputed SNPs were available, bringing the total number of SNPs to 2 366 856 genotyped and imputed autosomal SNPs. CRP values were natural log-transformed to assure model assumptions of normally distributed residuals. Genome-wide association analyses for ln(CRP) were conducted using multiple linear regression models as implemented in PLINK v1.07. Analysis was done under an additive genetic model, with one model adjusting for age, sex, BMI, T2D and the first two PCs of genotypes, whereas a second model adjusted for these covariates as well as smoking status and hypertension. An association *P*-value of $<5 \times 10^{-8}$ was considered genome-wide significant.

The replication in WA of our top-scoring GWAS findings was conducted in an independent set of unrelated subjects enrolled as part of the Africa America Diabetes Mellitus (AADM) study, described fully elsewhere (34,35). The AADM study is an international collaboration between American and African investigators to study the genetic epidemiology of T2D in WA, with the participants coming from two WA countries (Nigeria and Ghana). The selected SNPs were genotyped using Sequenom Homogenous MassEXTEND or iPLEX Gold SBE assays at the National Human Genome Research Institute (NHGRI). The quality control filters previously described for the discovery sample were applied to this data set. The association between each SNP and CRP level was assessed using similar regression models under an additive genetic model controlling for the same covariates as in the discovery sample; replication significance was set at a *P*-value of <0.05. The meta-analysis of the discovery and replication cohorts was conducted using the METAL software (36). The evidence of association was combined in two-sided *P*-values using a fixed effects model. The sample size of each study was used as weight and the sign of the beta value was used as the direction of the association of the minor allele. The heterogeneity test was performed to evaluate between-study heterogeneity and a *P*-value of χ^2 ($P < 0.05$) was used to declare significance (37).

To conduct *in silico* replication of previously published GWAS CRP signals (8,14,15) in AA [Howard University Family Study, (HUFS)], we used a two-step strategy. Of the 15 commonly reported SNPs, 12 were available (either

