

University of Groningen

In Silico Post Genome-Wide Association Studies Analysis of C-Reactive Protein Loci Suggests an Important Role for Interferons

Vaez, Ahmad; Jansen, Rick; Prins, Bram P.; Hottenga, Jouke-Jan; de Geus, Eco J. C.; Boomsma, Dorret I.; Penninx, Brenda W. J. H.; Nolte, Ilja M.; Snieder, Harold; Alizadeh, Behrooz Z.

Published in:
Circulation-Cardiovascular Genetics

DOI:
[10.1161/CIRCGENETICS.114.000714](https://doi.org/10.1161/CIRCGENETICS.114.000714)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Vaez, A., Jansen, R., Prins, B. P., Hottenga, J.-J., de Geus, E. J. C., Boomsma, D. I., Penninx, B. W. J. H., Nolte, I. M., Snieder, H., & Alizadeh, B. Z. (2015). In Silico Post Genome-Wide Association Studies Analysis of C-Reactive Protein Loci Suggests an Important Role for Interferons. *Circulation-Cardiovascular Genetics*, 8(3), 487-497. <https://doi.org/10.1161/CIRCGENETICS.114.000714>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

In Silico Post Genome-Wide Association Studies Analysis of C-Reactive Protein Loci Suggests an Important Role for Interferons

Ahmad Vaez, MD; Rick Jansen, PhD*; Bram P. Prins, MSc*; Jouke-Jan Hottenga, PhD; Eco J.C. de Geus, PhD; Dorret I. Boomsma, PhD; Brenda W.J.H. Penninx, PhD; Ilja M. Nolte, PhD; Harold Snieder, PhD†; Behrooz Z. Alizadeh, MD, MSc, PhD†

Background—Genome-wide association studies (GWASs) have successfully identified several single nucleotide polymorphisms (SNPs) associated with serum levels of C-reactive protein (CRP). An important limitation of GWASs is that the identified variants merely flag the nearby genomic region and do not necessarily provide a direct link to the biological mechanisms underlying their corresponding phenotype. Here we apply a bioinformatics-based approach to uncover the functional characteristics of the 18 SNPs that had previously been associated with CRP at a genome-wide significant level.

Methods and Results—In the first phase of in silico sequencing, we explore the vicinity of GWAS SNPs to identify all linked variants. In the second phase of expression quantitative trait loci analysis, we attempt to identify all nearby genes whose expression levels are associated with the corresponding GWAS SNPs. These 2 phases generate several relevant genes that serve as input to the next phase of functional network analysis. Our in silico sequencing analysis using 1000 Genomes Project data identified 7 nonsynonymous SNPs, which are in moderate to high linkage disequilibrium ($r^2 > 0.5$) with the GWAS SNPs. Our expression quantitative trait loci analysis, which was based on one of the largest single data sets of genome-wide expression probes ($n > 5000$) identified 23 significantly associated expression probes belonging to 15 genes (false discovery rate < 0.01). The final phase of functional network analysis revealed 93 significantly enriched biological processes (false discovery rate < 0.01).

Conclusions—Our post-GWAS analysis of CRP GWAS SNPs confirmed the previously known overlap between CRP and lipids biology. Additionally, it suggested an important role for interferons in the metabolism of CRP. (*Circ Cardiovasc Genet.* 2015;8:487-497. DOI: 10.1161/CIRCGENETICS.114.000714.)

Key Words: C-reactive protein ■ genome wide association study ■ interferon

C-reactive protein (CRP), a pentameric molecule, is the most widely studied inflammatory marker.¹ Elevated levels of serum CRP have been associated with increased risks of cancer,² type 2 diabetes mellitus,³ hypertension,⁴ coronary heart disease,⁵ stroke,⁶ bipolar disorder,⁷ and overall mortality.⁸ However, its causal contribution to the pathophysiology of chronic diseases remains controversial.⁹⁻¹² Serum levels of CRP are regulated by both genetic and environmental factors.^{11,13} Its heritability has been reported to range from 10% to 65%,¹⁴⁻¹⁷ and genome-wide association studies (GWASs) have successfully identified several genetic variants associated with

CRP levels.^{12,18} A recent meta-analysis of 25 GWAS studies, including $> 80\,000$ subjects, identified 18 CRP genetic variants at genome-wide significance.¹⁹

Clinical Perspective on p 497

One important limitation of GWASs is that the identified single nucleotide polymorphisms (SNPs) are not necessarily causally related to their associated traits or diseases. Many GWAS SNPs merely flag causal variants in their vicinity.²⁰ Hence, identifying associated SNPs by GWAS does not necessarily provide sufficient information on the biological mechanisms or pathways underlying their corresponding phenotype.

Received May 21, 2014; accepted February 18, 2015.

From the Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen (A.V., B.P.P., I.M.N., H.S., B.Z.A.); Department of Psychiatry, VU University Medical Center, Amsterdam (R.J., B.W.J.H.P.); and Neuroscience Campus Amsterdam, VU University and VU University Medical Center, Amsterdam (R.J., J.-J.H., E.J.C.d.G., D.I.B., B.W.J.H.P.), EMGO+ Institute, VU University and VU University Medical Center, Amsterdam (J.-J.H., E.J.C.d.G., D.I.B., B.W.J.H.P.), Department of Biological Psychology, VU University, Amsterdam, the Netherlands (J.-J.H., E.J.C.d.G., D.I.B.); and School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran (A.V.).

*Dr Jansen and B.P. Prins contributed equally to this work.

†Drs Snieder and Alizadeh contributed equally to this work.

The Data Supplement is available at <http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.114.000714/-DC1>.

Correspondence to Behrooz Z. Alizadeh, MD, MSc, PhD, Department of Epidemiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1 (9713 GZ), Groningen, the Netherlands. E-mail b.z.alizadeh@umcg.nl or Ahmad Vaez, MD, Department of Epidemiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1 (9713 GZ), Groningen, the Netherlands. E-mail a.vaez@umcg.nl

© 2015 American Heart Association, Inc.

Circ Cardiovasc Genet is available at <http://circgenetics.ahajournals.org>

DOI: 10.1161/CIRCGENETICS.114.000714

Therefore, after a successful GWAS study, it is essential to perform additional post-GWAS analyses to translate the GWAS findings represented by index SNPs into biological knowledge.²¹ For example, we previously demonstrated that serum protein levels are regulated by ribosomal functioning, proteasomal degradation, and immune-response signaling pathways, leading to a better functional understanding of the GWAS findings for serum protein levels.²² However, an in-depth post-GWAS analysis for CRP variants has not yet been performed,¹⁹ which means that CRP GWAS findings have been insufficiently translated into biological function. Consequently, the gain in knowledge on underlying mechanisms controlling CRP level has been limited. Given the clinical relevance of CRP as an established biomarker for many complex chronic disorders, an extended post-GWAS analysis of CRP variants may unravel new mechanisms, which will improve our understanding of the metabolism of CRP and its relevance to disease pathology.

Here we applied a bioinformatics-based approach to uncover the functional characteristics of the 18 CRP-associated variants.¹⁹ We first performed an *in silico* sequencing analysis using 1000 Genomes Project data²³ to identify nearby nonsynonymous coding variants. Second, we performed an expression quantitative trait loci (eQTL) analysis using a large data set of blood expression probes to find regulatory variants. Third, we integrated the findings of the abovementioned phases by performing a functional network analysis to unravel the underlying biological processes.

Methods

We followed a bioinformatics-based approach, including 3 distinct phases, each consisting of multiple steps as described later (Figure 1).

Phase I: In Silico Sequencing

Identifying Linked Variants

First, we converted the chromosome positions of the GWAS SNPs (gSNPs) from the National Center for Biotechnology Information Build 36 (Human Genome 18) to National Center for Biotechnology Information Build 37 (Human Genome 19) using the LiftOver tool from the University of California Santa Cruz (UCSC) Genome Project.²⁴ Then, we targeted regions of 1 Mb at either side of each gSNP, resulting in a mini-genome of 36 Mb. The appropriate Variant Call Format²⁵ file for each 2 Mb region was downloaded from the 1000 Genomes Project ftp server using the Tabix software package.²⁶ We used the data from the 1000 Genomes Project Full Phase 1, November 2010 release (using August 2010 alignments), including only the 283 subjects of European ancestry.²³ Subsequently, for each Variant Call Format file, the r^2 between the gSNP and all other biallelic SNPs residing within the corresponding 2 Mb area was calculated as a metric of linkage disequilibrium (LD) using VCFtools.²⁵ Only those SNPs in moderate to high ($r^2 > 0.50$) LD with the corresponding gSNP were used in the next step of the analysis (Figure 1).

