Journal of Genetics (2019) 98:20
https://doi.org/10.1007/s12041-019-1065-6

RESEARCH ARTICLE

Genomewide association study for C-reactive protein in Indians replicates known associations of common variants

GAURI PRASAD[1,2], ANIL K. GIRI[1,2], INDICO CONSORTIUM, ANALABHA BASU[3], NIKHIL TANDON[4]* and DWAIPAYAN BHARADWAJ[2,5]*

[1] *Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi 110 020, India*
[2] *Academy of Scientific and Innovative Research, CSIR-Institute of Genomics and Integrative Biology Campus, New Delhi 110 020, India*
[3] *National Institute of Biomedical Genomics, Netaji Subhas Sanatorium (Tuberculosis Hospital), Kalyani, West Bengal 741 251, India*
[4] *Department of Endocrinology and Metabolism, All India Institute of Medical Sciences, New Delhi 110 029, India*
[5] *Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi 110 067, India*
*For correspondence. E-mail: Nikhil Tandon, nikhil_tandon@hotmail.com; Dwaipayan Bharadwaj, db@jnu.ac.in.

Received 19 July 2018; revised 26 September 2018; accepted 22 October 2018; published online 26 February 2019

Abstract. Elevated C-reactive protein (CRP) serves as an independent biomarker for acute and chronic inflammation, and is also associated with metabolic diseases. Genomewide loci regulating CRP level in Indian population, a high-risk group for metabolic illness, is unexplored. Therefore, we aimed to discover common polymorphisms associated with plasma CRP level in 4493 Indians of Indo-European origin using genomewide association study. Genomewide strong associations of two known intronic variants in hepatocyte nuclear factor-1 α gene (*HNF1A*) were identified among Indian subjects. We also detected prior associations of several variants in/near metabolic and inflammatory process genes: *APOC1*, *LEPR*, *CRP*, *HNF4A*, *IL6R* and *APOE* with modest associations. This study confirms that Indians from Indo-European origin display similar core universal genetic factors for CRP levels.

Keywords. C-reactive protein; genomewide association study; Indians; metabolic diseases.

Introduction

C-reactive protein (CRP) is a major acute phase protein released from hepatocytes in response to any kind of stress in the body (Black *et al.* 2004). Elevated level of CRP independently confer risk for metabolic syndrome, cardiovascular disease and type-2 diabetes mellitus (T2DM), globally including Indians, a high-risk group for metabolic disorders (Ridker *et al.* 2004; Mahajan *et al.* 2009, 2012). Several factors such as age, gender, ethnicity, obesity and lifestyle affects CRP level (Ford *et al.* 2003; Khera *et al.* 2005; Rojo-Martinez *et al.* 2013), but genetic factors also account for substantial heritability (∼35–56%) (Pankow *et al.* 2001; Wessel *et al.* 2007).

Ample evidences exist for higher CRP level in Indians as compared to Europeans, Caucasians, Chinese and Malays (Chambers *et al.* 2001; Chandalia *et al.* 2003; Khoo *et al.* 2011). Most large-scale whole-genome assays were conducted to link circulating CRP level to genes and regulatory regions involved in metabolic, weight homoeostasis and immune pathways in European population (Benjamin *et al.* 2007; Reiner *et al.* 2008; Ridker *et al.* 2008; Elliott *et al.* 2009; Dehghan *et al.* 2011). Few genomewide

Members of the INDICO consortium are listed in the supplementary material.
DB designed the study and overall management of the project. GP, AKG and INDICO collected samples and conducted the experiments. GP and AKG analysed and interpreted data. GP drafted the manuscript. DB and AB critically revised the manuscript and checked statistical procedures. DB and NT approved the final version of the manuscript. All authors read and approved the final manuscript.

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12041-019-1065-6) contains supplementary material, which is available to authorized users.

association studies (GWAS) have included nonEuropean subjects (Okada *et al.* 2011; Reiner *et al.* 2012; Wu *et al.* 2012; Dorajoo *et al.* 2013; Kong and Lee 2013; Kocarnik *et al.* 2014). Some studies (Elliott *et al.* 2009; Dorajoo *et al.* 2013) identified genetic variants associated with CRP level inclusive of Indian subjects and identified known variants in *LEPR*, *IL6R*, *CRP*, *HNF1A*, *APOC1*, *APOE*, *GCKR*, *IL6* and *IL1F10*, but none of these Indian subjects were living in India. Indians have a different food habit, lifestyle patterns and genetic heterogeneity from other populations of the world (Tabassum *et al.* 2011, 2013; Giri *et al.* 2014), that may account for altered CRP level in resident Indians as compared to nonresident Indians. Such differences highlight the need for a separate genetic study of CRP level in this population. This may also help in understanding the genetics of complex disorders that are associated with CRP level (Plomin *et al.* 2009). Although, earlier we ourselves queried for known genetic determinants of CRP in Indians using the candidate gene-based approach (Mahajan *et al.* 2011). However, we did not look for contribution of genomewide variants that could explain population-specific genetic players regulating CRP.

The present study aim to investigate genomewide common variants associated with the plasma CRP level in a large sample size of nearly 4493 Indian subjects of Indo-European origin.

