REPORT

Allelic Heterogeneity at the *CRP* Locus Identified by Whole-Genome Sequencing in Multi-ancestry Cohorts

Laura M. Raffield,[1] Apoorva K. Iyengar,[1] Biqi Wang,[2] Sheila M. Gaynor,[3] Cassandra N. Spracklen,[1] Xue Zhong,[4] Madeline H. Kowalski,[5] Shabnam Salimi,[6] Linda M. Polfus,[7] Emelia J. Benjamin,[8,9,10] Joshua C. Bis,[11] Russell Bowler,[12] Brian E. Cade,[13,14] Won Jung Choi,[15] Alejandro P. Comellas,[16] Adolfo Correa,[17] Pedro Cruz,[18] Harsha Doddapaneni,[19] Peter Durda,[20] Stephanie M. Gogarten,[21] Deepti Jain,[21] Ryan W. Kim,[15] Brian G. Kral,[22,23] Leslie A. Lange,[24] Martin G. Larson,[2,10] Cecelia Laurie,[21] Jiwon Lee,[13] Seonwook Lee,[15] Joshua P. Lewis,[25] Ginger A. Metcalf,[19] Braxton D. Mitchell,[25,26] Zeineen Momin,[19] Donna M. Muzny,[19] Nathan Pankratz,[27] Cheol Joo Park,[15] Stephen S. Rich,[28] Jerome I. Rotter,[29] Kathleen Ryan,[25] Daekwan Seo,[15] Russell P. Tracy,[20,30] Karine A. Viaud-Martinez,[18] Lisa R. Yanek,[22] Lue Ping Zhao,[31,32] Xihong Lin,[3,33,34] Bingshan Li,[35] Yun Li,[1,5,36] Josée Dupuis,[2,10] Alexander P. Reiner,[37] Karen L. Mohlke,[1] Paul L. Auer,[38,*] TOPMed Inflammation Working Group, and NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

Whole-genome sequencing (WGS) can improve assessment of low-frequency and rare variants, particularly in non-European populations that have been underrepresented in existing genomic studies. The genetic determinants of C-reactive protein (CRP), a biomarker of chronic inflammation, have been extensively studied, with existing genome-wide association studies (GWASs) conducted in >200,000 individuals of European ancestry. In order to discover novel loci associated with CRP levels, we examined a multi-ancestry population (n = 23,279) with WGS (~38× coverage) from the Trans-Omics for Precision Medicine (TOPMed) program. We found evidence for eight distinct associations at the *CRP* locus, including two variants that have not been identified previously (rs11265259 and rs181704186), both of which are non-coding and more common in individuals of African ancestry (~10% and ~1% minor allele frequency, respectively, and rare or monomorphic in 1000 Genomes populations of East Asian, South Asian, and European ancestry). We show that the minor (G) allele of rs181704186 is associated with lower CRP levels and decreased transcriptional activity and protein binding *in vitro*, providing a plausible molecular mechanism for this African ancestry-specific signal. The individuals homozygous for rs181704186-G have a mean CRP level of 0.23 mg/L, in contrast to individuals heterozygous for rs181704186 with mean CRP of 2.97 mg/L and major allele homozygotes with mean CRP of 4.11 mg/L. This study demonstrates the utility of WGS in multi-ethnic populations to drive discovery of complex trait associations of large effect and to identify functional alleles in noncoding regulatory regions.

[1]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; [2]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; [3]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA; [4]Department of Medicine, Division of Genetic Medicine, Vanderbilt University, Nashville, TN 37232, USA; [5]Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA; [6]Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, MD 21201, USA; [7]Department of Preventive Medicine, Center for Genetic Epidemiology, University of Southern California, Los Angeles, CA 90089, USA; [8]Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA; [9]Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118, USA; [10]National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA 01702, USA; [11]Department of Medicine, Cardiovascular Health Research Unit, University of Washington, Seattle, WA 98101, USA; [12]Department of Medicine, Division of Pulmonary, Critical Care & Sleep Medicine, National Jewish Health, Denver, CO 80206, USA; [13]Department of Medicine, Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA 02115, USA; [14]Department of Medicine, Division of Sleep Medicine, Harvard Medical School, Boston, MA 02115, USA; [15]Macrogen USA, Rockville, MD 20850, USA; [16]Department of Medicine, Division of Pulmonary and Critical Care, University of Iowa, Iowa City, IA 52242, USA; [17]Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA; [18]Illumina Laboratory Services, Illumina Inc., San Diego, CA 92122, USA; [19]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [20]Department of Pathology & Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, VT 05446, USA; [21]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [22]GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [23]Division of Cardiology, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [24]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO 80045, USA; [25]Department of Medicine, Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD 21201, USA; [26]Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD 21201, USA; [27]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA; [28]Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; [29]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [30]Department of Biochemistry, Larner College of Medicine, University of Vermont, Burlington, VT 05446, USA; [31]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [32]School of Public Health, University of Washington, Seattle, WA 98195, USA; [33]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [34]Department of Statistics, Harvard University, Cambridge, MA 02138, USA; [35]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232, USA; [36]Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599, USA; [37]Department of Epidemiology, University of Washington, Seattle, WA 98195, USA; [38]Joseph J. Zilber School of Public Health, University of Wisconsin Milwaukee, Milwaukee, WI 53205, USA
*Correspondence: pauer@uwm.edu
https://doi.org/10.1016/j.ajhg.2019.12.002.

Whole-genome sequencing (WGS) data are being rapidly generated in deeply phenotyped cohorts or case-referent samples of complex disorders by projects such as the United Kingdom's 100,000 Genomes Project,[1] the National Institute of Mental Health's Whole Genome Sequencing for Psychiatric Disorders Consortium,[2] the National Human Genome Research Institute's Centers for Common Disease Genomics (CCDG) project (see Web Resources), and the National Heart, Lung, and Blood Institute's Trans-Omics for Precision Medicine (TOPMed) Program.[3] WGS resources can improve interrogation of low-frequency and rare variation associated with quantitative traits or clinical outcomes[4] compared to genotyping array-based studies. However, sample sizes remain modest compared to large-scale genome-wide association studies (GWASs).

WGS-based analysis may offer particular advantages for non-European populations currently underrepresented in GWASs, with ~95% of GWAS participants being of European or East Asian ancestry.[5] WGS can assess population-specific variants which are at very low frequency or absent in large European GWASs, including variants that are often poorly imputed with standard reference panels and genotyping arrays. Current imputation reference panels for non-European populations (notably 1000 Genomes phase 3, n = 5,008 haplotypes across 26 mostly non-European populations[6]) are also much smaller than resources like the Haplotype Reference Consortium (HRC) for European populations (n = 64,976 haplotypes),[7] making imputation of low-frequency variants more difficult. Along with discrepancies in imputation reference panel size, many genotyping arrays have poor genomic coverage in non-European populations.[8] Because WGS assesses the entire genome of each individual, the limitations of genotyping arrays and imputation reference panels are easily overcome, allowing better understanding of the genetic architecture of complex traits in non-European populations. Based on previous success in identifying novel coding low-frequency or population-specific variants for inflammatory biomarkers in sequencing-based analyses,[9,10] we evaluated the ability of WGS to identify additional high-impact non-coding variation for commonly assessed inflammation biomarker C-reactive protein (CRP).

CRP is an acute-phase protein synthesized in the liver and is often used as a biomarker for chronic low-grade inflammation. As such, its relationship to cardiovascular disease (CVD) has been well established by numerous epidemiological studies, though current analyses do not point to a causal relationship with CVD.[11,12] CRP has also been associated with inflammatory disorders,[13,14] type 2 diabetes,[15] and overall mortality,[16] and recent Mendelian randomization studies have pointed to a potential causal role in bipolar disorder and schizophrenia.[12]

CRP demonstrates substantial heritability in family-based studies (~30% in East Asians,[17] ~30%–40% in Europeans,[18–20] ~45% in African Americans[21]). CRP levels vary by race/ethnicity group with higher levels observed in individuals of African ancestry compared to European or East Asian ancestry.[22,23] The genetic architecture of CRP has been investigated in diverse populations by whole-exome sequencing (WES),[10] genome-wide association,[24–26] and fine-mapping studies imputed to various reference panels[27,28] in tens of thousands of samples. Most recently, the largest GWAS was conducted in up to 204,402 individuals of European ancestry, identifying 58 loci and explaining 7% of the trait variance.[12] Some studies have also reported population-specific variants associated with CRP levels.[27] Among reported loci, the locus surrounding the *CRP* (MIM: 123260) gene itself on chromosome 1 explains the largest portion of phenotypic variance (1.4%[12]), with multiple distinct signals reported and clear evidence of allelic heterogeneity across populations.[27,28] For example, using approximate conditional analysis, the most recent European GWAS analysis reported 13 signals at the *CRP* locus (including rs149520992, an intergenic variant with a minor allele frequency [MAF] of 1% in Europeans and rare in other populations),[12] and four distinct signals (shared across ancestry groups) were reported in the multi-ethnic fine-mapping effort from the Population Architecture using Genomics and Epidemiology (PAGE) study.[28] African-specific variant rs726640 or variants in linkage disequilibrium (LD) with it have also been reported in several previous studies.[26,27,29]

Using data from the NHLBI TOPMed WGS project, we sought to investigate the additional value of WGS (beyond whole-exome sequencing and imputed GWAS) for single-variant analysis in a set of 23,279 individuals predominantly of self-reported European, African American, East Asian, and Hispanic/Latino ancestry with measured CRP levels (Table S1). We identified association with CRP levels at eight known loci (*CRP, APOE* [MIM: 107741], *HNF1A* [MIM: 142410], *LEPR* [MIM: 601007], *GCKR* [MIM: 600842], *IL6R* [MIM: 147880], *IL1F10* [MIM: 615296], and *NLRP3* [MIM: 606416]) with $p < 1 \times 10^{-9}$ in an ancestry-pooled genome-wide single-variant analysis (Table 1, Figure S1). We also examined these eight CRP-associated loci separately in African American (n = 6,545) and European American (n = 15,065) participants (Table S2). In the European American analysis, at least one variant at each locus met the locus-wide significance threshold for association with CRP levels with the exception of the *NLRP3* locus. The African American analysis also demonstrated at least one locus-wide significant variant at all loci except *GCKR* and *LEPR*.

We performed stepwise conditional analyses at each of the eight loci by conditioning on the lead variant at each locus and then sequentially conditioning on each new lead variant until no variants met our locus-wide significance thresholds (Table 1). Stepwise conditional analyses were performed in ancestry pooled and stratified (self-reported European American- and African American-specific) analyses. We identified two conditionally distinct signals at *HNF1A* and eight at the *CRP* locus (Table 2, Figures 1, S2, and S3). The presence of multiple association signals

Table 1. Eight Loci Significantly Associated ($p < 1 \times 10^{-9}$) with C-Reactive Protein Levels in TOPMed

Locus	Lead Variant	Annotation	p Value	Beta	Effect Allele	TOPMed EAF Overall	TOPMed African American EAF	TOPMed European American EAF	After Conditioning on Lead Variant			
									New Lead Variant	p Value	2nd Signal Threshold	Total # Signals
LEPR	rs7516341	intronic	1.9E−19	−0.09	C	0.43	0.54	0.37	rs72683129	4.7E−05	4.7E−06	1
IL6R	rs4129267	intronic	5.0E−12	−0.07	T	0.33	0.14	0.40	rs149417774	2.7E−04	6.3E−06	1
CRP	rs7551731	intergenic	1.1E−65	−0.18	C	0.30	0.22	0.33	rs73024795	1.2E−42	2.4E−06	8
NLRP3	rs56188865	intronic	2.6E−11	−0.06	C	0.42	0.52	0.38	rs115695052	1.6E−05	4.5E−06	1
GCKR	rs1260326	missense, p.Leu446Pro (*GCKR*)	1.9E−13	−0.08	C	0.66	0.85	0.58	rs183628627	4.7E−04	6.7E−06	1
IL1F10	rs6734238	intergenic	8.4E−12	0.07	G	0.41	0.45	0.41	rs148498391	4.1E−04	6.2E−06	1
HNF1A	rs2243458	intronic	1.5E−33	−0.13	T	0.27	0.12	0.33	rs544759708	3.3E−06	4.3E−06	2
APOE	rs429358	missense, p.Cys130Arg (*APOE4*)	1.1E−65	−0.22	C	0.15	0.21	0.13	rs186472069	1.6E−05	4.7E−06	1

Significance threshold for identification of second signals calculated as p = (0.05/tested variants). EAF, effect allele frequency, for those in TOPMed CRP analysis.

at both *CRP* and *HNF1A* has been reported in previous studies, with at least two signals identified at both loci in a recent multi-ethnic fine-mapping effort (four signals at *CRP*, two signals at *HNF1A*)[28] and in the largest European meta-analysis (13 approximate conditional signals at *CRP* and 2 at *HNF1A*).[12] The eight identified signals at the *CRP* locus include low-frequency, exonic variants (rs1800947 [p.Leu184Leu] and rs553202904, a noncoding proxy for rs77832441 [p.Thr59Met]) and noncoding variants with much higher MAF in African ancestry individuals. These African American-driven signals include both known (rs73024795) and previously unreported (rs11265259, rs181704186) associations. In an unrelated subset (n = 17,371), these eight conditionally distinct signals explained 4.2% of variance in natural log transformed CRP (2.6% in European Americans, 6.0% in African Americans). When performing stepwise conditional analyses at the *CRP* locus separately by ancestry, five conditionally distinct signals were identified in African Americans alone and four conditionally distinct signals were identified in European Americans. Based on these results and with consideration of population-specific allele frequencies, four signals at *CRP* were driven primarily by African American individuals (rs73024795, rs11265259, rs181704186, rs2211321) and two by European Americans (rs553202904, rs12734907) (Table S3). The other two signals (rs7551731 and rs1800947) were shared between African Americans and European Americans.

To determine whether the association signals we observed at the *CRP* or *HNF1A* loci were tagging previously reported associations, we performed a separate conditional analysis by which we adjusted for all variants associated with CRP levels at the *CRP* or *HNF1A* loci in prior GWAS, fine-mapping, or exome-sequencing efforts (Tables S4 and S5). In this analysis, two African American-driven signals at *CRP* remained locus-wide significant including rs11265259 (signal "E"; $\beta = -0.32$, $p = 7.3 \times 10^{-18}$; African American MAF = 0.10) and rs181704186 (signal "H"; $\beta = -0.46$, $p = 3.0 \times 10^{-7}$; African American MAF = 0.01); both are rare or monomorphic in other ancestry populations, with no copies of the minor allele for either variant found in 1000 Genomes European, East Asian, or South Asian populations. We also note the unusually large effect size for rs181704186, with major allele homozygotes having mean CRP levels of 4.11 mg/L (similar to the overall TOPMed mean of 4.10 mg/L), heterozygotes, 2.97 mg/L, and minor allele homozygotes, 0.23 mg/L, respectively (Figure 2A). By contrast, the more common variant, rs11265259, has mean CRP levels of 4.10, 4.36, and 3.04 mg/L, respectively. LD in African Americans from TOPMed between rs11265259 and rs181704186 and known signals is listed in Table S6. After adjusting for known variants at the *HNF1A* locus (Table S5), both association signals were attenuated below the locus-wide significance threshold. We thus carried forward the two conditionally distinct *CRP* signals, and not the secondary signal at *HNF1A*, for further follow-up.

As both remaining *CRP* variant associations appeared to be distinct from any previously identified *CRP* locus variant association, we attempted to replicate these two signals using CRP measurements in African American women from the Women's Health Initiative (WHI) study (n = 7,108). The WHI participants had genotype data from an Affymetrix 6.0 array imputed to the TOPMed reference panel (freeze 5b, Michigan Imputation Server) but were not whole genome sequenced through TOPMed at the time of freeze 5b's release. Both variants were locus-wide significant (using the same $p = 2.47 \times 10^{-6}$ locus-wide threshold used in our TOPMed analysis in Table 2) in our independent WHI replication sample of African Americans (Table S7, rs11265259, $p = 6.1 \times 10^{-9}$, rs181704186, $p = 9.2 \times 10^{-11}$) with consistent direction

Table 2. Eight Conditionally Distinct Signals Associated with C-Reactive Protein Were Identified at the *CRP* Locus in TOPMed

Signal	Variant	Annotation	Beta	p Value	Effect Allele	TOPMed Overall EAF	TOPMed African American EAF	TOPMed European American EAF	1000 Genomes AFR EAF	1000 Genomes EUR EAF	Sequential Conditional p Value
A	rs7551731	intergenic	−0.18	1.1E−65	C	0.30	0.22	0.33	0.20	0.31	–
B	rs73024795	intergenic	0.36	5.0E−54	T	0.05	0.16	4.98E−04	0.18	N/A	1.2E−42
C	rs2211321	intergenic	−0.02	0.05	C	0.70	0.65	0.71	0.64	0.71	3.1E−27
D	rs553202904[a]	intergenic	−0.70	1.4E−12	G	0.002	3.82E-04	0.003	N/A	0.003	8.8E−17
E	rs11265259	intergenic	−0.18	8.9E−09	C	0.03	0.09	4.31E−04	0.10	N/A	9.3E−12
F	rs1800947	synonymous, p.Leu184Leu	−0.24	5.8E−26	G	0.05	0.01	0.06	0.002	0.05	9.2E−09
G	rs12734907	intergenic	0.08	1.5E−12	T	0.26	0.08	0.34	0.02	0.37	7.9E−10
H	rs181704186	intergenic	−0.61	3.9E−12	G	0.003	0.009	9.96E−05	0.01	N/A	1.0E−07

Abbreviations: AFR, African; EUR, European; N/A, not applicable (monomorphic). Letters correspond to the signals displayed in the LocusZoom plot in Figure 1. Beta, p value, and overall effect allele frequency are from TOPMed pooled ancestry analysis. EAF, effect allele frequency, for those in TOPMed CRP analysis.
[a]Proxy variant is missense, Thr59Met ($r^2 = 0.98$ in analyzed TOPMed samples)

of effect. This remained true when conditioning on all known variants from prior GWASs and exome-sequencing studies in Table S4 (rs11265259, $p = 8.7 \times 10^{-12}$, rs181704186, $p = 9.7 \times 10^{-6}$). These replication results in WHI provide evidence to the validity of these variants and show the utility of the TOPMed reference panel for imputation in non-European ancestry individuals.

We performed several *in silico* analyses to further characterize the putative functional regulatory mechanisms of these two variants. Both rs11265259 (located ~6 kb downstream of *CRP*, signal E) and rs181704186 (located ~37 kb upstream of *CRP*, signal H) have high Genomic Evolutionary Rate Profiling (GERP)[31] scores (7.08 for rs11265259, 7.45 for rs181704186), indicating sequence conservation across species. In addition, both variants are located in predicted enhancer regions based on ChromHMM[32] models in liver (Figures 2B and S4), where CRP is produced. Neither is in strong LD (defined as $r^2 > 0.8$) with any other variant sequenced in the TOPMed African American samples. Integrated functional annotation scores from FUN-LDA comparing all Roadmap Epigenomics project tissues were highest in adult liver for both variants (Table S8a), suggesting that liver is a likely tissue in which these variants play a functional role. The annotation score for rs181704186 was 1.0 in liver, the highest possible score. The highest score for rs11265259 was more modest (0.0746), suggesting weaker evidence of enhancer function for this variant. Concordant with these results, our cross-tissue annotation principal components analysis (see Supplemental Material and Methods) found that both rs181704186 and rs11265259 were in the top 10% for conservation (scores of 18.8 and 16.3, respectively), with rs181704186 also having high epigenetics and transcription factor binding scores (Table S8b). Neither *CRP* locus variant E nor H was colocalized with eQTLs from any tissue available in GTEx,[33] whole blood (eQTLGen browser[34]), or in a recent large adult liver eQTL analysis.[35] Curiously, however, the latter liver eQTL mega-analysis identified no *cis*-eQTL for *CRP*, despite the very high expression of *CRP* in the liver.[35] We do note, however, that existing eQTL datasets that include some African Americans (such as GTEx) are fairly small; greater sample sizes and increased genetic diversity of included participants are needed to better explore eQTL effects for ancestry specific or low frequency variants like rs181704186 and rs11265259. However, GeneHancer[36] did link the enhancer region containing rs181704186 to the *CRP* gene ("elite" enhancer-gene connection [interaction confidence score 10.61], reflecting both a high-likelihood enhancer and strong enhancer-gene link). In summary, rs181704186 in particular had strong functional annotation scores in a relevant tissue for CRP levels (liver), as well as a large effect size, making it an attractive candidate for functional follow-up.

Finally, because we observed multiple independent signals at the CRP locus, we attempted to jointly model these effects with the FINEMAP statistical fine-mapping approach. We ran FINEMAP separately on the African American (AA) and European American (EA) samples, assuming a maximum of 5 causal variants in AAs and 4 causal variants in EAs (based on the results from the ancestry-specific conditional analyses). The FINEMAP method identified 7 variants in the 95% credible set in AAs (see Table S9 for all variants in the credible sets, including AA conditional analysis lead rs11265259) and 26 variants in EAs, including conditional analysis lead variants rs2211320 and rs1800947. Interestingly, while rs11265259 was included in the 95% credible set in AAs, rs181704186 was not ($r^2 < 0.03$ with all 7 credible set variants). Nevertheless, we nominated the rs181704186 variant for experimental follow up based on the preponderance of annotation-based evidence detailed above.

We performed further *in vitro* functional assays to characterize the regulatory role of rs181704186. We cloned a 1141-bp element designed to capture the surrounding

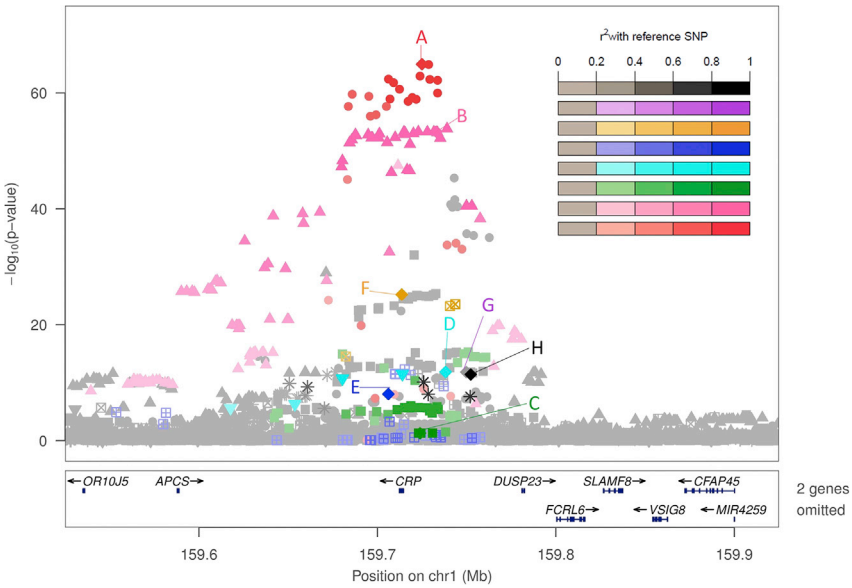


Figure 1. Eight Conditionally Distinct Signals Associated with C-Reactive Protein Were Identified at the *CRP* Locus in TOPMed

LocusZoom plot of −log10(p value) versus genomic location for all distinct signals at the *CRP* locus. Letters correspond to the list of conditionally distinct signals in Table 2. The lead variant for each conditionally distinct signal is indicated with a diamond, with other variants in linkage disequilibrium $r^2 > 0.2$ indicated in the colors used for each letter label and displayed on the legend at right, each with a different shape (for example, variants in close linkage disequilibrium with signal A (rs7551731) are displayed as red circles). Linkage disequilibrium is calculated using the same TOPMed samples included in our pooled ancestry C-reactive protein analyses.

regions of accessible chromatin and of cross-species conservation and containing each allele into a luciferase reporter vector in both orientations with respect to a minimal promoter (Table S10). Allele-specific clones of the reporter vector were transfected into the HepG2 hepatocyte/liver carcinoma cell line. Consistent with the GWAS direction of effect, the G allele associated with lower CRP levels was also associated with lower transcriptional activity in both the forward and reverse orientations (Figures 2C and S5A) than the A allele. *In vivo*, this likely reflects lower transcription of *CRP*, based on proximity and the GeneHancer links between this enhancer and the *CRP* transcription start site.[36] The cloned regulatory element appears to be a repressor, as the levels of transcriptional activity are lower than empty vector controls (Figure 2C).

We next performed an electrophoretic mobility shift assay (EMSA) to test the alleles of rs181704186 for differences in transcription factor binding (Figures 2E and S5B–S5D). We observed an allele-specific band at rs181704186-A (as indicated with an arrow; comparing lane 2 versus 7) that is competed away by a 40× excess of a probe containing the A allele (lane 3), but unaffected by probes containing the G allele (lane 4). The rs181704186 variant overlaps a CCAAT Enhancer Binding Protein Beta (CEBPB) binding site in ENCODE ChIP-seq experiments from HepG2 and HeLa cells, along with several other transcription factor binding proteins (Figure 2B). The rs181704186-G allele is predicted to disrupt the CEBPB motif, changing the position weight matrix log of the odds score from 14.8 to 2.9[17,18] (Figure 2D). CEBPB is a transcription factor known to be important for production of CRP in liver[37,38] and a strong candidate for contributing to the observed allelic differences in transcriptional activity. We attempted to supershift the EMSA DNA-protein complexes with antibodies to CEBPB. Incubation with an antibody targeting CEBPB showed a weaker band, which may represent a partially disrupted the A-allele-specific protein-DNA complex (lane 5). These allele-specific differences in protein binding are concordant with the transcriptional reporter assay and are suggestive that disruption of transcription factor binding at least partially mediates these regulatory effects, although further evidence is needed to determine the role of CEBPB and/or other transcription factors.

