

Published in final edited form as:

*Inflammation*. 2012 April ; 35(2): 574–583. doi:10.1007/s10753-011-9348-y.

## 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

Ying Wu<sup>1</sup>, Thomas W. McDade<sup>2,3</sup>, Christopher W. Kuzawa<sup>2,3</sup>, Judith Borja<sup>4</sup>, Yun Li<sup>1,5</sup>, Linda S. Adair<sup>6</sup>, Karen L. Mohlke<sup>1</sup>, and Leslie A. Lange<sup>1</sup>

<sup>1</sup> Department of Genetics, University of North Carolina, Chapel Hill, NC

<sup>2</sup> Department of Anthropology, Northwestern University, Evanston, IL

<sup>3</sup> Cells 2 Society: The Center for Social Disparities and Health at the Institute for Policy Research, Northwestern University, Evanston, IL

<sup>4</sup> USC-Office of Population Studies Foundation, University of San Carlos, Cebu City, Philippines

<sup>5</sup> Department of Biostatistics, University of North Carolina, Chapel Hill, NC

<sup>6</sup> Department of Nutrition, University of North Carolina, Chapel Hill, NC

### Abstract

Recent genome-wide association (GWA) studies have related several genetic loci, including *CRP*, *HNF1A* and *LEPR*, to circulating C-reactive protein (CRP) levels in populations of European ancestry. The genetic effects in other populations and across varying levels of exposure to a pathogenic environment, an important environmental factor associated with CRP, remain to be determined. We tested 2,073,674 SNPs for association with plasma CRP (limited to  $\leq 10$  mg/L) in 1,709 unrelated Filipino women from the Cebu Longitudinal Health and Nutrition Survey (CLHNS). The strongest evidence of association was observed with variants at *CRP* (rs876537,  $P = 1.4 \times 10^{-9}$ ) and *HNF1A* (rs7305618,  $P = 1.0 \times 10^{-8}$ ). Among other previously reported CRP associated loci, the *APOE*  $\epsilon 4$  haplotype was associated with decreased CRP level ( $P = 7.1 \times 10^{-4}$ ), and modest association was observed with *LEPR* (rs1892534,  $P = 0.076$ ), with direction of effects consistent with previous studies. The strongest signal at a locus not previously reported mapped to a gene desert region on chromosome 6q16.1 (rs1408282,  $P = 2.9 \times 10^{-6}$ ). Finally, we observed nominal evidence of interaction with exposure to a pathogenic environment for top main effect SNPs at *HNF1A* (rs7305618,  $P = 0.031$ ), *LEPR* (rs1892535,  $P = 0.030$ ) and 6q16.1 (rs1408282,  $P = 0.046$ ). Our findings demonstrate convincing evidence that genetic variants in *CRP* and *HNF1A* contribute to plasma CRP in Filipino women, and provide the first evidence that exposure to a pathogenic environment may modify the genetic influence at the *HNF1A*, *LEPR* and 6q16.1 loci on plasma CRP level.

### Keywords

C-reactive protein; genome-wide association; Filipino women; gene-environment interaction

---

Send correspondence to: Leslie A. Lange, Department of Genetics, University of North Carolina, 5112 Genetic Medicine Building, Chapel Hill, NC 27599-7264, USA, leslie\_lange@med.unc.edu.

Competing interest

The authors declare that they have no competing interests.

## Background

C-reactive protein (CRP), initially recognized as a hallmark of inflammation and the acute-phase response, is a powerful marker for risk of cardiovascular disease (CVD), type 2 diabetes and metabolic syndrome [1]. Family studies have estimated that inter-individual variation in blood CRP is 35–40% heritable [2]. Recent GWA studies have identified several genetic loci significantly associated ( $P < 5 \times 10^{-8}$ ) with plasma CRP. Seven loci identified in the Women's Genome Health Study (WGHS), including signals at *CRP*, *HNF1A*, *IL6R*, *APOE*, *LEPR*, *GCKR* and 12q23.2 accounted for 10.1% of the variation in CRP levels in Caucasian women [3]. Further confirmatory evidence was provided by the Pharmacogenomics and Risk of Cardiovascular Disease (PARC) study that variants at *CRP*, *HNF1A* and *APOE* were associated with plasma CRP in Europeans [4]. However, it remains to be determined what loci are most strongly associated with CRP levels in populations of non-European ancestry.

The genetic variants identified by prior GWA studies explain only a small fraction of the total variation believed to reflect genetic effects on CRP level. Both gene-gene and gene-environment interactions may underlie complex phenotypes [5] [6]. According to a 2004 World Health Organization report, infectious diseases still account for more than 30% of all mortality in Southeast Asia [7]. In the Philippines, respiratory infection ranked among the top causes of mortality in 2006 [8]. As a major source of inflammatory stimuli, exposure to a pathogenic environment results in elevated levels of CRP. Earlier investigations in CLHNS samples showed convincing evidence of the role of exposure to a pathogenic environment in predicting plasma CRP levels in Filipinos [9–10]. The burden of environmental inflammatory stimuli in CLHNS samples provides an opportunity to explore whether exposure to a pathogenic environment interacts with genetic variants to influence CRP concentrations.

In light of the important roles of both genetic variation and exposure to a pathogenic environment in predicting circulating CRP levels, the primary purposes of this study were to conduct a GWA scan to identify loci associated with plasma CRP levels in Filipino women and to examine whether the genetic contributions to CRP levels interact with exposure to a pathogenic environment in this population.

## Methods

### Study population and data collection

The study sample consisted of 1,798 healthy Filipino women from the CLHNS, an on-going community-based study of mother-child pairs that began in 1983. The study population, design and protocols for this longitudinal cohort have been previously described [11]. Data for this paper come from the 2005 CLHNS survey on the mothers. Written informed consent was obtained from all participants, and study protocols were approved by the University of North Carolina Institute Review Board for the Protection of Human Subjects. Anthropometric measurements and comprehensive data on household demographics, income levels, environmental quality and health behaviors were collected through in-home interviews administered by trained staff (data available online at <http://www.cpc.unc.edu/projects/cebu>). Overnight fasting blood samples were obtained at the 2005 survey and collected into EDTA-coated tubes. Plasma CRP concentrations were measured by a high-sensitivity immunoturbidimetric method (Synchron LX20, lower detection limit: 0.1 mg/L; Beckman Coulter, Fullerton, CA) with between-assay coefficient of variations (CVs)  $< 7.6$  across the assay range [9]. Seventy-seven participants with CRP levels  $> 10$  mg/L were excluded from the initial analysis. The general characteristics of 1,709 samples with CRP levels  $\leq 10$  mg/L are presented in Supplementary Table 1.