Materials and methods

Study population

Total 4493 apparently healthy subjects, who were volunteers of Indian Diabetes Consortium (INDICO) (INdian DIabetes Consortium 2011), residing in and around Delhi were enrolled for the current study. Blood samples from individuals were collected after overnight fasting. Subjects are well characterized for biochemical and anthropometric measures as mentioned previously (Mahajan *et al.* 2009). Plasma CRP was measured using an ELISA kit (Biocheck, USA) and Cobas Integra 400 Plus (Roche Diagnostics, Mannheim, Germany) via immunoturbidimetric assay. CRP values are reported in mg/L.

This study was approved by human research ethics committees of All India Institute of Medical Sciences and CSIR-Institute of Genomics and Integrative Biology, Delhi, India. This study was carried out in accordance with principles of Helsinki Declarations. Prior informed written consent was obtained from all study subjects before their participation in the study.

Discovery phase

Discovery phase (stage I) subjects were genotyped as part of the T2DM GWAS conducted in our laboratory using Illumina Human 610-quad bead chips (Tabassum *et al.* 2013). Normoglycemic subjects, who were controls in T2DM GWAS, were chosen for the current study. Data for the subjects were generated as described previously (Tabassum *et al.* 2013). Stringent sample and single-nucleotide polymorphism (SNP) quality controls were followed. Subjects with call rate <0.95, sex disparities and low/overt heterozygosity (3 standard deviation (SD) away from mean value) were excluded. Further, SNPs with minor allele frequency (MAF) <0.01 were removed. Among SNPs with MAF>0.01, those with a call rate of <0.99 and Hardy–Weinberg equilibrium (HWE) P value $< 1 \times 10^{-6}$ were excluded. Mitochondrial and sex chromosome SNPs were also discarded. SNP pruning was performed using –indep-pairwise command with r^2 of 0.2 and a window size of 50 using PLINK v1.07 (Purcell *et al.* 2007). Identity-by-descent analysis was performed using 118,496 independent pruned SNPs. Related subjects were detected based on Pi-hat score >0.1875 and member of pair with a lower sample call rate was removed. Principal component analysis was used to detect population outliers. First 10 principal components were used to identify 29 subjects as population outliers (6 SD away from mean value) that were removed. We limited our analysis on samples with CRP<10 mg/L not taking lipid lowering drugs. CRP values were inverse normalized using R (http://www.r-project.org/). Finally, association of 519,533 autosomal SNPs with transformed CRP values were assessed in 1097 participants using the linear regression model in PLINK. Age, sex, body mass index (BMI), smoking status, alcohol consumption status, tobacco chewing status and first three principal components were used as covariates in the model. Quantile–quantile (QQ) analysis was used to compare the theoretical and observed distribution of P values using the qqman package in R (Turner 2014).

Validation phase and meta-analysis

Stage I signals obtained for association with CRP level ($P<10^{-4}$) besides previously known variants were queried in independent 3396 normoglycemic individuals for association with CRP trait. Selected SNP variants for CRP were genotyped along with other variants for ongoing GWAS for multiple other biochemical traits through Illumina Golden Gate technology (Shen *et al.* 2005). Overall, 204 samples (6%) were genotyped as replicates to assess experimental accuracy (replication error <0.01 was observed). Samples having poor call rate (<90%) were discarded. SNPs with GenTran score <0.6, cluster separation score <0.4 and call rate <90% were excluded. SNPs with MAF<0.01 and HWE $P < 1 \times 10^{-6}$ were also discarded prior to genetic analysis.

Moreover, we also evaluated the combined association status of five *cis*-acting (proximal to CRP gene) and 10 *trans*-acting variants (distal to CRP gene) that were genotyped in stage I of the current study and previously

genotyped in another independent Indo-European sample set by our group (Mahajan *et al.* 2011). *Cis*-acting variants were rs3093077, rs4131568, rs2794520, rs2592887 and rs12093699 in the *CRP* gene. *Trans*-acting variants were rs1892534, rs12753193, rs753947 (*LEPR*), rs4129267 (*IL6R*), rs2464196, rs1169300, rs735396 (*HNF1A*), rs7953 249 (*HNF1A-AS1*), rs2075650 (*APOE*) and rs10778213 (*ASCL1-C12orf42*). We used 385 subjects from our previous study as replication cohort for all these variants except for variant rs1892534. For variant rs1892534, we used 1026 subjects from previous study as replication cohort. Meta-analysis was conducted with METAL using the fixed effect inverse variance method (Willer *et al.* 2010).

Power calculation

The statistical power of the study was calculated using Quanto software (http://biostats.usc.edu/software). Additive model of inheritance was assumed for a range of allele frequencies 0.01–0.50. Two-tailed test at a significance level of 0.05 and effect sizes ranging from 0.0001 to 0.66, obtained from the literature were used for power calculation, and it was found that the present study is sufficiently powered to detect association of genetic variants (figure 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/). Average CRP level of 2.45 mg/L and SD of 2.15 mg/L were used to calculate the power of the study. MAF was calculated from combined data for discovery and replication phase subjects.