Identifying Linked Nonsynonymous SNPs

All these SNPs in LD with any of the gSNPs were annotated by ANNOVAR software²⁷ and then filtered in a stepwise manner. First, the SNPs were annotated to distinguish exonic variants from other variant types (intronic, intergenic, etc.). Nonexonic variants were excluded from further analyses. The remaining SNPs were annotated again to distinguish synonymous from nonsynonymous exonic SNPs, and synonymous SNPs were excluded. As a further step, the nonsynonymous SNPs (nsSNPs) were then characterized for their damaging effect on the corresponding protein using Sorting Intolerant From Tolerant (SIFT)²⁸ and Polymorphism Phenotyping (PolyPhen)²⁹

prediction scores. Their scores were obtained from Ensembl release 71 (accessed June 8, 2013).³⁰ Whenever multiple scores were available for a single nsSNP, we selected the most damaging prediction scores as the smallest SIFT and the largest PolyPhen scores. These scores are just provided as Data Supplement about linked variants and hence, were not used in the downstream analyses.

In Silico Pleiotropy Analysis

To extend our knowledge of the possible function of the 18 CRP-associated loci, we sought to identify any trait or outcome associated with these 18 loci. Thus, for all gSNPs, as well as all SNPs in LD ($r^2 > 0.80$) with any of the gSNPs, we checked for genome-wide significant ($P < 5 \times 10^{-8}$) pleiotropic effects on other complex traits or diseases identified in previous GWAS studies as listed in the National Human Genome Research Institute GWAS Catalog (Catalog of Published Genome-Wide Association Studies)³¹ using ANNOVAR software (accessed June 13, 2013).²⁷ However, as shown in Figure 1, the results of this step were not used in the downstream analyses, but were indeed used in the final interpretation of the results.

Phase II: eQTL Analysis

The data set of genome-wide expression probes and gene expression measurements have been described in more detail elsewhere.^{32,33}

Subjects

The 2 parent projects that supplied data for the eQTL analysis are large-scale longitudinal studies: the Netherlands Study of Depression and Anxiety³⁴ and the Netherlands Twin Registry.³⁵ The Netherlands Study of Depression and Anxiety and the Netherlands Twin Registry studies were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, and all subjects provided written informed consent. The sample used for eQTL analysis after quality control consisted of 5071 subjects, 3109 the Netherlands Twin Registry (from 1571 families: 614 dizygotic twin pairs, 1 monozygotic triplet, 668 monozygotic twin pairs, 394 siblings, and 148 unrelated subjects), and 1962 the Netherlands Study of Depression and Anxiety participants. The age of the participants ranged from 17 to 88 years (mean 38, SD 13), and 65% of the sample was female.³²

Blood Sampling, RNA Extraction, and Measurements

Venous blood samples were drawn in the morning after an overnight fast. Heparinized whole blood samples were transferred within 20 minutes of sampling into PAXgene Blood RNA tubes (Qiagen) and stored at -20°C . Gene expression assays were conducted at the Rutgers University Cell and DNA Repository (<http://www.rucdr.org>). Samples were hybridized to Affymetrix U219 arrays containing 530467 probes summarized in 49293 probesets. All probes are 25 bases in length and designed to be perfect match complements to a designated transcript. Array hybridization, washing, staining, and scanning were performed in an Affymetrix GeneTitan System per the manufacturer's protocol. Gene expression data were required to pass standard Affymetrix quality control metrics (Affymetrix expression console) before further analysis. Probes that did not map uniquely to Human Genome 19 or that contained a polymorphic SNP (dbSNP137 common with minor allele frequency > 0.01) were removed for downstream analysis, resulting in 423201 probes, summarized in 44241 probesets, targeting 18238 unique genes. Probeset expression values were obtained using robust multiarray average normalization implemented in Affymetrix Power Tools (APT, v 1.12.0). Samples with low average correlation with other samples and samples with incorrect sex-chromosome expression were removed.³²

Genotype Data

DNA extraction has been described earlier.³⁶ Genotyping was done on multiple chip platforms for several partly overlapping subsets of participants. The following platforms were used: Affymetrix Perlegen 5.0, Illumina 370, Illumina 660, Illumina Omni Express 1 mol/L, and Affymetrix 6.0. After array-specific data analysis, genotype calls were made with the platform-specific software (Genotyper, Beadstudio). The extensive genotyping quality control steps and 1000 Genomes

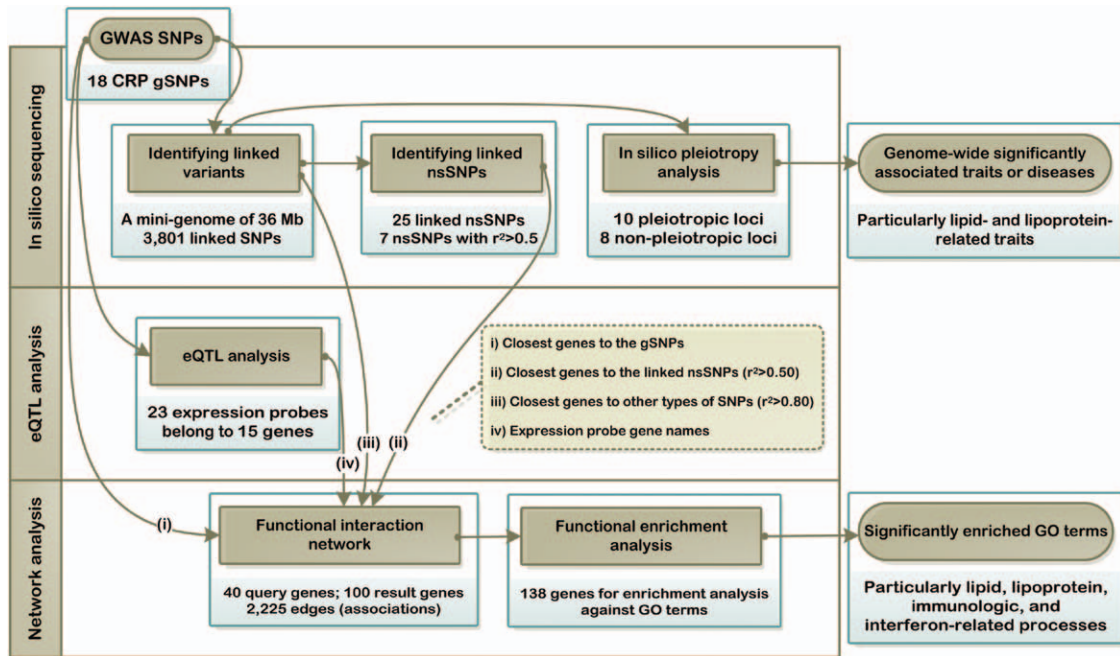


Figure 1. Flow diagram of the steps of CRP post-GWAS analysis. The inner grey boxes show the components of the pipeline, whereas the outer blue boxes show the main results of post-GWAS analysis of 18 genome-wide significantly associated CRP SNPs. CRP indicates C-reactive protein; eQTL, expression quantitative trait loci; GO, Gene Ontology; GWAS, genome-wide association study; gSNP, GWAS SNPs; nsSNP, nonsynonymous SNPs; and SNP, single nucleotide polymorphism.

imputation procedures are described in the Data Supplement Text S1. Genotypes were coded into dosage format and filtered at minor allele frequency >0.01 and imputation quality of $R^2 > 0.30$ for eQTL analysis.

eQTL Analysis

Inverse quantile normal transformation was applied to the individual probeset data to obtain normal distributions. The transformed probeset data were then residualized with respect to the covariates sex, age, body mass index, smoking status, several technical covariates, and 3 principal components (PCs) from the genotype data. Genotype PCs were constructed using pruned GWAS data after removing ethnic outliers as described earlier.³⁷ The residualized probeset data were subjected to a principal component analysis to remove the first 50 PCs to adjust the gene expression levels for nongenetic variation, as proposed by Fehrmann et al. They have shown that removing expression PCs drastically increases the number of eQTLs.³⁸ We observe the same phenomenon in our data. Removing expression PCs has become a standard procedure in many eQTL studies.^{33,39} Probesets at <1 Mb distance from the gSNPs were selected for eQTL analysis as follows: for each probeset–gSNP combination at maximally 1 Mb distance, a linear mixed model was fitted with expression level as dependent variable, genotype as fixed effect, and family ID and zygosity as random effects to account for family and twin relations.⁴⁰ Mixed models and resulting *P* values were computed using the function lmer from the lme4 R package (<http://CRAN.R-project.org/package=lme4>). To correct for multiple testing, false discovery rate (FDR) was computed using all *P* values from each probeset–gSNP combination at maximally 1 Mb distance using the function p.adjust from the stats R package, and any signal with $FDR < 0.01$ was considered significant. The appropriate gene names of those significantly associated expression probes were then used in the next step as a set of prioritized biological candidate genes (Figure 1).