Using data from the TOPMed program, we report two low-frequency, population-specific variants that are associated with circulating CRP levels. Prior studies of genotypes imputed to the 1000 Genomes reference panels have not detected these associations. The best powered CRP GWAS to date included only individuals of European ancestry,[12] a population for which these variants would not have been detectable given their very low frequency. Notably, a recent study from the PAGE consortium included CRP as an exemplary quantitative trait, with data from 8,349 African Americans with CRP, genotyped on the Multi-Ethnic Genotyping Array (MEGA) and imputed to 1000 Genomes Phase 3. Neither variant was observed to be associated with CRP, despite detailed examination of secondary signals in a larger pooled sample size than available here for African Americans (and in a sample including some of the same African American participants, notably from WHI, as in our discovery and replication cohorts). This suggests that the use of a genotyping array developed to more equitably capture global genetic variation and subsequent imputation to the 1000 Genomes reference panel may still miss some population-specific variant associations that can be identified using WGS. In WHI our CRP-associated variants can be well imputed using TOPMed as a reference panel (imputation quality $r^2 \geq 0.9$); the TOPMed reference panel has ~20× larger sample size than 1000 Genomes Phase 3, and increased imputation quality is expected in African Americans based on previous work.[39] Imputation quality is only modestly attenuated in WHI using 1000 Genomes

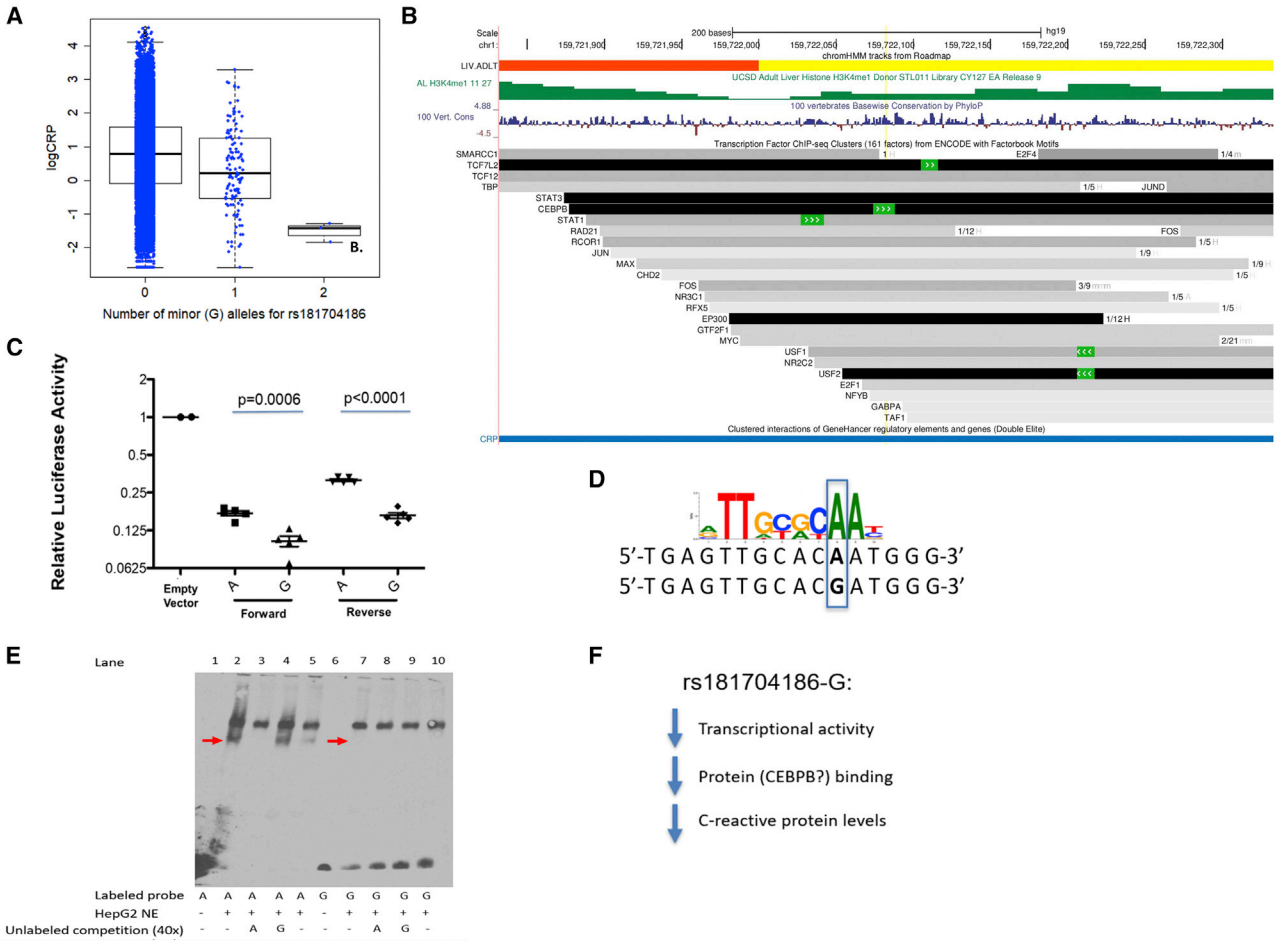


Figure 2. Regulatory Role of Low-Frequency, African Ancestry-Specific Variant rs181704186
(A) Boxplot of natural log-transformed CRP values by allele for rs181704186 (for 23,157 major allele homozygotes, 119 heterozygotes, and 3 minor allele homozygotes).
(B) Genome browser plot for rs181704186, chromHMM annotation in adult liver (yellow, enhancer; yellow, enhancer; red, transcription start site) from RoadMap Epigenomics, H3K4me1 signal from adult liver, 100 vertebrates basewise conservation by PhyloP, transcription factor ChIP-seq clusters from ENCODE (161 factor version, motifs highlighted in green, proportion cell types detected/total number of cell types assayed displayed). We also display GeneHancer's connection of the region containing this variant to *CRP*. No other variants have linkage disequilibrium $r^2 \geq 0.8$ with lead variant rs181704186.
(C) Luciferase assay demonstrating reduced transcriptional activity for the G allele, which is also associated with lower CRP levels. Blue lines indicate the groups compared for each listed p value.
(D) Disrupted CEBPB transcription factor binding motif position weight matrix from Kheradpour and Kellis[30] (CEBPB-disc1, with blue box highlighting position changed by rs181704186).
(E) Differential protein binding for A and G allele in EMSA assay. EMSA with biotin-labeled probes containing the A or G allele of rs181704186 shows an allele-specific band (lane 2 versus 7, indicated with red arrows) that is competed away by 40-fold excess of unlabeled probe containing the A allele (lane 3), but unaffected by a 40-fold excess of probe containing the G allele (lane 4). Incubation with an antibody targeting CEBPB partially disrupts the A-allele-specific protein-DNA complex (lane 5). NE, nuclear extract.
(F) Summary of direction of effect of rs181704186-G.

Phase 3 as a reference panel (imputation quality $r^2 \geq 0.75$), but this still leads to weaker association for rs11265259 in particular using 1000 Genomes imputation, likely due to a reduction in effective sample size (product of sample size and r^2). Concurrent association analysis in both sequenced and imputed data (using the largest relevant sequencing dataset, such as TOPMed, as a reference panel) may be a powerful strategy for discovering low-frequency and rare variant associations with many complex traits, particularly in non-European populations.[39]

Our results using WGS and replicated with TOPMed imputed data exemplify the value of WGS in individuals of diverse genetic ancestry. Despite having only 10% of the sample size of the largest European GWAS meta-analysis to date, the genetic diversity and accurate genotype calls for low frequency and rare variants in our multi-ancestry study afforded us the ability to detect additional population-specific association signals, including a low-frequency variant with a large effect size. These association signals add to our knowledge of the extensive allelic heterogeneity and diversity of the *CRP* genomic region, which

contains a number of shared and population-specific coding and regulatory alleles.[10,12,28] Ultimately, finer dissection of the functional alleles at the *CRP* locus may have consequences for understanding the biology of acute or chronic inflammation or the causal role of CRP in inflammation-related complex disorders. To determine whether the two replicated African-specific CRP-associated variants (rs11265259 and rs181704186) have downstream clinical consequences, we performed a phenome-wide association study (pheWAS) in the BioVU biobank. No phenotype associations were statistically significant at a Bonferroni adjusted level. Though this result may be a consequence of small sample size or sub-optimal imputation quality, it is largely consistent with previous studies that have failed to find a large number of clinical outcomes that correlate with CRP-associated variants.[12]

A primary goal of many human genetics studies is to identify the causal allele that underlies the association with a human trait or disease. As such, the value of deep sequencing data on hundreds of thousands of individuals from diverse genetic backgrounds should not be understated. Our results demonstrate the potential for WGS analysis to discover genetic signals, including conditionally distinct, low-frequency signals at known loci. Limitations of our current analysis include the modest sample size, particularly for ancestry groups other than European and African Americans, and the focus on single-variant tests only. As larger sample sizes become available, further study of aggregate tests for very rare variants and structural variation is warranted. Future studies from TOPMed and other large WGS efforts integrating both sequencing data and dense imputation, along with interrogation of rich functional annotation databases and higher-throughput cellular assays, will continue to clarify the role of genetic variation on complex traits.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.12.002.

Acknowledgments

Analysis of CRP variants was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (RO1 DK072193 and U01 DK105561). A.K.I. and K.L.M. were supported by RO1 DK072193. L.M.R. was supported by T32 HL129982. C.N.S. was supported by American Heart Association Postdoctoral Fellowship 15POST24470131 and 17POST33650016. E.J.B. was supported by HHSN2682015000001I, N01-HC 25195, RO1 HL64753, R01 HL076784, and R01 AG028321. B.E.C. was supported by K01 HL135405. M.H.K., Y.L., and A.P.R. were supported by R01 HL129132. A.C. was supported by HHSN268201800010, HHSN268201800011, HHSN268201800012, HHSN268201800013, HHSN268201800015, and HHSN268201800015. J.P.L. was supported by R01 HL137922. R.P.T. was supported by R01 HL120854. J.D. was supported by R01 HL128914. P.L.A. was supported by R01 HL132947. B.L. was supported by U01HG009086. S.S. was supported by K01AG059898.

Declaration of Interests

The authors declare no competing interests.

Received: October 4, 2019
Accepted: December 2, 2019
Published: December 26, 2019

Web Resources

AuthorArranger, https://authorarranger.nci.nih.gov/#/
Centers for Common Disease Genomics (CCDG), https://ccdg.rutgers.edu/
ENCORE, https://encore.sph.umich.edu/
eQTLGen, https://www.eqtlgen.org/index.html
GTEx, https://www.gtexportal.org/home/
OASIS, https://edn.som.umaryland.edu/OASIS/
OMIM, https://www.omim.org/
TOPMed Methods, https://www.nhlbiwgs.org/topmed-whole-genome-sequencing-project-freeze-5b-phases-1-and-2

References

1. The NIHR BioResource on behalf of the 100000 Genomes Project. (2019). Whole-genome sequencing of rare disease patients in a national healthcare system. bioRxiv. https://doi.org/10.1101/507244.
2. Sanders, S.J., Neale, B.M., Huang, H., Werling, D.M., An, J.-Y., Dong, S., Abecasis, G., Arguello, P.A., Blangero, J., Boehnke, M., et al.; Whole Genome Sequencing for Psychiatric Disorders (WGSPD) (2017). Whole genome sequencing in psychiatric disorders: the WGSPD consortium. Nat. Neurosci. *20*, 1661–1668.
3. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. bioRxiv. https://doi.org/10.1101/563866.
4. Lappalainen, T., Scott, A.J., Brandt, M., and Hall, I.M. (2019). Genomic Analysis in the Age of Human Genome Sequencing. Cell *177*, 70–84.
5. Popejoy, A.B., and Fullerton, S.M. (2016). Genomics is failing on diversity. Nature *538*, 161–164.
6. The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. Nature *526*, 68–74.
7. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al.; Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.
8. Wojcik, G.L., Fuchsberger, C., Taliun, D., Welch, R., Martin, A.R., Shringarpure, S., Carlson, C.S., Abecasis, G., Kang, H.M., Boehnke, M., et al. (2018). Imputation-Aware Tag SNP Selection To Improve Power for Large-Scale, Multi-ethnic Association Studies. G3 (Bethesda) *8*, 3255–3267.
9. Polfus, L.M., Raffield, L.M., Wheeler, M.M., Tracy, R.P., Lange, L.A., Lettre, G., Miller, A., Correa, A., Bowler, R.P., Bis, J.C., et al. (2019). Whole genome sequence association with E-

selectin levels reveals Loss-of-function variant in African Americans. Hum. Mol. Genet. *28*, 515–523.
10. Schick, U.M., Auer, P.L., Bis, J.C., Lin, H., Wei, P., Pankratz, N., Lange, L.A., Brody, J., Stitziel, N.O., Kim, D.S., et al.; Cohorts for Heart and Aging Research in Genomic Epidemiology; and National Heart, Lung, and Blood Institute GO Exome Sequencing Project (2015). Association of exome sequences with plasma C-reactive protein levels in >9000 participants. Hum. Mol. Genet. *24*, 559–571.
11. Prins, B.P., Abbasi, A., Wong, A., Vaez, A., Nolte, I., Franceschini, N., Stuart, P.E., Guterriez Achury, J., Mistry, V., Bradfield, J.P., et al.; PAGE Consortium; International Stroke Genetics Consortium; Systemic Sclerosis consortium; Treat OA consortium; DIAGRAM Consortium; CARDIoGRAMplusC4D Consortium; ALS consortium; International Parkinson's Disease Genomics Consortium; Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium; CKDGen consortium; GERAD1 Consortium; International Consortium for Blood Pressure; Schizophrenia Working Group of the Psychiatric Genomics Consortium; and Inflammation Working Group of the CHARGE Consortium (2016). Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. PLoS Med. *13*, e1001976.
12. Ligthart, S., Vaez, A., Võsa, U., Stathopoulou, M.G., de Vries, P.S., Prins, B.P., Van der Most, P.J., Tanaka, T., Naderi, E., Rose, L.M., et al.; LifeLines Cohort Study; and CHARGE Inflammation Working Group (2018). Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. Am. J. Hum. Genet. *103*, 691–706.
13. Markatseli, T.E., Voulgari, P.V., Alamanos, Y., and Drosos, A.A. (2011). Prognostic factors of radiological damage in rheumatoid arthritis: a 10-year retrospective study. J. Rheumatol. *38*, 44–52.
14. Gaitonde, S., Samols, D., and Kushner, I. (2008). C-reactive protein and systemic lupus erythematosus. Arthritis Rheum. *59*, 1814–1820.
15. Wang, X., Bao, W., Liu, J., Ouyang, Y.-Y., Wang, D., Rong, S., Xiao, X., Shan, Z.-L., Zhang, Y., Yao, P., and Liu, L.G. (2013). Inflammatory markers and risk of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care *36*, 166–175.
16. Zacho, J., Tybjaerg-Hansen, A., and Nordestgaard, B.G. (2010). C-reactive protein and all-cause mortality–the Copenhagen City Heart Study. Eur. Heart J. *31*, 1624–1632.
17. Austin, M.A., Zhang, C., Humphries, S.E., Chandler, W.L., Talmud, P.J., Edwards, K.L., Leonetti, D.L., McNeely, M.J., and Fujimoto, W.Y. (2004). Heritability of C-reactive protein and association with apolipoprotein E genotypes in Japanese Americans. Ann. Hum. Genet. *68*, 179–188.
18. Pankow, J.S., Folsom, A.R., Cushman, M., Borecki, I.B., Hopkins, P.N., Eckfeldt, J.H., and Tracy, R.P. (2001). Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. Atherosclerosis *154*, 681–689.
19. Vickers, M.A., Green, F.R., Terry, C., Mayosi, B.M., Julier, C., Lathrop, M., Ratcliffe, P.J., Watkins, H.C., and Keavney, B. (2002). Genotype at a promoter polymorphism of the interleukin-6 gene is associated with baseline levels of plasma C-reactive protein. Cardiovasc. Res. *53*, 1029–1034.
20. Schnabel, R.B., Lunetta, K.L., Larson, M.G., Dupuis, J., Lipinska, I., Rong, J., Chen, M.-H., Zhao, Z., Yamamoto, J.F., Meigs, J.B., et al. (2009). The relation of genetic and environmental factors to systemic inflammatory biomarker concentrations. Circ Cardiovasc Genet *2*, 229–237.
21. Fox, E.R., Benjamin, E.J., Sarpong, D.F., Rotimi, C.N., Wilson, J.G., Steffes, M.W., Chen, G., Adeyemo, A., Taylor, J.K., Samdarshi, T.E., and Taylor, H.A., Jr. (2008). Epidemiology, heritability, and genetic linkage of C-reactive protein in African Americans (from the Jackson Heart Study). Am. J. Cardiol. *102*, 835–841.
22. Khera, A., McGuire, D.K., Murphy, S.A., Stanek, H.G., Das, S.R., Vongpatanasin, W., Wians, F.H., Jr., Grundy, S.M., and de Lemos, J.A. (2005). Race and gender differences in C-reactive protein levels. J. Am. Coll. Cardiol. *46*, 464–469.
23. Lakoski, S.G., Cushman, M., Palmas, W., Blumenthal, R., D'Agostino, R.B., Jr., and Herrington, D.M. (2005). The relationship between blood pressure and C-reactive protein in the Multi-Ethnic Study of Atherosclerosis (MESA). J. Am. Coll. Cardiol. *46*, 1869–1874.
24. Wu, Y., McDade, T.W., Kuzawa, C.W., Borja, J., Li, Y., Adair, L.S., Mohlke, K.L., and Lange, L.A. (2012). Genome-wide association with C-reactive protein levels in CLHNS: evidence for the CRP and HNF1A loci and their interaction with exposure to a pathogenic environment. Inflammation *35*, 574–583.
25. Okada, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Kamatani, Y., Hosono, N., Tsunoda, T., Matsuda, K., Tanaka, T., Kubo, M., et al. (2011). Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. Hum. Mol. Genet. *20*, 1224–1231.
26. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514–518.
27. Reiner, A.P., Beleza, S., Franceschini, N., Auer, P.L., Robinson, J.G., Kooperberg, C., Peters, U., and Tang, H. (2012). Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. Am. J. Hum. Genet. *91*, 502–512.
28. Kocarnik, J.M., Richard, M., Graff, M., Haessler, J., Bien, S., Carlson, C., Carty, C.L., Reiner, A.P., Avery, C.L., Ballantyne, C.M., et al. (2018). Discovery, fine-mapping, and conditional analyses of genetic variants associated with C-reactive protein in multiethnic populations using the Metabochip in the Population Architecture using Genomics and Epidemiology (PAGE) study. Hum. Mol. Genet. *27*, 2940–2953.
29. Doumatey, A.P., Chen, G., Tekola Ayele, F., Zhou, J., Erdos, M., Shriner, D., Huang, H., Adeleye, J., Balogun, W., Fasanmade, O., et al. (2012). C-reactive protein (CRP) promoter polymorphisms influence circulating CRP levels in a genome-wide association study of African Americans. Hum. Mol. Genet. *21*, 3063–3072.
30. Kheradpour, P., and Kellis, M. (2014). Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. Nucleic Acids Res. *42*, 2976–2987.
31. Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., Sidow, A.; and NISC Comparative Sequencing Program (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. *15*, 901–913.
32. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40*, D930–D934.

33. Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., et al.; GTEx Consortium (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat. Genet. *50*, 956–967.
34. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S., et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv. https://doi.org/10.1101/447367.
35. Strunz, T., Grassmann, F., Gayán, J., Nahkuri, S., Souza-Costa, D., Maugeais, C., Fauser, S., Nogoceke, E., and Weber, B.H.F. (2018). A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver. Sci. Rep. *8*, 5865.
36. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) *2017*, bax028.
37. Wang, T.M., Hsieh, S.C., Chen, J.W., and Chiang, A.N. (2013). Docosahexaenoic acid and eicosapentaenoic acid reduce C-reactive protein expression and STAT3 activation in IL-6-treated HepG2 cells. Mol. Cell. Biochem. *377*, 97–106.
38. Tsukada, J., Yoshida, Y., Kominato, Y., and Auron, P.E. (2011). The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation. Cytokine *54*, 6–19.
39. Kowalski, M.H., Qian, H., Hou, Z., Rosen, J.D., Tapia, A.L., Shan, Y., Jain, D., Argos, M., Arnett, D.K., Avery, C., et al. (2019). Use of ~100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. bioRxiv. https://doi.org/10.1101/683201.

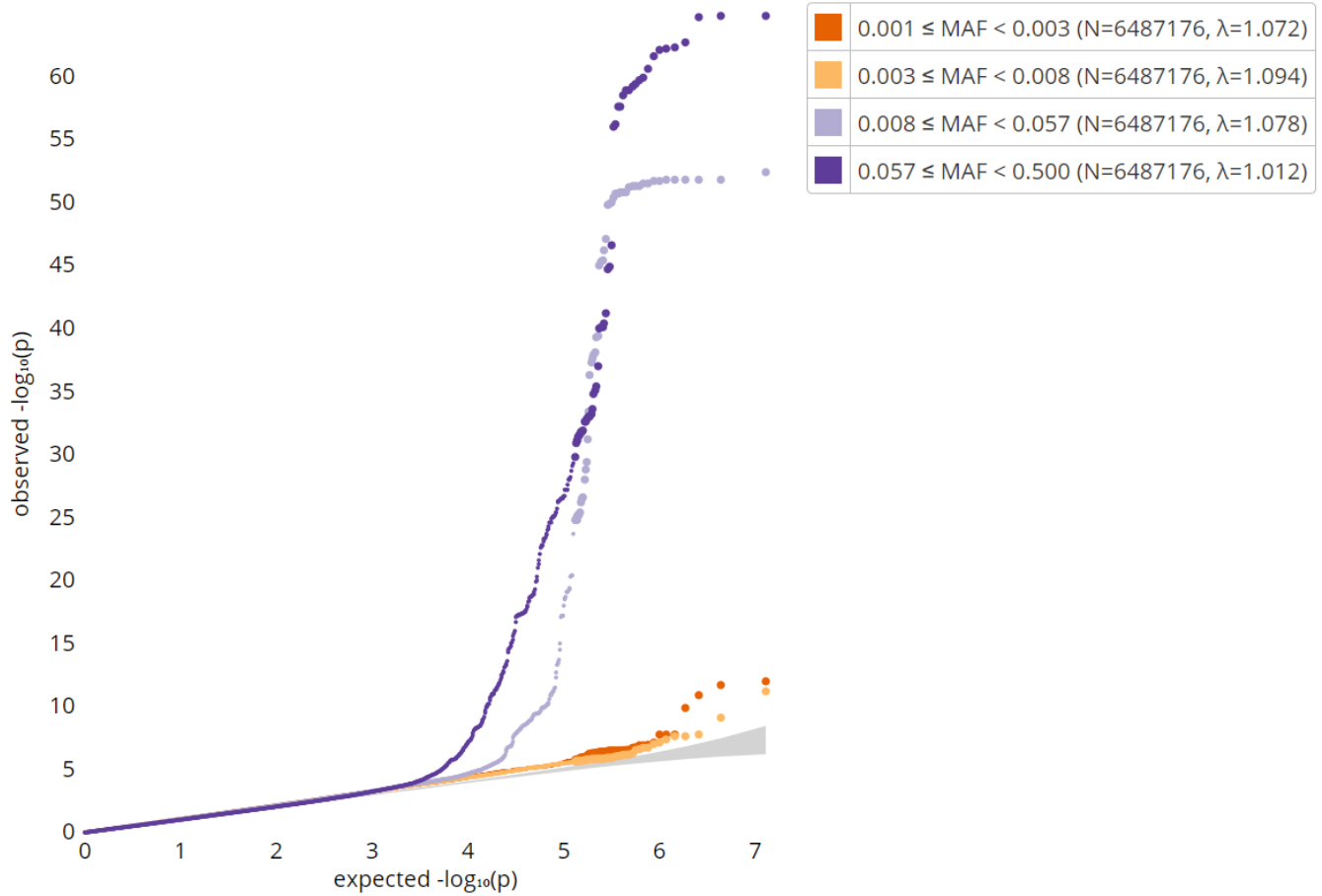
The American Journal of Human Genetics, Volume *106*

Supplemental Data

Allelic Heterogeneity at the *CRP* Locus Identified by Whole-Genome Sequencing in Multi-ancestry Cohorts

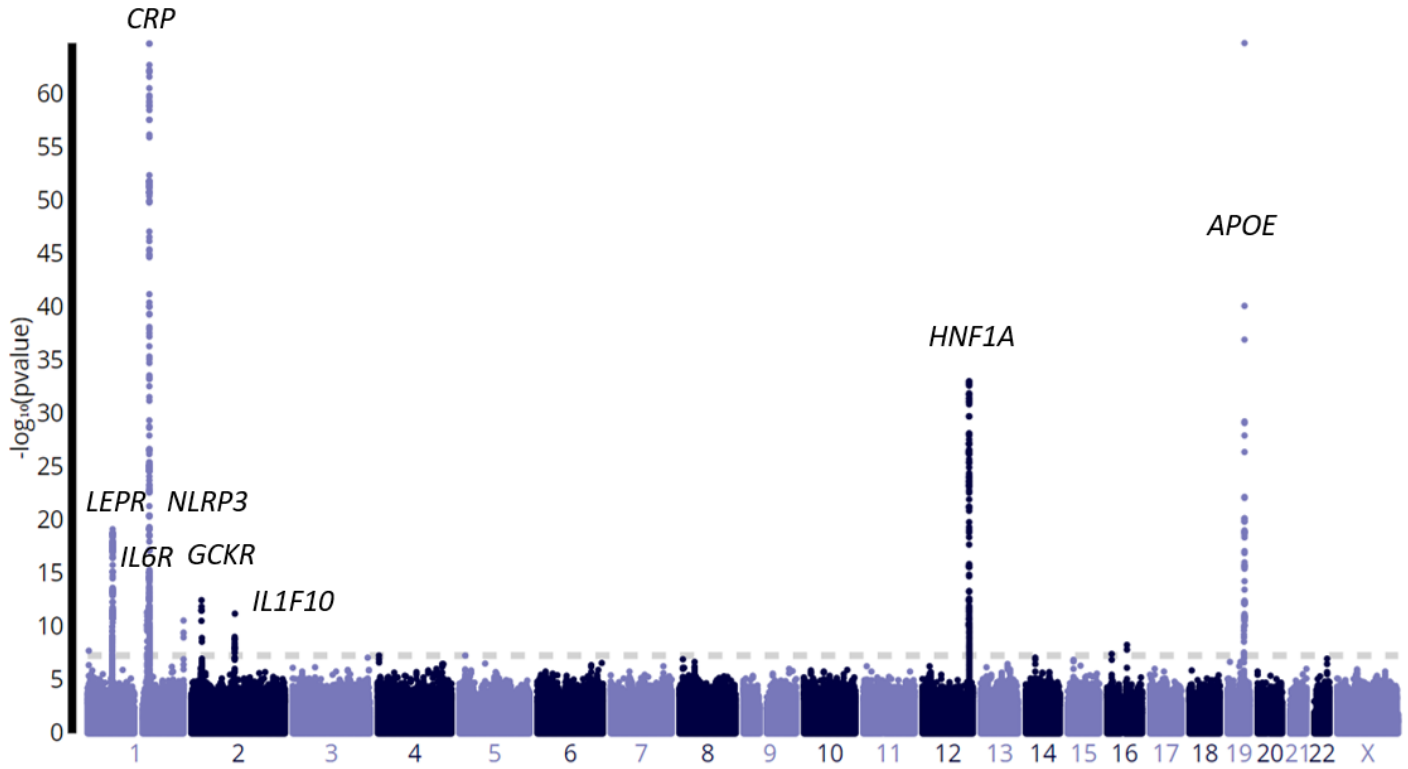
Laura M. Raffield, Apoorva K. Iyengar, Biqi Wang, Sheila M. Gaynor, Cassandra N. Spracklen, Xue Zhong, Madeline H. Kowalski, Shabnam Salimi, Linda M. Polfus, Emelia J. Benjamin, Joshua C. Bis, Russell Bowler, Brian E. Cade, Won Jung Choi, Alejandro P. Comellas, Adolfo Correa, Pedro Cruz, Harsha Doddapaneni, Peter Durda, Stephanie M. Gogarten, Deepti Jain, Ryan W. Kim, Brian G. Kral, Leslie A. Lange, Martin G. Larson, Cecelia Laurie, Jiwon Lee, Seonwook Lee, Joshua P. Lewis, Ginger A. Metcalf, Braxton D. Mitchell, Zeineen Momin, Donna M. Muzny, Nathan Pankratz, Cheol Joo Park, Stephen S. Rich, Jerome I. Rotter, Kathleen Ryan, Daekwan Seo, Russell P. Tracy, Karine A. Viaud-Martinez, Lisa R. Yanek, Lue Ping Zhao, Xihong Lin, Bingshan Li, Yun Li, Josée Dupuis, Alexander P. Reiner, Karen L. Mohlke, Paul L. Auer, TOPMed Inflammation Working Group, and NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

Figure S1a: QQ-plot of association analysis for C-reactive protein in TOPMed.