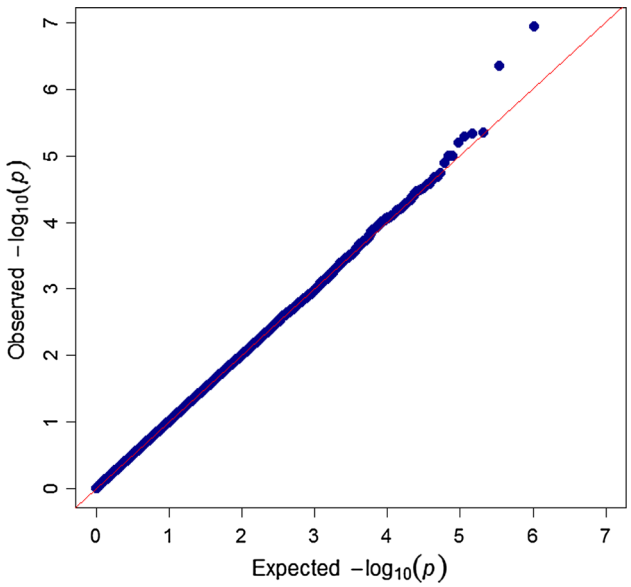

Figure 1. QQ plot comparing the calculated distribution of $-log_{10}P$ values observed for association of SNPs with CRP level in stage I with theoretical distribution expected under the null hypothesis of no association with the phenotype. The genomic inflation factor (λ) was calculated to be 1.003.

Phenotypic variance explained by SNPs

To calculate proportion of heritability explained by the identified SNPs, GCTA tool (Yang *et al.* 2011) was used. We performed restricted maximum likelihood (REML) analysis followed by One GRM (–grm option) for quantitative traits. Inverse normalized CRP values were used as input for the phenotype. Analysis was adjusted for categorical covariates; sex, smoking status, alcohol consumption status and tobacco chewing status, and continuous covariates such as age, BMI and first three principal components. Summary result of REML analysis reported phenotypic variance of 4.07% by all the identified SNPs in Indian population.

Pathway analysis of significant genes

To comprehend the metabolic pathways involving CRP regulating genes, pathway analysis was performed with significant genes (meta-analysis P <0.05) using GeneMANIA (Montojo *et al.* 2010).

Results and discussion

Genomewide analysis for circulating CRP

The present study is the first GWAS for circulating CRP level in Indians that ascertains the role of previously described variants for CRP regulation. Detailed characteristics of study population are presented in table 1. Figure 1 demonstrates the QQ plot, where the observed P values show good agreement with null hypothesis. Genomic inflation factor (λ) was computed to be 1.003 indicating no effect of type-1 errors on inflation of the observed P values. Figure 2 depicts the association of P values for

Table 1. Descriptive characteristics of the study participants.

Feature	Stage I	Stage II
N (M/F)	1097 (560/537)	3396 (1781/1615)
FPG (mg/dL)	87.2	89.8
Median (IQR)	79.8–93.7	82.5–97.9
BMI (kg/m^2)	24.25	26.02
Median (IQR)	21.1–27.7	23.0–29.1
Age (years)	50	50
Median (IQR)	44–60	43–60
Smoking (%)	24.5	16.1
Alcoholic (%)	14.3	14.1
Tobacco consumers (%)	12.9	7.3
CRP (mg/L)	1.34	1.84
Median (IQR)	0.69–3.04	0.96–3.63

Data are presented as median values (interquartile ranges). *N*, number of subjects; IQR, interquartile range; FPG, fasting plasma glucose; BMI, body mass index; CRP, C-reactive protein.

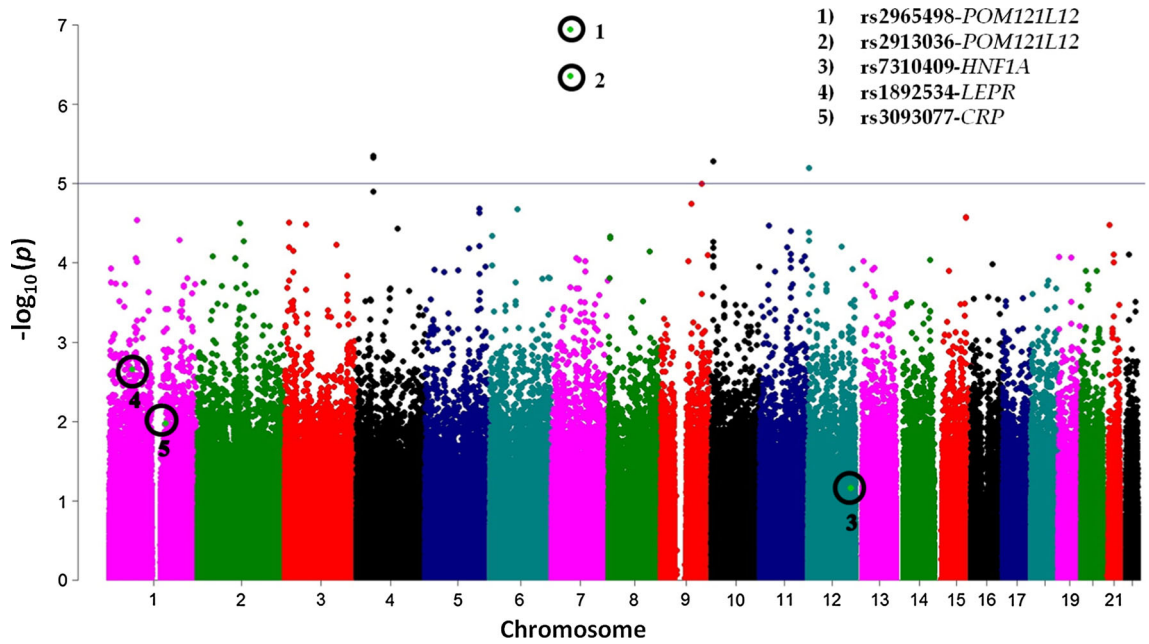

Figure 2. Manhattan plot for the SNPs associated with CRP level in stage I. Genomic positions for respective SNPs are shown in the *x*-axis (NCBI built 37).