As a further step, for each locus with significant eQTL signal of $FDR < 0.01$, we also identified the most significantly associated eQTL SNP (eSNP) for the corresponding transcript. We then performed conditional analyses to see if the gSNP is independently associated with the expression level. For conditional eQTL analysis, the transformed probeset data were residualized with respect to the corresponding eSNP before applying the mixed model. These eSNPs were not used in the downstream analysis (Figure 1).

Phase III: Network Analysis

Functional Interaction Network

To construct a functional association interaction network, we applied the GeneMANIA algorithm together with its large set of accompanying functional association data on coexpression, physical interaction, genetic interaction, shared protein domains, colocalization, and predicted association networks. This data set comprises 286 extended association networks.⁴¹

We combined 4 biologically prioritized candidate gene sets into a single query gene set, which was used as input for the interaction network analysis: (1) closest genes to the gSNPs, (2) closest genes to the nsSNPs in high LD ($r^2 > 0.50$) with the corresponding gSNP, (3) closest genes to other types of SNPs in very high LD ($r^2 > 0.80$) with the corresponding gSNP, and (4) expression probe gene names significantly ($FDR < 0.01$) associated with gSNPs based on the eQTL analysis (Figure 1). We used different LD thresholds for nsSNPs than other types of SNPs as nsSNPs are more likely to be functionally important and also are more likely to reside within a lower frequency spectrum. Consequently, nsSNPs may be in modest LD with common gSNPs. Therefore, we used a more lenient LD threshold for nsSNPs ($r^2 > 0.50$) to ensure not to miss potentially functional variants with modest frequency and a standard LD threshold of $r^2 > 0.80$ for other types of SNPs.

Next, we constructed a weighted composite functional association network using the Cytoscape software platform,⁴² extended by the GeneMANIA plugin.⁴³ We selected all available networks option with a 100-gene output (accessed July 15, 2013).

Functional Enrichment Analysis

All the genes in the composite network, either from the query or the resulting gene sets, were then used for functional enrichment analysis against Gene Ontology terms (GO terms) to identify the most relevant GO terms using the same plugin.⁴³ Each GO annotation has an evidence code indicating the type of experimental or computational support for that association, for example, inferred from reviewed computational analysis (RCA) or inferred from electronic annotation (IEA). The first one (RCA) points to those predictions based on computational analyses of experimental data sets like protein–protein interaction or expression data. The latter (IEA) points

to computationally assigned evidence codes, which have not been reviewed by a curator to verify their accuracy (<http://www.geneontology.org/GO.evidence.shtml>).⁴⁴ IEA is the least reliable, but the most prevalent evidence code, that is, about 47% of all of the human GO annotations are based on IEA codes (accessed July 26, 2013). As both RCA and IEA annotations are solely based on computational predictions, the functional enrichment analysis was only performed against GO term annotations with non-IEA and non-RCA evidence codes to avoid circularity.⁴⁴ We considered any GO term with FDR <0.01 as significantly and those GO terms with FDR between 0.01 and 0.1 as suggestively enriched. We then used the RamiGO R package⁴⁵ for the visualization of significant GO terms within the appropriate GO tree.

Results

Here, we followed a bioinformatics-based approach as summarized in Figure 1. We included the 18 SNPs that showed genome-wide significant association with CRP in the study by Dehghan et al¹⁹ (Table 1).

Phase I: In Silico Sequencing

In this phase, we aimed to explore thoroughly the genomic area around the 18 gSNPs to identify nearby nsSNPs as potentially functional variants. We used 1000 Genomes Project data as the most detailed catalogue of human genetic variation.²³ The mini-genome of 36 Mb contains 167 003 SNPs. Of these, 3801 SNPs are in LD with the nearby gSNP at $r^2 > 0.10$, of which only 48 are exonic, including 25 nsSNPs (Table I in the Data Supplement). Of the nsSNPs, 9 map to the same gene and 16 map to other genes than the gSNPs. Please note that Tables I–III in the Data Supplement provide a thorough description of the vicinity of gSNPs by applying a liberal cutoff of $r^2 > 0.1$.

Table 1. The 18 Genome-Wide Associated CRP SNPs Used as Primary Input to the Post-GWAS Analysis

No. of gSNP	SNP ID	Chr	Position	Alleles	
1	rs2794520	1	159678816	C	T
2	rs4420638	19	45422946	A	G
3	rs1183910	12	121420807	G	A
4	rs4420065	1	66161461	T	C
5	rs4129267	1	154426264	C	T
6	rs1260326	2	27730940	T	C
7	rs12239046	1	247601595	T	C
8	rs6734238	2	113841030	A	G
9	rs9987289	8	9183358	A	G
10	rs10745954	12	103483094	A	G
11	rs1800961	20	43042364	C	T
12	rs340029	15	60894965	C	T
13	rs10521222	16	51158710	C	T
14	rs12037222	1	40064961	G	A
15	rs13233571	7	72971231	C	T
16	rs2847281	18	12821593	A	G
17	rs6901250	6	117114025	G	A
18	rs4705952	5	131839618	G	A

The SNPs are ordered according to the significance of their association with CRP in the meta-GWAS article. Alleles indicates ensembl reference/alternative alleles; Chr, chromosome; CRP, C-reactive protein; gSNP, GWAS SNP; GWAS, genome-wide association study; SNP, single nucleotide polymorphism; position, chromosome position build 37.

These results are considered as complementary information. However, only 7 of the nsSNPs are in moderate to high LD ($r^2 > 0.5$) with the gSNPs and were used in the downstream analyses (Figure 1). The nsSNPs were then characterized for their deleterious effect on the corresponding protein function using 2 different tools, SIFT²⁸ and PolyPhen.²⁹ Interestingly, 8, 6, 4, and 10 of the nsSNPs are considered as damaging according to SIFT alone, PolyPhen alone, both SIFT and PolyPhen, or any of the 2 prediction scores, respectively (Figure 2 drawn by Circos⁴⁶; Tables I and II in the Data Supplement).

In silico pleiotropy analysis of all gSNPs, as well as all SNPs that are in LD with their nearby gSNP, identified several genome-wide significantly ($P < 5 \times 10^{-8}$) associated traits or diseases other than CRP that had already been reported in previous GWAS studies as listed in the GWAS catalog.³¹ By considering all gSNPs and only their highly linked variants ($r^2 > 0.80$), 10 loci had effects on other traits, whereas 8 loci, including the *CRP* locus itself, did not show any pleiotropic effect. The locus harboring *GCKR* was the most pleiotropic region, having reported GWAS associations with a variety of metabolic-related traits. Most of the identified traits are metabolic-related traits, particularly lipid- and lipoprotein-related traits, for example, cholesterol, high-density lipoprotein, low-density lipoprotein, and triglyceride levels (Figure 3; Table III in the Data Supplement).

Phase II: eQTL Analysis

In this phase, we aimed to perform an eQTL analysis to determine whether the gSNPs affect CRP levels through regulating gene expression levels. Here we used a large data set of genome-wide expression probes in peripheral blood consisting of 5071 subjects. The eQTL analysis identified 23 expression probes that were significantly associated with 8 gSNPs at FDR < 0.01. The 23 expression probes belong to 15 genes, of which 4 are the same genes and 11 are different genes from those mapping to the corresponding gSNPs. Those expression probe gene names were then used in the next step as a set of prioritized biological candidate genes (Figure 1).