Based on collection of multiple proxy measures of the likelihood of exposure to infectious microbes, a pathogen score was constructed using the mean value of five interviewer-assigned variables, each scored on a 3-point scale (0 = low exposure, 1 = moderate and 2 = high): (1) cleanliness of the food preparation area; (2) means of garbage disposal; (3) presence of excrement near the house; (4) level of garbage; and (5) excrement present in the neighborhood surrounding the household [9]. The pathogen score was negatively correlated with socio-economic status as measured by household income (corr:  $-0.198$ ,  $P < 0.0001$ ) and household assets (corr:  $-0.333$ ,  $P < 0.0001$ ) and with waist circumference (corr:  $-0.106$ ,  $P < 0.0001$ ) (Supplementary Table 2). In addition, a dichotomous variable (0, 1) was constructed based on whether participants reported any symptoms of infection at the time of blood collection. Symptoms included runny nose, cough, fever, diarrhea, sore throat and the more general categories of flu, cold and sinusitis.

### GWA SNP genotyping and imputation

SNP genotyping was conducted using the Affymetrix Genome-wide Human SNP Array 5.0 and quality control was performed, both as previously described [12]. Briefly, samples with  $<97\%$  genotyping call rate were excluded, as were 81 members of estimated first-degree relative pairs. SNPs were excluded for call rate  $< 90\%$ , deviation from Hardy-Weinberg equilibrium ( $P < 10^{-6}$ ),  $\geq 3$  discrepancies among duplicate pairs, Mendelian inheritance errors among five CEPH trios and/or CEPH sample genotype discrepancies with HapMap. Genotype imputation was conducted using MACH [13] and 352,264 genotyped SNPs that were polymorphic in both HapMap phase II CEU and CHB+JPT samples. After excluding SNPs with  $r^2 \leq 0.3$  or estimated minor allele frequency (MAF)  $\leq 0.01$ , we obtained a final imputed data set of 1,697,660 SNPs. Missing genotype data for the 352,264 directly genotyped SNPs were also imputed. In addition, 23,750 directly genotyped SNPs non-polymorphic in both HapMap populations but with MAF  $> 0.01$  in the CLHNS samples were also analyzed, for a total of 2,073,674 SNPs tested for association with plasma CRP level.

Additional imputation within flanking regions of previously reported loci was performed based on haplotypes created from the 1000 Genomes Project pilot release (June, 2010) of CEU+CHB+JPT samples. Regional plots were created using LocusZoom [14].

### Supplementary genotyping

Due to the unavailability in our genotyped and imputed data, candidate SNPs in the *APOE* gene (rs429358 and rs7412) were genotyped by TaqMan allelic discrimination (Applied Biosystems, Foster City, CA). Two additional assays representing *CRP* triallelic SNP rs3091244 were designed as described [15]. The triallelic genotypes were assigned as shown in Supplementary Table 3, and samples with ambiguous genotypes were genotyped by Sanger sequencing. For each assay, the genotyping success rate was  $>98.9\%$  and no discrepancies among 80 duplicate pairs were observed.

### Statistical analysis

CRP values were natural log-transformed after adding the constant 0.10 to satisfy model assumptions of normally distributed residuals, conditional on the covariates of age in the year 2005, household assets, natural log-transformed household income, number of previous pregnancies (as three categories: 0–4, 5–10,  $\geq 11$ ), infectious status, pathogenic score, and 7 principal components of genetic variation (Supplementary Table 4). Additional traits were tested for association at SNPs demonstrating association evidence for CRP. These traits included waist circumference, BMI, triglycerides, HDL cholesterol, LDL cholesterol, total cholesterol and homocysteine, among which traits were natural log-transformed except waist circumference. Array Studio software version 3.2 (OmicSoft, Morrisville, NC) was used to

perform the initial GWA analyses, and SAS version 9.2 (SAS Institute, Cary, NC) was used to perform additional analyses for association between selected SNPs and CVD related traits. A multivariable linear regression model assuming an additive genetic model was used for association tests. The triallelic SNP rs3091244 was tested for association as a biallelic SNP considering either the SNP ‘A’ allele or the ‘T’ allele to be the effect allele, in which CC genotype was coded 0, AC or CT were coded 1, and AA, AT or TT coded 2 under an additive genetic model. The effects of association with each additional copy of the SNP ‘A’ allele (CC, CT or TT = 0, AC or AT = 1 and AA = 2) or the ‘T’ allele (AA, AC or CC = 0, AT or CT = 1 and TT = 2) were also evaluated in two separate analyses. In conditional analysis, we tested the additive genotype effects for each of the SNPs within the flanking region at each locus in individual linear regression models with the same covariates as above plus the genotype of the index SNP.

Haplotype analyses were performed using the “haplo.stat” R package. Haplotypes and haplotype frequencies were estimated using the R function “haplo.em”. The association between haplotypes and CRP was assessed using the R function “haplo.glm”. An additive model was assumed, in which the regression coefficient  $\beta$  represents the expected change in CRP level with each additional copy of the specific haplotype compared to the reference haplotype. The most common T-C (*APOE*  $\epsilon$ 3) haplotype of rs429358 and rs7412 was set as the reference haplotype. The R function “haplo.score” was used to compute the global score statistics to test the overall association between haplotypes and CRP. The same covariates used for genotype analysis were also applied in the haplotype analysis models.

The seven SNPs displaying the strongest evidence for association at each of the seven loci were included in interaction analyses between genotype and exposure to a pathogenic environment. The intensity of exposure to a pathogenic environment was dichotomized into a lower exposure group (pathogen score  $< 0.5$ ,  $n = 877$ ) and a higher exposure group (pathogen score  $\geq 0.5$ ,  $n = 832$ ), using the median pathogen score as the cut-off point.  $P$  values for genotype by pathogenic score interaction  $< 0.05$  were considered to be significant.