Table 2. Association status of previously reported variants genotyped in stage II with circulating CRP level (P value $< 1 \times 10^{-3}$) in Indian subjects.

SNP	CHR	BP (hg19)	Gene/nearest gene	Al/A2	Stage II	
					Effect size (95% CI)	P value
rs1183910	12	121420807	*HNF1A*	A/G	−0.27 (−0.36 to −0.17)	5.03×10^{-8}
rs4420638	19	45422946	*APOC1*	G/A	−0.38 (−0.53 to −0.24)	3.41×10^{-7}

Association results were obtained from genotyped data in 3396 independent subjects of stage II.

519,533 SNPs among 1097 subjects in stage I. Variants rs2965498 (effect size = 1.91, P value = 1.14×10^{-7}) and rs2913036 (effect size = 1.64, P value = 4.48×10^{-7}) lying 261 and 250 kb upstream of POM121 trans-membrane nucleoporin-like 12 (*POM121L12*), respectively featured the strongest novel signals in stage I (figure 2). Both these variants were in strong linkage disequilibrium (LD) with each other in stage I with $r^2 = 0.79$. However, these signals did not replicate in stage II.

Signals with stage I $P<10^{-4}$ along with previously reported variants for CRP level were followed for stage II in 3396 subjects. Among the previously reported variants taken for replication, rs1183910 (*HNF1A*) passed near genomewide significance ($P = 5.03 \times 10^{-8}$) followed by strong association of rs4420638 near *APOC1* gene ($P = 3.41 \times 10^{-7}$) (table 2).

Meta-analysis in 4493 subjects (1097 subjects in stage I and 3396 subjects in stage II) identified *HNF1A* variant, rs7310409 ($P = 5.28 \times 10^{-11}$) as the top signal that crossed genomewide significance (table 3). This was followed by strong replications of other known variants, rs1892534 ($P = 5.87 \times 10^{-5}$) and rs3093077 ($P = 2.59 \times 10^{-4}$) in/near *LEPR* and *CRP* as discovered in our previous study (Mahajan *et al.* 2011).

Discovered associations of genomewide significant *HNF1A* variants, rs7310409 and rs1183910 in Indian population are consistent with earlier findings in several different populations (Reiner *et al.* 2008; Ridker *et al.* 2008; Elliott *et al.* 2009; Dehghan *et al.* 2011; Okada *et al.* 2011; Wu *et al.* 2012; Dorajoo *et al.* 2013; Kong and Lee 2013; Kocarnik *et al.* 2014). Our study was effectively powered to mark the GWAS level association of *HNF1A* variants for the first time in Indian population. This finding is further supported by prior large-scale genetic studies conducted in Asian-Indian subjects by the London Life Sciences Population cohort (LOLIPOP) (CRP effect (%) = −13.6, $P = 1.2 \times 10^{-17}$ for rs1183910) (Elliott *et al.* 2009), and South-Asian Indians of Singapore Indian Eye Study (SINDI) (beta = 0.038, $P = 0.003$ for rs7310409 and beta = 0.046, $P = 3.96 \times 10^{-4}$ for rs1183910) (Dorajoo *et al.* 2013), where strong associations of *HNF1A* variants rs1183910 and rs7310409 were detected.

Figure 3 depicts all the three genotypes (homozygous protective, heterozygous and homozygous risk genotype)

Table 3. Variants associated with circulating CRP level in meta-analysis ($P < 1 \times 10^{-3}$) among Indian subjects.

SNP	CHR	BP (hg19)	Gene/ nearest gene	Al/A2	Stage I		Stage II		Meta-analysis				
					Effect size (95% CI)	*P* value	Effect size (95% CI)	*P* value	Effect size (95% CI)	*P* value	Dir	Het I Sq	Het-*P* val
rs7310409*	12	121424861	*HNF1A*	G/A	0.35 (−0.028 to 0.73)	0.07	0.31 (0.21 to 0.41)	3.15×10^{-10}	0.31 (0.22 to 0.41)	5.28×10^{-11}	++	0	0.83
rs1892534‡	1	66105944	*LEPR*	A/G	−0.59 (−0.97 to −0.22)	2.21×10^{-3}	−0.26 (−0.43 to −0.092)	2.38×10^{-3}	−0.31 (−0.47 to −0.16)	5.87×10^{-5}	−−	31	0.23
rs3093077†	1	159679636	*CRP*	G/A	0.73 (0.17 to 1.30)	0.01	0.63 (0.16 to 1.10)	9.12×10^{-3}	0.67 (0.31 to 1.03)	2.59×10^{-4}	++	0	0.78

Association results were obtained from genotyped data in 1097 subjects from stage I and *3396 independent subjects of current stage II or †385 independent subjects of previous study or ‡1026 independent subjects of previous study as replication cohort. Variant rs7310409 is a previously known signal taken in replication despite its higher *P* value ($> 10^{-4}$) in stage I. Het-*P* val; *P* value for heterogeneity in effect sizes in meta-analysis; Het-I Sq; Chi-square value for heterogeneity test.