Additionally, we identified the 23 SNPs that were most significantly associated with the corresponding expression probes (eSNPs; Figure 4; Table IV in the Data Supplement). eQTL analysis of the gSNPs conditional on the corresponding eSNPs revealed that for the majority of expression probes, the corresponding gSNP is not independently associated with expression levels, that is, the observed effect of gSNPs on expression probes are mostly explained by the eSNPs (Table IV in the Data Supplement).

Phase III: Network Analysis

In this phase, we generated a list of biologically prioritized candidate genes based on the findings of phases I and II as input for the construction of a functional interaction network as detailed in the methods section. Four sets of query genes were combined to create the final input list of prioritized genes for the functional interaction network analysis (Figure 1). After removing duplicate entries, the combined query gene set contained 40 genes. Two genes (*LOC157273* and *PPIEL*) could not be found in any of the available interaction resources, resulting in a final list of 38

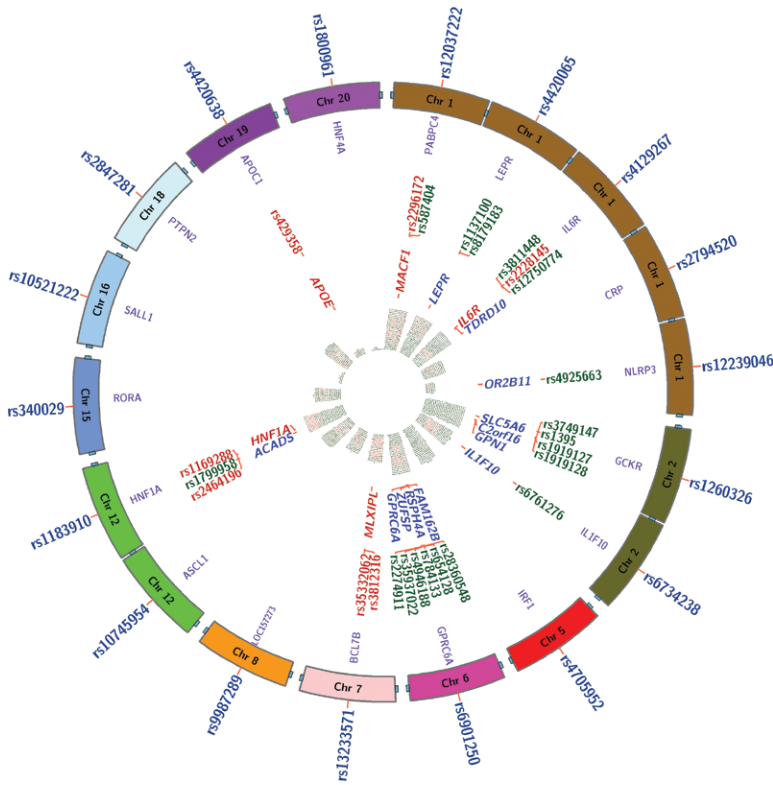


Figure 2. Results of in silico sequencing (drawn by Circos).⁴⁶ It illustrates the map of nsSNPs within the 2 Mb vicinity of 18 CRP-associated SNPs. The rings from outermost to innermost represent (a) 18 CRP-associated SNPs (gSNPs), (b) genomic regions of 2 Mb surrounding each gSNP, (c) closest genes to the gSNPs, (d) 25 nsSNPs in LD with the gSNPs, (e) closest genes to the nsSNPs, (f) 3801 SNPs in LD with the gSNP at $r^2 > 0.10$. The red color in rings d, e, and f indicates moderate to high LD ($r^2 > 0.50$) with the corresponding gSNP. CRP indicates C-reactive protein; GWAS, genome-wide association study; LD, linkage disequilibrium; gSNP, GWAS SNPs; nsSNP, nonsynonymous SNPs; and SNP, single nucleotide polymorphism.

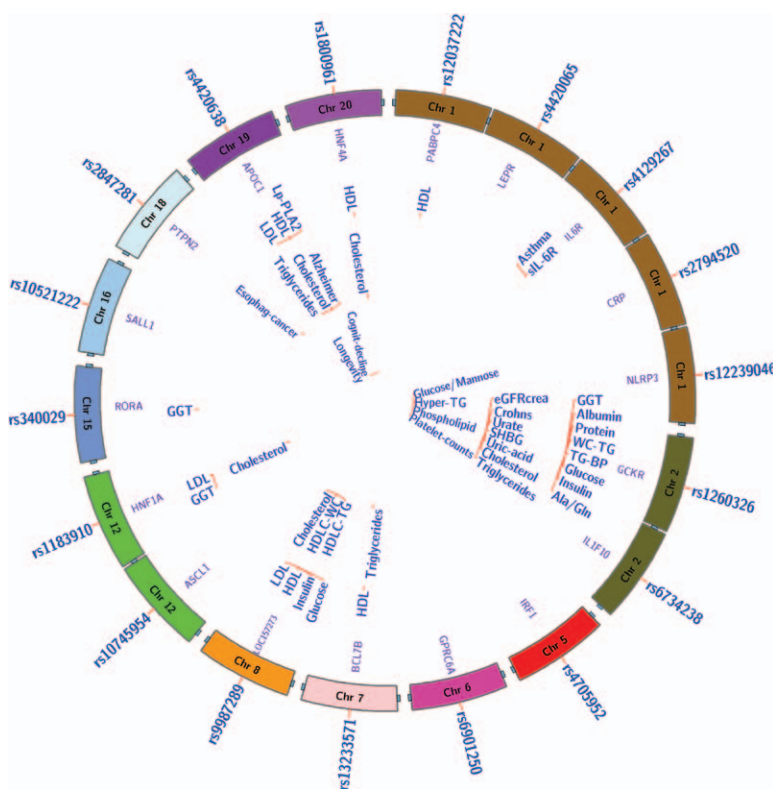


Figure 3. Results of in silico pleiotropy analysis. The 3 innermost rings show complex traits or diseases other than CRP, identified in previous GWAS studies to be genome-wide significantly associated with any of the gSNPs or their highly linked variants ($r^2 > 0.80$); Ala/Gln indicates alanine/glutamine; Cognit-decline, cognitive decline; CRP, C-reactive protein; eGFRcrea, estimated glomerular filtration rate by serum creatinine; Esophag-cancer, esophageal cancer; GGT, gamma glutamyl transferase; gSNP, GWAS SNPs; GWAS, genome-wide association study; HDL, high-density lipoprotein; HDLC-TG, HDL cholesterol-triglycerides; HDLC-WC, HDL cholesterol-waist circumference; Hyper-TG, hypertriglyceridemia; LDL, low-density lipoprotein; Lp-PLA2, lipoprotein-associated phospholipase A2; SHBG, sex hormone-binding globulin; sIL-6R, soluble interleukin-6 receptor; SNP, single nucleotide polymorphism; TG-BP, triglycerides-blood pressure; and WC-TG, waist circumference-triglycerides; for full trait or disease names, please see Table III in the Data Supplement.

genes (Table 2). The final composite association network contained those 38 query genes, as well as the output gene set, that is, the 100 genes connected to the query gene set. Altogether these were connected with 2225 associations, also known as edges (Figure I and Table V in the Data

Supplement). All the genes in the composite network were then used for functional enrichment analysis against GO terms,⁴⁷ which revealed 93 significantly ($FDR < 0.01$) and 79 suggestively ($0.1 < FDR < 0.01$) enriched terms (Table VI in the Data Supplement). The majority of enriched terms