Observed versus expected $-\log_{10}$ p-values for all variants included in the pooled ancestry C-reactive protein analysis on ENCORE, stratified by minor allele frequency (MAF) bin, with genomic inflation λ for each bin.

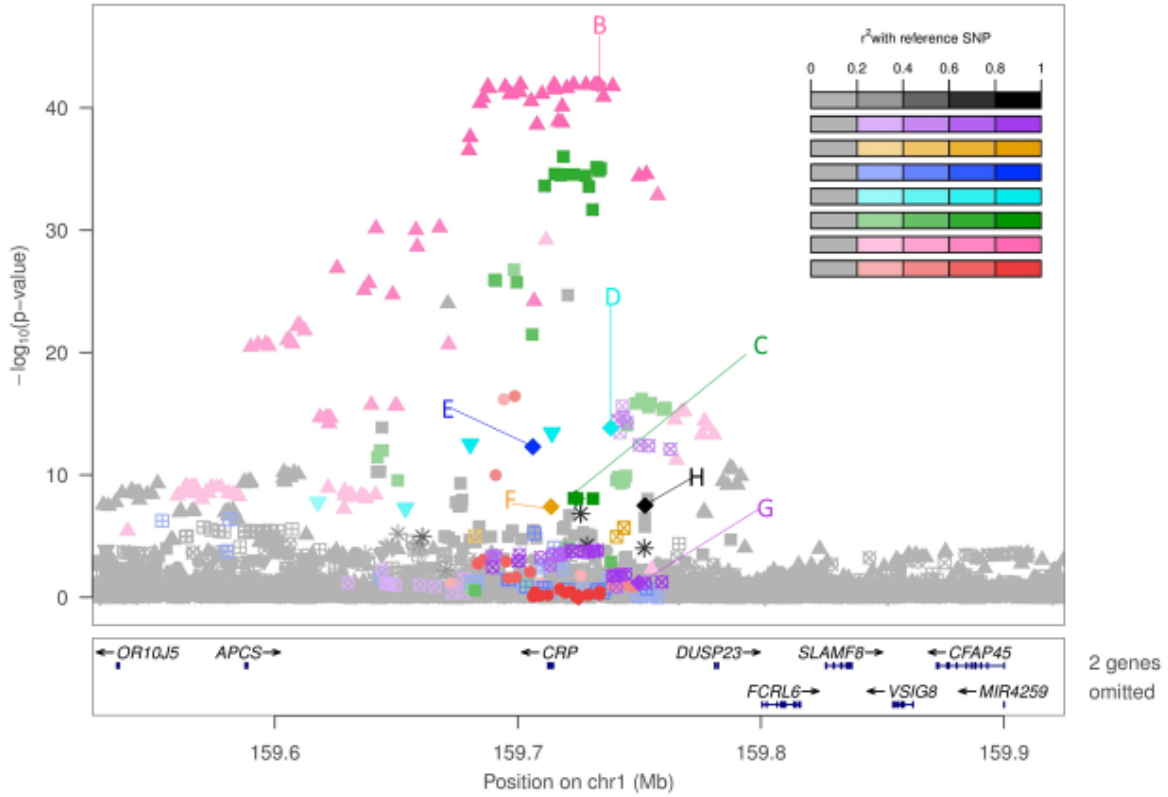
Figure S1b: Manhattan plot of CRP association signals in TOPMed.



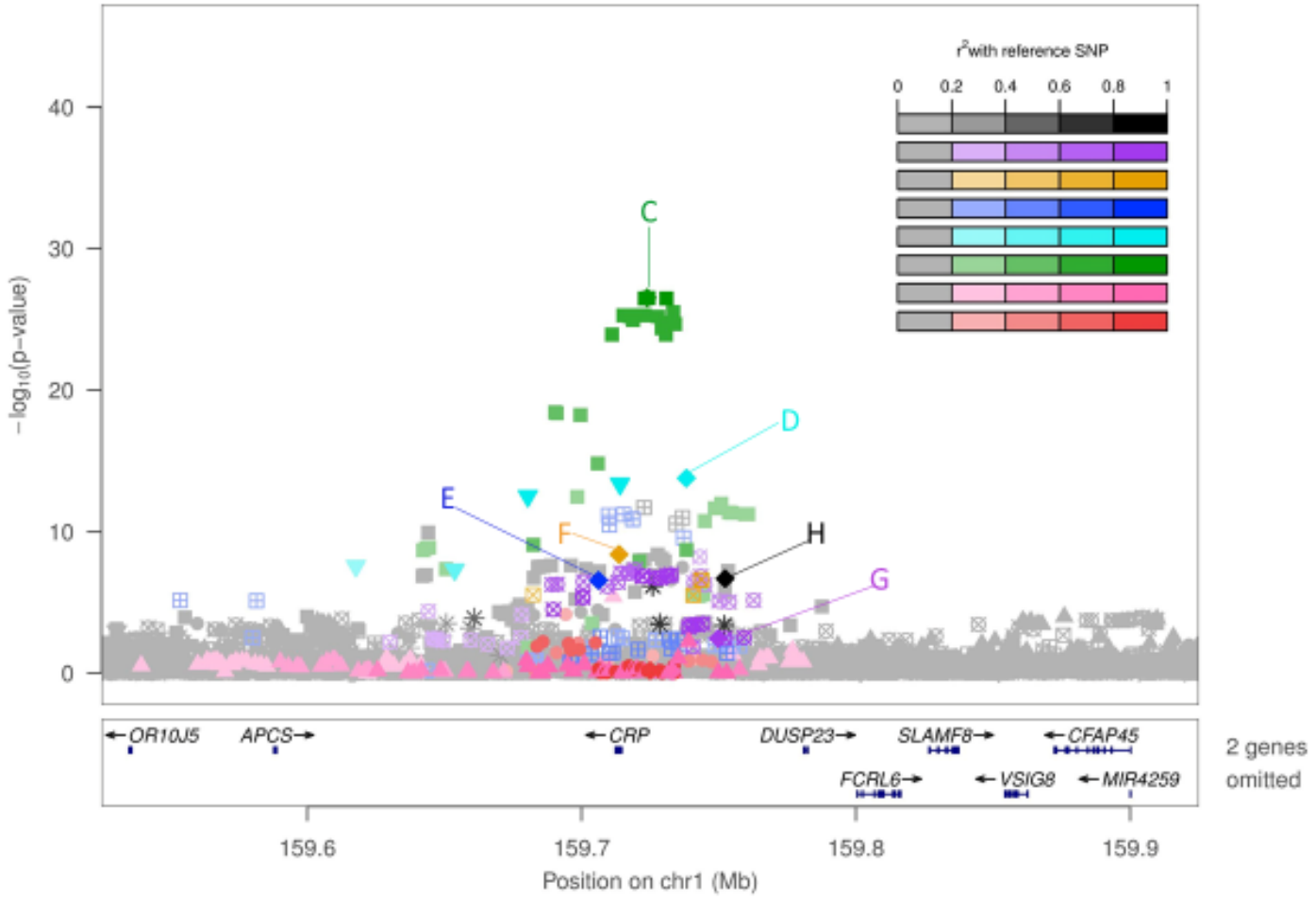
Y axis displays -$\log_{10}$ p-values for all variants included in the pooled ancestry C-reactive protein analysis on ENCORE, with the *x* axis displaying chromosomal position.

Figure S2: LocusZoom plots for sequential conditional analysis results at *CRP* locus, as well as plot of *CRP* locus adjusting for all previously identified *CRP* locus variants. For each plot, linkage disequilibrium is calculated using the same TOPMed samples included in our ancestry pooled C-reactive protein analyses.

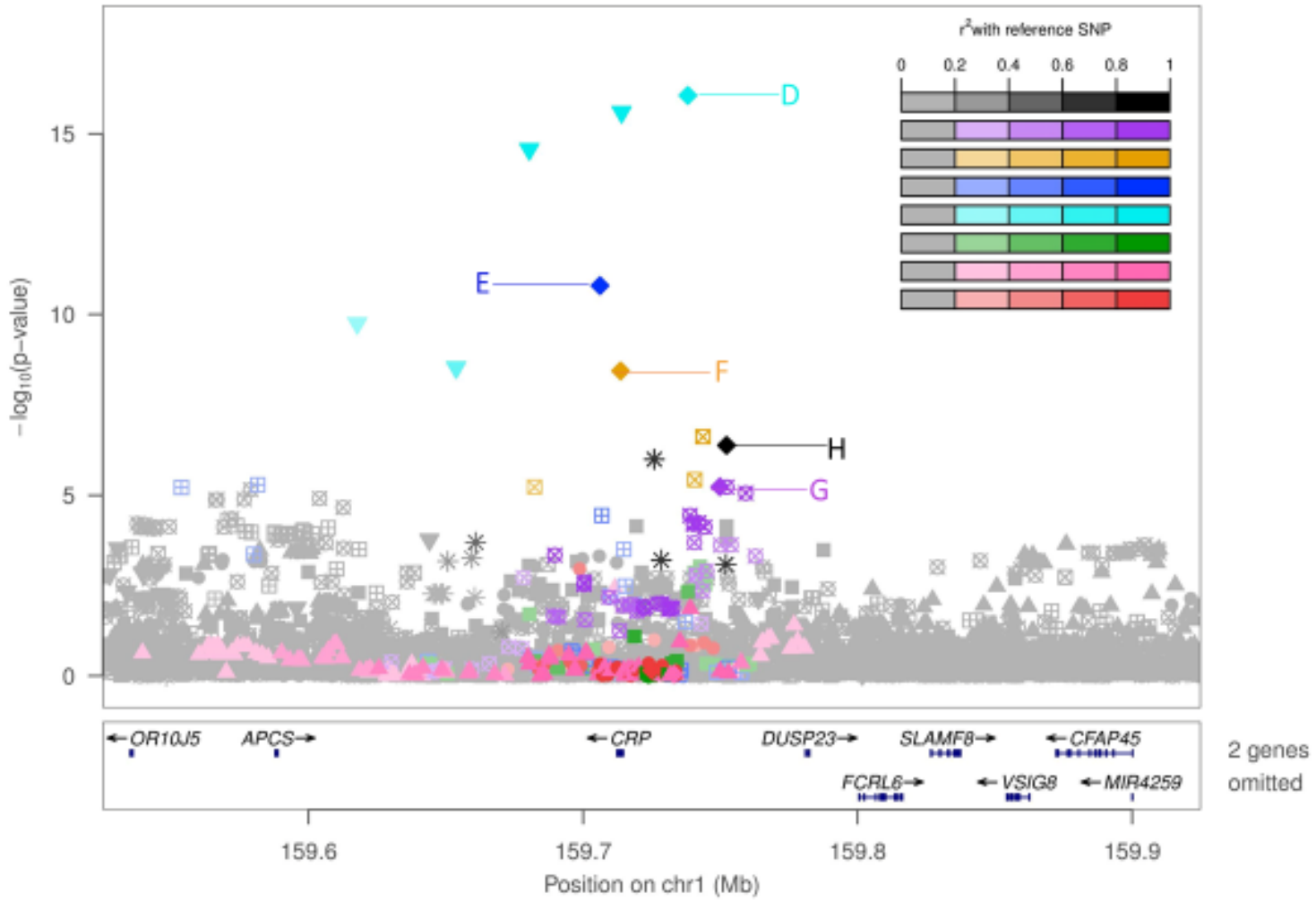
a. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731 (lead variant rs73024795). Letters in this and subsequent figures correspond to the list of conditionally distinct signals in Table 2. All plots display -log10(p-value) versus genomic location for all distinct signals subsequent to ones conditioned on, using order from Table. The lead variant for each conditionally distinct signal is indicated with a diamond, with other variants in linkage disequilibrium r^2>0.2 indicated in the colors used for each letter label and displayed on the legend at right, each with a different shape (for example, variants in close linkage disequilibrium with signal B (rs73024795) are displayed as pink triangles).



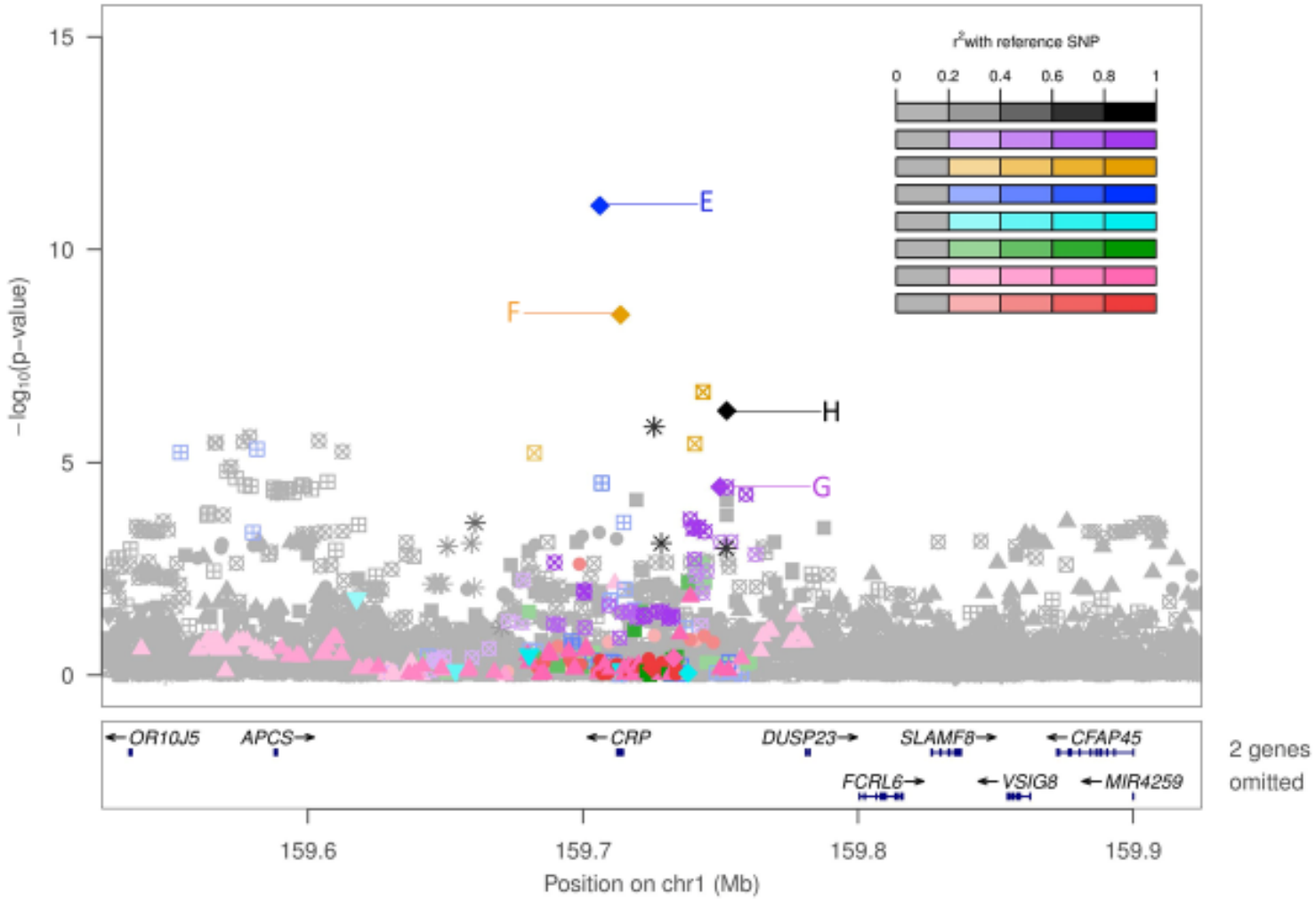
b. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731 and rs73024795 (lead variant rs2211321).



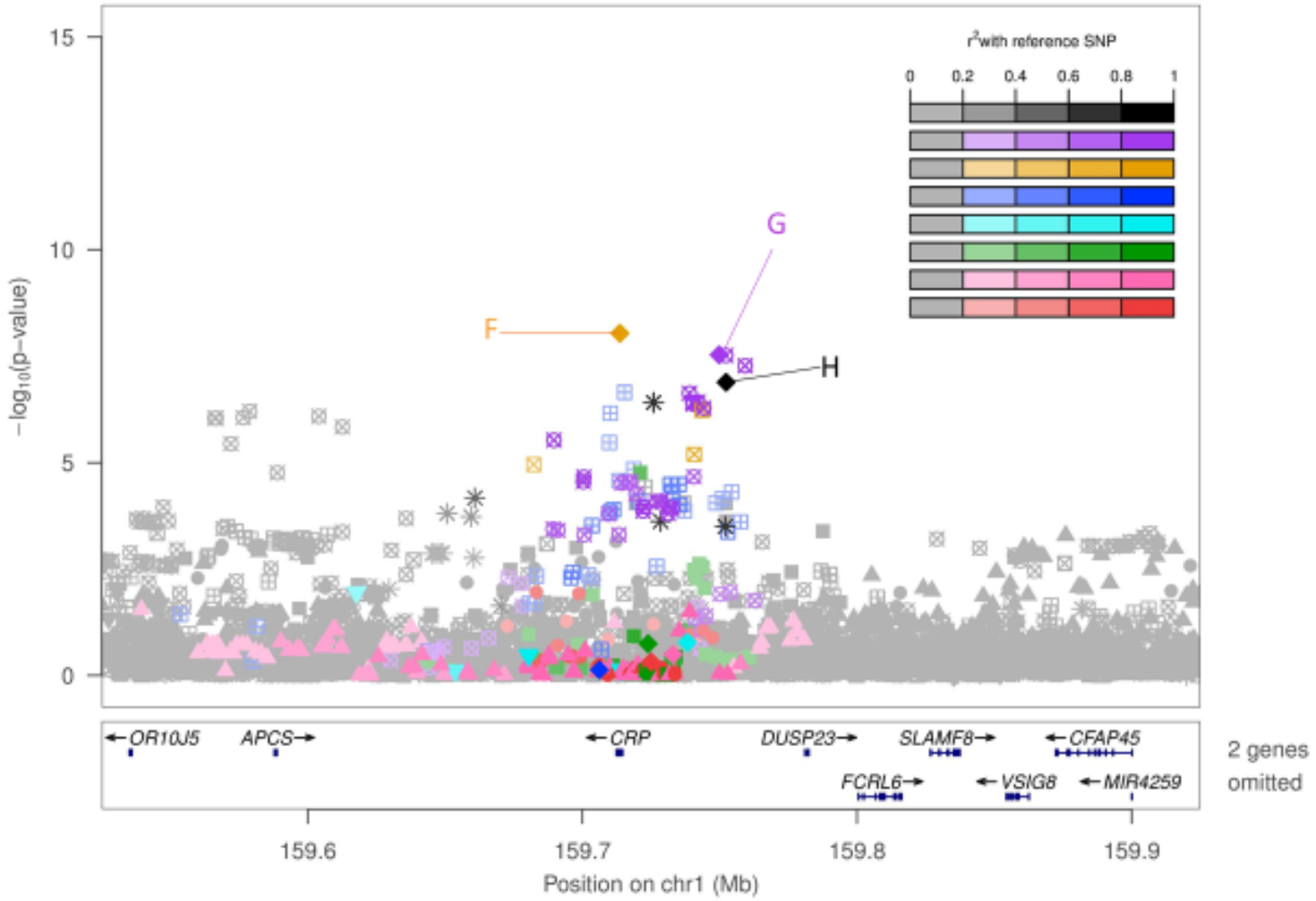
c. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, and rs2211321 (lead variant rs553202904).



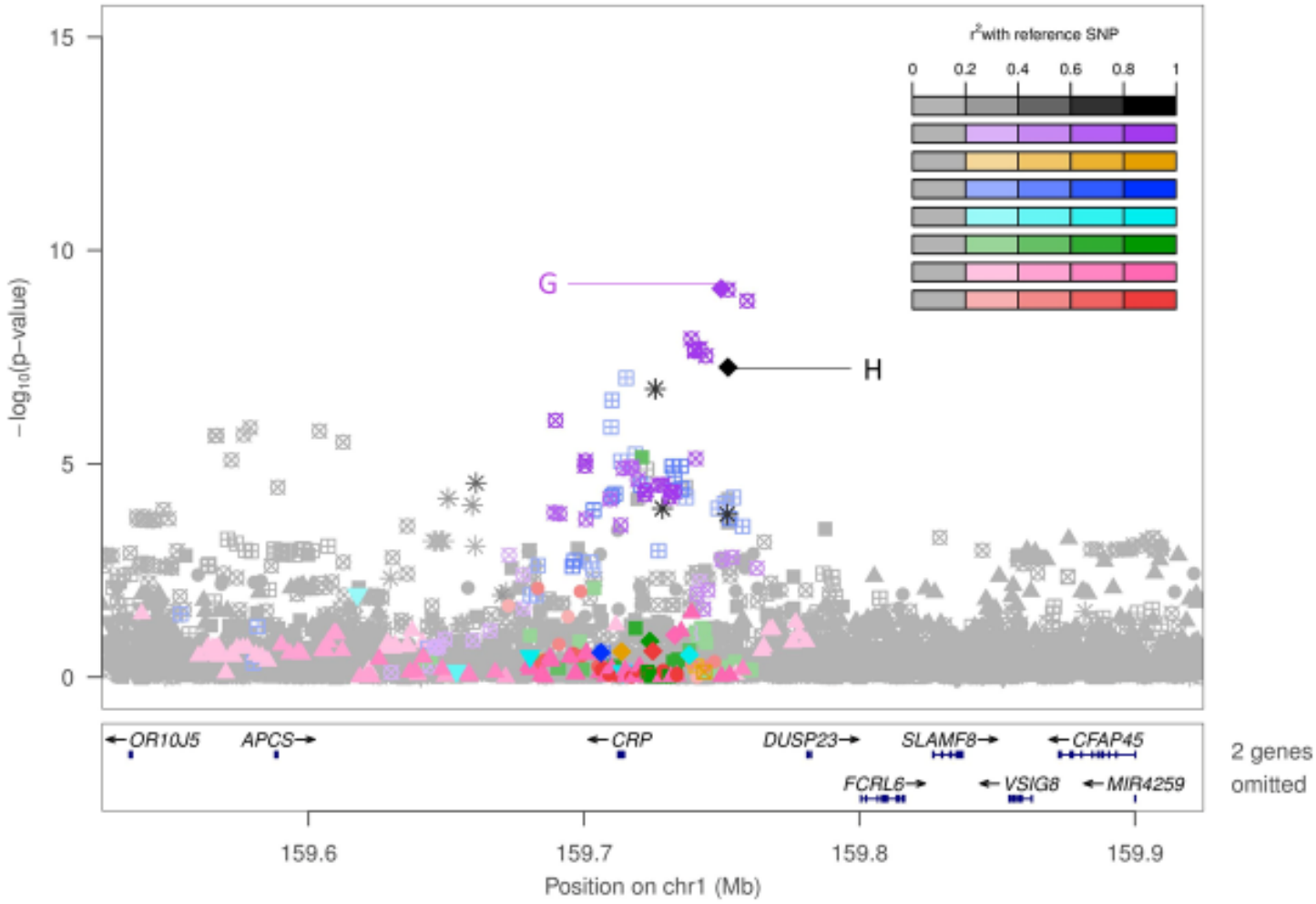
d. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, and rs553202904 (lead variant rs11265259).