genotyped or imputed) in this study and we directly checked their association statistics and *P*-values in our AA data set. As a second step, we used LD to find additional markers in LD with the reported SNPs. We queried a 500 kb window centered on the reported SNP for the SNPs in LD ($r^2 \geq 0.3$) with the query SNP as previously described by Ramos *et al.* ([38]) and then obtained association statistics for the resulting SNPs. The appropriate HapMap reference population for the source GWAS (CEU for European, GIH for Indian Asian) was used to determine LD ([8]). To correct for multiple testing of correlated test statistics for SNPs in LD, the number of independent tests was estimated by computing the number of effective degrees of freedom (df) using the spectrally decomposed covariance matrix ([39]). *P*-values were adjusted for the number of independent tests in the region. *P*-values for replication were considered significant if the adjusted $P < 0.05$.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Clyne, B. and Olshaker, J.S. (1999) The C-reactive protein. *J. Emerg. Med.*, **17**, 1019–1025.
2. Ganapathi, M.K., May, L.T., Schultz, D., Brabenec, A., Weinstein, J., Sehgal, P.B. and Kushner, I. (1988) Role of interleukin-6 in regulating synthesis of C-reactive protein and serum amyloid A in human hepatoma cell lines. *Biochem. Biophys. Res. Commun.*, **157**, 271–277.
3. Castell, J.V., Gomez-Lechon, M.J., David, M., Andus, T., Geiger, T., Trullenque, R., Fabra, R. and Heinrich, P.C. (1989) Interleukin-6 is the major regulator of acute phase protein synthesis in adult human hepatocytes. *FEBS Lett.*, **242**, 237–239.
4. Ridker, P.M. (2009) C-reactive protein: eighty years from discovery to emergence as a major risk marker for cardiovascular disease. *Clin. Chem.*, **55**, 209–215.
5. Weinhold, B. and Ruther, U. (1997) Interleukin-6-dependent and -independent regulation of the human C-reactive protein gene. *Biochem. J.*, **327**, 425–429.
6. Fox, E.R., Benjamin, E.J., Sarpong, D.F., Rotimi, C.N., Wilson, J.G., Steffes, M.W., Chen, G., Adeyemo, A., Taylor, J.K., Samdarshi, T.E. *et al.* (2008) Epidemiology, heritability, and genetic linkage of C-reactive protein in African Americans (from the Jackson Heart Study). *Am. J. Cardiol.*, **102**, 835–841.
7. Florez, H., Castillo-Florez, S., Mendez, A., Casanova-Romero, P., Larreal-Urdaneta, C., Lee, D. and Goldberg, R. (2006) C-reactive protein is elevated in obese patients with the metabolic syndrome. *Diabetes Res. Clin. Pract.*, **71**, 92–100.
8. Elliott, P., Chambers, J.C., Zhang, W., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund, P., Engert, J.C., Bennett, D. *et al.* (2009) Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA*, **302**, 37–48.
9. Ridker, P.M., Cushman, M., Stampfer, M.J., Tracy, R.P. and Hennekens, C.H. (1997) Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N. Eng. J. Med.*, **336**, 973–979.
10. Yeh, E.T. (2005) High-sensitivity C-reactive protein as a risk assessment tool for cardiovascular disease. *Clin. Cardiol.*, **28**, 408–412.
11. Lee, C.C., You, N.C., Song, Y., Hsu, Y.H., Manson, J., Nathan, L., Tinker, L. and Liu, S. (2009) Relation of genetic variation in the gene coding for C-reactive protein with its plasma protein concentrations: findings from the Women's Health Initiative Observational Cohort. *Clin. Chem.*, **55**, 351–360.
12. Dehghan, A., Kardys, I., de Maat, M.P., Uitterlinden, A.G., Sijbrands, E.J., Bootsma, A.H., Stijnen, T., Hofman, A., Schram, M.T. and Witteman, J.C. (2007) Genetic variation, C-reactive protein levels, and incidence of diabetes. *Diabetes*, **56**, 872–878.
13. Hage, F.G. and Szalai, A.J. (2007) C-reactive protein gene polymorphisms, C-reactive protein blood levels, and cardiovascular disease risk. *J. Am. Coll. Cardiol.*, **50**, 1115–1122.
14. Reiner, A.P., Barber, M.J., Guan, Y., Ridker, P.M., Lange, L.A., Chasman, D.I., Walston, J.D., Cooper, G.M., Jenny, N.S., Rieder, M.J. *et al.* (2008) Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am. J. Hum. Genet.*, **82**, 1193–1201.
15. Ridker, P.M., Pare, G., Parker, A., Zee, R.Y., Danik, J.S., Buring, J.E., Kwiatkowski, D., Cook, N.R., Miletich, J.P. and Chasman, D.I. (2008) Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am. J. Hum. Genet.*, **82**, 1185–1192.
16. Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P. *et al.* (2011) Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*, **123**, 731–738.
17. Okada, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Kamatani, Y., Hosono, N., Tsunoda, T., Matsuda, K., Tanaka, T., Kubo, M. *et al.* (2011) Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. *Hum. Mol. Genet.*, **20**, 1224–1231.
18. Wu, Y., McDade, T.W., Kuzawa, C.W., Borja, J., Li, Y., Adair, L.S., Mohlke, K.L. and Lange, L.A. (2011) Genome-wide Association with C-Reactive Protein Levels in CLHNS: evidence for the CRP and HNF1A loci and their interaction with exposure to a pathogenic environment. *Inflammation*, doi:10.1007/s10753-011-9348-y
19. Patel, D.A., Srinivasan, S.R., Xu, J.H., Li, S., Chen, W. and Berenson, G.S. (2006) Distribution and metabolic syndrome correlates of plasma C-reactive protein in biracial (black-white) younger adults: the Bogalusa Heart Study. *Metab. Clin. Exp.*, **55**, 699–705.
20. Rosenberg, N.A., Huang, L., Jewett, E.M., Szpiech, Z.A., Jankovic, I. and Boehnke, M. (2010) Genome-wide association studies in diverse populations. *Nat. Rev. Genet.*, **11**, 356–366.
21. Crawford, D.C., Sanders, C.L., Qin, X., Smith, J.D., Shephard, C., Wong, M., Witrak, L., Rieder, M.J. and Nickerson, D.A. (2006) Genetic variation is associated with C-reactive protein levels in the Third National Health and Nutrition Examination Survey. *Circulation*, **114**, 2458–2465.