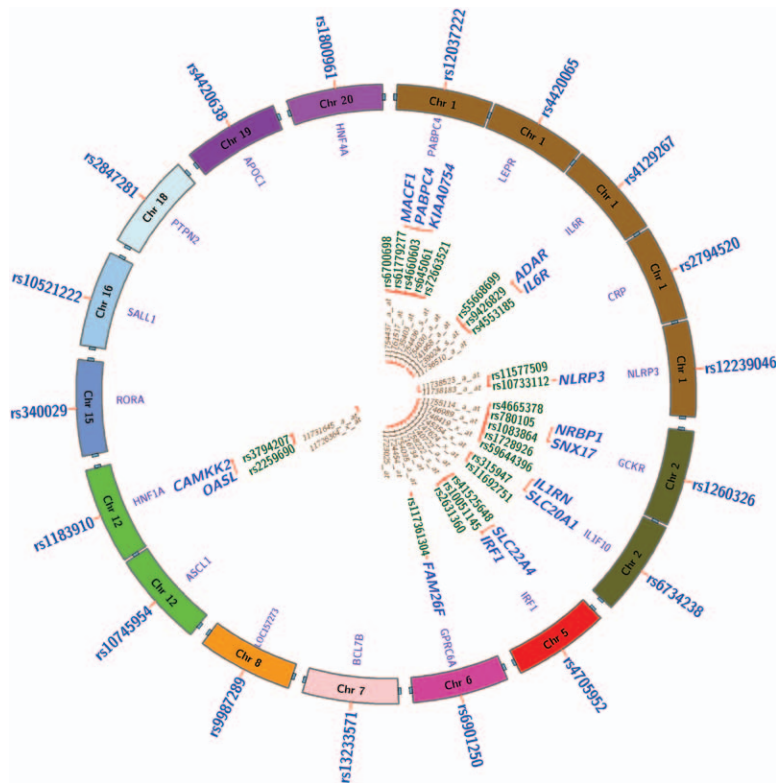


Figure 4. Results of eQTL analysis. The 3 inner-most rings represent (d) gene names of significantly associated expression probes, (e) the most significantly associated eQTL SNPs (eSNPs) for the corresponding expression probes, (f) expression probes significantly associated with gSNPs. eQTL indicates expression quantitative trait loci; GWAS, genome-wide association study; gSNP, GWAS SNPs; and SNP, single nucleotide polymorphism.

can be broadly categorized into 2 major groups: (1) terms related to immunologic processes, cytokines, and especially interferons and (2) terms related to lipids and lipoprotein metabolism.

Thirty-three of the 93 significantly enriched terms belong to the first category, of which 7 have an $FDR < 5 \times 10^{-15}$: cytokine-mediated signaling pathway (GO:0019221, $FDR = 9.47 \times 10^{-37}$), type I interferon-mediated signaling pathway (GO:0060337, $FDR = 1.05 \times 10^{-34}$), cellular response to type I interferon (GO:0071357, $FDR = 1.05 \times 10^{-34}$), response to type I interferon (GO:0034340, $FDR = 1.22 \times 10^{-34}$), interferon- γ -mediated signaling pathway (GO:0060333, $FDR = 5.78 \times 10^{-16}$), response to interferon- γ (GO:0034341, $FDR = 5.78 \times 10^{-16}$), cellular response to interferon- γ (GO:0071346, $FDR = 1.69 \times 10^{-15}$). Figure 5 visualizes these 7 terms within their corresponding GO tree. Ten out of 33 significantly enriched terms of this first category are specifically related to interferons (Table VI in the Data Supplement). Forty-three of 93 significantly enriched terms belong to the second category, that is, they are all related to the metabolism of fatty acids (eg, GO:0042304: regulation of fatty acid biosynthetic process, $FDR = 1.01 \times 10^{-4}$), triglycerides (eg, GO:0070328: triglyceride homeostasis, $FDR = 7.93 \times 10^{-5}$), cholesterol (eg, GO:0042632: cholesterol homeostasis, $FDR = 2.71 \times 10^{-4}$), and especially lipoproteins (eg, GO:0034361: very-low-density lipoprotein particle, $FDR = 3.45 \times 10^{-5}$; Table VI in the Data Supplement).

Discussion

In the present study, we performed a post-GWAS analysis of 18 genome-wide significantly associated CRP SNPs. This strategy yielded new information on biological processes involved in CRP metabolism.

Here we shed light on the genomic context of the vicinity of gSNPs in 2 steps. We first investigated the nearby genomic region to identify all linked variants, with emphasis on nsSNPs as potentially functional variants. A strength of this approach is the use of r^2 as a metric of LD rather than predefined physical distance. Although nsSNPs have a high likelihood to be functional, they may constitute only a small fraction of the mechanisms involved. Therefore, we included all SNP types into the analyses. In the second step, that is, the eQTL analysis, we identified any nearby gene whose expression level is associated with its corresponding gSNP. Here we used one of the largest single data sets of genome-wide expression probes in peripheral blood currently available worldwide of >5000 samples, which was analyzed by a stringent statistical approach. The 2 steps identified several relevant genes that were jointly used as the input to the next step, that is, the functional network analysis. The strength of this approach is including the genes from the eQTL analysis in the functional network analysis, as we think these genes are at least as important as those genes to which gSNPs or their linked variants map. This approach has added value to stand-alone eQTL results as they are translated to biological insights in a broader context through integration to other data domains.

In the next step, we constructed a functional association interaction network followed by functional enrichment analysis against GO terms. Such an interaction network is considered to represent cofunctionality of the connected genes.⁴¹ The large data set of functional association data that is used contains not only coexpression data, but also physical interaction, genetic interaction, shared protein domains, colocalization, and predicted association networks. As a result, the constructed interaction network is a composite based on these

Table 2. Biologically Prioritized Candidate Gene Set Used as the Input Query to the Network Analysis

No. of gSNP	Gene Name	Ensembl Gene ID	Query Gene Set
1	CRP	ENSG00000132693	i; iii
2	APOC1	ENSG00000130208	i; iii
2	APOE	ENSG00000130203	ii
2	APOC1P1	ENSG00000214855	iii
3	HNF1A	ENSG00000135100	i; ii; iii
3	CAMKK2	ENSG00000110931	iv
3	OASL	ENSG00000135114	iv
4	LEPR	ENSG00000116678	i; iii
5	IL6R	ENSG00000160712	i; ii; iii; iv
5	ADAR	ENSG00000160710	iv
6	GCKR	ENSG00000084734	i; iii
6	NRBP1	ENSG00000115216	iv
6	SNX17	ENSG00000115234	iv
7	NLRP3	ENSG00000162711	i; iii; iv
8	IL1F10	ENSG00000136697	i; iii
8	IL1RN	ENSG00000136689	iii; iv
8	SLC20A1	ENSG00000144136	iv
9	LOC157273	ENSG00000254235	i; iii
10	ASCL1	ENSG00000139352	i; iii
10	C12orf42	ENSG00000179088	iii
11	HNF4A	ENSG00000101076	i
12	RORA	ENSG00000069667	i; iii
13	SALL1	ENSG00000103449	i; iii
14	PABPC4	ENSG00000090621	i; iii; iv
14	MACF1	ENSG00000127603	ii; iv
14	HEYL	ENSG00000163909	iii
14	PPIEL	ENSG00000243970	iii
14	BMP8A	ENSG00000183682	iii
14	KIAA0754	ENSG00000255103	iv
15	BCL7B	ENSG00000106635	i; iii
15	MLXIPL	ENSG00000009950	ii; iii
15	BAZ1B	ENSG00000009954	iii
15	TBL2	ENSG00000106638	iii
16	PTPN2	ENSG00000175354	i; iii
17	GPRC6A	ENSG00000173612	i; iii
17	RFX6	ENSG00000185002	iii
17	FAM162B	ENSG00000183807	iii
17	FAM26F	ENSG00000188820	iv
18	IRF1	ENSG00000125347	i; iv
18	SLC22A4	ENSG00000197208	iv

The query gene set includes the following: (i) closest genes to the 18 gSNPs, (ii) closest genes to the nsSNPs in high LD ($r^2 > 0.50$) with the corresponding gSNP, (iii) closest genes to other types of SNPs in very high LD ($r^2 > 0.80$) with the corresponding gSNP, and (iv) expression probe gene names significantly associated with gSNPs (FDR < 0.01) based on the eQTL analysis. The combined query gene set contained 40 genes, of which, 2 genes, *LOC157273* and *PPIEL*, could not be found in any of the interaction resources. The order of genes follows the order of gSNPs in Table 1. eQTL indicates expression quantitative trait loci; FDR, false discovery rate; GWAS, genome-wide association study; LD, linkage disequilibrium; gSNP, GWAS SNPs; nsSNP, nonsynonymous SNPs; and SNP, single nucleotide polymorphism.

different data sources.⁴¹ As described in the methods section, the functional enrichment analysis is performed against GO terms after excluding those annotations with computer-generated inferred from RCA and IEA evidence codes. Thus, about half of the GO annotations are disregarded to avoid circularity and to obtain more robust results (<http://www.geneontology.org/GO.evidence.shtml>).⁴⁴

Our post-GWAS analysis of CRP GWAS SNPs eventually yielded a range of enriched biological processes after several intermediate steps. Some processes like acute-phase response or acute inflammatory response with significant FDR values are expected and appropriate terms for CRP providing confidence in our results. Interestingly, about one third of the significantly enriched terms were related to immunologic processes, cytokines, and interferons. Even more interesting, 10 of the significantly enriched terms, including 6 of the top most significant ones, are those pointing to the biology of interferons. In particular, type I interferon associated biological processes are highlighted with 3 significant enriched terms with $FDR < 1 \times 10^{-30}$.