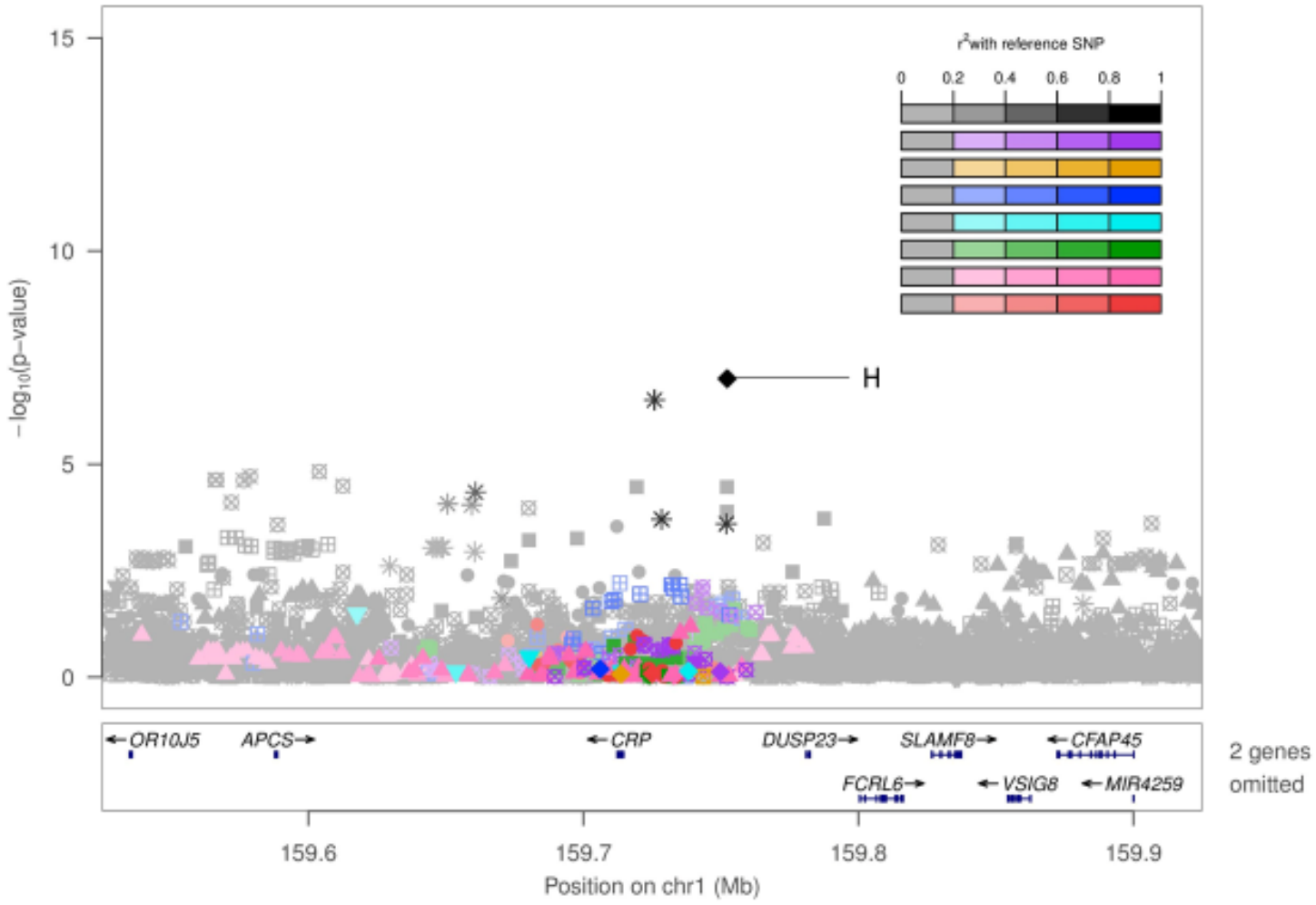
e. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, and rs11265259 (lead variant rs1800947).



f. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, rs11265259, and rs1800947 (lead variant rs12734907).



g. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, rs11265259, rs1800947, and rs12734907 (lead variant rs181704186).



h. Ancestry pooled analysis conditioned on all previously known variants from GWAS and exome sequencing studies. Only signals E and H are labelled, as these are the only signals still reaching our locus-wide significance threshold (as listed in Table 1).

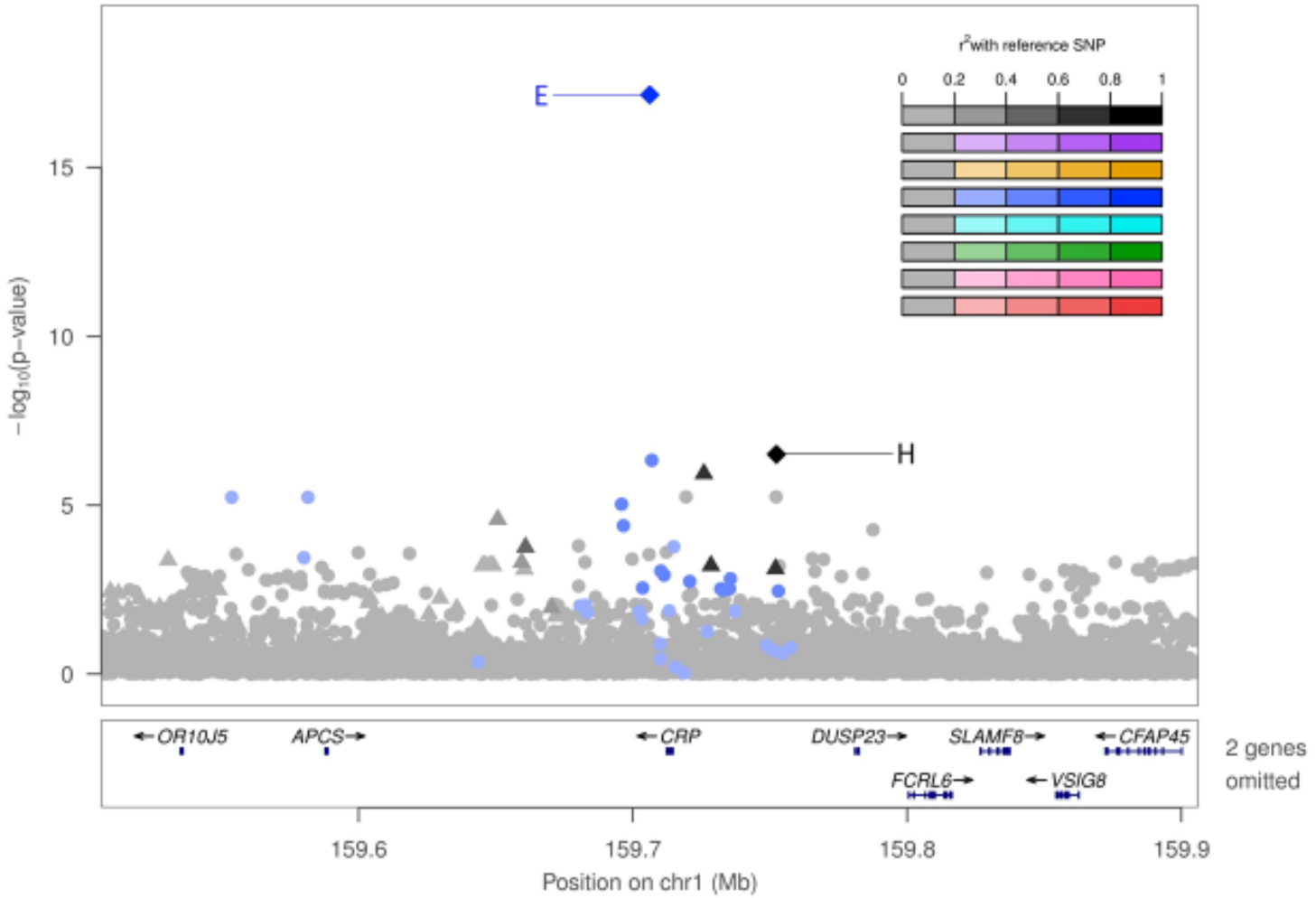
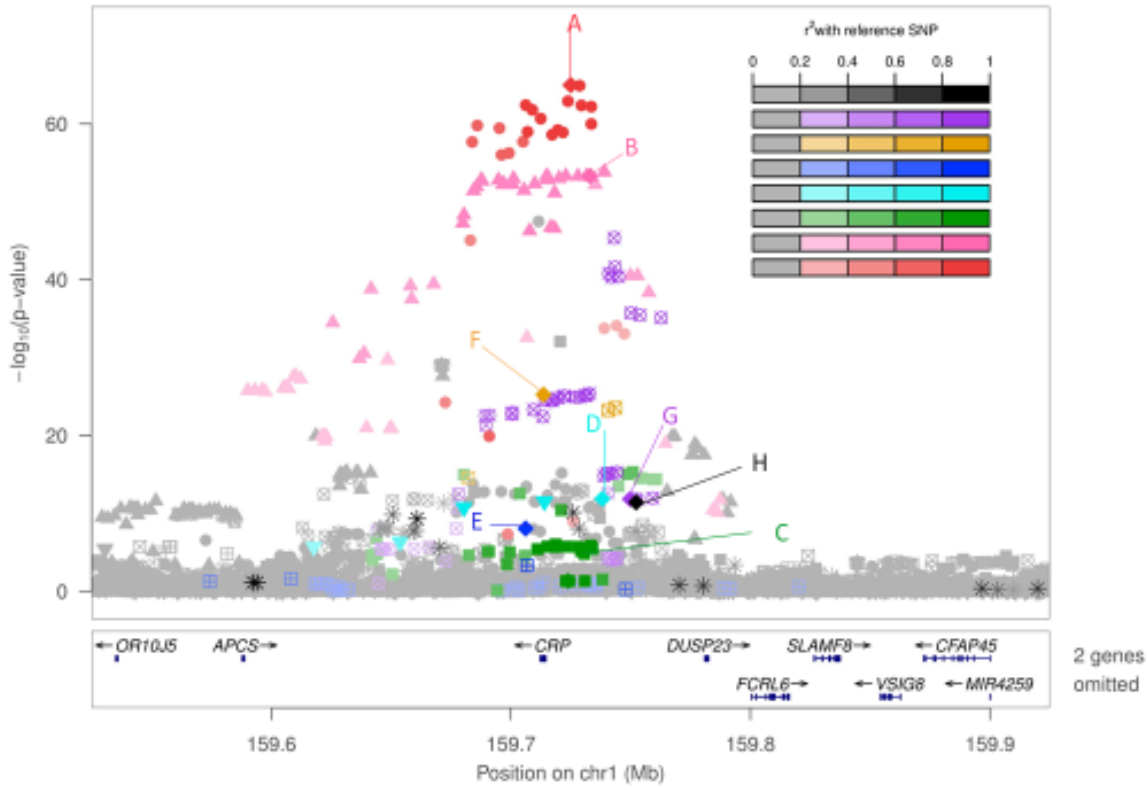
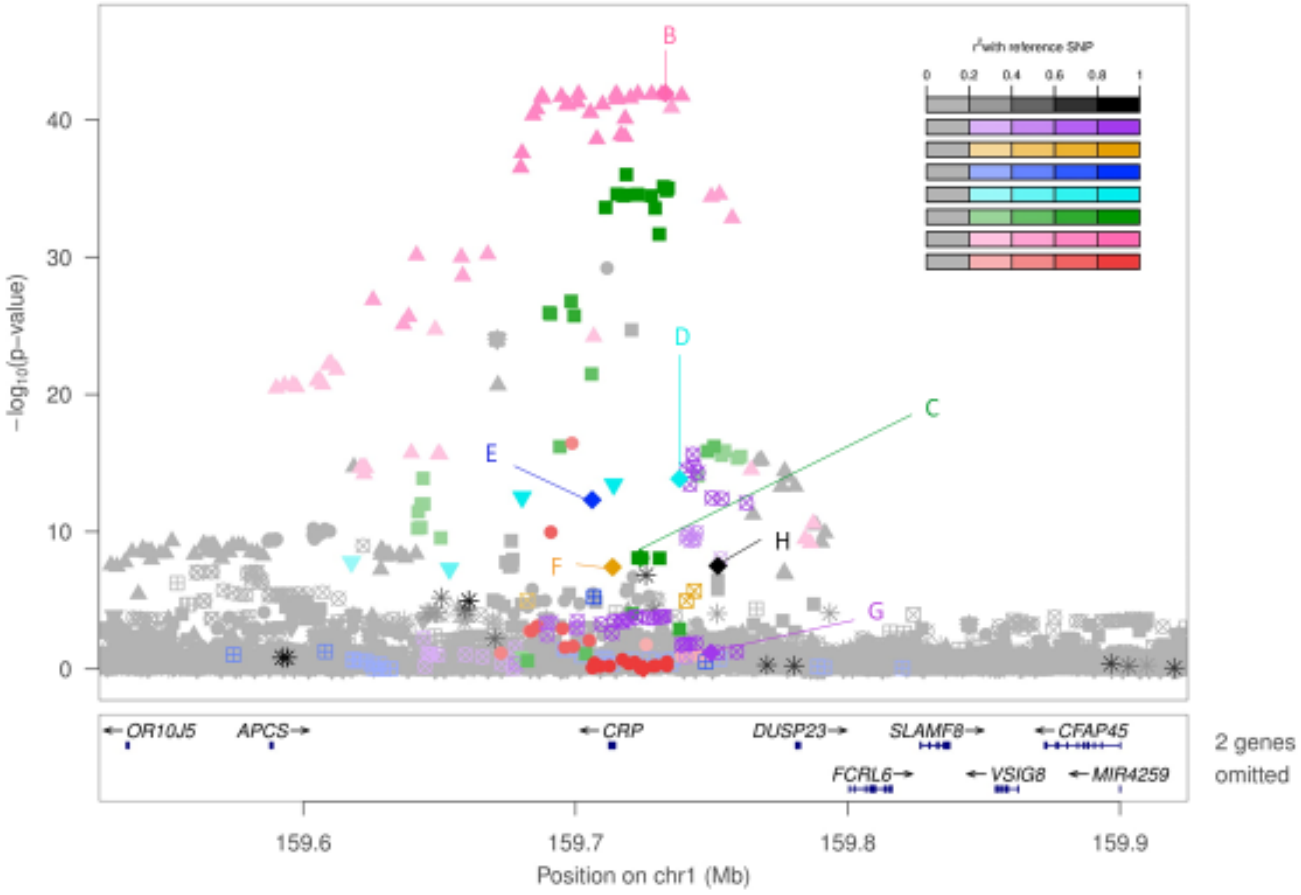


Figure S3: LocusZoom plots for sequential conditional analysis results at *CRP* locus, as well as plot of *CRP* locus adjusting for all previously identified *CRP* locus variants, with ancestry stratified LD reference panels.

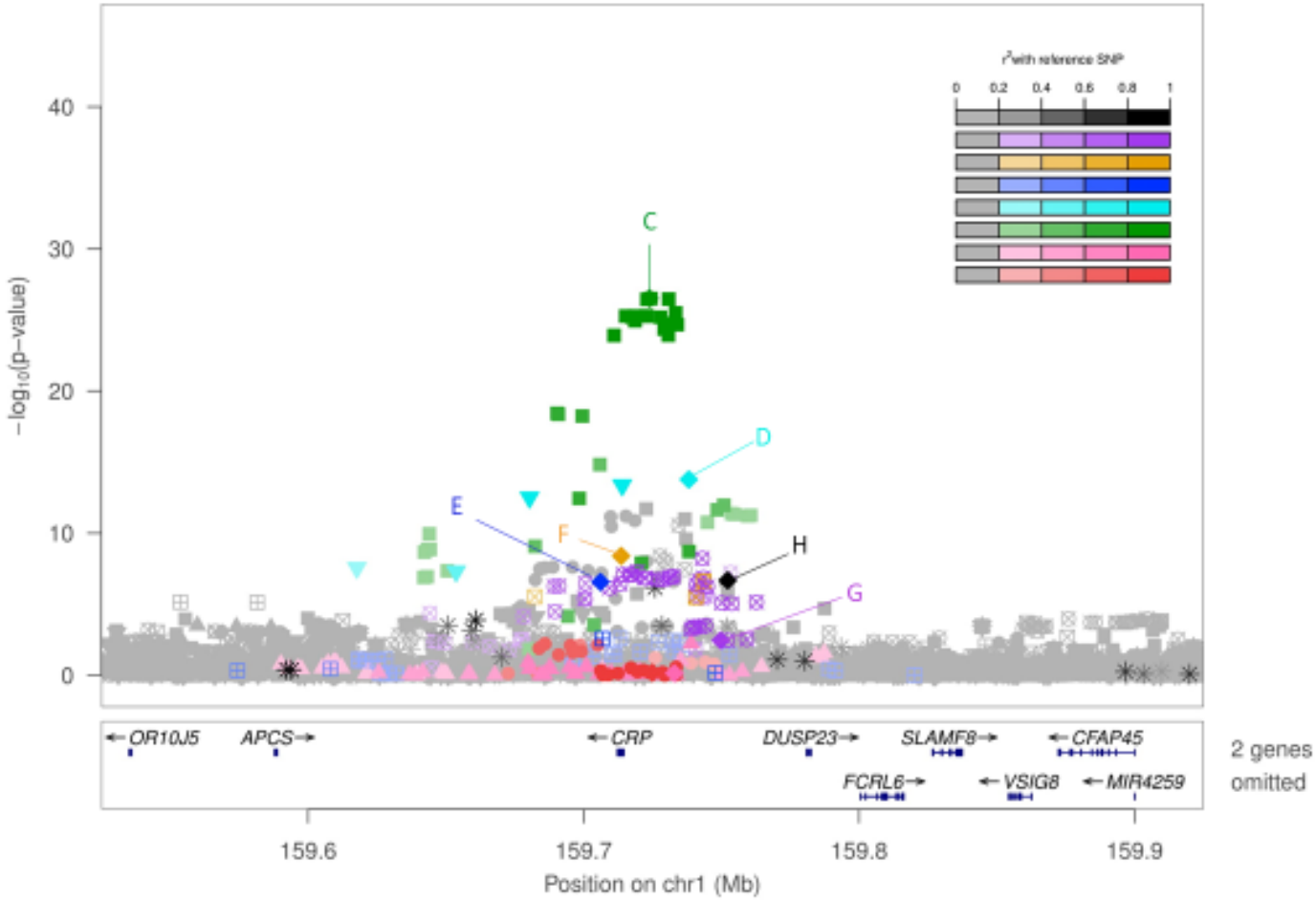
a. In ancestry pooled analysis, LocusZoom plot of association results (lead variant rs7551731). Letters in this and subsequent figures correspond to the list of conditionally distinct signals in Table 2. All plots display -log10(p-value) versus genomic location for all distinct signals subsequent to ones conditioned on, using order from Table. The lead variant for each conditionally distinct signal is indicated with a diamond, with other variants in linkage disequilibrium r^2>0.2 indicated in the colors used for each letter label and displayed on the legend at right, each with a different shape (for example, variants in close linkage disequilibrium with signal B (rs73024795) are displayed as pink triangles). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



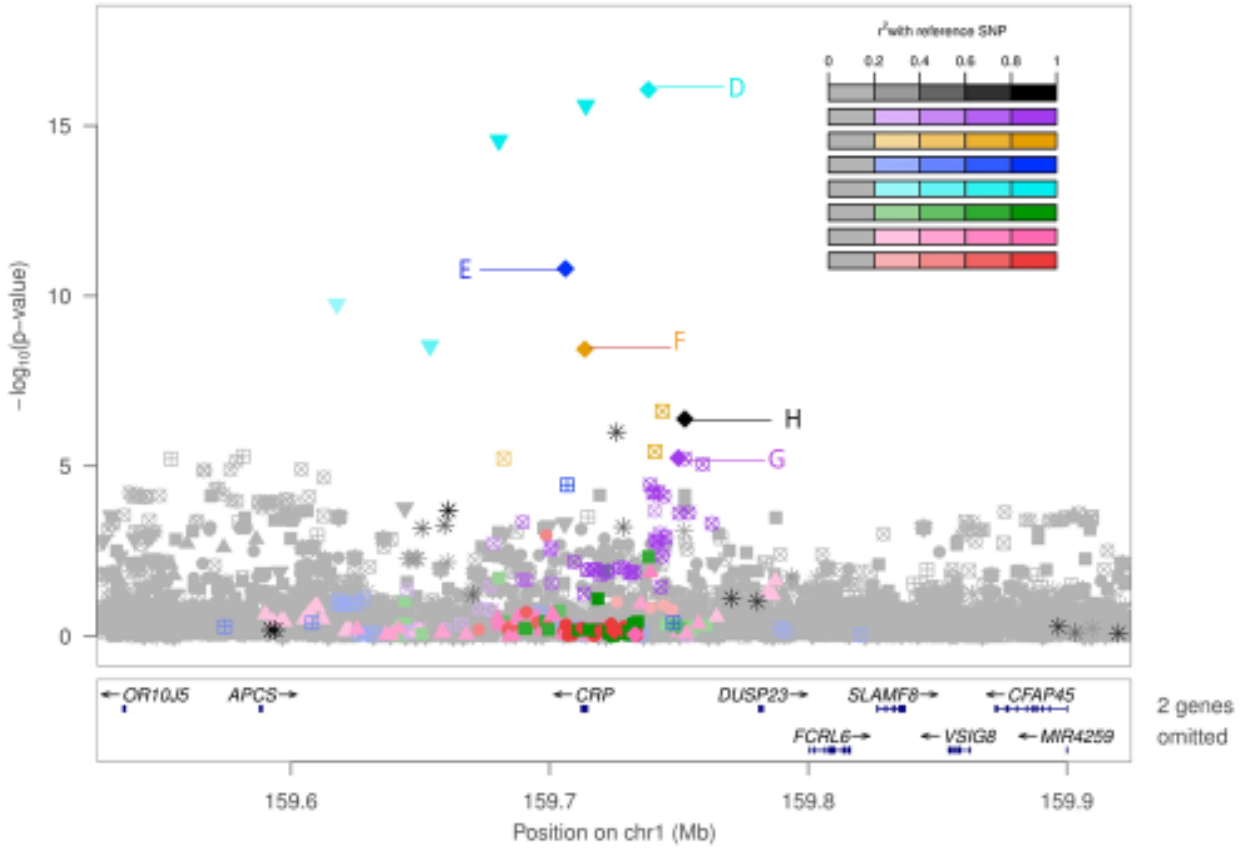
b. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731 (lead variant rs73024795). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



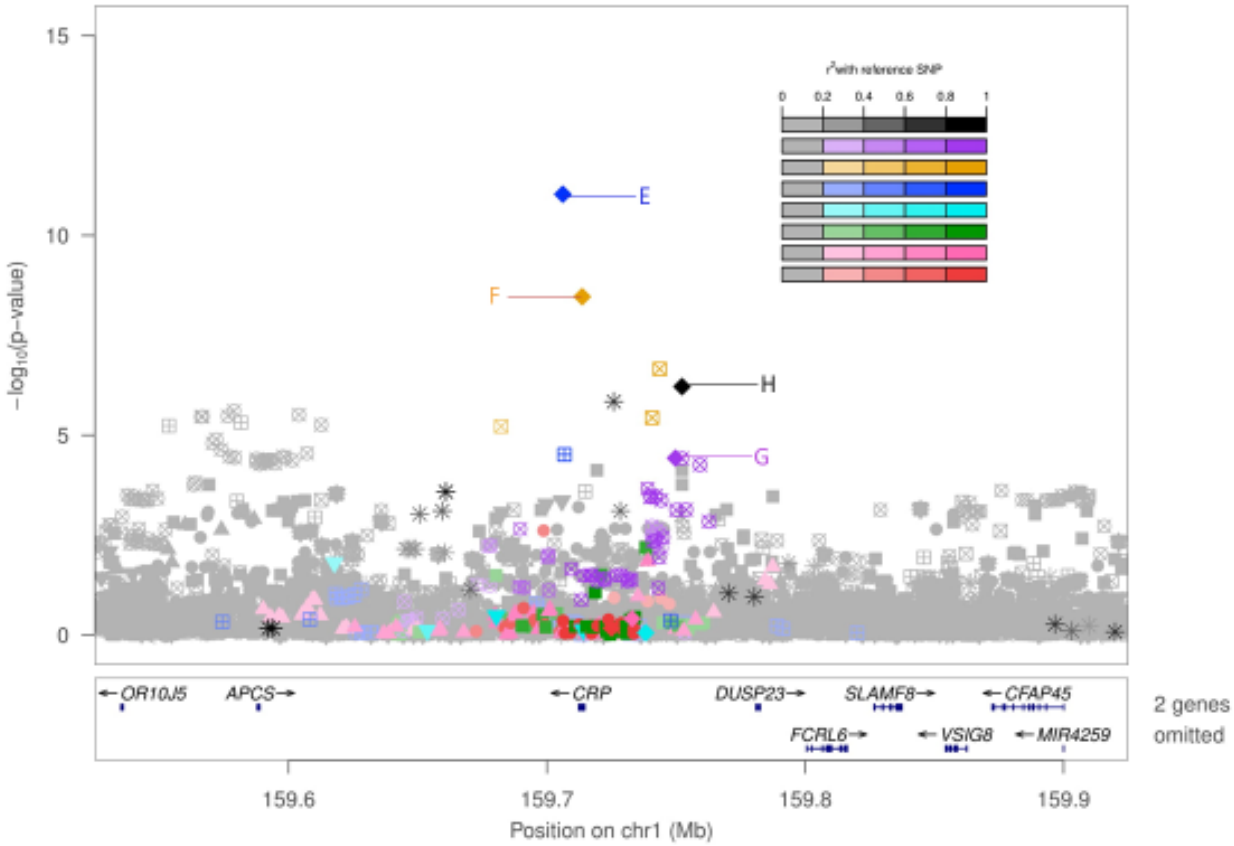
c. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731 and rs73024795 (lead variant rs2211321). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