## Results

Tests of association between 2,073,674 SNPs and plasma CRP in 1,709 unrelated CLHNS mothers resulted in 28 SNPs at three loci that showed evidence of association at a threshold of  $P < 5 \times 10^{-6}$  (Figure 1). The observed genomic control inflation factor ( $\lambda_{GC}$ ) was 1.02, suggesting no substantial population stratification. Consistent with the findings of previous GWA study in populations of European ancestry, the most compelling associations were observed at *CRP* (best  $P = 1.4 \times 10^{-9}$  for rs876537) and *HNF1A* (best  $P = 1.0 \times 10^{-8}$  for rs7305618) (Table 1, Figure 1). Conditional analysis for SNPs within the *CRP* locus suggested that the *CRP* SNPs likely represented a single signal for plasma CRP in CLHNS mothers (all conditioned  $P > 0.03$ , Supplementary Table 5A). Conditioning on the *HNF1A* variant rs7305618, we observed no evidence for a secondary signal at the *HNF1A* locus (all conditioned  $P > 0.08$ , Supplementary Table 5B). Additionally, a suggestive signal at a locus not previously reported to be associated with CRP mapped to a gene desert region on chromosome 6q16.1 (rs1408282, best  $P = 1.4 \times 10^{-6}$ , Table 1). Conditional analysis revealed that there were no other SNPs located in the region with independent evidence for association with plasma CRP (all conditioned  $P > 0.36$ , Supplementary Table 5C).

We next tested whether adjustment for waist circumference, a known predictor of CRP level [9–10], would affect the observed genetic associations. The waist-adjusted association became stronger for the *CRP* (rs876537,  $P = 3.4 \times 10^{-11}$ ) and *HNF1A* (rs7305618,  $P = 4.0 \times 10^{-10}$ ) loci (Table 1), suggesting that accounting for waist circumference removed an independent source of variation in CRP levels. Results were consistent in an analysis that

also included the 77 additional study participants that were excluded in the initial analysis due to CRP levels > 10 mg/L (data not shown).

We next examined whether additional CRP-associated loci identified in European populations were replicated in the CLHNS, using a definition of replication as  $P < 0.05$  and consistent direction of effect. Among the SNPs that were both previously identified [3] and available in CLHNS samples, the widely reported *CRP* gene variant rs1205 ( $P = 8.5 \times 10^{-9}$ ) provided convincing evidence for association with CRP level (Table 2). The triallelic *CRP* variant rs3091244 shown previously to affect *CRP* promoter activity and influence the gene transcription [16] displayed significant association in the same direction in CLHNS samples ( $\beta = 0.257$ ,  $P = 5.2 \times 10^{-7}$ , tested as a biallelic SNP considering the allele A or T to be the effect allele). Specifically, the SNP ‘A’ allele and ‘T’ allele displayed consistent direction of association, and were significantly associated with increased CRP levels when we examined the effect of the A and T allele separately in two analyses (A allele:  $\beta = 0.245$ ,  $P = 3.8 \times 10^{-5}$ ; T allele:  $\beta = 0.183$ ,  $P = 0.023$ ). The frequency of the rs3091244 alleles differ between CLHNS (C:0.79, A:0.13, T: 0.08) and predominantly European-descended participants in the Framingham Heart Study (C:0.62, A:0.07, T: 0.31) [17]. Reciprocal conditional analysis suggested that despite not representing an independent effect, the SNP rs876537 was likely to be a stronger signal compared to rs3091244 at the *CRP* locus ( $P$  for rs876537 conditional on rs3091244 =  $6.8 \times 10^{-4}$ ;  $P$  for rs3091244 conditional on rs876537 = 0.12). We also provided evidence for the association with the two previously studied nonsynonymous SNPs in the *HNF1A* gene ( $P = 1.4 \times 10^{-7}$  for Ile27Leu and  $7.5 \times 10^{-6}$  for Ser486Asn). Additionally, significant association was observed for the intronic SNPs rs1169286 ( $P = 5.1 \times 10^{-7}$ ) and rs7310409 ( $P = 1.6 \times 10^{-6}$ ) within the first intron of *HNF1A*, consistent with previous studies for these SNPs [4]. For the *LEPR*, *IL6R*, *GCKR* and chromosome 12q23.2 loci, our study had > 89% power, using a significance threshold of  $\alpha = 0.05$  to detect the reported percentage of total variation explained by each additional copy of the effect allele [3]. The signal at *LEPR* did not meet our replication threshold; however, it showed modest evidence of association (rs1892534,  $P = 0.076$ ) with a consistent direction of effect as reported previously. We did not replicate associations with other reported GWA study SNPs at *IL6R* (rs4129267,  $P = 0.24$ ), *GCKR* (rs1260326,  $P = 0.20$ ), and chromosomal region 12q23.2 (rs10778213,  $P = 0.63$ ), although the directions of effect were consistent with the previous report for *IL6R* and *GCKR* SNPs.

As the *APOE* SNPs rs769449 and rs2075650 reported by the WGHS[3] were not genotyped or well-imputed in CLHNS samples, we investigated for association between CRP level and the *APOE* variants rs429358 and rs7412 that comprise the *APOE*  $\epsilon 2$ ,  $\epsilon 3$ , and  $\epsilon 4$  haplotypes [18] (Table 3). The *APOE* SNP rs429358 was significantly associated with CRP levels ( $P = 4.8 \times 10^{-4}$ ) under an additive model. Given the small number of 15 minor C allele homozygotes, we also performed a test assuming a dominant model and observed that the CRP levels were significantly lower in C allele carriers ( $P = 1.0 \times 10^{-4}$ ). Haplotype analysis based on a score statistic provided compelling evidence for an overall association between *APOE* haplotypes and CRP level (Global  $P = 2.6 \times 10^{-3}$ ). The *APOE*  $\epsilon 4$  haplotype with a frequency of 0.085 was significantly associated with lower CRP ( $P = 7.1 \times 10^{-4}$ ), whereas the most common haplotype *APOE*  $\epsilon 3$  with an estimated frequency of 0.779 was modestly associated with elevated CRP ( $P = 0.01$ ) (Table 3). Further analysis that evaluated the effect on CRP level of each additional copy of the specific haplotype compared with the homozygote reference haplotype (*APOE*  $\epsilon 3$ ) suggested that the *APOE*  $\epsilon 4$  haplotype was significantly associated with decreased level of CRP ( $\beta = -0.269$ ,  $P = 5.9 \times 10^{-4}$ ).