The link between interferons and CRP has not been well established, probably because the measurement of interferons is complicated by their short half-lives. Although few studies have addressed the direct link between CRP and interferons and although this link has not been appreciated as a potential mechanism underlying the biology of CRP, our finding is in fact amply supported by those few in vitro and clinical observations. An in vitro observation by Enocsson et al showed that interferon- α , the main representative of the type I interferon family, inhibits CRP secretion in a dose-dependent fashion mediated by the type I interferon receptor.⁴⁸ Furthermore, although CRP levels are highly associated with most inflammatory states, as CRP level is a well-known metric for the detection and evaluation of many inflammatory diseases,¹⁰ elevated CRP levels correlate poorly with those inflammatory conditions that are characterized by high levels of interferon- α , such as systemic lupus and viral infections.⁴⁹⁻⁵⁴ This observation is in line with the abovementioned in vitro observation that increased levels of interferon- α suppress CRP levels.⁴⁸ Likewise, there are yet unexplained phenomena in lupus patients as there is a 10- to 50-fold increased risk of myocardial infarction,^{55,56} whereas there is no association between cardiovascular disease and CRP levels in these patients.⁵³ This lack of an association is unexpected because CRP is an established risk factor for coronary heart disease.^{5,6} Moreover, in lupus patients, lack of correlation between interleukin-6, the main stimulant of CRP secretion, and CRP has been reported.⁵⁷ These related observations may be explained by the fact that lupus patients are known to have a high level of interferon- α and that interferon- α is an inhibitor of CRP secretion. Another line of evidence comes from infectious diseases. In viral infections, in contrast to bacterial infections, there is generally a mild, poorly correlated increase of CRP level, making CRP a widely used diagnostic tool in distinguishing viral from bacterial infections.⁵⁴ This can be explained by the notion that patients with viral infections have high levels of interferon- α ⁴⁹ and by an inverse relation between interferon- α and CRP levels.⁴⁸ Finally, although the analysis had started with CRP gSNPs, it interestingly returned 4 significantly enriched terms

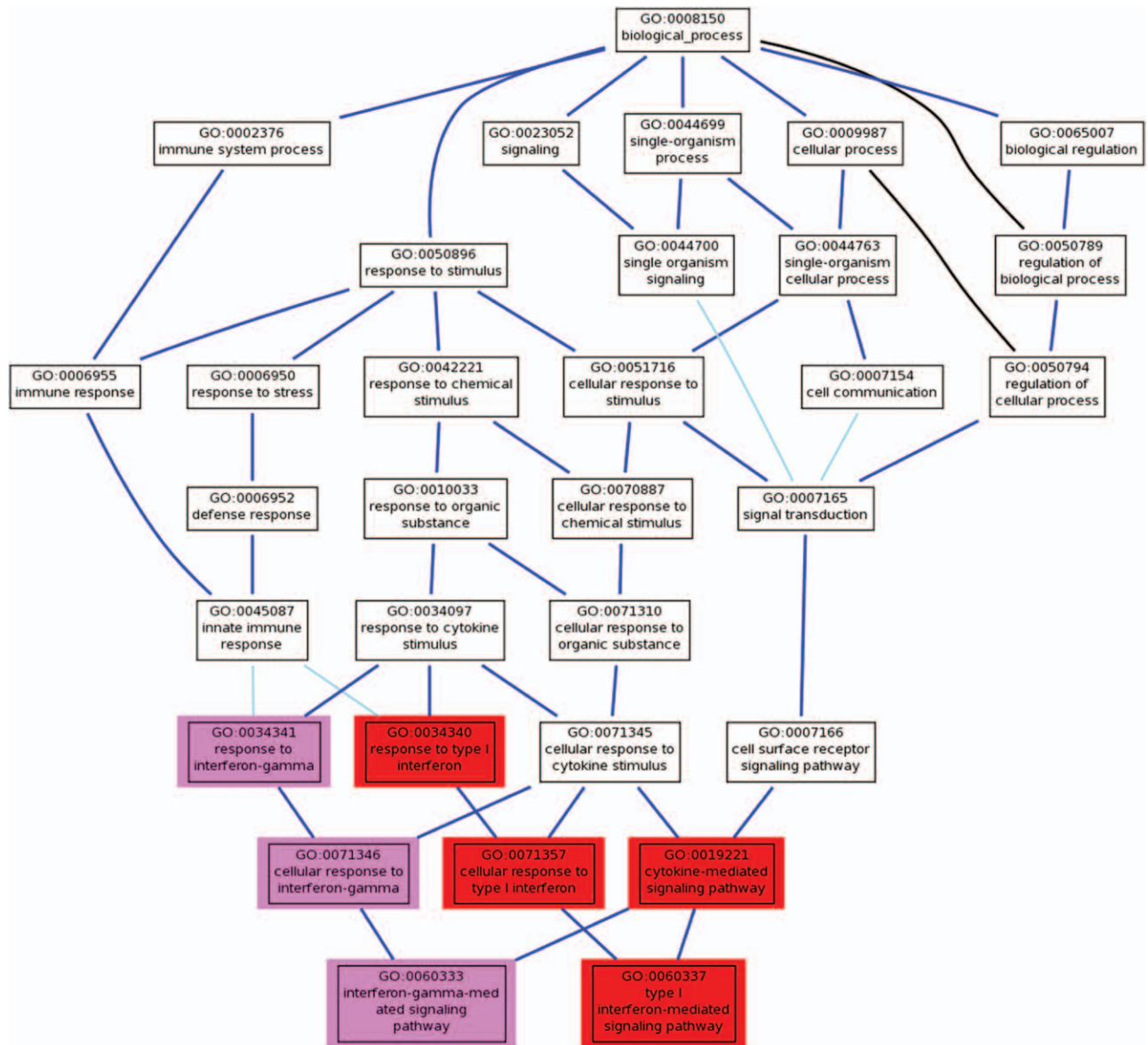


Figure 5. The most significantly enriched GO terms with $FDR < 5 \times 10^{-15}$. They are visualized as highlighted boxes within their corresponding GO tree, as red for those with $FDR < 1 \times 10^{-30}$ and purple for those with $5 \times 10^{-15} < FDR < 1 \times 10^{-30}$. The relations between the boxes have standard colors: black (regulates), blue (is_a), or light blue (part_of) (<http://www.geneontology.org/GO.ontology-ext.relations.shtml>).⁴⁵ FDR indicates false discovery rate; and GO, Gene Ontology.

specifically related to defense responses to viruses (Table VI in the Data Supplement). Considering the blunted response of CRP levels to viral infections,⁵⁴ this unexpected finding once again suggests an important role of interferon- α in CRP metabolism.

The in silico pleiotropy analysis revealed several pleiotropic effects between CRP gSNPs and other metabolic traits, particularly lipid- and lipoprotein-related traits. These results show strong concordance with those from our functional network analysis, as about half of the significantly enriched GO terms point to biological processes related to lipids and lipoproteins metabolism. These findings are also fully in line with existing knowledge of overlap between the biology of CRP and lipids with metabolism of both CRP and lipids related to the liver. Further, CRP levels are significantly associated with

weight, waist-circumference, body mass index, cholesterol, triglycerides, low-density lipoprotein (weakly) and negatively associated with high-density lipoprotein concentrations.^{6,58-61} Both CRP and lipids are well-known risk factors for coronary heart disease.⁶¹ Thus, our results show extensive genetic overlap between CRP and lipid metabolism, although the exact mechanisms underlying these significant associations remain to be elucidated.

In early 2010, Dickson et al suggested that observed GWAS associations between a common SNP and trait of interest can be explained by multiple rare variants at the locus in LD with that SNP, so-called synthetic associations.⁶² However, there are several lines of evidence indicating that GWAS associations are rarely caused by synthetic associations with rare variants.⁶³⁻⁶⁵ Later on, Visscher and colleagues state that

instead the combined evidence supports a highly polygenic model of disease susceptibility which is built on causal variants across the entire range of the allele-frequencies.⁶⁶ Hence, our approach of including all gSNPs, as well as their linked SNPs and eQTL results, is more consistent with the polygenic model than with the synthetic association model.