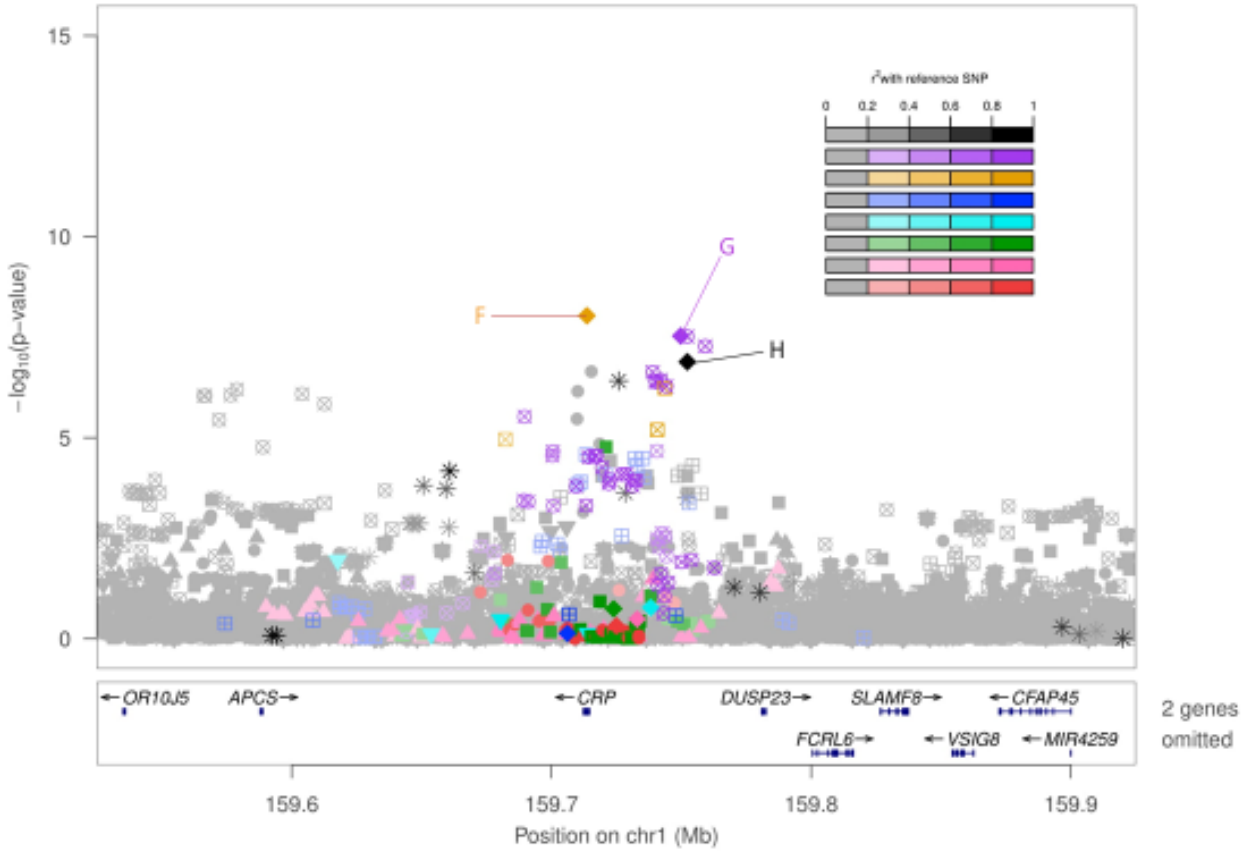
d. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, and rs2211321 (lead variant rs553202904). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



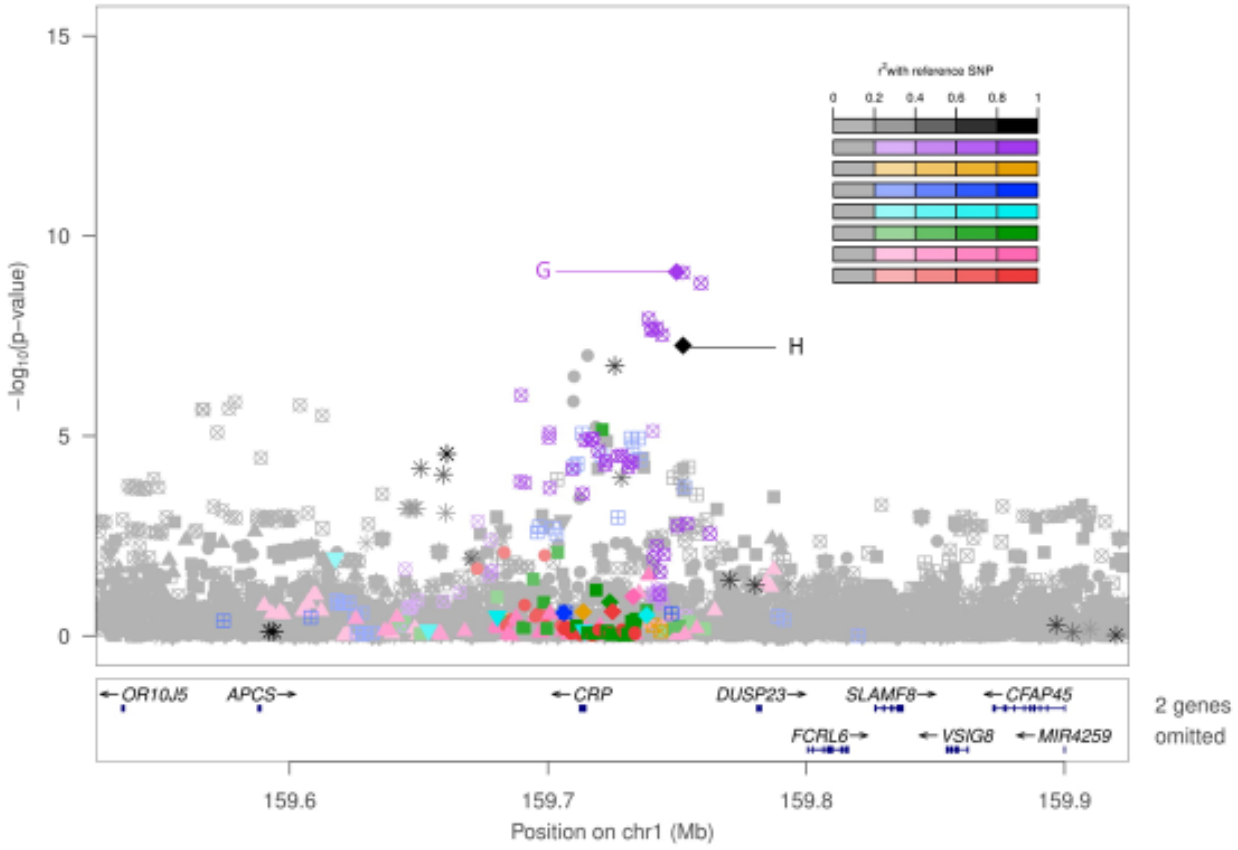
e. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, and rs553202904 (lead variant rs11265259). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



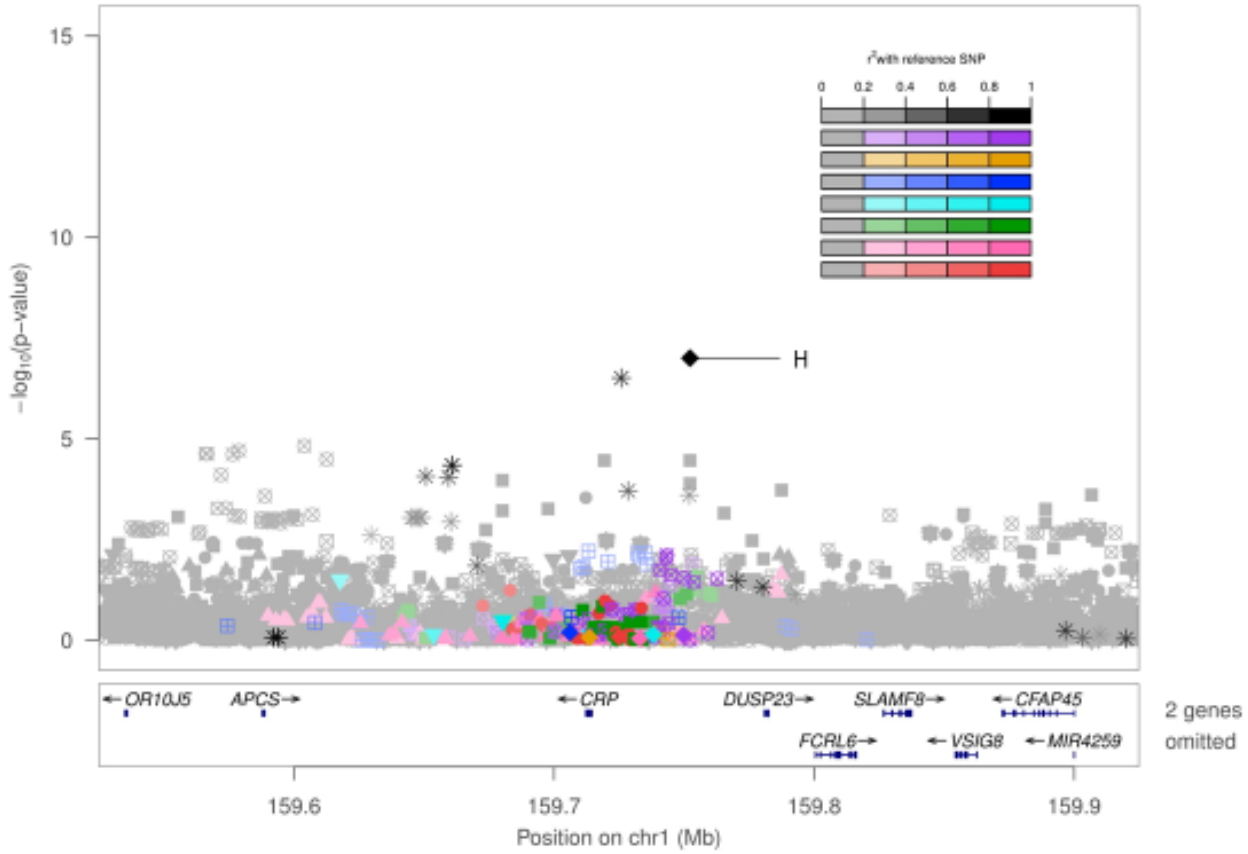
f. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, and rs11265259 (lead variant rs1800947). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



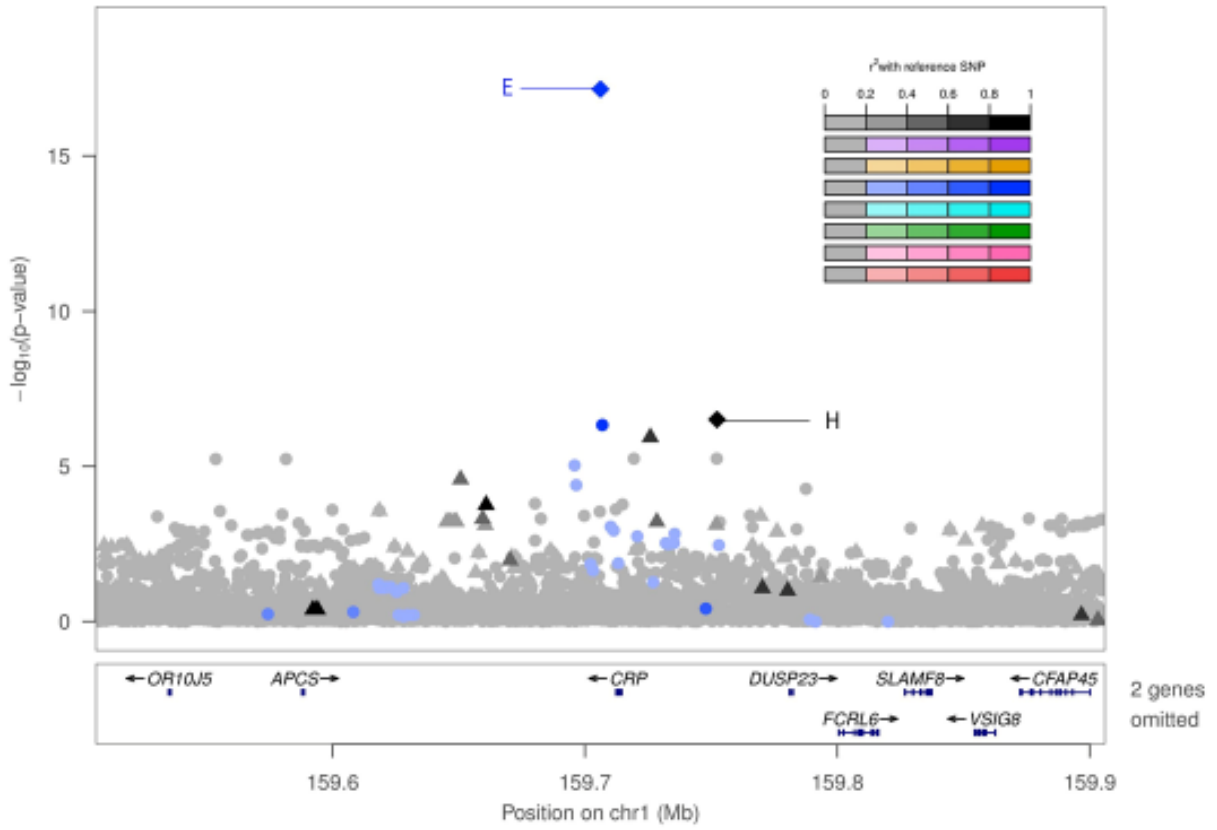
g. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, rs11265259, and rs1800947 (lead variant rs12734907). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis.



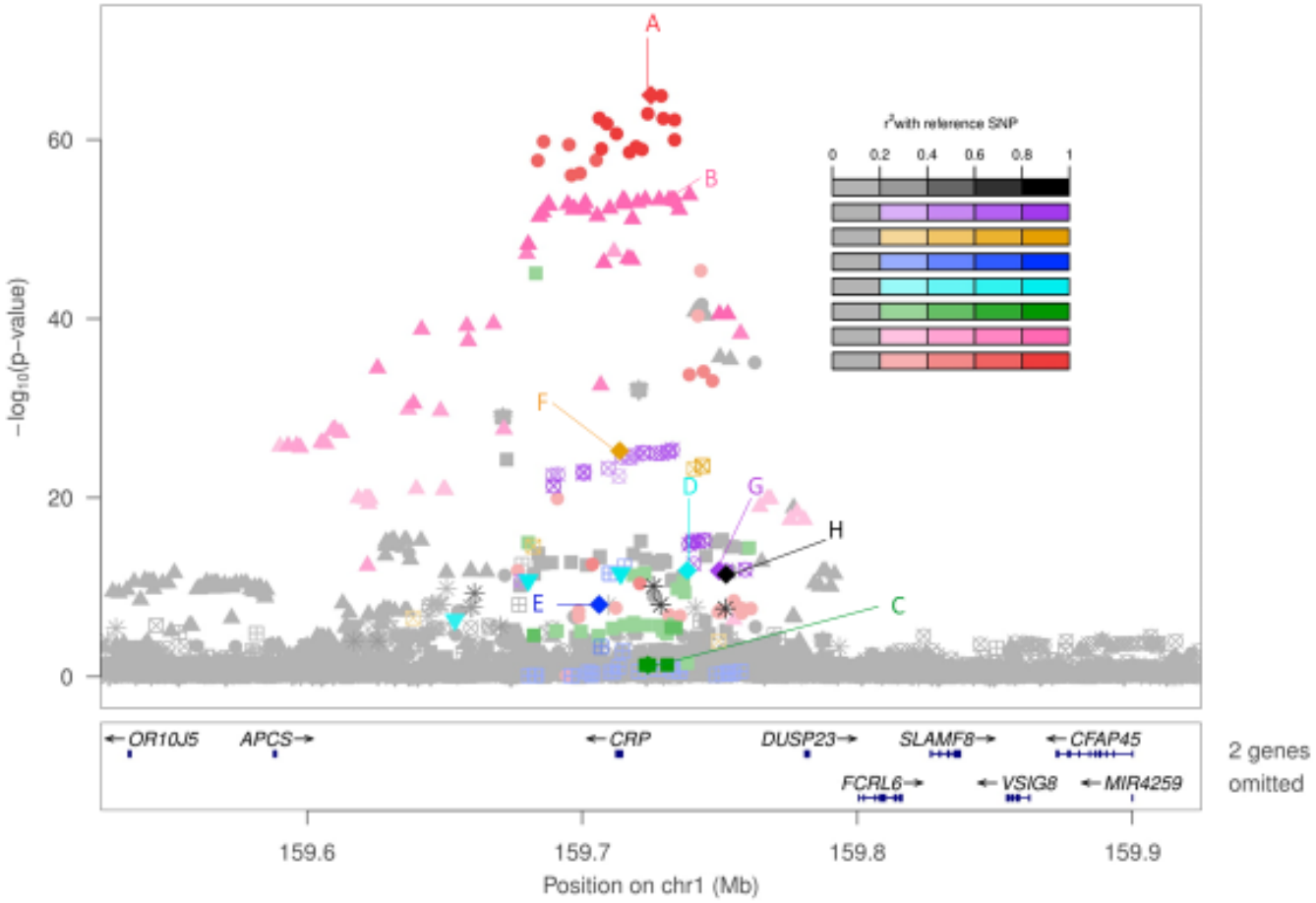
h. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, rs11265259, rs1800947, and rs12734907 (lead variant rs181704186). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



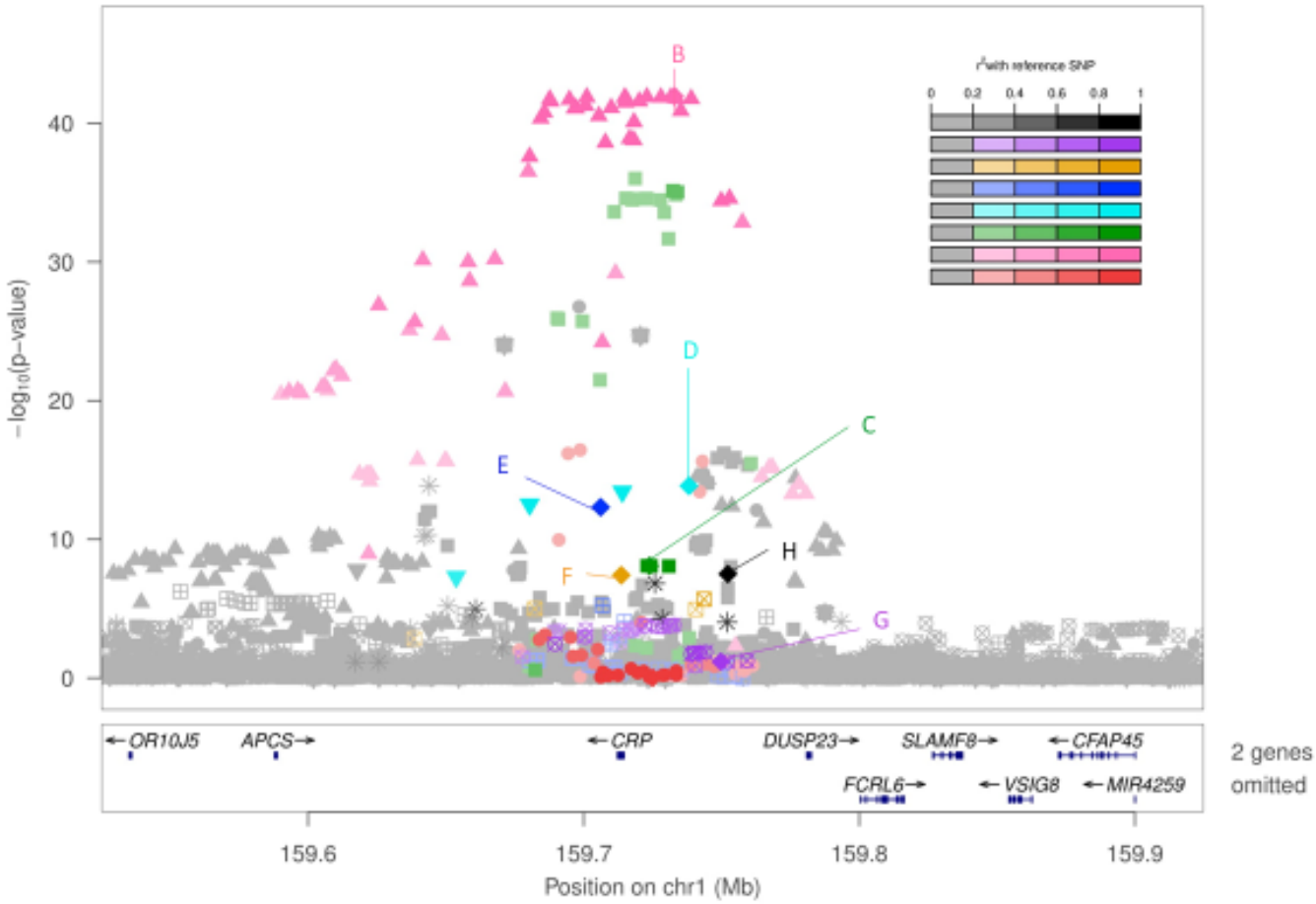
i. Ancestry pooled analysis conditioned on all previously known variants from GWAS and exome sequencing studies. Only signals E and H are labelled, as these are the only signals still reaching our locus-wide significance threshold (as listed in Table 1). Linkage disequilibrium is calculated based on European American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



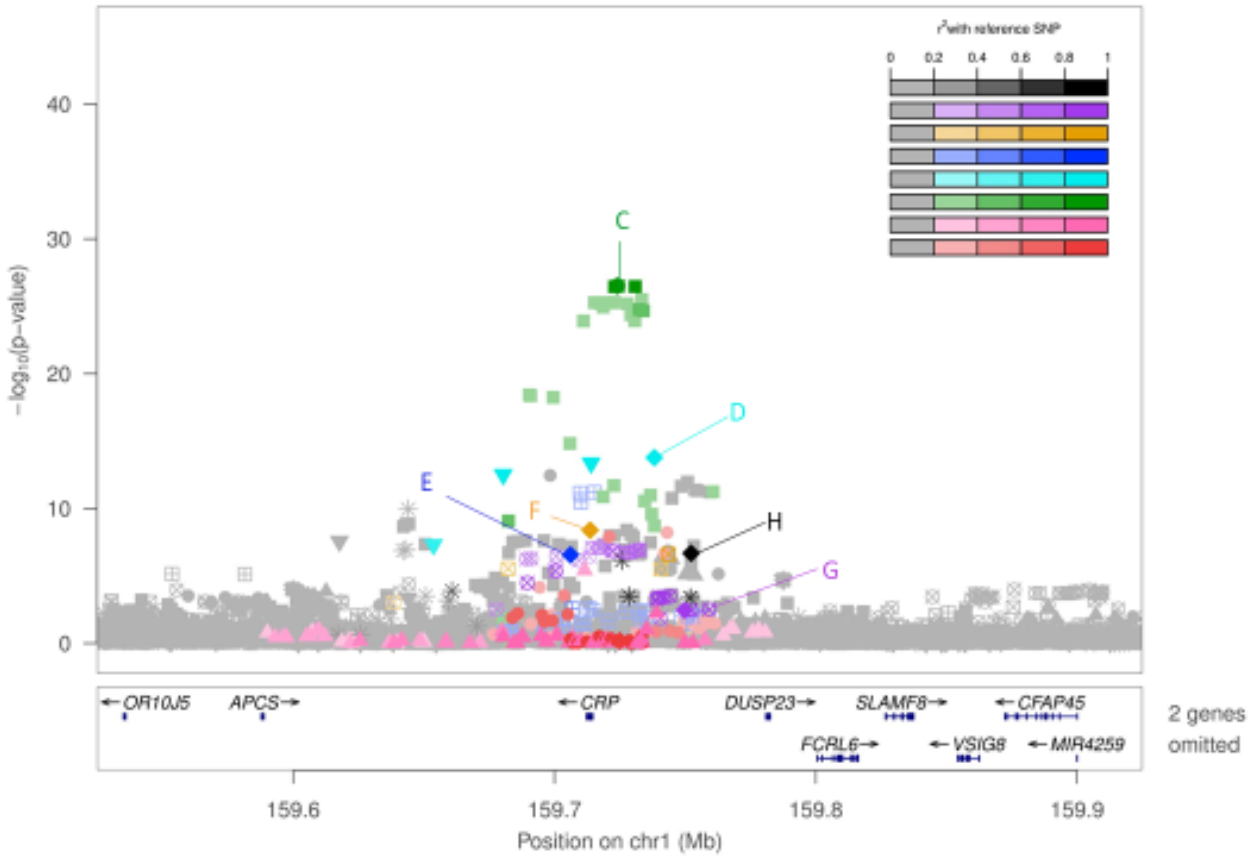
j. In ancestry pooled analysis, LocusZoom plot of association results (lead variant rs7551731). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