Additional variation within flanking regions of the CRP associated loci was investigated by assessing the association of SNPs imputed based on 1000 Genomes Project Pilot data. We observed a non-HapMap SNP rs2592902 ( $P = 7.3 \times 10^{-10}$ ) showing slightly stronger

association compared to rs876537 ( $P = 2.0 \times 10^{-9}$ ), the lead SNP at the *CRP* locus (Supplementary Figure 1). Reciprocal conditional analysis provided no evidence for independent effects (conditional  $P$  for rs876537 = 0.59, conditional  $P$  for rs2592902 = 0.13). Further, conditioning on rs876537, none of the remaining 220 SNPs in this 200 kb gene region (chr1: 157,841,557–158,041,557) showed evidence for a secondary signal at the *CRP* locus (conditional  $P > 0.005$ ). Additionally, the comprehensive imputation of 283 SNPs in a 200 kb flanking region (chr12: 119,787,315–119,987,315) of *HNFA* index SNP rs7305618 ( $P = 5.2 \times 10^{-8}$ ) revealed stronger evidence of association for CRP level with the *HNFA* promoter SNP rs2255531 ( $P = 2.9 \times 10^{-8}$ ). The findings based on the reciprocal conditional analysis provided little evidence for their independence (conditional  $P$  for rs7305618 = 0.35, conditional  $P$  for rs2255531 = 0.15). In models accounting for rs7305618, the strength of the association with the remaining SNPs were greatly attenuated (conditional  $P > 0.024$ ), suggesting no separate signal at the *HNFA* locus for plasma CRP. Further analysis using the 1000 Genome Project imputed SNPs within 2 Mb expansive flanking regions of the previously reported loci including *LEPR*, *IR6R*, *APOE*, *GCKR* and 12q23.2 did not produce compelling evidence for additional *CRP* associated signals in CLHNS samples (data not shown).

We next tested whether the intensity of exposure to a pathogenic environment modifies the effects of genotypes on plasma CRP in Filipino women. Among eight SNPs at seven previously reported loci and at the suggestive signal at 6q16.1, nominally significant interactions were detected between genotype and pathogen score on plasma CRP for the SNPs *HNFA* rs7305618 ( $P = 0.031$ ), *LEPR* rs1892534 ( $P = 0.030$ ) and 6q16.1 rs1408282 ( $P = 0.046$ ) (Table 4). In a secondary analysis stratifying by pathogen score (see Methods) above ( $n=832$ ) or below ( $n=877$ ) the median value, the estimated increase in log-CRP level for each additional T allele for the *HNFA* SNP rs7305618 was 0.348 ( $P = 1.3 \times 10^{-7}$ ) in individuals with higher pathogen score but was only 0.181 ( $P = 0.0063$ ) in those with lower pathogen score. Similar findings were detected for the SNP rs1408282 at 6q16.1 with strong and significant association predominantly detected in individuals with higher pathogen score ( $\beta = 0.593$ ,  $P = 1.5 \times 10^{-6}$ ) compared to those with lower pathogen score ( $\beta = 0.226$ ,  $P = 0.057$ ). We observed that the genetic influence of *LEPR* variant rs1892534 on plasma CRP is stronger in individuals under lower exposure to a pathogenic environment ( $\beta = 0.249$ ,  $P = 0.0047$ ) than those having higher exposure ( $\beta = 0.068$ ,  $P = 0.44$ ). Additionally, we found no evidence for genotype by waist circumference interaction on plasma CRP in Filipino women (All  $P$  for interaction  $> 0.20$ ).

As plasma CRP is highly correlated with several traits related to CVD, we further explored whether the CRP-associated SNPs were also significantly associated with quantitative traits including BMI, waist circumference, lipid profile, and homocysteine. Strong associations were observed for the *APOE* variant rs729358 with LDL-C ( $P = 6.5 \times 10^{-7}$ ) and with total cholesterol levels ( $P = 7.2 \times 10^{-5}$ ) (Table 5). Significant evidence for association was also found for *GCKR* variant rs1260326 with triglyceride level ( $P = 0.0064$ ), for *CRP* variant rs3091244 with triglycerides ( $P = 0.036$ ) and with LDL cholesterol ( $P = 0.048$ ). In an exploratory analysis to assess whether the associations between CRP associated SNPs and CVD related traits were mediated by CRP concentrations, adjustment for plasma CRP did not attenuate results (Supplementary Table 6).

## Discussion

In this GWA study of plasma CRP in Filipino women, we observed evidence of association ( $P < 5 \times 10^{-8}$ ) with SNPs at *CRP* and *HNFA*, provided supporting evidence for SNPs at *APOE* and *LEPR*, and detected suggestive evidence at a novel locus on chromosome 6. The five loci together explain an estimated proportion of CRP variance of 5.6% in CLHNS

women. We further observed that the associations between CRP concentrations and *HNF1A* genetic variants were stronger in women with a higher level of exposure to a pathogenic environment.