Figure 3. CRP level variation with respect to genotypes observed for genomewide associated SNPs in *HNF1A* loci. Average CRP level in individuals (*y*-axis) according to different genotypes (*x*-axis) for identified GWAS SNPs in current study has been plotted. Homozygous protective genotype with no CRP increasing allele, heterozygous genotype and homozygous risk genotype with two CRP increasing allele have been denoted as 0, 1 and 2, respectively.

observed in the two GWAS variants that strongly correlate with the mean CRP level in individuals harbouring respective genotypes ($r = 0.99$ for rs7310409 and rs1183910).

Other regulatory features of identified loci

Utilizing publically accessible HaploReg and RegulomeDB databases (Boyle *et al.* 2012; Ward and Kellis 2012), other functional regulatory features for GWAS signals and their LD SNPs were extracted. These SNPs have either promoter (H3K9ac and H3K4me3) or enhancer histone marks (H3K27ac and H3K4me1) or both, and DNAse hypersensitivity in several tissues including liver (table 1 in electronic supplementary material). Other CRP-associated variants were also found to harbour similar regulatory features (table 2 in electronic supplementary material). *HNF1A* is considerably expressed in liver and functions as a transcription factor. It has been shown to bind promoter regions of CRP gene and regulate CRP synthesis from liver cells (Li and Goldman 1996). Additionally, mutations in this gene accounts for substantial cases with maturity-onset diabetes of the young (MODY-3) that occurs during the early stages of life (<25 years) characterized by autosomal dominant inheritance and insulin secretory defects (Fajans *et al.* 2001). Based on the overlapping analysis with publically available datasets, observed *HNF1A* variants and their LD partners showed the presence of several regulatory features such as histone marks, DNAse I hypersensitive sites and binding sites for various proteins and factors of transcription machinery in liver and other tissues. This essentially implicates functional role of these variants in regulating *HNF1A* transcription and expression. Altered *HNF1A* expression could also potentially alter the expression of its target genes. Since, Indians are at high risk for metabolic syndrome, where earlier we demonstrated CRP levels as

Table 4. Association status of known GWAS signals for CRP in Indian population.

SNP	Gene/nearest gene	Indians				Source			
		Genotyping stage	MAF	Effect size	*P* value	MAF	Effect size	*P* value	Reference
rs1183910	*HNF1A*	Stage II	0.40 (A)	−0.27	5.03×10^{-8}	NR (A)	−0.15	2×10^{-124}	Dehghan *et al.* (2011)
rs4420638	*APOC1*	Stage II	0.12 (G)	−0.38	3.41×10^{-7}	0.10 (G)	−0.14	3×10^{-7}	Okada *et al.* (2011)
rs9987289	*LOC157273*	Stage II	0.11 (A)	−0.14	0.07	NR (A)	0.07	3×10^{-13}	Dehghan *et al.* (2011)
rs7310409*	*HNF1A*	Stages I and II	0.45 (G)	0.31	5.28×10^{-11}	NR (G)	0.15	7×10^{-17}	Ridker *et al.* (2008)
rs1892534‡	*LEPR*	Stages I and II	0.47 (A)	−0.31	5.87×10^{-5}	0.38 (A)	−0.17	7×10^{-21}	Ridker *et al.* (2008)
rs3093077†	*CRP*	Stages I and II	0.12 (G)	0.67	2.59×10^{-4}	NR (G)	0.25	7.72×10^{-11}	Ridker *et al.* (2008)
rs1800961*	*HNF4A*	Stages I and II	0.02 (A)	−0.48	2.06×10^{-3}	NR (A)	−0.09	2.00×10^{-9}	Dehghan *et al.* (2011)
rs4129267†	*IL6R*	Stages I and II	0.30 (A)	−0.39	4.09×10^{-3}	NR (A)	−0.08	2×10^{-48}	Dehghan *et al.* (2011)
rs12753193†	*LEPR*	Stages I and II	0.46 (G)	−0.34	5.28×10^{-3}	0.37 (G)	−0.15	3.27×10^{-17}	Ridker *et al.* (2008)
rs4131568†	*CRP*	Stages I and II	0.26 (A)	−0.37	5.29×10^{-3}	NR (T)	0.07	2.45×10^{-7}	Kocarnik *et al.* (2014)
rs157580*	*APOE*	Stages I and II	0.46 (G)	0.11	0.02	NA	NA	NA	NA
rs7539471†	*LEPR*	Stages I and II	0.24 (A)	−0.27	0.05	0.31 (A)	0.12	1.87×10^{-10}	Ridker *et al.* (2008)
rs2794520†	*CRP*	Stages I and II	0.30 (A)	−0.25	0.06	NR (A)	−0.16	2×10^{-186}	Dehghan *et al.* (2011)
rs2464196†	*HNF1A*	Stages I and II	0.42 (A)	−0.18	0.12	0.29 (A)	−0.14	3.32×10^{-13}	Ridker *et al.* (2008)
rs1169300†	*HNF1A*	Stages I and II	0.42 (A)	−0.16	0.19	0.29 (A)	−0.15	1.15×10^{-13}	Ridker *et al.* (2008)
rs2592887†	*CRP*	Stages I and II	0.34 (A)	−0.17	0.19	NR (A)	−0.16	1.04×10^{-18}	Ridker *et al.* (2008)
rs7953249†	*HNF1A-AS1*	Stages I and II	0.42 (A)	0.15	0.19	NR (A)	0.13	6.98×10^{-13}	Ridker *et al.* (2008)
rs2075650†	*APOE*	Stages I and II	0.10 (G)	−0.24	0.2	0.12 (G)	0.12	2×10^{-21}	Dorajoo *et al.* (2013)
rs780094*	*GCKR*	Stages I and II	0.23 (A)	0.07	0.23	NR (A)	0.14	7×10^{-15}	Ridker *et al.* (2008)
rs12093699†	*CRP*	Stages I and II	0.20 (A)	−0.17	0.24	NR (G)	0.13	2.15×10^{-11}	Ridker *et al.* (2008)
rs735396†	*HNF1A*	Stages I and II	0.46 (A)	0.13	0.25	NR (A)	0.15	5.39×10^{-15}	Ridker *et al.* (2008)
rs1260326*	*GCKR*	Stages I and II	0.23 (A)	0.06	0.31	NR (T)	0.07	5×10^{-40}	Dehghan *et al.* (2011)
rs10778213†	*ASCL1-C12orf42*	Stages I and II	0.49 (G)	0.001	0.99	NR (G)	−0.12	1.16×10^{-10}	Ridker *et al.* (2008)