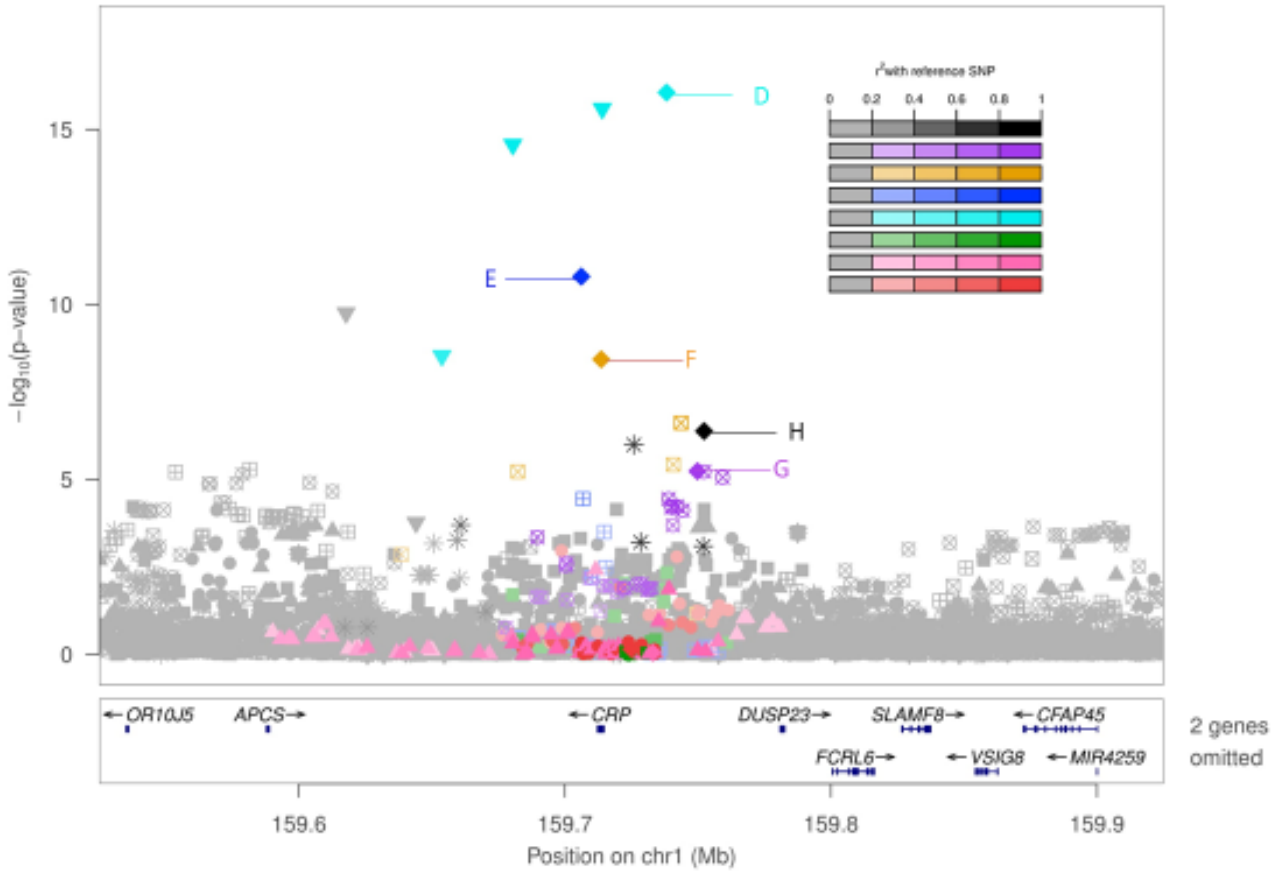
k. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731 (lead variant rs73024795). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



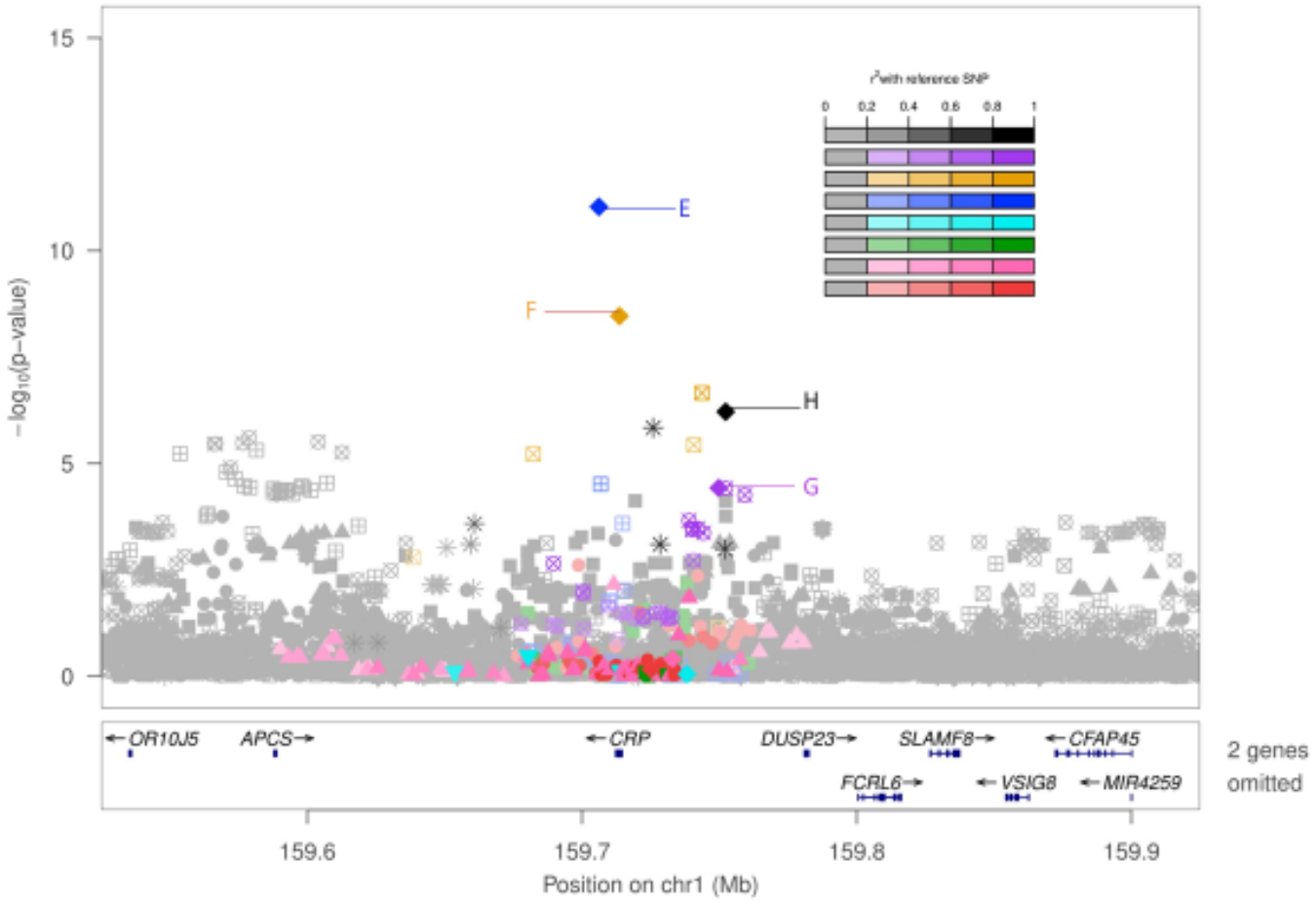
I. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731 and rs73024795 (lead variant rs2211321). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



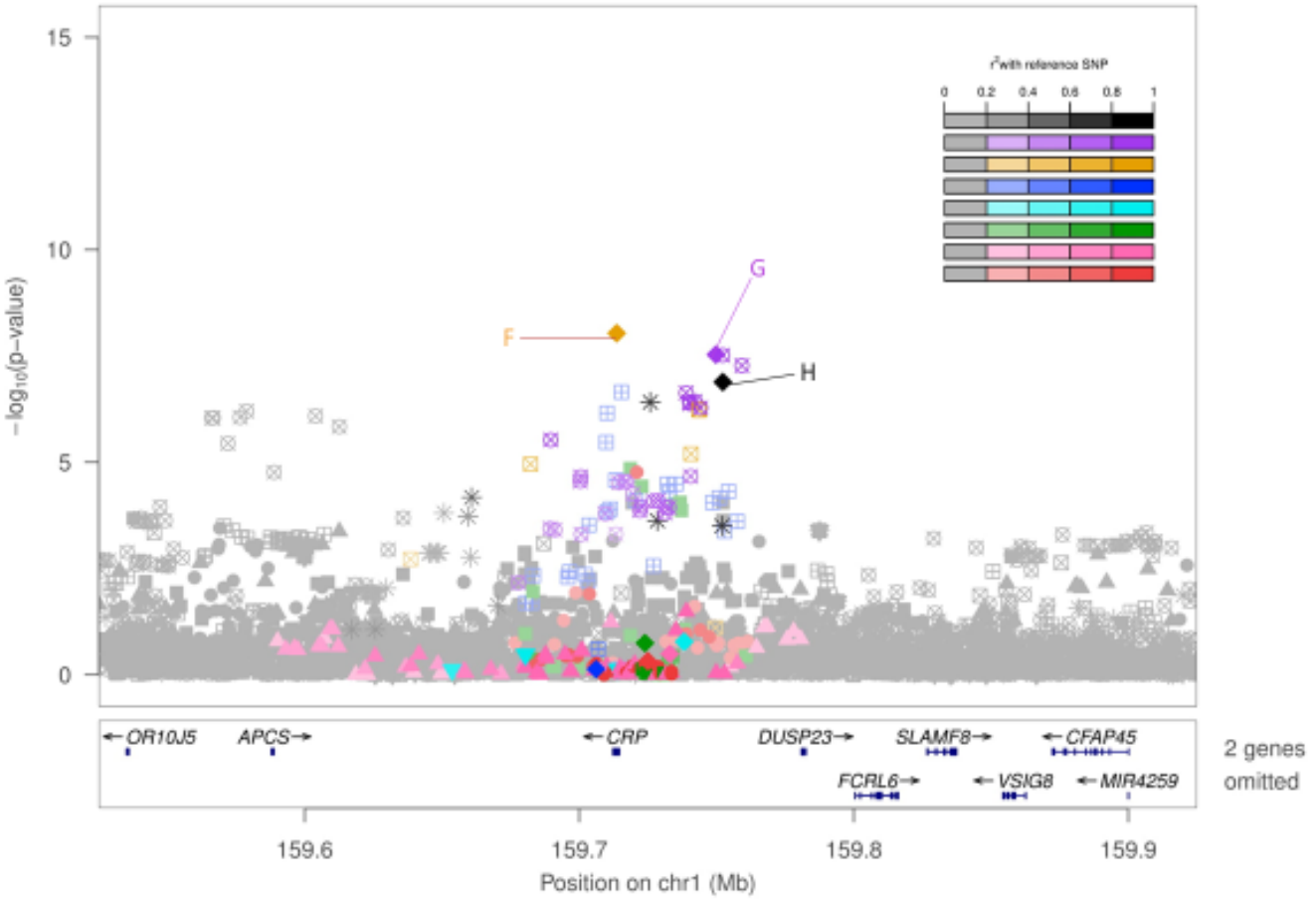
m. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, and rs2211321 (lead variant rs553202904). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



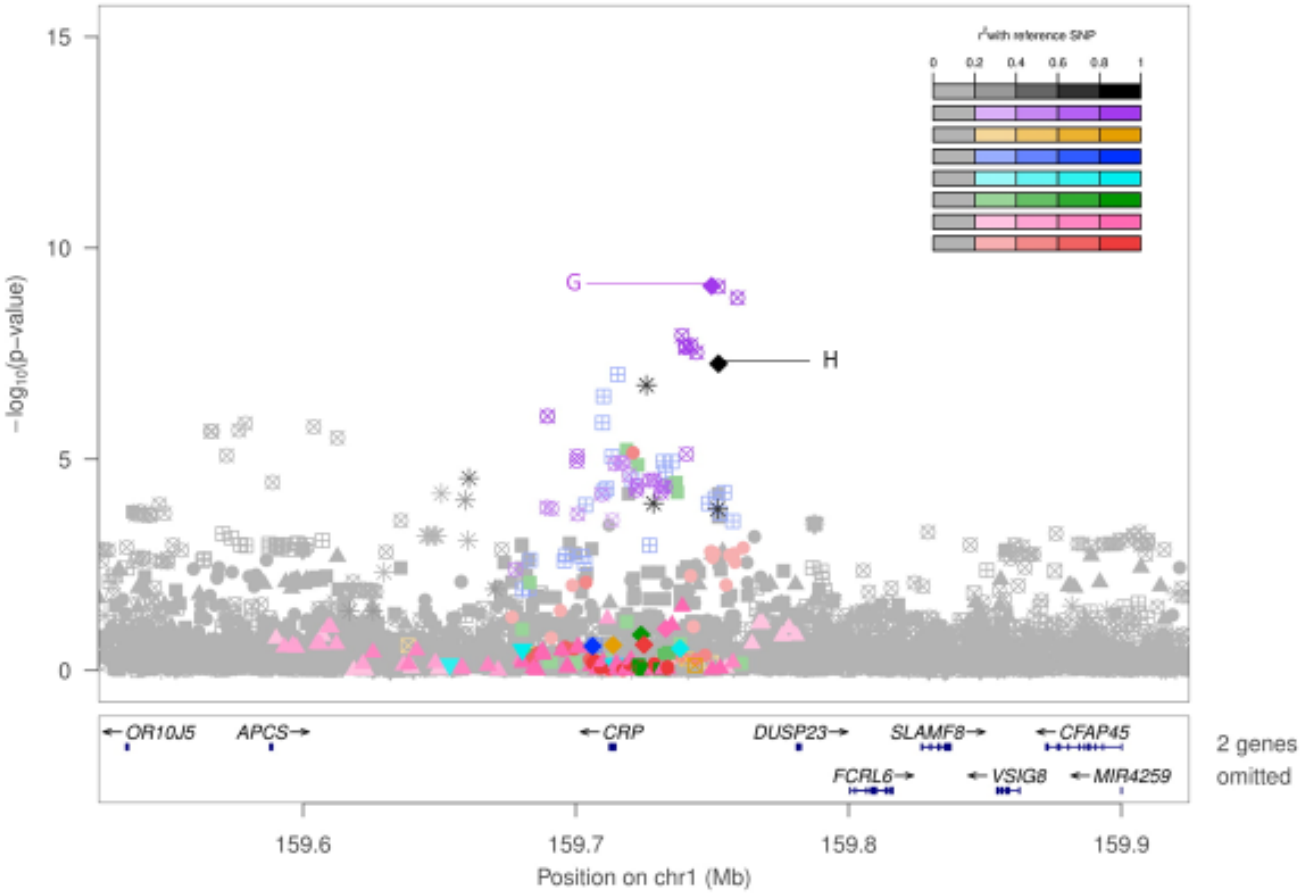
n. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, and rs553202904 (lead variant rs11265259). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



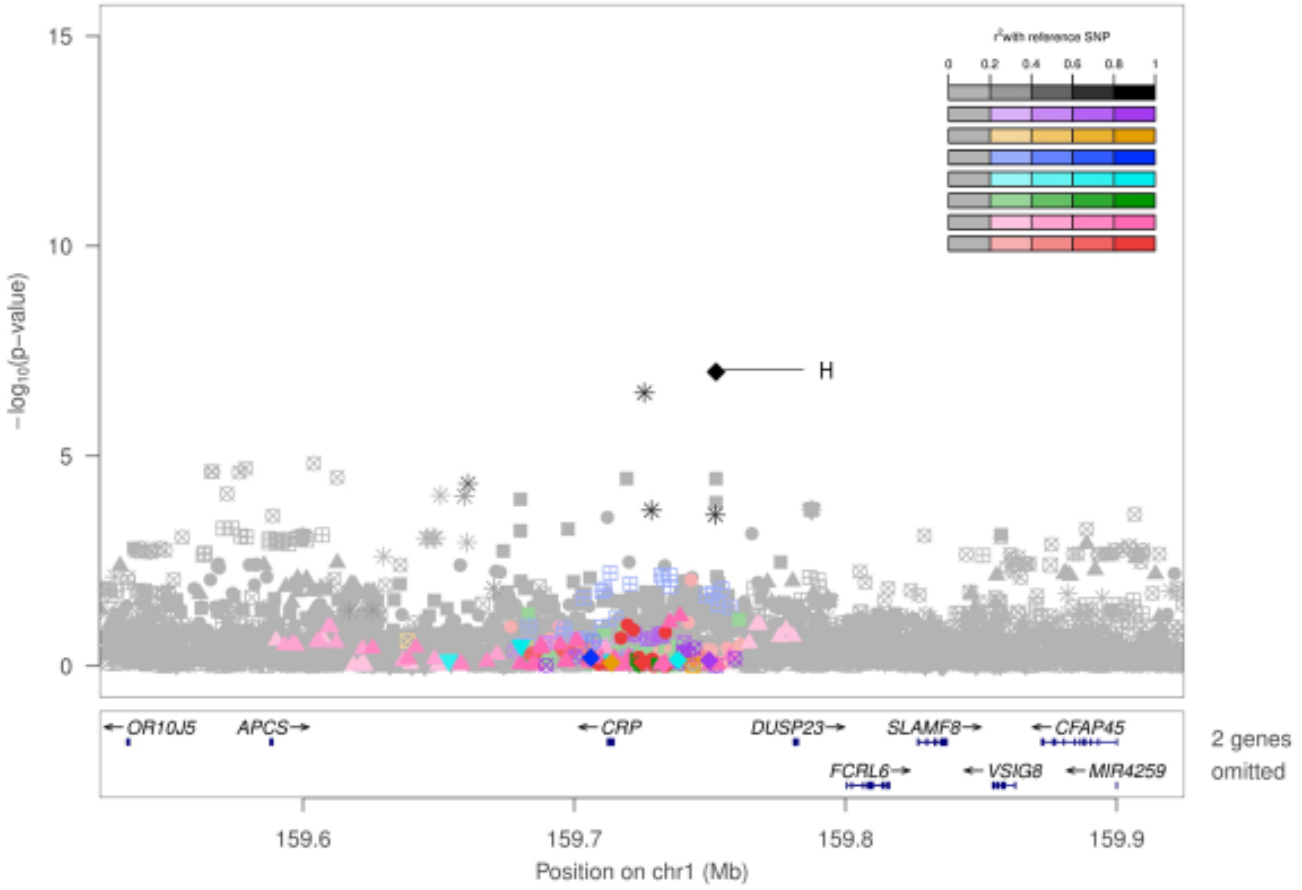
o. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, and rs11265259 (lead variant rs1800947). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



p. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, rs11265259, and rs1800947 (lead variant rs12734907). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis.



q. In ancestry pooled analysis, LocusZoom plot of association results conditioned on rs7551731, rs73024795, rs2211321, rs553202904, rs11265259, rs1800947, and rs12734907 (lead variant rs181704186). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis; association statistics are from pooled analysis.



r. Ancestry pooled analysis conditioned on all previously known variants from GWAS and exome sequencing studies. Only signals E and H are labelled, as these are the only signals still reaching our locus-wide significance threshold (as listed in Table 1). Linkage disequilibrium is calculated based on African American participants in TOPMed CRP analysis; association statistics are from pooled analysis.

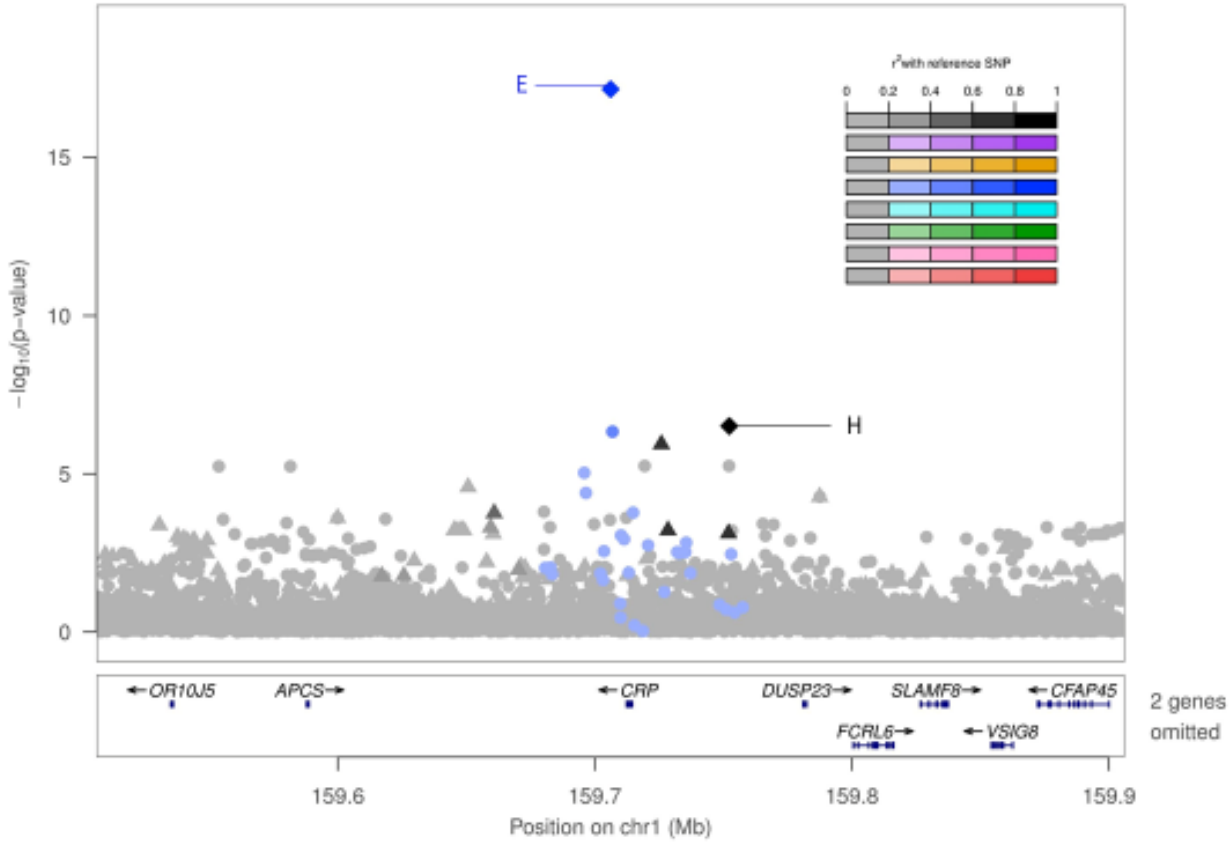


Figure S4: **Functional annotation information for rs11265259.** Genome browser plot for rs11265259, displaying UCSC genes, chromHMM annotation in adult liver (yellow=enhancer, green=weak transcription, red=transcription start site) from RoadMap Epigenomics, H3K4me1 signal from adult liver, 100 vertebrates basewise conservation by PhyloP, and transcription factor ChIP-seq clusters from ENCODE (161 factor version, motifs highlighted in green, proportion cell types detected/ total number of cell types assayed displayed). Unlike in the plot for rs181704186, we did not display GeneHancer due to lack of any relevant signals. No other variants have linkage disequilibrium r2 ≥ 0.8 with lead variant rs11265259.

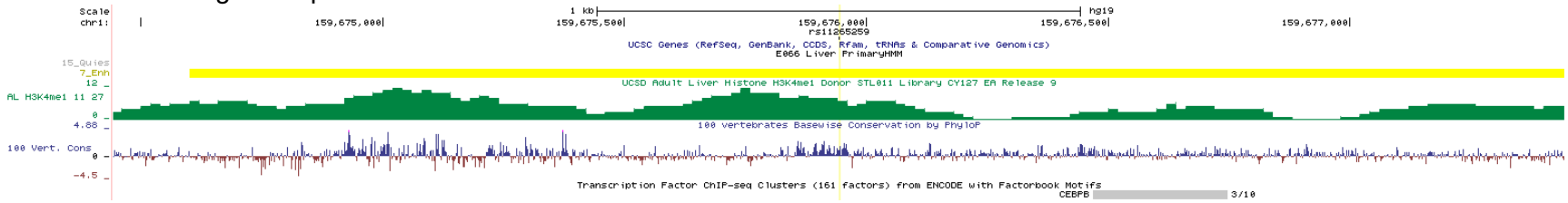
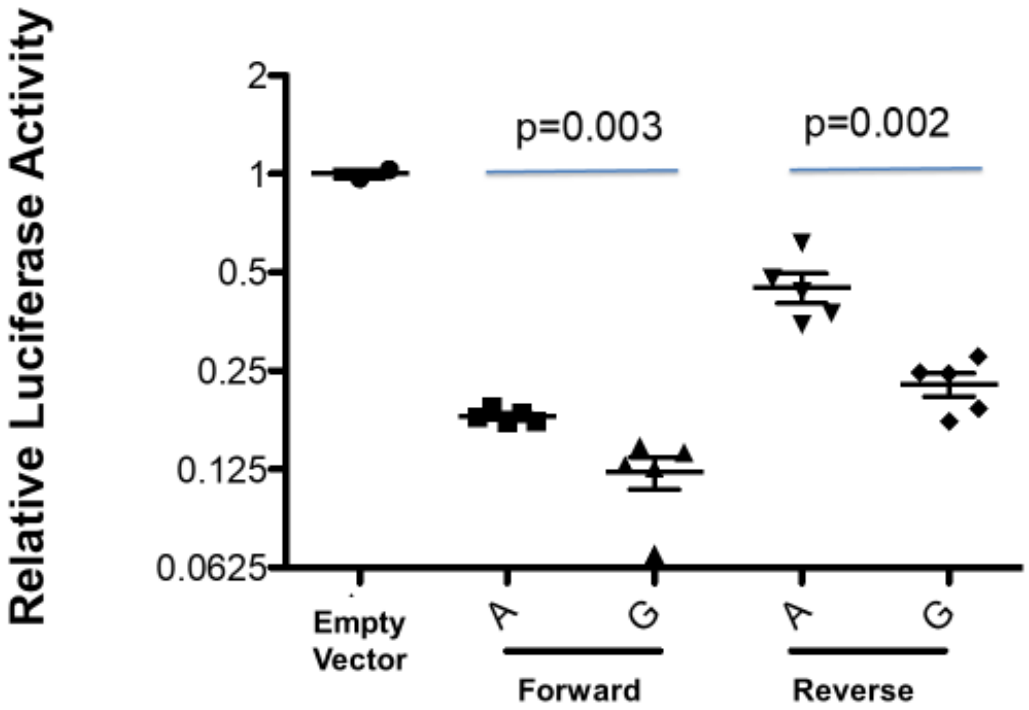
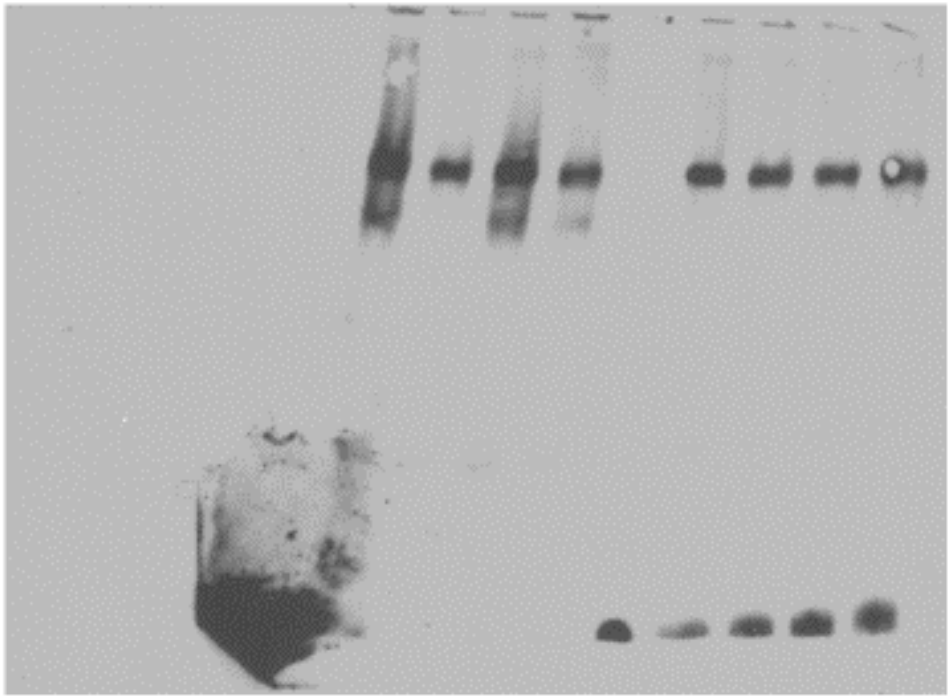


Figure S5: Additional Information for Functional Assays.

a. Additional luciferase assay for rs181704186. Blue lines indicate the groups compared for each listed p-value.