Our strongest association result mapped to the *CRP* locus on chromosome 1, and explained ~2.0% of the total variation in plasma CRP in CLHNS samples. Encoding the protein CRP, the gene *CRP* has been widely investigated as a candidate gene for circulating CRP level. Several *CRP* variants including the promoter triallelic SNP rs3091244 [16–17], synonymous SNP rs1800947 [19–20], 3' UTR variant rs1130864 [20–21] and 3' flanking variant rs1205 [22–23] have been reproducibly reported to associate with circulating CRP. The Framingham Heart Study (FHS), based on 3,301 predominantly European-descended participants, revealed that rs3091244 was the only SNP that remained significantly associated with CRP level after accounting for correlation with other SNPs. An *in vitro* study further suggested its functional importance by demonstrating that the transcription factor binding occurred only when the variant T allele was present [16]. In our study, we observed strong evidence of the association with the triallelic SNP rs3091244 ( $P = 5.2 \times 10^{-7}$ ). However, the T allele suggested to be functionally relevant, only displayed modest association with increased CRP level in CLHNS samples ( $P = 0.021$ ), and the lower frequency of the T allele in Filipinos (T: 0.08) compared to that in Framingham (T: 0.31) suggested a difference in genetic background on the *CRP* gene promoter region between populations. As the lead SNP showing the strongest association with plasma CRP in our GWA study, the variant rs876537 residing 7 kb downstream of the *CRP* gene represents a stronger signal for CRP than the triallelic SNP rs3091244 based on our findings of reciprocal conditional analysis (conditioned  $P$  for rs3091244 = 0.12,  $P$  for rs876437 =  $6.8 \times 10^{-4}$ , Supplementary Table 5A). Additional investigation of the association of SNPs imputed based on 1000 Genome Project Pilot data within the 200 kb flanking region of the SNP rs876537, together with the analysis results from conditioning on rs876537 revealed no separate signal for CRP level at the *CRP* locus in CLHNS (Supplementary Figure 1). Of note, the strongest CRP associated SNPs imputed from either HapMap or 1000 Genomes Project Pilot data were both mainly clustered in the 3'-UTR or 3' flanking region of *CRP* gene (chr1: 157,919,000–157,949,000). As these gene regions are involved in post-transcriptional processes by regulating mRNA stability, mRNA localization and translational efficiency, SNPs in this region may affect protein production as well as circulating CRP levels. Thus, fine-mapping in this region may help to locate the functional variants that contribute to the variation in CRP levels.

We confirmed in CLHNS the findings from studies of European origin individuals [3–4] showing significant association between CRP level and variants at the *HNF1A* locus. The proportion of CRP variation explained by *HNF1A* was 1.9% in our study, higher than the previously reported 1.1% in Europeans [3]. The protein HNF1 $\alpha$  is shown to bind directly to the *CRP* promoter to activate gene transcription and to promote cytokine-driven *CRP* gene expression [24]. Given the complex role of HNF1 $\alpha$  in modulating the expression of target genes like *CRP*, common variants at *HNF1A* locus are suggested to alter the gene and protein function via multiple mechanisms [25]. The missense variants rs1169288 (Ile27Leu) and rs2464196 (Ser486Asn) were previously reported to show association with CRP level and suggested to be functionally relevant [25]. We observed significant association with these two variants in CLHNS women ( $P = 1.4 \times 10^{-7}$  and  $7.5 \times 10^{-6}$  for Ile27Leu and Ser482Asn, respectively), but conditional analysis suggested that the lead SNP in this region (rs7305618) represents a stronger signal at the *HNF1A* locus, thus providing little evidence for Ile27Leu (conditional  $P = 0.0033$  for rs7305618 and 0.089 for Ile27Leu) or Ser482Asn (conditional  $P = 0.0002$  for rs7305618 and 0.78 for Ile27Leu) to be functional. Using data imputed based on 1000 Genome Project, we observed that the strongest signals in the

CLHNS samples for CRP at the *HNFA* locus were located near the promoter and 5' flanking region of the gene (Data not shown).

We did not observe evidence for association between CRP level and the *IL6R*, *GCKR*, and 12q23.2 loci reported in WGHS, although the estimated directions of effect were the same for *IL6R* and *GCKR*. Further analysis using the 1000 Genome Project imputed SNPs within 2 Mb flanking gene regions of these reported loci did not provide evidence for additional CRP associated signals. As we had > 89% power to detect the reported percentage of total variation explained by each additional copy of the effect allele [3] in CLHNS, the inconsistent findings between studies may be attributed to an overestimation of the original associations owing to the “winner’s curse” [26], or differing genetic or environmental backgrounds between populations.

This current study provided evidence for interaction between a gene and exposure to a pathogenic environment on CRP level in CLHNS women. Pathogen exposure was determined by interviewer assessment of several measures that serve as proxies for exposure to infectious microbes (see Methods). As suggested, taking account of gene by environment interactions may help elucidate the potential environmental factors that affect only a subgroup of genetically susceptible individuals and provide insight into the biological mechanisms underlying complex diseases [27]. Pathogen exposure is a strong determinant for elevated CRP concentration, as CRP has the capacity to recognize pathogens and participates in the systemic response to acute inflammatory stimulus by recruiting and activating complement, opsonizing pathogens, and promoting activity of phagocytic cells [28]. Prior findings of a strong correlation between exposure to a pathogenic environment and CRP in the CLHNS sample provided impetus to further examine gene-by-pathogen interactions on CRP level [9–10]. Our findings suggest that the genetic effects of *HNFA* (rs7305618), *LEPR* (rs1892534) and 6q16.1 (rs1408282) on plasma CRP differ with exposure to a pathogenic environment and that the exposure may modify the genetic susceptibility to elevated CRP level in Filipino women.

Prior CLHNS studies noted that CRP levels were substantially lower in the Philippines than in the United States [9–10]. The shared genetic variants associated with CRP concentrations identified in prior GWA scans, and in the Filipinos studied here [3–4] suggest that the variation in CRP level across ethnicities is not explained by differences in genetic background across populations, but largely by environmental and developmental factors. Interestingly, the recent finding that higher level of microbial exposures in infancy predicted lower CRP levels in adulthood among the offspring of the mothers studied here suggests that developmental plasticity is an important influence on the development of anti-pathogen defenses and could help explain the low CRP levels in the Philippines relative to the US and elsewhere [29].

In conclusion, we performed a genome-wide association for plasma CRP in a population of Filipino women. We provide confirmatory evidence for the association of previously identified loci at *CRP* and *HNFA*, which suggests shared genetic influences on CRP level between this sample in the Philippines and the European-descended populations studied in past GWA scans. To our knowledge, this study is the first to report an interaction between exposure to a pathogenic environment and genetic variants on plasma CRP. These data suggest that different exposure intensity to a pathogenic environment may modify the effect of the alleles associated with plasma CRP level. On the basis of the interactive pattern observed here, further studies in larger populations and investigation into potential biological mechanisms is warranted.



## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank the Office of Population Studies Foundation research and data collection teams and the study participants who generously provided their time for this study.

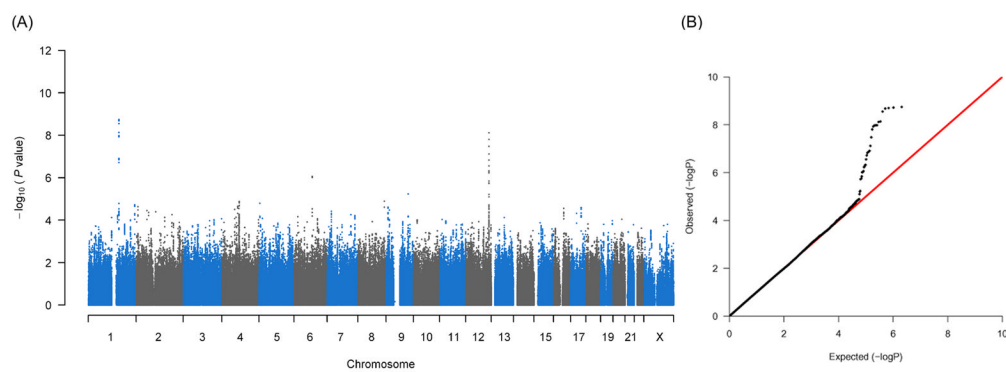
### Sources of Funding

This work was supported by National Institutes of Health grants DK078150, TW05596, HL085144, RR20649, ES10126, and DK56350.

## References

- Hage FG, Szalai AJ. C-reactive protein gene polymorphisms, C-reactive protein blood levels, and cardiovascular disease risk. *J Am Coll Cardiol.* 2007; 50:1115–1122. [PubMed: 17868801]
- Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH, Tracy RP. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis.* 2001; 154:681–689. [PubMed: 11257270]
- Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP, Chasman DI. 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.* 2008; 82:1185–1192. [PubMed: 18439548]
- Reiner AP, Barber MJ, Guan Y, Ridker PM, Lange LA, Chasman DI, Walston JD, Cooper GM, Jenny NS, Rieder MJ, Durda JP, Smith JD, Novembre J, Tracy RP, Rotter JI, Stephens M, Nickerson DA, Krauss RM. Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am J Hum Genet.* 2008; 82:1193–1201. [PubMed: 18439552]
- Cordell HJ. Genome-wide association studies: Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet.* 2009; 10:392–404. [PubMed: 19434077]
- Moore JH, Williams SM. Epistasis and its implications for personal genetics. *Am J Hum Genet.* 2009; 85:309–320. [PubMed: 19733727]
- WHO. The World Health Report 2004 --changing histroy. Geneva: World Health Orgnization; 2004.
- WHO. Mortality country fact sheet 2006. Geneva: World Health Orgnization; 2006.
- McDade TW, Rutherford JN, Adair L, Kuzawa C. Adiposity and pathogen exposure to a pathogenic environment predict C-reactive protein in Filipino women. *J Nutr.* 2008; 138:2442–2447. [PubMed: 19022970]
- McDade TW, Rutherford JN, Adair L, Kuzawa C. Population differences in associations between C-reactive protein concentration and adiposity: comparison of young adults in the Philippines and the United States. *Am J Clin Nutr.* 2009; 89:1237–1245. [PubMed: 19225115]
- Adair LS, Popkin BM, Akin JS, Guilkey DK, Gultiano S, Borja J, Perez L, Kuzawa CW, McDade T, Hindin MJ. Cohort Profile: The Cebu Longitudinal Health and Nutrition Survey. *Int J Epidemiol.* 2010 [Epub ahead of print].
- Lange LA, Croteau-Chonka DC, Marvelle AF, Qin L, Gaulton KJ, Kuzawa CW, McDade TW, Wang Y, Li Y, Levy S, Borja JB, Lange EM, Adair LS, Mohlke KL. Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. *Hum Mol Genet.* 2010; 19:2050–2058. [PubMed: 20154341]
- Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology.* 2010; 34:816–834. [PubMed: 21058334]
- Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26:2336–2337. [PubMed: 20634204]