Despite using one of the largest single data sets of genome-wide expression probes for eQTL analysis, it contained only blood expression probes. This limitation may have affected the list of associated genes. A similar approach but using a large data set of tissue-specific expression data, particularly liver cells of healthy individuals, may better reveal the associated gene expressions. However, to the best of our knowledge, such a homogenous large data set of liver cells from healthy individuals does not exist yet. Furthermore, if there is cryptic relatedness among our subjects, it is possible that our eQTL results might be slightly biased. However, our population is relatively outbred and known relationships among subjects were taken into account in the analysis. Under these circumstances, Voight and Pritchard suggest that the bias is expected to be negligible.⁶⁷ Our functional enrichment analysis was done using GO terms; one may suggest a more extended approach by including other annotation sources like KEGG and Reactome pathways. However, as these resources only contain a limited number of pathways, it is unlikely this would have affected our main conclusions.

Finally, the results of this in silico study need to be followed up by further in vitro, in vivo, and epidemiological studies. The association of interferon- α with coronary heart disease and other CRP-associated traits or diseases, as well as the association of CRP gSNPs or CRP genetic risk scores with clinical conditions like systemic lupus, are yet to be investigated. These results also highlight the need and potential for a GWAS on serum levels of interferon- α . Finally, although those CRP gSNPs are based on a large meta-GWAS, including >80000 subjects, the explained variance in CRP level by all those 18 gSNPs is only around 5%.¹⁹ To further unravel the underlying genetic mechanisms controlling CRP levels, a larger meta-GWAS on CRP is needed to find additional common variants, whereas other approaches, such as meta-analyses of exome chip data, will be needed to find variants of lower frequency affecting serum levels of CRP.

In summary, in this in silico study, we followed a bioinformatics-based approach aiming to translate CRP GWAS signals into biological insights. Our post-GWAS analysis of CRP GWAS SNPs reemphasizes the previously known overlap between the biology of CRP and lipids. Additionally, it suggests an important role for interferons in the metabolism of CRP.

Accession numbers: Gene expression and genotype data used for this study will be available at dbGaP, accession number phs000486.v1.p1 (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000486.v1.p1).

Sources of Funding

The Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR) were funded by the Netherlands Organization for Scientific Research (MagW/ZonMW grants 904-61-090, 985-10-002, 904-61-193, 480-04-004, 400-05-717,

912-100-20; Spinozapremie 56-464-14192; Geestkracht program grant 10-000-1002); the Center for Medical Systems Biology (NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure, VU University's Institutes for Health and Care Research and Neuroscience Campus Amsterdam, NBIC/BioAssist/RK (2008.024); the European Science Foundation (EU/QLRT-2001-01254); the European Community's Seventh Framework Program (FP7/2007–2013); ENGAGE (HEALTH-F4-2007–201413); and the European Science Council (ERC, 230374). Gene-expression data were funded by the US National Institute of Mental Health (RC2 MH089951) as part of the American Recovery and Reinvestment Act of 2009. NESDA and NTR were further supported by funding from the Neuroscience Campus Amsterdam and through a NWO-VICI grant (number 91811602) and funding from ZonMW Middelgroot (911-09-032).

Disclosures

None.

References

- Nordestgaard BG. Does elevated C-reactive protein cause human atherothrombosis? Novel insights from genetics, intervention trials, and elsewhere. *Curr Opin Lipidol*. 2009;20:393–401. doi: 10.1097/MOL.0b013e3283307bfe.
- Allin KH, Bojesen SE, Nordestgaard BG. Baseline C-reactive protein is associated with incident cancer and survival in patients with cancer. *J Clin Oncol*. 2009;27:2217–2224. doi: 10.1200/JCO.2008.19.8440.
- Dehghan A, Kardys I, de Maat MP, Uitterlinden AG, Sijbrands EJ, Bootsma AH, et al. Genetic variation, C-reactive protein levels, and incidence of diabetes. *Diabetes*. 2007;56:872–878. doi: 10.2337/db06-0922.
- Sesso HD, Buring JE, Rifai N, Blake GJ, Gaziano JM, Ridker PM. C-reactive protein and the risk of developing hypertension. *JAMA*. 2003;290:2945–2951. doi: 10.1001/jama.290.22.2945.
- Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, et al. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med*. 2004;350:1387–1397. doi: 10.1056/NEJMoa032804.
- Kaptoge S, Di Angelantonio E, Lowe G, Pepys MB, Thompson SG, Collins R, et al. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet*. 2010;375:132–140.
- De Berardis D, Conti CM, Campanella D, Carano A, Scali M, Valchera A, et al. Evaluation of C-reactive protein and total serum cholesterol in adult patients with bipolar disorder. *Int J Immunopathol Pharmacol*. 2008;21:319–324.
- Harris TB, Ferrucci L, Tracy RP, Corti MC, Wacholder S, Ettinger WH Jr, et al. Associations of elevated interleukin-6 and C-reactive protein levels with mortality in the elderly. *Am J Med*. 1999;106:506–512.
- Tremblay J. Genetic determinants of C-reactive protein levels in metabolic syndrome: a role for the adrenergic system? *J Hypertens*. 2007;25:281–283. doi: 10.1097/HJH.0b013e328013dc13.
- Umbarino D, Zeng L. Is C reactive protein expression affected by local microenvironment? *Heart*. 2013;99:514–515. doi: 10.1136/heartjnl-2012-303436.
- Danik JS, Ridker PM. Genetic determinants of C-reactive protein. *Curr Atheroscler Rep*. 2007;9:195–203.
- Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, et al. Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA*. 2009;302:37–48. doi: 10.1001/jama.2009.954.
- Kathiresan S, Larson MG, Vasan RS, Guo CY, Gona P, Keaney JF Jr, et al. 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. doi: 10.1161/CIRCULATIONAHA.105.591271.
- Saunders CL, Gulliford MC. Heritabilities and shared environmental effects were estimated from household clustering in national health survey data. *J Clin Epidemiol*. 2006;59:1191–1198. doi: 10.1016/j.jclinepi.2006.02.015.
- Rahman I, Bennet AM, Pedersen NL, de Faire U, Svensson P, Magnusson PK. Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res Hum Genet*. 2009;12:286–294. doi: 10.1375/twin.12.3.286.