Association results presented in Indians were obtained from genotyped data in 1097 subjects from stage I; *3396 independent subjects of current stage II; †385 independent subjects of previous study; ‡1026 independent subjects of previous study as replication cohort. Signals with $P<0.05$ have been represented in bold.

a good predictor for detecting Indian individuals with metabolic syndrome (Mahajan *et al.* 2012). *HNF1A* might be a key regulator of CRP and therefore strongly associate with CRP levels in high-risk Indian population. Identified *HNF1A* variants together contributed for 1.11% to variance in CRP level in Indians. Strong association ($\sim 10^{-7}$) near apolipoprotein gene *APOC1* (explained variance = 0.95%) of our study was also in agreement with studies linking this region with CRP level (Elliott *et al.* 2009; Okada *et al.* 2011).

Association status of known GWAS signals for CRP in Indian population

We evaluated the association status of previously described variants of CRP in the GWAS catalogue (Welter *et al.* 2014) and literature among Indians and other ethnic populations (table 4). The effect sizes were converted to uniform unit by absolute conversion factor before comparison. Reported variants associated with CRP level ($P<0.05$) in Indians were rs1183910 and rs7310409 in *HNF1A*, rs4420638 near *APOC1*, rs1892534 and rs12753193 in *LEPR*, rs3093077 and rs4131568 in *CRP*, rs1800961 in *HNF4A*, rs4129267 in *IL6R* and rs157580 in *APOE*. Majority of these variants have higher effect sizes in Indians compared to other ethnicities. Among the known significant variants detected in our population ($P<0.05$), we observed directional consistency for effect sizes across present study and source study for all loci except for rs4131568 at *CRP*. Identification of known signals in genetically heterogeneous Indian population reflects that major variants associated with circulating CRP level are substantially similar across the globe. Therefore, genes identified for CRP regulation in Indians were observed to be genes of metabolic and immune-function relevance consistent with earlier studies in other populations (Reiner *et al.* 2008; Ridker *et al.* 2008; Elliott *et al.* 2009; Dehghan *et al.* 2011; Okada *et al.* 2011; Wu *et al.* 2012; Dorajoo *et al.* 2013; Kong and Lee 2013; Kocarnik *et al.* 2014). Since CRP levels have been found to differ in individuals with different life style (exercise, smoking, food habit etc.) (Rojo-Martinez *et al.* 2013), we also evaluated the effects of discovered loci in the presence of endogenous environmental factors. Here, we considered the effects of identified loci in smokers, alcoholics, tobacco chewers and nonvegetarians in comparison to their contemporaries of healthy habits in Indian subjects (table 3 in electronic supplementary material). We observed strong reduction in effect size of *LEPR* variant rs12753193 among nonsmokers (effect size = 0.48) in comparison with smokers (effect size = 0.86). Further, the effect size of this variant was higher in alcoholics (effect size = 0.91) than nonalcoholics (effect size = 0.53). *LEPR* variant rs1892534 had higher effect size in alcoholics (effect size = 0.9) than nonalcoholics (effect size = 0.59). This variant also had higher effect size in nonvegetarians (effect size = 1.3) than vegetarians (effect size = 0.47). *HNF1A* variant rs1183910 showed notable difference in effect sizes for the CRP level among alcoholics (effect size = 0.38) versus nonalcoholics (effect size = 0.26) and nonvegetarians (effect size = 0.53) versus vegetarians (effect size = 0.19). Similarly effect size of *IL6R* variant rs4129267 was higher in tobacco chewers (effect size = 1.09) than nontobacco chewers (effect size = 0.56). Further, variant rs7310409 in *HNF1A* showed different effect sizes among nonvegetarians (effect size = 0.53) and vegetarians (effect size = 0.26) with higher effect size in the former group. These results suggest that an inherent environment in India might also influence the effect of identified genetic loci for CRP in Indians. All identified variants contributed to nearly 4% of the variance in the CRP level in Indian population that indicates major contribution of several other genetic and epigenetic players in modulating circulating CRP level in this population.