15. Morita A, Nakayama T, Doba N, Hinohara S, Mizutani T, Soma M. Genotyping of triallelic SNPs using TaqMan PCR. *Mol Cell Probes*. 2007; 21:171–176. [PubMed: 17161935]
16. Szalai AJ, Wu J, Lange EM, McCrory MA, Langefeld CD, Williams A, Zakharkin SO, George V, Allison DB, Cooper GS, Xie F, Fan Z, Edberg JC, Kimberly RP. Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *J Mol Med*. 2005; 83:440–447. [PubMed: 15778807]
17. Kathiresan S, Larson MG, Vasani RS, Guo CY, Gona P, Keaney JF Jr, Wilson PW, Newton-Cheh C, Musone SL, Camargo AL, Drake JA, Levy D, O'Donnell CJ, Hirschhorn JN, Benjamin EJ. Contribution of clinical correlates and 13 C-reactive protein gene polymorphisms to interindividual variability in serum C-reactive protein level. *Circulation*. 2006; 113:1415–1423. [PubMed: 16534007]
18. Judson R, Brain C, Dain B, Windemuth A, Ruano G, Reed C. New and confirmatory evidence of an association between APOE genotype and baseline C-reactive protein in dyslipidemic individuals. *Atherosclerosis*. 2004; 177:345–351. [PubMed: 15530909]
19. Suk HJ, Ridker PM, Cook NR, Zee RY. Relation of polymorphism within the C-reactive protein gene and plasma CRP levels. *Atherosclerosis*. 2005; 178:139–145. [PubMed: 15585211]
20. Kovacs A, Green F, Hansson LO, Lundman P, Samnegard A, Boquist S, Ericsson CG, Watkins H, Hamsten A, Tornvall P. A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. *Atherosclerosis*. 2005; 178:193–198. [PubMed: 15585218]
21. Brull DJ, Serrano N, Zito F, Jones L, Montgomery HE, Rumley A, Sharma P, Lowe GD, World MJ, Humphries SE, Hingorani AD. Human CRP gene polymorphism influences CRP levels: implications for the prediction and pathogenesis of coronary heart disease. *Arterioscler Thromb Vasc Biol*. 2003; 23:2063–2069. [PubMed: 12842840]
22. Miller DT, Zee RY, Suk Danik J, Kozlowski P, Chasman DI, Lazarus R, Cook NR, Ridker PM, Kwiatkowski DJ. Association of common CRP gene variants with CRP levels and cardiovascular events. *Ann Hum Genet*. 2005; 69:623–638. [PubMed: 16266402]
23. Lange LA, Carlson CS, Hindorf LA, Lange EM, Walston J, Durda JP, Cushman M, Bis JC, Zeng D, Lin D, Kuller LH, Nickerson DA, Psaty BM, Tracy RP, Reiner AP. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. *JAMA*. 2006; 296:2703–2711. [PubMed: 17164456]
24. Nishikawa T, Hagihara K, Serada S, Isobe T, Matsumura A, Song J, Tanaka T, Kawase I, Naka T, Yoshizaki K. Transcriptional complex formation of c-Fos, STAT3, and hepatocyte NF-1 alpha is essential for cytokine-driven C-reactive protein gene expression. *J Immunol*. 2008; 180:3492–3501. [PubMed: 18292576]
25. Reiner AP, Gross MD, Carlson CS, Bielinski SJ, Lange LA, Fornage M, Jenny NS, Walston J, Tracy RP, Williams OD, Jacobs DR Jr, Nickerson DA. Common coding variants of the HNF1A gene are associated with multiple cardiovascular risk phenotypes in community-based samples of younger and older European-American adults: the Coronary Artery Risk Development in Young Adults Study and The Cardiovascular Health Study. *Circ Cardiovasc Genet*. 2009; 2:244–254. [PubMed: 20031592]
26. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*. 2003; 33:177–182. [PubMed: 12524541]
27. Thomas D. Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet*. 2010; 11:259–272. [PubMed: 20212493]
28. Black S, Kushner I, Samols D. C-reactive Protein. *J Biol Chem*. 2004; 279:48487–48490. [PubMed: 15337754]
29. McDade TW, Rutherford J, Adair L, Kuzawa CW. Early origins of inflammation: microbial exposures in infancy predict lower levels of C-reactive protein in adulthood. *Proc Biol Sci*. 2010; 277:1129–1137. [PubMed: 20007176]



**Figure 1.**  
Genome-wide association with plasma CRP in CLHNS  
(A) Manhattan plot of results from the CLHNS genome-wide association study of plasma CRP level  
(B) Quantile-Quantile plot of observed and expected  $P$  values for association with plasma CRP level

Table 1

Loci associated with plasma CRP in CLHNS women

SNP	Closest Gene	Chr	Position	Effect Allele	Non-Effect Allele	Minor Allele: Frequency	Baseline Model*		Model 2 <sup>†</sup>	
							$\beta$ (SE)	P value	$\beta$ (SE)	P value
rs876537	CRP	1	157,941,557	C	T	C: 0.43	0.288 (0.047)	1.4E-09	0.288 (0.043)	3.4E-11
rs7305618	HNF1A	12	119,887,315	T	C	C: 0.48	0.267 (0.046)	1.0E-08	0.267 (0.042)	4.0E-10
rs1408282	6q16.1	6	93,908,973	A	G	A: 0.10	0.412 (0.085)	1.4E-06	0.367 (0.078)	2.9E-06

Plasma CRP level was increased by 0.10 and then natural log transformed to satisfy model assumptions.  $\beta$  values correspond to the change in log-transformed CRP with each additional copy of the effect allele;

\* Baseline model: adjusted for age, household assets, number of previous pregnancies, household income (natural log transformed), pathogen score, infectious status and PC1-7 (n = 1,709);

<sup>†</sup> Model 2: further adjusted for waist circumference based on baseline Model (n=1,696)

**Table 2**  
SNP association with plasma CRP in CLHNS for previously reported variants

SNP	Nearby Gene	Annotation	Chr	Effect Allele	Non-effect Allele	Previous Studies			CLHNS			
						Effect Allele Freq	Reported $\beta$	Reported P	Reference	Effect Allele Freq	$\beta$ (SE)	P
rs1205	CRP	3' UTR	1	C	T	0.66	0.20	1.7E-26	(3)	0.46	0.257 (0.044)	8.5E-09
rs3091244	CRP	Promoter	1	A/T	C	A:0.07 T:0.31	0.20	6.2E-28	(3)	A:0.13 T:0.08	0.257 (0.051)	5.2E-07
rs1169288	HNF1A	Ile27Leu	12	A	C	0.72	0.12	2.6E-07	(4)	0.66	0.265 (0.050)	1.4E-07
rs2464196	HNF1A	Ser486Asn	12	G	A	0.69	0.14	3.3E-13	(3)	0.55	0.198 (0.044)	7.5E-06
rs1169286	HNF1A	Intron	12	T	C	0.61	0.10	1.6E-06	(4)	0.56	0.246 (0.049)	5.1E-07
rs7310409	HNF1A	Intron	12	G	A	0.57	0.15	6.8E-17	(3)	0.67	0.223 (0.046)	1.6E-06
rs429358	APOE	Cys130Arg	19	T	C	0.79	ND	3.3E-08	(18)	0.92	0.272 (0.078)	4.8E-04
rs1892534	LEPR	Downstream	1	C	T	0.67	0.17	6.5E-21	(3)	0.15	0.110 (0.062)	0.076
rs4129267	IL6R	Intron	1	C	T	0.67	0.10	2.0E-08	(3)	0.71	0.068 (0.058)	0.24
rs1260326	GCKR	Leu446Pro	2	T	C	0.43	0.16	3.6E-14	(3)	0.42	0.061 (0.048)	0.20
rs10778213	12q23.2	----	12	C	T	0.48	0.12	1.2E-10	(3)	0.80	0.027 (0.054)	0.63