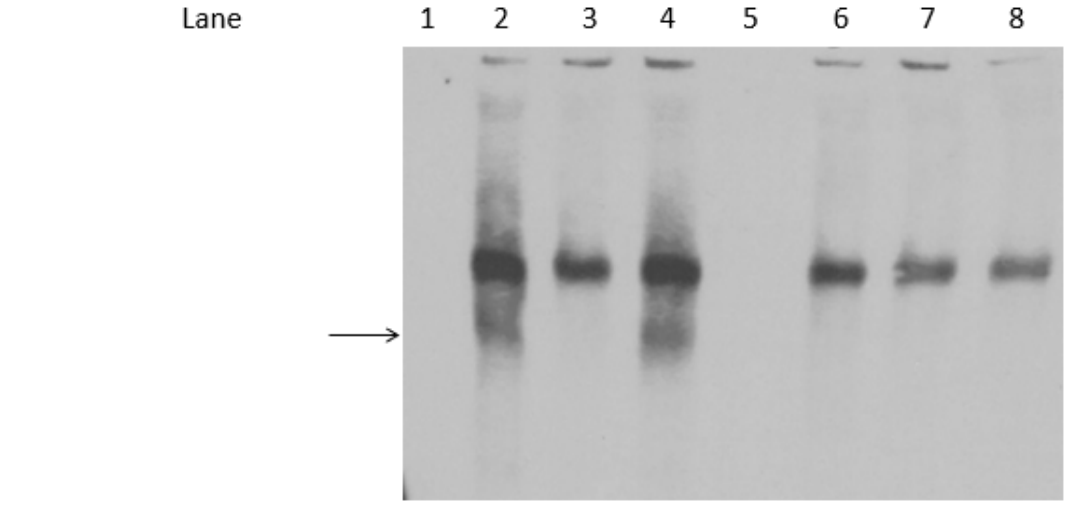


b. Full EMSA from main text Figure 2. NE, nuclear extract.



Lane	1	2	3	4	5	6	7	8	9	10
Labeled probe	A	A	A	A	A	G	G	G	G	G
HepG2 NE	-	+	+	+	+	-	+	+	+	+
Unlabeled competition (40x)	-	-	A	G	-	-	-	A	G	-
CEBPB antibody	-	-	-	-	+	-	-	-	-	+

c. Additional replicate of EMSA for rs181704186. EMSA with biotin-labeled probes containing the A or G allele of rs181704186 shows an allele-specific band (arrow; lane 2 versus 6) that is competed away by 40-fold excess of unlabeled probe containing the A allele (lane 3), but unaffected by a 40-fold excess of probe containing the G allele (lane 4). NE, nuclear extract.



d. Full EMSA from Figure S4c.

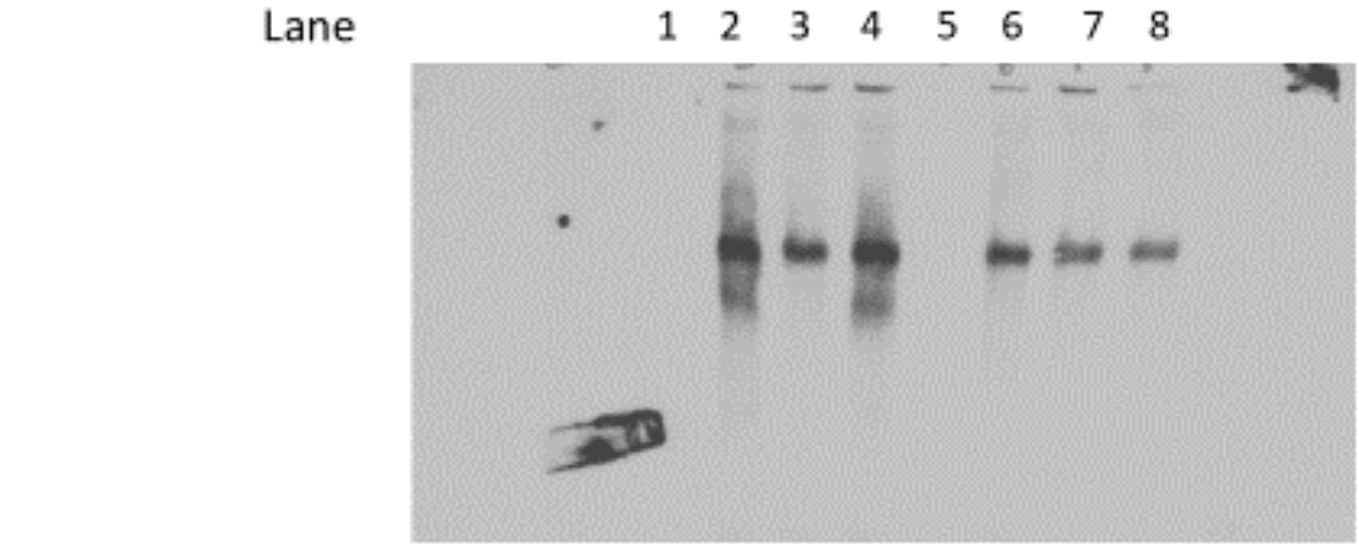


Table S1: Cohort demographic characteristics and C-reactive protein assays.

Study	N	Mean (SD) Age	% female	Mean (SD) C-reactive Protein	Self-Reported Ancestry	Case/control status (if used for sample selection)	Assay
Atherosclerosis Risk in Communities (ARIC)	2,433	63.7 (5.6)	51.1%	4.3 (6.2)	95.9% EA, 4.1% AA	Case 5.5% (VTE or atrial fibrillation), control 94.5%	BNII analyzer (Siemens Healthcare Diagnostics, Deerfield, Illinois)[1]
Cleveland Family Study (CFS)	570	42.3 (18.7)	56.0%	4.4 (6.1)	42.1% EA, 54.9% AA, 3.0% Multiple	NA	Dade Behring BNII nephelometer
Framingham Heart Study (FHS)	3,151	53.0 (14.3)	53.5%	3.5 (5.1)	EA	NA	Dade Behring BN 100 High Sensitivity CRP Agent, Dade Behring CardioPhase hsCRP, Roche cobas c501 CRP High Sensitivity Assay
Genetic Epidemiology of COPD (COPDGene)	504	63.6 (8.7)	51.0%	4.8 (5.8)	EA	NA	Myriad RBM custom multiplex[2]
Genetic Studies of Atherosclerosis Risk (GeneSTAR)	1,525	43.0 (11.9)	60.3%	2.9 (3.1)	55.6% EA, 44.4% AA	NA	Dako and E80C
Jackson Heart Study (JHS)	3,035	55.5 (12.8)	61.9%	5.0 (7.3)	AA	NA	Immunoturbidimetric CRP-Latex assay (Kamiya Biomedical Company, Seattle, WA) using a Hitachi 911 analyzer (Roche Diagnostics, Indianapolis, IN)[3]
Multi-Ethnic Study of Atherosclerosis (MESA) and MESA Family	4,289	61.1 (9.8)	51.5%	3.5 (4.9)	38.7% EA, 27.2% AA, 21.5% HL, 12.6% AS	NA	BNII nephelometer (Dade-Behring)[4]

Old Order Amish	988	49.4 (16.1)	50.1%	2.0 (3.4)	EA	NA	Nephelometry
Women's Health Initiative (WHI)	6,784	67.2 (6.7)	100.0%	4.9 (6.3)	18.5% AA, 78.7% EA, 3.9% HL, 0.7% AS 0.7 AI/AN, 0.2% Other	Case 47.1% (stroke and VTE), control 52.9%	Multiple assays, including BNII Nephlometer, DiaSorin, hs-immunotechnique - Behring analyzer (Denka Seiken; Niigata, Japan), Immulite Analyzer, a Roche Modular P Chemistryanalyzer, SPQ High Sensitivity ReagentHitachi Analyzer

Note that some individuals identify both as of African American or European American ancestry and as Hispanic/Latino.
Abbreviations: EA, European American; AA, African American; HL, Hispanic/Latino; AS, East Asian; AI/AN American Indians/Alaska Natives; VTE, venous thromboembolism; SD, standard deviation; NA, not applicable

Table S2a: Lead variants and number of distinct signals in European American specific association analysis

Locus	Lead Variant	Annotation	P-value	Beta	Effect Allele Frequency	Effect Allele	Number of distinct signals	Sequential Conditioning Lead Variants
LEPR	rs6588153	intronic	6.5E-16	-0.10	0.37	A	1	
IL6R	rs61812598	intronic	1.8E-07	-0.06	0.40	A	1	
CRP	rs2211320	intergenic	7.7E-39	-0.16	0.32	A	4	rs4425982, rs553202904, rs1800947
NLRP3	rs10157379	intronic	1.1E-05	0.05	0.62	T	0	
GCKR	rs1260326	missense, p.Leu446Pro (*GCKR*)	1.2E-12	-0.09	0.58	C	1	
IL1F10	rs28648961	intergenic	1.1E-07	0.07	0.40	A	1	
HNF1A	rs11065384	intronic	1.0E-25	0.13	0.67	C	1	
APOE	rs429358	missense, p.Cys130Arg (*APOE4*)	1.6E-42	-0.24	0.13	C	1	

Analysis was performed at 8 loci (500 kb ± genome-wide significant variants) identified in the pooled ancestry analysis. We used the same locus-wide thresholds as pooled ancestry analysis in Table 1.

Table S2b: Lead variants and number of distinct signals in African American-specific association analysis

Locus	Lead Variant	Annotation	P-value	Beta	Effect Allele Frequency	Effect Allele	Number of distinct signals	Sequential Conditioning Lead Variants
LEPR	rs112200619	intronic	1.9E-05	-0.66	0.003	C	0	
IL6R	rs4129267	intronic	1.2E-06	-0.13	0.14	T	1	
CRP	rs112563958	intergenic	8.6E-43	0.32	0.17	T	5	rs4428887, rs3122014, rs11265259, rs181704186
NLRP3	rs56188865	intronic	1.1E-07	-0.09	0.52	C	1	
GCKR	rs556974380	intergenic	8.0E-04	0.50	0.003	C	0	
IL1F10	rs6734238	intergenic	2.2E-06	0.08	0.45	G	1	
HNF1A	rs1169284	intronic	3.6E-06	-0.10	0.24	C	1	
APOE	rs429358	missense, p.Cys130Arg (*APOE4*)	1.3E-21	-0.21	0.21	C	1	

Analysis was performed at 8 loci (500 kb ± genome-wide significant variants) identified in the pooled ancestry analysis. We used the same locus-wide thresholds as pooled ancestry analysis in Table 1.

Table S3: Results from African American and European American stratified analyses for eight signals detected at *CRP* locus in pooled ancestry results (unconditioned).

Signal	Variant	Beta EA	P-value EA	EAF EA	Beta AA	P-value AA	EAF AA	LD with EA leads	LD with AA leads
A	rs7551731	-0.16	2.5E-38	0.33	-0.21	2.5E-23	0.22	r^2=0.93 with rs2211320, lead signal	r^2=0.89 with rs4428887, second signal
B	rs73024795	NA	NA	0.0005	0.33	2.5E-42	0.16		r^2=0.70 with rs112563958, lead signal
C	rs2211321	0.03	5.2E-02	0.71	-0.11	4.9E-09	0.65	r^2=0.98 with rs4425982, fourth signal	r^2=0.36 with rs3122014, third signal
D	rs553202904	-0.71	8.2E-12	0.0031	NA	NA	0.0004	Third signal	
E	rs11265259	NA	NA	0.0004	-0.17	7.3E-08	0.09		Fourth signal
F	rs1800947	-0.24	1.7E-22	0.06	-0.29	4.6E-04	0.01	Second signal	
G	rs12734907	0.10	5.8E-16	0.34	-0.04	0.27	0.08		
H	rs181704186	NA	NA	0.0001	-0.59	3.9E-10	0.01		Fifth signal

We also list the linkage disequilibrium (if r^2>0.2) between the eight *CRP* locus lead variants from the ancestry pooled analysis with the lead *CRP* locus variants identified in ancestry stratified *CRP* locus conditional analyses Five locus-wide significant variants were identified at the *CRP* locus in African Americans, and four locus-wide significant variants were identified at *CRP* in European Americans. Abbreviations: AA, African American; EA, European American; EAF, effect allele frequency; LD, linkage disequilibrium with ancestry stratified conditional analysis results, from European (EUR) (for EA individuals) or African (AFR) (for AA individuals) 1000 Genomes phase 1; NA, not applicable (did not meet 0.1% minor allele frequency threshold). Effect alleles are listed in Table 2.

Table S4: Previous genome-wide significant variants at the C-reactive protein (*CRP*) locus used for conditional analyses

Variant	Position (Chr 1; GRCh38)	Reference	Included in conditional analysis?	r^2<0.9 with other previously identified variants
rs3027012	159,204,333	[5]	Yes	No
rs56288844	159,330,024	[5]	Yes	No
rs6695494	159,603,761	[5]	Yes	No
rs149520992	159,697,727	[5]	Yes	No
rs72698571	159,701,146	[5]	Yes	No
rs12029262	159,709,406	[5]	Yes	No
rs3091244	159,714,875	[5-7]	No (FAIL variant-adjusted for in sensitivity analysis)	-
rs2246469	159,721,022	[5]	Yes	No
rs141729353	159,734,040	[5]	Yes	Yes-kept, removed LD proxies
rs11265263	159,740,727	[5]	Yes	No
rs4131568	159,752,266	[5]	Yes	No
rs3845624	159,248,476	[8]	Yes	No
rs16842484	159,677,134	[6]	Yes	No
rs12093699	159,678,198	[9]	Yes	No
rs10494326	159,679,910	[10]	Yes	No
rs2592887	159,683,149	[6]	Yes	No
rs726640	159,685,728	[11]	Yes	No
rs2592902	159,685,936	[12]	Yes	Yes-removed
rs876537	159,705,143	[13]	Yes	Yes- kept, removed LD proxies
rs16842559	159,706,381	[14]	Yes	Yes-removed
rs2794520	159,709,026	[15]	Yes	Yes- kept, removed LD proxies
rs1800947	159,713,648	[5; 16]	Yes	No
rs77832441	159,714,024	[16]	Yes	No
rs3093059	159,715,346	[17]	Yes	Yes- kept, removed LD proxies
rs1341665	159,721,769	[8]	Yes	Yes- removed
rs2808634	159,722,783	[10]	Yes	Yes-removed
rs7553007	159,728,759	[10]	Yes	Yes- removed
rs11265260	159,730,249	[18]	Yes	Yes- removed

Previously identified variants were identified through review of the literature (particularly [5; 19]). A previously reported tri-allelic variant and the lead *CRP* locus SNP from a multi-ethnic PAGE fine-mapping effort, rs3091244, failed the variant quality filter in TOPMed. We adjusted for the variant calls that were available in a sensitivity analysis, additionally adjusting for all previously identified *CRP* locus variants, and signals E and H from Table 2 remained unchanged (β= -0.32, p= 7.09 x 10^{-18} for rs11265259, β= -0.47, p= 2.89 x 10^{-7} for rs181704186). This variant is also common across ancestry groups and not in high LD with either signal E or H in African ancestry individuals from TOPMed or 1000 Genomes.

We also performed a conditional analysis adjusting only for previously identified *CRP* locus variants with linkage disequilibrium $r^2<0.9$ (as assessed in AA and EA ancestry samples used in TOPMed CRP analysis) with any other previously identified *CRP* variant to prevent potential problems with collinearity. Both signals E and H from Table 2 were still significant in the pooled analysis (β= -0.29, p= 3.57×10^{-16} for rs11265259, β= -0.47, p= 2.55×10^{-7} for rs181704186), and in African Americans alone (β= -0.28, p= 3.13×10^{-13} for rs11265259, β= -0.49, p= 4.64×10^{-6} for rs181704186).

Abbreviations: LD, linkage disequilibrium.

Table S5: Previous genome-wide significant variants at the HNF1 Homeobox A (*HNF1A*) locus used for conditional analyses.

Variant	Position (Chr 12; GRCh38)	Reference	Included in conditional analysis?	r^2<0.9 with other previously identified variants
rs1039302	120,798,455	[20]	Yes	No
rs2650000	120,951,159	[21]	Yes	No
rs7305618	120,965,129	[13]	Yes	No
rs7953249	120,965,921	[22]	Yes	No
rs7979473	120,982,457	[6; 10]	No-FAIL variant	
rs1183910	120,983,004	[23]	Yes	No
rs2393791	120,986,153	[24]	Yes	Yes-kept, removed LD proxies
rs7310409	120,987,058	[5; 17]	Yes	Yes-removed
rs2259816	120,997,784	[10]	Yes	Yes-kept, removed LD proxies
rs1169310	121,001,630	[18]	Yes	Yes-removed
rs2259883	121,024,336	[5]	Yes	No

Previously identified variants were identified through review of the literature (particularly [5; 19]). One previously reported variant was not available for conditional analysis at the *HNF1A* locus (rs7979473[10]), as it failed variant quality filters. However, since both *HNF1A* signals in our analysis were attenuated to non-significance (post-conditioning lead variant rs544759708, $p= 2.69 \times 10^{-5}$, $\beta= -0.46$) even without adjusting for this variant, we did not pursue further conditional analysis adjusting for fail variants. We also did not condition on secondary signal rs2243616 from [6], as it did not meet a conventional genome-wide significance threshold.

We did perform a conditional analysis adjusting only for previously identified *HNF1A* locus variants with linkage disequilibrium $r^2<0.9$ (as assessed in AA and EA ancestry samples used in TOPMed CRP analysis) with any other previously identified *HNF1A* variant to prevent potential problems with collinearity. Results were unchanged (post-conditioning lead variant rs544759708, $p= 2.69 \times 10^{-5}$, $\beta= -0.46$).

Abbreviations: LD, linkage disequilibrium.

Table S6: Linkage disequilibrium between rs11265259 and rs181704186 and previously reported *CRP* locus variants, as well as other signals from TOPMed conditional analysis, in African Americans from the TOPMed CRP analysis

SNP_A	SNP_B	r^2	D'	Source
1:159706154_T/C_rs11265259	1:159724989_T/C_rs7551731	0.028	1	Signal A, our paper
1:159706154_T/C_rs11265259	1:159732996_C/T_rs73024795	0.018	1	Signal B, our paper
1:159706154_T/C_rs11265259	1:159723932_T/C_rs2211321	0.051	1	Signal C, our paper
1:159706154_T/C_rs11265259	1:159738205_A/G_rs553202904	3.63E-05	1	Signal D, our paper
1:159706154_T/C_rs11265259	1:159706154_T/C_rs11265259	1	1	Signal E, our paper
1:159706154_T/C_rs11265259	1:159713648_C/G_rs1800947	0.001	1	Signal F, our paper
1:159706154_T/C_rs11265259	1:159749804_A/T_rs12734907	0.008	1	Signal G, our paper
1:159706154_T/C_rs11265259	1_159204333_T_C	0.004	1	rs3027012
1:159706154_T/C_rs11265259	1_159248476_A_C	0.013	0.97	rs3845624
1:159706154_T/C_rs11265259	1_159330024_A_G	1.75E-04	1	rs56288844
1:159706154_T/C_rs11265259	1_159603761_T_C	0.007	0.31	rs6695494
1:159706154_T/C_rs11265259	1_159677134_C_T	0.114	0.72	rs16842484
1:159706154_T/C_rs11265259	1_159678198_A_G	0.127	0.77	rs12093699
1:159706154_T/C_rs11265259	1_159679910_T_C	0.020	0.97	rs10494326
1:159706154_T/C_rs11265259	1_159683149_T_C	0.086	0.90	rs2592887
1:159706154_T/C_rs11265259	1_159685728_A_G	0.018	0.98	rs726640
1:159706154_T/C_rs11265259	1_159685936_T_G	0.029	0.96	rs2592902
1:159706154_T/C_rs11265259	1_159697727_T_C	1.01E-06	0.09	rs149520992
1:159706154_T/C_rs11265259	1_159701146_T_C	9.21E-04	0.77	rs72698571
1:159706154_T/C_rs11265259	1_159705143_T_C	0.025	1	rs876537
1:159706154_T/C_rs11265259	1_159706381_C_T	0.003	0.97	rs16842559
1:159706154_T/C_rs11265259	1_159709026_T_C	0.026	1	rs2794520
1:159706154_T/C_rs11265259	1_159709406_C_G	0.004	1	rs12029262
1:159706154_T/C_rs11265259	1_159714024_A_G	3.63E-05	1	rs77832441
1:159706154_T/C_rs11265259	1_159715346_G_A	0.271	1	rs3093059
1:159706154_T/C_rs11265259	1_159721022_A_G	0.051	0.98	rs2246469
1:159706154_T/C_rs11265259	1_159721769_A_G	0.026	1	rs1341665
1:159706154_T/C_rs11265259	1_159722783_T_C	0.017	1	rs2808634
1:159706154_T/C_rs11265259	1_159728759_A_G	0.028	1	rs7553007
1:159706154_T/C_rs11265259	1_159730249_G_A	0.007	1	rs11265260
1:159706154_T/C_rs11265259	1_159734040_C_T	0.021	1	rs141729353
1:159706154_T/C_rs11265259	1_159740727_A_C	0.003	1	rs11265263
1:159706154_T/C_rs11265259	1_159752266_T_C	0.008	1	rs4131568

1:159752293_A/G_rs181704186	1:159724989_T/C_rs7551731	0.027	0.96	Signal A, our paper
1:159752293_A/G_rs181704186	1:159732996_C/T_rs73024795	0.001	0.93	Signal B, our paper
1:159752293_A/G_rs181704186	1:159723932_T/C_rs2211321	0.003	0.85	Signal C, our paper
1:159752293_A/G_rs181704186	1:159738205_A/G_rs553202904	0.002	0.19	Signal D, our paper
1:159752293_A/G_rs181704186	1:159706154_T/C_rs11265259	0.001	1	Signal E, our paper
1:159752293_A/G_rs181704186	1:159713648_C/G_rs1800947	8.74E-06	0.003	Signal F, our paper
1:159752293_A/G_rs181704186	1:159749804_A/T_rs12734907	0.001	1	Signal G, our paper
1:159752293_A/G_rs181704186	1_159204333_T_C	3.49E-04	1	rs3027012
1:159752293_A/G_rs181704186	1_159248476_A_C	0.001	1	rs3845624
1:159752293_A/G_rs181704186	1_159330024_A_G	1.31E-04	0.02	rs56288844
1:159752293_A/G_rs181704186	1_159603761_T_C	0.007	0.78	rs6695494
1:159752293_A/G_rs181704186	1_159677134_C_T	0.003	0.93	rs16842484
1:159752293_A/G_rs181704186	1_159678198_A_G	0.003	0.92	rs12093699
1:159752293_A/G_rs181704186	1_159679910_T_C	0.002	1	rs10494326
1:159752293_A/G_rs181704186	1_159683149_T_C	0.009	0.96	rs2592887
1:159752293_A/G_rs181704186	1_159685728_A_G	0.001	0.86	rs726640
1:159752293_A/G_rs181704186	1_159685936_T_G	0.022	0.92	rs2592902
1:159752293_A/G_rs181704186	1_159697727_T_C	2.41E-04	0.04	rs149520992
1:159752293_A/G_rs181704186	1_159701146_T_C	1.78E-05	0.35	rs72698571
1:159752293_A/G_rs181704186	1_159705143_T_C	0.028	0.92	rs876537
1:159752293_A/G_rs181704186	1_159706381_C_T	9.77E-05	0.02	rs16842559
1:159752293_A/G_rs181704186	1_159709026_T_C	0.028	0.93	rs2794520
1:159752293_A/G_rs181704186	1_159709406_C_G	0.198	0.94	rs12029262
1:159752293_A/G_rs181704186	1_159714024_A_G	0.002	0.19	rs77832441
1:159752293_A/G_rs181704186	1_159715346_G_A	0.003	1	rs3093059
1:159752293_A/G_rs181704186	1_159721022_A_G	0.014	0.94	rs2246469
1:159752293_A/G_rs181704186	1_159721769_A_G	0.029	0.96	rs1341665
1:159752293_A/G_rs181704186	1_159722783_T_C	0.002	1	rs2808634
1:159752293_A/G_rs181704186	1_159728759_A_G	0.027	0.96	rs7553007
1:159752293_A/G_rs181704186	1_159730249_G_A	2.57E-05	0.01	rs11265260
1:159752293_A/G_rs181704186	1_159734040_C_T	0.002	1	rs141729353
1:159752293_A/G_rs181704186	1_159740727_A_C	2.37E-04	1	rs11265263
1:159752293_A/G_rs181704186	1_159752266_T_C	7.39E-04	1	rs4131568

Table S7: Women's Health Initiative (WHI) replication analysis

Variant	Beta	P-value	Effect Allele	Effect Allele Frequency	Beta, Post Conditioning	P-value, Post Conditioning	Effect Allele Frequency, 1000G	Beta, 1000G	P-value, 1000G	Beta, Post Conditioning, 1000G	P-value, Post Conditioning, 1000G
rs11265259	-0.18	6.1×10^{-9}	C	0.08	-0.26	8.7×10^{-12}	0.08	-0.17	1.3×10^{-7}	-0.23	2.1×10^{-9}
rs181704186	-0.58	9.2×10^{-11}	G	0.009	-0.45	9.7×10^{-6}	0.008	-0.62	7.1×10^{-11}	-0.50	2.2×10^{-6}

Conditional analysis was done conditioning on all variants in Table S4. Imputation was performed to TOPMed freeze 5b, using the Michigan Imputation Server. Imputation quality was r^2=0.94 and r^2=0.95 for rs181704186 and r^2=0.92 and r^2=0.90 for rs11265259, in two separately analyzed subsets of WHI Affymetrix 6.0 data (one in participants overlapping the Population Architecture using Genomics and Epidemiology (PAGE) MEGA array study (n= 4685) which would have been included in [19], the rest (n=2423) with Affymetrix data only). Imputation quality for all included variants in the conditional analysis was ≥0.6. For the 1000G phase 3 imputation, imputation quality was r^2=0.83 and r^2=0.86 for rs181704186 and r^2=0.77 and r^2=0.81 for rs11265259. Results from the subsets were meta-analyzed with metal (2011-03-25 version).