Pathway enrichment analysis

Pathway analysis of significant genes in meta-analysis ($P<0.05$) demonstrated significant enrichment of genes involved in pancreas development (table 4 in electronic supplementary material). Inclusive of current study samples, earlier we demonstrated that higher CRP levels were associated with T2DM in BMI-independent manner (Mahajan *et al.* 2009). Current study also detected significant enrichment of CRP-associated genes in pancreas development that probably signify shared genetic heritability for pathophysiology of CRP level and T2DM that is adult-onset and most common form of diabetes across the globe including India. This is in agreement with the hypothesis that complex disorders are outcomes of precursor quantitative traits (Plomin *et al.* 2009). It implicates essentiality of studying quantitative trait genetics to better understand mechanistic heterogeneity of complex disorders across different populations.

In summary, our study is the first GWAS conducted for circulatory CRP levels in Indians. Our study confirmed known associations of common variants at genomewide scale in a large sample size of Indo-European population.

Acknowledgements

The authors are thankful to all the study participants and the members of INdian DIabetes COnsortium (INDICO) for their support in generation of data. This study was supported by grants provided by the Council of Scientific and Industrial Research (CSIR), Government of India through Centre for Cardiovascular and Metabolic Disease Research (CARDIOMED) project (Grant no. BSC0122-(13)). This study was also partially funded by the Department of Science and Technology, Government of India through PURSE II CDST/SR/PURSE PHASE II/11. GP

and AKG acknowledge University Grants Commission (UGC), Government of India for Senior Research Fellowship.

References

Benjamin E. J., Dupuis J., Larson M. G., Lunetta K. L., Booth S. L., Govindaraju D. R. *et al.* 2007 Genome-wide association with select biomarker traits in the Framingham heart study. *BMC Med. Genet.* **8**, S11.

Black S., Kushner I. and Samols D. 2004 C-reactive protein. *J. Biol. Chem.* **279**, 48487–48490.

Boyle A. P., Hong E. L., Hariharan M., Cheng Y., Schaub M. A., Kasowski M. *et al.* 2012 Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797.

Chambers J. C., Eda S., Bassett P., Karim Y., Thompson S. G., Gallimore J. R. *et al.* 2001 C-reactive protein, insulin resistance, central obesity, and coronary heart disease risk in Indian Asians from the United Kingdom compared with European whites. *Circulation* **104**, 145–150.

Chandalia M., Cabo-Chan A. V., Devaraj S., Jialal I., Grundy S. M. and Abate N. 2003 Elevated plasma high-sensitivity C-reactive protein concentrations in Asian Indians living in the United States. *J. Clin. Endocrinol. Metab.* **88**, 3773–3776.

Dehghan A., Dupuis J., Barbalic M., Bis J. C., Eiriksdottir G., Lu C. *et al.* 2011 Meta-analysis of genome-wide association studies in >80,000 subjects identifies multiple loci for C-reactive protein level. *Circulation* **123**, 731–738.

Dorajoo R., Li R., Ikram M. K., Liu J., Froguel P., Lee J. *et al.* 2013 Are C-reactive protein associated genetic variants associated with serum levels and retinal markers of microvascular pathology in Asian populations from Singapore? *PLoS One* **8**, e67650.

Elliott P., Chambers J. C., Zhang W., Clarke R., Hopewell J. C., Peden J. F. *et al.* 2009 Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA* **302**, 37–48.

Fajans S. S., Bell G. I. and Polonsky K. S. 2001 Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N. Engl. J. Med.* **345**, 971–980.

Ford E. S., Giles W. H., Myers G. L., Rifai N., Ridker P. M. and Mannino D. M. 2003 C-reactive protein concentration distribution among US children and young adults: findings from the National Health and Nutrition Examination Survey, 1999–2000. *Clin. Chem.* **49**, 1353–1357.

Giri A. K., Khan N. M., Basu A., Tandon N., Scaria V. and Bharadwaj D. 2014 Pharmacogenetic landscape of clopidogrel in north Indians suggest distinct interpopulation differences in allele frequencies. *Pharmacogenomics* **15**, 643–653.

INdian DIabetes Consortium 2011 INDICO: the development of a resource for epigenomic study of Indians undergoing socioeconomic transition. *Hugo J.* **5**, 65–69.

Khera A., McGuire D. K., Murphy S. A., Stanek H. G., Das S. R., Vongpatanasin W. *et al.* 2005 Race and gender differences in C-reactive protein levels. *J. Am. Coll. Cardiol.* **46**, 464–469.