Effect allele: the allele that is associated with elevated CRP levels;

**Table 3**

CRP association with *APOE* SNPs and haplotypes

SNP	Genotype	n	LS means (SE)	P (add)	P (dom)
rs429358	TT	1415	0.031 (0.050)		
	TC	252	-0.310 (0.087)	4.8E-04	1.0E-04
	CC	15	0.104 (0.326)		
rs7412	CC	1307	-0.011 (0.051)		
	CT	340	-0.011 (0.079)	0.86	0.95
	TT	26	-0.132 (0.250)		
Haplotype	rs429358-rs7412	Frequency	Hap-score	P	
<i>APOE</i> ε2	T-T	0.116	-0.23	0.81	
<i>APOE</i> ε3	T-C	0.799	2.55	0.01	
<i>APOE</i> ε4	C-C	0.085	-3.39	7.1E-04	

The "haplo.score" function implemented in the "haplo.stats" R package was used to compute score statistics to test association between haplotypes and plasma CRP. Hap-score is the score statistic for the specific haplotype. The global *P* value for association between haplotypes and CRP level was  $2.6 \times 10^{-3}$ .

**Table 4**

Interaction between genetic variants and pathogen exposure on plasma CRP in 1,709 CLHNS individuals

SNP	Gene	Low pathogen exposure		High pathogen exposure		<i>P</i> (inter)
		$\beta$ (SE)	<i>P</i>	$\beta$ (SE)	<i>P</i>	
rs876537	<i>CRP</i>	0.220 (0.067)	0.0010	0.345 (0.067)	3.6E-07	0.18
rs7305618	<i>HNFLA</i>	0.181 (0.066)	0.0063	0.348 (0.065)	1.3E-07	0.031
rs429358	<i>APOE</i>	0.333 (0.108)	0.0022	0.256 (0.112)	0.022	0.85
rs1892534	<i>LEPR</i>	0.249 (0.088)	0.0047	0.068 (0.089)	0.44	0.030
rs4129267	<i>IL6R</i>	0.010 (0.081)	0.90	0.144 (0.082)	0.079	0.097
rs1260326	<i>GCKR</i>	0.066 (0.066)	0.32	0.053 (0.069)	0.45	0.69
rs10778213	12q23.2	0.062 (0.077)	0.42	0.010 (0.077)	0.90	0.72
rs1408282	6q16.1	0.226 (0.119)	0.057	0.593 (0.123)	1.5E-06	0.046

Pathogen score < 0.5 (median) was defined as low pathogen exposure (n = 877, 51.3%) while the pathogen score  $\geq$  0.5 was defined as high pathogen exposure (n = 832, 48.7%); *P* (inter): *P* for interaction between genotype and pathogen score on CRP level.

Table 5

Associations between SNPs and other cardiovascular traits in CLHNS

SNP ( <i>Gene</i> )	BMI	Waist circumference	TG	HDLc	LDLc	Chol	Hcy
rs876537 ( <i>CRP</i> )	$\beta$ (SE)	-0.163 (0.393)	-0.027 (0.018)	0.016 (0.009)	0.014 (0.011)	0.007 (0.008)	-0.005 (0.009)
	<i>P</i>	0.68	0.14	0.088	0.20	0.34	0.55
rs3091244 ( <i>CRP</i> )	$\beta$ (SE)	-0.497 (0.421)	-0.041 (0.020)	0.010 (0.010)	0.023 (0.012)	0.010 (0.008)	0.008 (0.010)
	<i>P</i>	0.24	0.036	0.31	0.048	0.22	0.38
rs7305618 ( <i>HNF1A</i> )	$\beta$ (SE)	0.068 (0.385)	0.008 (0.018)	-0.005 (0.009)	-0.019 (0.011)	-0.013 (0.007)	-0.012 (0.007)
	<i>P</i>	0.86	0.65	0.61	0.065	0.090	0.10
rs429358 ( <i>APOE</i> )	$\beta$ (SE)	0.610 (0.641)	-0.022 (0.030)	0.021 (0.015)	-0.086 (0.017)	-0.049 (0.012)	-0.012 (0.014)
	<i>P</i>	0.34	0.45	0.18	6.5E-07	7.2E-05	0.43
rs1892534 ( <i>LEPR</i> )	$\beta$ (SE)	0.189 (0.512)	-0.029 (0.024)	0.004 (0.012)	-0.005 (0.014)	-0.006 (0.010)	0.002 (0.012)
	<i>P</i>	0.71	0.23	0.76	0.74	0.52	0.88
rs4129267 ( <i>IL6R</i> )	$\beta$ (SE)	0.061 (0.476)	0.031 (0.022)	-0.004 (0.011)	-0.004 (0.013)	-0.010 (0.009)	0.013 (0.011)
	<i>P</i>	0.90	0.15	0.71	0.75	0.27	0.22
rs1260326 ( <i>GCKR</i> )	$\beta$ (SE)	0.051 (0.392)	0.049 (0.018)	0.008 (0.009)	0.000 (0.011)	0.008 (0.008)	-0.008 (0.009)
	<i>P</i>	0.90	0.0064	0.42	0.95	0.28	0.38
rs10778213 ( <i>J2q23.2</i> )	$\beta$ (SE)	-0.266 (0.447)	0.008 (0.021)	-0.002 (0.011)	-0.008 (0.012)	-0.002 (0.009)	-0.003 (0.010)
	<i>P</i>	0.55	0.71	0.88	0.52	0.86	0.73
rs1408282 ( <i>6q16.1</i> )	$\beta$ (SE)	1.117 (0.704)	-0.020 (0.033)	-0.010 (0.017)	0.005 (0.019)	-0.001 (0.014)	0.012 (0.016)
	<i>P</i>	0.11	0.55	0.54	0.79	0.92	0.45

All traits were natural log-transformed except waist circumference;

TG: triglyceride; HDLc: HDL-Cholesterol; LDLc: LDL-Cholesterol; Chol: total cholesterol; Hcy: homocysteine;

Adjusted for age, household assets, household income (natural log transformed), number of previous pregnancies, pathogen score, infectious status, and PC1-7;