16. Su S, Miller AH, Snieder H, Bremner JD, Ritchie J, Maisano C, et al. Common genetic contributions to depressive symptoms and inflammatory markers in middle-aged men: the Twins Heart Study. *Psychosom Med*. 2009;71:152–158. doi: 10.1097/PSY.0b013e31819082ef.
17. Neijts M, van Dongen J, Klufft C, Boomsma DI, Willemsen G, de Geus EJ. Genetic architecture of the pro-inflammatory state in an extended twin-family design. *Twin Res Hum Genet*. 2013;16:931–940. doi: 10.1017/thg.2013.58.
18. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, et al. 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. doi: 10.1016/j.ajhg.2008.03.015.
19. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*. 2011;123:731–738. doi: 10.1161/CIRCULATIONAHA.110.948570.
20. Wang X, Prins BP, Söber S, Laan M, Snieder H. Beyond genome-wide association studies: new strategies for identifying genetic determinants of hypertension. *Curr Hypertens Rep*. 2011;13:442–451. doi: 10.1007/s11906-011-0230-y.
21. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*. 2011;43:513–518. doi: 10.1038/ng.840.
22. Franceschini N, van Rooij FJ, Prins BP, Feitosa MF, Karakas M, Eckfeldt JH, et al. Discovery and fine mapping of serum protein loci through trans-ethnic meta-analysis. *Am J Hum Genet*. 2012;91:744–753. doi: 10.1016/j.ajhg.2012.08.021.
23. 1000 Genomes Project Consortium; Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–1073. doi: 10.1038/nature09534.
24. Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, et al. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res*. 2013;41(Database issue):D64–D69. doi: 10.1093/nar/gks1048.
25. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al.; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–2158. doi: 10.1093/bioinformatics/btr330.
26. Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics*. 2011;27:718–719. doi: 10.1093/bioinformatics/btq671.
27. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164. doi: 10.1093/nar/gkq603.
28. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073–1081. doi: 10.1038/nprot.2009.86.
29. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–249. doi: 10.1038/nmeth0410-248.
30. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, et al. Ensembl 2012. *Nucleic Acids Res*. 2012;40(Database issue):D84–D90. doi: 10.1093/nar/gkr991.
31. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106:9362–9367. doi: 10.1073/pnas.0903103106.
32. Jansen R, Batista S, Brooks AI, Tischfield JA, Willemsen G, van Grootheest G, et al. Sex differences in the human peripheral blood transcriptome. *BMC Genomics*. 2014;15:33. doi: 10.1186/1471-2164-15-33.
33. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, et al. Heritability and genomics of gene expression in peripheral blood. *Nat Genet*. 2014;46:430–437. doi: 10.1038/ng.2951.
34. Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P, et al.; NESDA Research Consortium. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res*. 2008;17:121–140. doi: 10.1002/mpr.256.
35. Boomsma DI, de Geus EJ, Vink JM, Stubbe JH, Distel MA, Hottenga JJ, et al. Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet*. 2006;9:849–857. doi: 10.1375/183242706779462426.
36. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet*. 2008;16:335–342. doi: 10.1038/sj.ejhg.5201979.
37. Abdellaoui A, Hottenga JJ, de Knijff P, Nivard MG, Xiao X, Scheet P, et al. Population structure, migration, and diversifying selection in the Netherlands. *Eur J Hum Genet*. 2013;21:1277–1285. doi: 10.1038/ejhg.2013.48.
38. Fehrmann RS, Jansen RC, Veldink JH, Westra HJ, Arends D, Bonder MJ, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet*. 2011;7:e1002197. doi: 10.1371/journal.pgen.1002197.
39. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45:1238–1243. doi: 10.1038/ng.2756.
40. Visscher PM, Benyamin B, White I. The use of linear mixed models to estimate variance components from data on twin pairs by maximum likelihood. *Twin Res*. 2004;7:670–674. doi: 10.1375/1369052042663742.
41. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*. 2008;9 Suppl 1:S4. doi: 10.1186/gb-2008-9-s1-s4.
42. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, et al. A travel guide to Cytoscape plugins. *Nat Methods*. 2012;9:1069–1076. doi: 10.1038/nmeth.2212.
43. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, et al. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics*. 2010;26:2927–2928. doi: 10.1093/bioinformatics/btq562.
44. Mostafavi S, Morris Q. Combining many interaction networks to predict gene function and analyze gene lists. *Proteomics*. 2012;12:1687–1696. doi: 10.1002/pmic.201100607.
45. Schröder MS, Gusenleitner D, Quackenbush J, Culhane AC, Haibe-Kains B. RamiGO: an R/Bioconductor package providing an AmiGO visualize interface. *Bioinformatics*. 2013;29:666–668. doi: 10.1093/bioinformatics/bts708.
46. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–1645. doi: 10.1101/gr.092759.109.
47. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25–29. doi: 10.1038/75556.
48. Enocsson H, Sjöwall C, Skogh T, Eloranta ML, Rönnblom L, Wetterö J. Interferon-alpha mediates suppression of C-reactive protein: explanation for muted C-reactive protein response in lupus flares? *Arthritis Rheum*. 2009;60:3755–3760. doi: 10.1002/art.25042.
49. Theofilopoulos AN, Baccala R, Beutler B, Kono DH. Type I interferons (alpha/beta) in immunity and autoimmunity. *Annu Rev Immunol*. 2005;23:307–336. doi: 10.1146/annurev.immunol.23.021704.115843.
50. Lech M, Rommele C, Anders HJ. Pentraxins in nephrology: C-reactive protein, serum amyloid P and pentraxin-3. *Nephrol Dial Transplant*. 2013;28:803–811. doi: 10.1093/ndt/gfs448.
51. Becker GJ, Waldburger M, Hughes GR, Pepys MB. Value of serum C-reactive protein measurement in the investigation of fever in systemic lupus erythematosus. *Ann Rheum Dis*. 1980;39:50–52.
52. Honig S, Gorevic P, Weissmann G. C-reactive protein in systemic lupus erythematosus. *Arthritis Rheum*. 1977;20:1065–1070.
53. Nikpour M, Gladman DD, Ibañez D, Urowitz MB. Variability and correlates of high sensitivity C-reactive protein in systemic lupus erythematosus. *Lupus*. 2009;18:966–973. doi: 10.1177/0961203309105130.
54. Sasaki K, Fujita I, Hamasaki Y, Miyazaki S. Differentiating between bacterial and viral infection by measuring both C-reactive protein and 2'-5'-oligoadenylate synthetase as inflammatory markers. *J Infect Chemother*. 2002;8:76–80. doi: 10.1007/s101560200010.
55. Manzi S, Meilahn EN, Rairie JE, Conte CG, Medsger TA Jr, Jansen-McWilliams L, et al. Age-specific incidence rates of myocardial infarction and angina in women with systemic lupus erythematosus: comparison with the Framingham Study. *Am J Epidemiol*. 1997;145:408–415.
56. Esdaile JM, Abrahamowicz M, Grodzicky T, Li Y, Panaritis C, du Berger R, et al. Traditional Framingham risk factors fail to fully account for accelerated atherosclerosis in systemic lupus erythematosus. *Arthritis Rheum*. 2001;44:2331–2337.
57. Gabay C, Roux-Lombard P, de Moerloose P, Dayer JM, Vischer T, Guerne PA. Absence of correlation between interleukin 6 and C-reactive protein blood levels in systemic lupus erythematosus compared with rheumatoid arthritis. *J Rheumatol*. 1993;20:815–821.
58. Mendall MA, Patel P, Ballam L, Strachan D, Northfield TC. C reactive protein and its relation to cardiovascular risk factors: a population based cross sectional study. *BMJ*. 1996;312:1061–1065.

59. Kraja AT, Province MA, Arnett D, Wagenknecht L, Tang W, Hopkins PN, et al. Do inflammation and procoagulation biomarkers contribute to the metabolic syndrome cluster? *Nutr Metab (Lond)*. 2007;4:28. doi: 10.1186/1743-7075-4-28.
60. Sakkinen PA, Wahl P, Cushman M, Lewis MR, Tracy RP. Clustering of procoagulation, inflammation, and fibrinolysis variables with metabolic factors in insulin resistance syndrome. *Am J Epidemiol*. 2000;152:897–907.
61. Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *N Engl J Med*. 2002;347:1557–1565. doi: 10.1056/NEJMoa021993.
62. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol*. 2010;8:e1000294. doi: 10.1371/journal.pbio.1000294.
63. Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol*. 2011;9:e1000580. doi: 10.1371/journal.pbio.1000580.
64. Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol*. 2011;9:e1000579. doi: 10.1371/journal.pbio.1000579.
65. Orozco G, Barrett JC, Zeggini E. Synthetic associations in the context of genome-wide association scan signals. *Hum Mol Genet*. 2010;19(R2):R137–R144. doi: 10.1093/hmg/ddq368.
66. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90:7–24. doi: 10.1016/j.ajhg.2011.11.029.
67. Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. *PLoS Genet*. 2005;1:e32. doi: 10.1371/journal.pgen.0010032.

CLINICAL PERSPECTIVE

Genome-wide association studies (GWAS) have successfully identified genetic variants associated with complex traits or diseases. One important limitation of GWAS is that the identified variants merely flag the nearby genomic region and do not necessarily provide insight into the biological mechanisms underlying the investigated phenotype. A large GWAS on serum levels of C-reactive protein (CRP) had successfully identified 18 single nucleotide polymorphisms associated with serum levels of CRP at genome-wide significance. However, those CRP GWAS findings had been insufficiently translated to biological knowledge. Here, we applied an efficient integrated pipeline of sequential bioinformatics-based approaches for post-GWAS analysis of the 18 CRP single nucleotide polymorphisms. Our in silico analyses eventually yielded enrichment of biological processes (1) confirming the previously known overlap between the biology of CRP and lipids and (2) suggesting an important role for interferons in the metabolism of CRP. Although CRP levels are highly associated with most inflammatory states, elevated CRP levels correlate poorly with those inflammatory conditions that are characterized by high levels of interferons, such as systemic lupus and viral infections. Furthermore, there is a 10- to 50-fold increased risk of myocardial infarction in lupus patients, whereas there is no association between cardiovascular disease and CRP levels in these patients. This lack of an association is unexpected because CRP is an established risk factor for coronary heart disease. These odd clinical observations may be explained by the suggested role of interferons in CRP metabolism because interferon- α can serve as an inhibitor of CRP secretion.