Khoo C. M., Sairazi S., Taslim S., Gardner D., Wu Y., Lee J. *et al.* 2011 Ethnicity modifies the relationships of insulin resistance, inflammation, and adiponectin with obesity in a multiethnic Asian population. *Diabetes Care* **34**, 1120–1126.

Kocarnik J. M., Pendergrass S. A., Carty C. L., Pankow J. S., Schumacher F. R., Cheng I. *et al.* 2014 Multiancestral analysis of inflammation-related genetic variants and c-reactive protein in the population architecture using genomics and epidemiology study. *Circ. Cardiovasc. Genet.* **7**, 178–188.

Kong M. and Lee C. 2013 Genetic associations with C-reactive protein level and white blood cell count in the KARE study. *Int. J. Immunogenet.* **40**, 120–125.

Li S. P. and Goldman N. D. 1996 Regulation of human C-reactive protein gene expression by two synergistic IL-6 responsive elements. *Biochemistry* **35**, 9060–9068.

Mahajan A., Tabassum R., Chavali S., Dwivedi O. P., Bharadwaj M., Tandon N. *et al.* 2009 High-sensitivity C- reactive protein levels and type 2 diabetes in urban North Indians. *J. Clin. Endocrinol. Metab.* **94**, 2123–2127.

Mahajan A., Tabassum R., Chavali S., Dwivedi O. P., Chauhan G., Ghosh S. *et al.* 2011 Common variants in CRP and LEPR influence high sensitivity C-reactive protein levels in North Indians. *PLoS One* **6**, e24645.

Mahajan A., Jaiswal A.,Tabassum R., Podder A., Ghosh S., Madhu S. V. *et al.* 2012 Elevated level of C-reactive protein as a risk factor for metabolic syndrome in Indians. *Atherosclerosis* **220**, 275–281.

Montojo J., Zuberi K., Rodriguez H., Kazi F., Wright G., Donaldson S. L. *et al.* 2010 GeneMANIA cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* **26**, 2927–2928.

Okada Y., Takahashi A., Ohmiya H., Kumasaka N., Kamatani Y., Hosono N. *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.

Pankow J. S., Folsom A. R., Cushman M., Borecki I. B., Hopkins P. N., Eckfeldt J. H. *et al.* 2001 Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis* **154**, 681–689.

Plomin R., Haworth C. M. and Davis O. S. 2009 Common disorders are quantitative traits. *Nat. Rev. Genet.* **10**, 872–878.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D. *et al.* 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575.

Reiner A. P., Barber M. J., Guan Y., Ridker P. M., Lange L. A., Chasman D. I. *et al.* 2008 Polymorphisms of the *HNF1A* gene encoding hepatocyte nuclear factor-1α are associated with C-reactive protein. *Am. J. Hum. Genet.* **82**, 1193–1201.

Reiner A. P., Beleza S., Franceschini N., Auer P. L., Robinson J. G., Kooperberg C. *et al.* 2012 Genome-wide association and population genetic analysis of c-reactive protein in African American and Hispanic American women. *Am. J. Hum. Genet.* **91**, 502–512.

Ridker P. M., Wilson P. W. F. and Grundy S. M. 2004 Should C-reactive protein be added to metabolic syndrome and to assessment of global cardiovascular risk? *Circulation* **109**, 2818–2825.

Ridker P. M., Pare G., Parker A., Zee R. Y., Danik J. S., Buring J. E. *et al.* 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.

Rojo-Martinez G., Soriguer F., Colomo N., Calle A., Goday A., Bordiú E. *et al.* 2013 Factors determining high-sensitivity C-reactive protein values in the Spanish population. Di@bet.es study. *Eur. J. Clin. Invest.* **43**, 1–10.

Shen R., Fan J. B., Campbell D., Chang W., Chen J., Doucet D. *et al.* 2005 High-throughput SNP genotyping on universal bead arrays. *Mutat. Res.* **573**, 70–82.

Tabassum R., Mahajan A., Chauhan G., Dwivedi O. P., Dubey H., Sharma V. *et al.* 2011 No association of TNFRSF1B variants with type 2 diabetes in Indians of Indo-European origin. *BMC Med. Genet.* **12**, 110.

Tabassum R., Chauhan G., Dwivedi O. P., Mahajan A., Jaiswal A., Kaur I. *et al.* 2013 Genome-wide association study for

type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. *Diabetes* **62**, 977–986.

Turner S. D. 2014 qqman: an R package for visualizing GWAS results using Q–Q and Manhattan plots. BioRxiv.

Ward L. D. and Kellis M. 2012 HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934.

Welter D., MacArthur J., Morales J., Burdett T., Hall P., Junkins H. *et al.* 2014 The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006.

Wessel J., Moratorio G., Rao F., Mahata M., Zhang L., Greene W. *et al.* 2007 C-reactive protein, an 'intermediate phenotype' for inflammation: human twin studies reveal heritability, association with blood pressure and the metabolic syndrome, and the influence of common polymorphism at catecholaminergic/beta-adrenergic pathway loci. *J. Hypertens.* **25**, 329–343.

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

Wu Y., McDade T. W., Kuzawa C. W., Borja J., Li Y., Adair L. S. *et al.* 2012 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* **35**, 574–583.

Yang J., Lee S. H., Goddard M. E. and Visscher P. M. 2011 GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82.

Corresponding editor: H. A. Ranganath