22. Carlson, C.S., Aldred, S.F., Lee, P.K., Tracy, R.P., Schwartz, S.M., Rieder, M., Liu, K., Williams, O.D., Iribarren, C., Lewis, E.C. *et al.* (2005) Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *Am. J. Hum. Genet.*, **77**, 64–77.

23. Shriner, D., Adeyemo, A., Gerry, N.P., Herbert, A., Chen, G., Doumatey, A., Huang, H., Zhou, J., Christman, M.F. and Rotimi, C.N. (2009) Transferability and fine-mapping of genome-wide associated loci for adult height across human populations. *PLoS ONE*, **4**, e8398.

24. Crawford, D.C., Yi, Q., Smith, J.D., Shephard, C., Wong, M., Witrak, L., Livingston, R.J., Rieder, M.J. and Nickerson, D.A. (2006) Allelic spectrum of the natural variation in CRP. *Hum. Genet.*, **119**, 496–504.

25. Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E. *et al.* (2007) Replicating genotype-phenotype associations. *Nature*, **447**, 655–660.

26. Greene, C.S., Penrod, N.M., Williams, S.M. and Moore, J.H. (2009) Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE*, **4**, e5639.

27. Charles, B.A., Shriner, D., Doumatey, A., Chen, G., Zhou, J., Huang, H., Herbert, A., Gerry, N.P., Christman, M.F., Adeyemo, A. *et al.* (2011) A genome-wide association study of serum uric acid in African Americans. *BMC Med. Genomics*, **4**, 17.

28. Chen, G., Ramos, E., Adeyemo, A., Shriner, D., Zhou, J., Doumatey, A.P., Huang, H., Erdos, M.R., Gerry, N.P., Herbert, A. *et al.* (2011) UGT1A1 is a major locus influencing bilirubin levels in African Americans. *Eur. J. Hum. Genet.*, doi:10.1038/ejhg.2011.206.

29. Zabaneh, D. and Balding, D.J. (2010) A genome-wide association study of the metabolic syndrome in Indian Asian men. *PLoS ONE*, **5**, e11961.

30. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Allen, H.L., Lindgren, C.M., Luan, J.A., Magi, R. *et al.* (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.*, **42**, 937–948.

31. Adeyemo, A., Gerry, N., Chen, G., Herbert, A., Doumatey, A., Huang, H., Zhou, J., Lashley, K., Chen, Y., Christman, M. *et al.* (2009) A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet.*, **5**, e1000564.

32. Gao, X. and Starmer, J.D. (2008) AWclust: point-and-click software for non-parametric population structure analysis. *BMC Bioinf.*, **9**, 77.

33. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.

34. Rotimi, C.N., Dunston, G.M., Berg, K., Akinsete, O., Amoah, A., Owusu, S., Acheampong, J., Boateng, K., Oli, J., Okafor, G. *et al.* (2001) In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. *Ann. Epidemiol.*, **11**, 51–58.

35. Rotimi, C.N., Chen, G., Adeyemo, A.A., Furbert-Harris, P., Parish-Gause, D., Zhou, J., Berg, K., Adegoke, O., Amoah, A., Owusu, S. *et al.* (2004) A genome-wide search for type 2 diabetes susceptibility genes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. *Diabetes*, **53**, 838–841.

36. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.

37. Higgins, J.P., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *BMJ*, **327**, 557–560.

38. Ramos, E., Chen, G., Shriner, D., Doumatey, A., Gerry, N.P., Herbert, A., Huang, H., Zhou, J., Christman, M.F., Adeyemo, A. *et al.* (2011) Replication of genome-wide association studies (GWAS) loci for fasting plasma glucose in African-Americans. *Diabetologia*, **54**, 783–788.

39. Bretherton, C.S., Widmann, M., Dymnikov, V.P., Wallace, J.M. and Bladé, I. (1999) The effective number of spatial degrees of freedom of a time-varying field. *J. Clim.*, **12**, 1990–2009.