Table S8a: FUN-LDA tissue specific annotation scores for *CRP* locus variants (signals E and H from Table 2), with top two tissues as well as any additional tissues with an annotation score >0.9 listed.

Epigenome Dataset	Full Name	Score	SNP	Label
E066	Liver	0.0746	rs11265259	CRP, Signal E
E042	Primary T helper 17 cells PMA-I stimulated	0.00006		
E066	Liver	1	rs181704186	CRP, Signal H
E023	Mesenchymal Stem Cell Derived Adipocyte Cultured Cells	0.99977		
E117	HeLa-S3 Cervical Carcinoma Cell Line	0.99907		
E025	Adipose Derived Mesenchymal Stem Cell Cultured Cells	0.98627		
E030	Primary neutrophils from peripheral blood	0.9787		

Table S8b: Annotation PCs for CRP locus variants (signals E and H from Table 2)

rsID	Epigenetics	Conservation	Protein Function	Negative Selection	Distance to Nearest Coding Variant	Mutation Density	Transcription Factor
rs11265259	6.1	18.8	3.0	2.2	1.1	5.0	9.2
rs181704186	10.0	16.3	3.0	3.3	0.3	6.4	18.0

As further described in the methods, we used a novel, multi-dimensional annotation pipeline to derive annotation PCs from individual functional annotations in the following categories: epigenetics, conservation, protein function, negative selection, distance to coding variants, mutation density, and transcription factor binding. Variant-specific annotation PCs are given as the PHRED-scaled scores from the first principal component of the category's individual annotations.

Table S9a: 95% Credible Set Variants for *CRP* locus in European Americans, derived using FINEMAP.

Variants in Credible Set (4 causal variant setting)	
1:159727120_G/C_rs3116653	1:159685936_G/T_rs2592902
1:159728695_C/T_rs3116651	1:159690923_GA/G_rs60702037
1:159731554_C/T_rs3116655	1:159695286_TAA/T_rs3039321
1:159732697_G/A_rs12727021	1:159696131_C/G_rs2808624
1:159743672_A/G_rs74596724	1:159699194_C/T_rs11265257
1:159744970_G/A_rs4656848	1:159705143_C/T_rs876537
1:159748522_A/G_rs4255379	1:159713301_G/A_rs1130864
1:159750926_G/A_rs4420078	1:159713648_C/G_rs1800947
1:159753330_T/C_rs6677719	1:159716693_G/A_rs3116636
1:159753731_G/A_rs4656849	1:159716703_A/G_rs3116635
1:159754730_T/C_rs11265268	1:159719533_T/C_rs3122012
1:159759547_G/C_rs4433388	1:159722054_C/CT_rs35625772
1:159760689_G/T_rs7418263	1:159723815_G/A_rs2211320

Table S9b: 95% Credible Set Variants for *CRP* locus in African Americans, derived using FINEMAP.

Variants in Credible Set (5 causal variant setting)
1:159718685_C/T_rs2808633
1:159723031_G/A_rs2794518
1:159723932_C/T_rs2211321
1:159730897_G/A_rs10797053
1:159741019_G/C_rs10437340
1:159743435_A/T_rs12083620
1:159758337_T/C_rs11265269

Table S10: Oligonucleotide sequences for functional assays. Forward and reverse oligonucleotides indicate forward or reverse directions (5'-3') with respect to the genome.

PCR primers for reporter assays	**Sequence (5'-3')**	**Region (hg19)**
rs181704186 Forward	TTCATGGGGCAGATGATACA	chr1:159,721,514-159,722,654
rs181704186 Reverse	GGCATGTTGTCTTGCAGGTA	
Oligonucleotide sequences for EMSAs		
rs181704186 Forward	AGTTGCACA/GATGGGAGG	chr1:159,722,075-159,722,091
rs181704186 Reverse	CCTCCCATT/CGTGCAACT	

Supplemental Materials and Methods

Statistical Analysis

Our analysis included 23,279 individuals [average age 59.2 years; 32% male; predominantly European (64.7%) and African American (28.1%) ancestry] from nine cohorts. Inverse-normalized natural log-transformed CRP values were assessed. Models were adjusted for sex, age, study, and self-reported ancestry, as well as ten cross-cohort ancestry principal components calculated from 228,497 LD pruned variants (r^2<0.1 across all freeze 5b individuals) with a minor allele frequency >1%. For each cohort, basic demographic, self-reported ancestry sub-group, and assay information is displayed in Table S1. Individuals with raw CRP levels of zero or residual values more than 3 standard deviations outside the mean were excluded.

We analyzed variants and indels with a minor allele frequency >0.1% (corresponding to a minor allele count >46 in our pooled ancestry sample) using WGS data from freeze 5b (see https://www.nhlbiwgs.org/topmed-whole-genome-sequencing-project-freeze-5b-phases-1-and-2 and preprint at[25] for sequencing and variant calling methods). We used EPACTS (v3.3.3) on the University of Michigan ENCORE server for initial analyses with EMMAX to control for sample relatedness. Stepwise conditional analysis at each identified locus was performed locally using the same EPACTS version. Loci were declared significant at a threshold of $p<1x10^{-9}$ based on estimated number of independent tests for whole genome sequencing data[26]. Within an identified locus (500 kb on each side of any variant with $p<1 \times 10^{-9}$), we defined a Bonferroni corrected p-value threshold based on the number of tested variants. This threshold is conservative given the correlation between variants within a locus, but increases our confidence in the robustness of identified distinct signals. We then performed stepwise conditional analysis to define the number of conditionally distinct signals at each locus, defined as the number of rounds of conditional analysis needed to have no variants within the locus with a p-values lower than the locus threshold. We also performed analyses adjusting for variants previously attaining genome-wide significance at the *CRP* (Table S4) and *HNF1A* loci (Table S5). Conditional analyses at the *CRP* locus were visualized using LocusZoom, using linkage disequilibrium calculated from included TOPMed subjects[27].

We performed statistical fine-mapping with FINEMAP (v1.3.1)[28] using marginal test statistics and TOPMed derived LD reference panels in unrelated EA and AA participants from TOPMed (selected from individuals included in TOPMed CRP analysis using PC-air, n=4,442 AAs, n=11,397 EAs). We chose FINEMAP specifically because it permits a large number of potential causal variants at a locus. At the CRP locus, we input a maximum of 5 causal variants in our sample of AAs and 4 causal variants in EAs, based on the number of conditionally distinct signals from stepwise conditional analysis.

Annotation

For signals E and H, we used the FUN-LDA program (functional effect prediction for noncoding variants, using a latent Dirichlet allocation model) to identify the tissue type in which they were most likely to have a functional effect[29]. This program gives each variant a score ranging from 0 to 1 (with higher scores indicating variants that are most likely to be functional) based on epigenetic assays evidence, with scores available for 127 cell types and tissues in Roadmap Epigenomics. FUN-LDA scores are derived by summing posterior probabilities for active functional classes (such as promoters and enhancers), based on histone modifications and quantitative DNAse hypersensitivity.

We also used a novel, multi-dimensional annotation pipeline, which derives annotation PCs from individual functional annotations in the following categories: epigenetics, conservation, protein function, negative selection, distance to coding variants, mutation

density, and transcription factor binding. Variant-specific annotation PCs are given as the PHRED-scaled scores from the first principal component of the category's individual annotations. Values greater than 10 thus represent variants in the top 10% of a given annotation category. We report the epigenetic PC calculated from individual annotations percent GC within +/- 75bp window, percent CpG within +/- 75bp window, maximum Encode H3K4me1 level, maximum Encode H3K4me2 level, maximum Encode H3K4me3 level, maximum Encode H3K9ac level, maximum Encode H3K9me3 level, maximum Encode H3K27ac level, maximum Encode H3K27me3 level, maximum Encode H3K36me3, maximum Encode H3K79me2 level, maximum Encode H4K20me1 level, maximum Encode EncodeH2AFZ level, ReMap count of binding transcription factors, ReMap count of binding transcription factors for cell line combinations, distance to nearest Transcribed Sequence Start (TSS), and distance to nearest Transcribed Sequence End (TSE). We also report conservation PC calculated from the neutral evolution score of GERP++, rejected Substitution score of GERP++, primate PhastCons conservation score, mammalian PhastCons conservation score, vertebrate PhastCons conservation score, primate PhyloP score, mammalian PhyloP score, and the vertebrate PhyloP score. All annotations are drawn from CADD database[30]. Finally, we assessed whether lead CRP-associated variants were colocalized (r^2>0.8 in EUR or AFR populations; 1000 Genomes phase 3) with lead variants from eQTL signals from GTEx (all tissues[31]) or eQTLGen (whole blood)[32] (https://www.eqtlgen.org/index.html) or a recent large adult liver eQTL analysis[33]. We also examined GeneHancer for enhancer/gene pairings, as determined based on scores for tissue co-expression correlation between genes and enhancer RNAs, enhancer-targeted transcription factor genes, eQTLs for variants within enhancers, and promoter capture Hi-C[34].

Replication

African American individuals with Affymetrix 6.0 data from the WHI[35] SHARe resource (dbGaP phs000386.v7.p3) were imputed using TOPMed freeze 5b as a reference panel. We then performed association analysis for inverse-normalized natural log-transformed CRP in the *CRP* region using the EMMAX test in EPACTS v3.2.6, adjusting for an empirical kinship matrix and age. We also performed an additional analysis adjusting for known *CRP* locus variants from GWAS and exome sequencing analyses.

PheWAS

We additionally performed a follow-up phenome-wide association study (pheWAS) for rs181704186 and rs11265259 in BioVU. BioVU is the biobank of Vanderbilt University Medical Center (VUMC) that houses de-identified DNA samples linked to phenotypic data derived from electronic health records (EHRs) system of VUMC. DNA samples were genotyped with genome-wide arrays including the Multi-Ethnic Global **(**MEGA) array, and the genotype data were imputed into the HRC reference panel using the Michigan imputation server. In total, 1815 phecodes (i.e., groupings of ICD codes into clinically similar diseases or traits) were tested for association in up to 5275 African Americans. Association between each binary phecode and a SNP was assessed using logistic regression, while adjusting for covariates of age, gender, genotyping array type/batch and 10 principal components of ancestry.

Functional Assays

Cell Culture HepG2 human liver carcinoma cells were cultured in MEM-alpha (Corning) supplemented with 10% FBS, 2 mM L-glutamine, and 1 mM sodium pyruvate. Cells were maintained at 37°C in 5% CO_2.

Transcriptional Reporter Assays Oligonucleotide primers (Table S10) containing KpnI and XhoI restriction sites were designed to PCR-amplify a 1,141-bp region (GRCh37/hg19 –chr1:159,721,514 – 159,722,654) surrounding rs181704186. A DNA segment from an individual homozygous for rs181704186-A was amplified, digested with KpnI and XhoI, and ligated into the minimal promoter-containing luciferase reporter vector pGL4.23 (Promega). The constructs were altered to create vectors containing the low-frequency rs181704186-G allele using the QuikChange site-directed mutagenesis kit (Stratagene). Isolated clones were sequenced for genotype and fidelity.

1.3×10^5 HepG2 cells per well were seeded in 24-well plates. Cells were co-transfected using Lipofectamine 3000 (Life Technologies) with five independent pGL4.23 constructs and *Renilla* luciferase vector phRL-TK (Promega) to control for transfection efficiency. 48 hours post-transfection, cells were lysed with Passive Lysis Buffer (Promega) and measured for luciferase activity using the Dual-Luciferase Reporter Assay system (Promega) as directed and previously described[36]. Reporter assays were repeated on a second separate day and yielded comparable results.

Electrophoretic Mobility Shift Assays Nuclear protein was extracted from HepG2 cells using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific). Biotinylated and unlabeled 17-bp oligonucleotide probes (Table S10) were designed centered around rs181704186 and annealed as previously described[36]. Electrophoretic Mobility Shift Assays (EMSAs) were performed using the LightShift Chemiluminescent EMSA Kit (Thermo Fisher Scientific). 20 uL binding reactions containing 10 ug nuclear protein, 400 fmol labeled probe, 1x binding buffer, and 50 ng/uL poly(dI-dC) were incubated at room temperature for 25 minutes. For competition reactions, 40-fold excess of unlabeled probe was incubated in the reaction for 15 minutes prior to addition of the labeled probe, followed by 25 minutes of incubation. Gel electrophoresis and transfer, wash, and detection steps were performed as previously described[37]. EMSAs were carried out on a second separate day and yielded comparable results.

Supplemental Acknowledgements
Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for "NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study" (phs000974) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-12S2). WGS for "NHLBI TOPMed: The Jackson Heart Study" (phs000964) was performed at the University of Washington Northwest Genomics Center (HHSN268201100037C). WGS for "NHLBI TOPMed: Genetic Epidemiology of COPD (COPDGene) in the TOPMed Program" (phs000951) was performed at the Broad Institute (HHSN268201500014C) and the University of Washington Northwest Genomics Center [3R01HL089856-08S, Phase 2 at The Broad Institute]. WGS for "NHLBI TOPMed: Genetics of Cardiometabolic Health in the Amish" (phs000956) was performed at the Broad Institute (3R01HL121007-01S1). WGS for "NHLBI TOPMed: Trans-Omics for Precision Medicine Whole Genome Sequencing Project: ARIC" (phs001211) was performed at Baylor and the Broad Institute [3R01HL092577-06S1 (The Broad Institute, AFGen), HHSN268201500015C (Baylor, VTE), 3U54HG003273-12S2 (Baylor, VTE)]. WGS for "NHLBI TOPMed: GeneSTAR (Genetic Study of Atherosclerosis Risk)" (phs001218) was performed at Macrogen (3R01HL112064-04S1), Illumina (R01HL112064), and the Broad Institute [HHSN268201500014C (The Broad Institute, AA_CAC)].

WGS for “NHLBI TOPMed: MESA and MESA Family AA-CAC” (phs001416) was performed at the Broad Institute [3U54HG003067-13S1 (MESA, TOPMed supplement to NHGRI), HHSN268201500014C (The Broad Institute, AA_CAC)]. WGS for “NHLBI TOPMed: Women's Health Initiative (WHI)” (phs001237) was performed at the Broad Institute (HHSN268201500014C). WGS for “NHLBI TOPMed: The Cleveland Family Study (WGS)” (phs000954) was performed at the University of Washington Northwest Genomics Center (3R01HL098433-05S1). Centralized read mapping and genotype calling along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A listing of WHI investigators can be found at: https://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf.

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I/HHSN26800001) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS.

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. MESA Family is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, and by the National Center for Research Resources, Grant UL1RR033176. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The COPDGene project described was supported by Award Number U01 HL089897, U01 HL089856, R01 HL095432, and R01 HL129937 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The

COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. A full listing of COPDGene investigators can be found at: http://www.copdgene.org/directory

GeneSTAR was supported by the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, HL112064, HL11006, HL118356) and by a grant from the National Institutes of Health/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center.

The Framingham Heart Study inflammation work was supported by HHSN268201500001I, N01-HC 25195, 75N92019D00031, 1RO1 HL64753, R01 HL076784, and 1 R01 AG028321.

The Cleveland Family Study has been supported by National Institutes of Health grants [R01-HL046380, KL2-RR024990, R35-HL135818, HL 046389, HL113338, R35HL135818 and R01-HL113338].

The TOPMed component of the Amish Research Program was supported by NIH grants R01 HL121007, U01 HL072515, and R01 AG18728.

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute.

We used the OASIS (Omics Analysis, Search and Information System) tool for initial exploration of our TOPMed association results for C-reactive protein. OASIS is a web-based platform for mining and visualizing omics association analysis results (and thus enabling the transformation of massive volumes of “results data” into a more complete understanding of biology). OASIS resources, video library and contact information available from https://edn.som.umaryland.edu/OASIS/.

Supplementary References

1. Whelton, S.P., Roy, P., Astor, B.C., Zhang, L., Hoogeveen, R.C., Ballantyne, C.M., and Coresh, J. (2013). Elevated high-sensitivity C-reactive protein as a risk marker of the attenuated relationship between serum cholesterol and cardiovascular events at older age. The ARIC Study. American Journal of Epidemiology 178, 1076-1084.

2. Sun, W., Kechris, K., Jacobson, S., Drummond, M.B., Hawkins, G.A., Yang, J., Chen, T.-h., Quibrera, P.M., Anderson, W., Barr, R.G., et al. (2016). Common Genetic Polymorphisms Influence Blood Biomarker Measurements in COPD. PLoS genetics 12, e1006011.
3. Fox, E.R., Benjamin, E.J., Sarpong, D.F., Rotimi, C.N., Wilson, J.G., Steffes, M.W., Chen, G., Adeyemo, A., Taylor, J.K., Samdarshi, T.E., et al. (2008). Epidemiology, heritability, and genetic linkage of C-reactive protein in African Americans (from the Jackson Heart Study). Am J Cardiol 102, 835-841.
4. Mann, D.M., Shimbo, D., Cushman, M., Lakoski, S., Greenland, P., Blumenthal, R.S., Michos, E.D., Lloyd-Jones, D.M., and Muntner, P. (2013). C-reactive protein level and the incidence of eligibility for statin therapy: the multi-ethnic study of atherosclerosis. Clinical cardiology 36, 15-20.
5. Ligthart, S., Vaez, A., Võsa, U., Stathopoulou, M.G., de Vries, P.S., Prins, B.P., Van der Most, P.J., Tanaka, T., Naderi, E., Rose, L.M., et al. (2018). Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. The American Journal of Human Genetics 103, 691-706.
6. Kocarnik, J.M., Richard, M., Graff, M., Haessler, J., Bien, S., Carlson, C., Carty, C.L., Reiner, A.P., Avery, C.L., Ballantyne, C.M., et al. (2018). Discovery, fine-mapping, and conditional analyses of genetic variants associated with C-reactive protein in multiethnic populations using the Metabochip in the Population Architecture using Genomics and Epidemiology (PAGE) study. Human molecular genetics 27, 2940-2953.
7. Ridker, P., Pare, G., Parker, A., Zee, R., Danik, J., Buring, J., Kwiatkowski, D., Cook, N., Miletich, J., and Chasman, D. (2008). Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. American journal of human genetics 82, 1185-1192.
8. Naitza, S., Porcu, E., Steri, M., Taub, D.D., Mulas, A., Xiao, X., Strait, J., Dei, M., Lai, S., Busonero, F., et al. (2012). A genome-wide association scan on the levels of markers of inflammation in Sardinians reveals associations that underpin its complex regulation. PLoS genetics 8, e1002480.
9. Melzer, D., Perry, J.R., Hernandez, D., Corsi, A.M., Stevens, K., Rafferty, I., Lauretani, F., Murray, A., Gibbs, J.R., Paolisso, G., et al. (2008). A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS genetics 4, e1000072.
10. Reiner A , P., Beleza, S., Franceschini, N., Auer P , L., Robinson J , G., Kooperberg, C., Peters, U., and Tang, H. (2012). Genome-wide Association and Population Genetic Analysis of C-Reactive Protein in African American and Hispanic American Women. Am J Hum Genet 91, 502-512.
11. Doumatey, A.P., Chen, G., Tekola Ayele, F., Zhou, J., Erdos, M., Shriner, D., Huang, H., Adeleye, J., Balogun, W., Fasanmade, O., et al. (2012). C-reactive protein (CRP) promoter polymorphisms influence circulating CRP levels in a genome-wide association study of African Americans. Hum Mol Genet 21, 3063-3072.
12. Williams, S.R., Hsu, F.C., Keene, K.L., Chen, W.M., Nelson, S., Southerland, A.M., Madden, E.B., Coull, B., Gogarten, S.M., Furie, K.L., et al. (2016). Shared genetic susceptibility of vascular-related biomarkers with ischemic and recurrent stroke. Neurology 86, 351-359.
13. Wu, Y., McDade, T.W., Kuzawa, C.W., Borja, J., Li, Y., Adair, L.S., Mohlke, K.L., and Lange, L.A. (2012). Genome-wide association with C-reactive protein levels in CLHNS: evidence for the CRP and HNF1A loci and their interaction with exposure to a pathogenic environment. Inflammation 35, 574-583.

14. Dorajoo, R., Li, R., Ikram, M.K., Liu, J., Froguel, P., Lee, J., Sim, X., Ong, R.T., Tay, W.T., Peng, C., et al. (2013). Are C-reactive protein associated genetic variants associated with serum levels and retinal markers of microvascular pathology in Asian populations from Singapore? PloS one 8, e67650.
15. Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P., et al. (2011). Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation 123, 731-738.
16. Schick, U.M., Auer, P.L., Bis, J.C., Lin, H., Wei, P., Pankratz, N., Lange, L.A., Brody, J., Stitziel, N.O., Kim, D.S., et al. (2015). Association of exome sequences with plasma C-reactive protein levels in >9000 participants. Hum Mol Genet 24, 559-571.
17. Okada, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Kamatani, Y., Hosono, N., Tsunoda, T., Matsuda, K., Tanaka, T., Kubo, M., et al. (2011). Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. Hum Mol Genet 20, 1224-1231.
18. Reiner, A.P., Barber, M.J., Guan, Y., Ridker, P.M., Lange, L.A., Chasman, D.I., Walston, J.D., Cooper, G.M., Jenny, N.S., Rieder, M.J., et al. (2008). Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. Am J Hum Genet 82, 1193-1201.
19. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature 570, 514-518.
20. Lowe, J.K., Maller, J.B., Pe'er, I., Neale, B.M., Salit, J., Kenny, E.E., Shea, J.L., Burkhardt, R., Smith, J.G., Ji, W., et al. (2009). Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. PLoS genetics 5, e1000365.
21. Kocarnik, J.M., Pendergrass, S.A., Carty, C.L., Pankow, J.S., Schumacher, F.R., Cheng, I., Durda, P., Ambite, J.L., Deelman, E., Cook, N.R., et al. (2014). Multiancestral analysis of inflammation-related genetic variants and C-reactive protein in the population architecture using genomics and epidemiology study. Circulation Cardiovascular genetics 7, 178-188.
22. Kim, D.K., Cho, M.H., Hersh, C.P., Lomas, D.A., Miller, B.E., Kong, X., Bakke, P., Gulsvik, A., Agusti, A., Wouters, E., et al. (2012). Genome-wide association analysis of blood biomarkers in chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine 186, 1238-1247.
23. Elliott, P., Chambers, J.C., Zhang, W., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund, P., Engert, J.C., Bennett, D., et al. (2009). Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. Jama 302, 37-48.
24. Vinayagamoorthy, N., Hu, H.J., Yim, S.H., Jung, S.H., Jo, J., Jee, S.H., and Chung, Y.J. (2014). New variants including ARG1 polymorphisms associated with C-reactive protein levels identified by genome-wide association and pathway analysis. PloS one 9, e95866.
25. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. bioRxiv, 563866.
26. Pulit, S.L., de With, S.A., and de Bakker, P.I. (2017). Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. Genetic epidemiology 41, 145-151.
27. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336-2337.

28. Benner, C., Spencer, C.C., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32, 1493-1501.
29. Backenroth, D., He, Z., Kiryluk, K., Boeva, V., Pethukova, L., Khurana, E., Christiano, A., Buxbaum, J.D., and Ionita-Laza, I. (2018). FUN-LDA: A Latent Dirichlet Allocation Model for Predicting Tissue-Specific Functional Effects of Noncoding Variation: Methods and Applications. American journal of human genetics 102, 920-942.
30. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic acids research 47, D886-d894.
31. Gamazon, E.R., Segre, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., et al. (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat Genet 50, 956-967.
32. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S., et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv, 447367.
33. Strunz, T., Grassmann, F., Gayan, J., Nahkuri, S., Souza-Costa, D., Maugeais, C., Fauser, S., Nogoceke, E., and Weber, B.H.F. (2018). A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver. Scientific reports 8, 5865.
34. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database : the journal of biological databases and curation 2017.
35. The Women's Health Initiative Study Group. (1998). Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials 19, 61-109.
36. Fogarty, M.P., Cannon, M.E., Vadlamudi, S., Gaulton, K.J., and Mohlke, K.L. (2014). Identification of a Regulatory Variant That Binds FOXA1 and FOXA2 at the CDC123/CAMK1D Type 2 Diabetes GWAS Locus. PLoS genetics 10, e1004633.
37. Roman, T.S., Cannon, M.E., Vadlamudi, S., Buchkovich, M.L., Wolford, B.N., Welch, R.P., Morken, M.A., Kwon, G.J., Varshney, A., Kursawe, R., et al. (2017). A Type 2 Diabetes-Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer at the ADCY5 Locus. Diabetes 66, 2521-2530.

TOPMed Banner Authors

Name	Institution(s)	Primary Department	Institution City	Institution State	Zip Code	Country
Abe, Namiko	New York Genome Center		New York	New York	10013	US
Abecasis, Gonçalo	University of Michigan		Ann Arbor	Michigan	48109	US
Albert, Christine	Massachusetts General Hospital		Boston	Massachusetts	02114	US

Almasy, Laura	Children's Hospital of Philadelphia, University of Pennsylvania		Philadelphia	Pennsylvania	19104	US
Alonso, Alvaro	Emory University		Atlanta	Georgia	30322	US
Ament, Seth	University of Maryland		Baltimore	Maryland	21201	US
Anderson, Peter	University of Washington		Seattle	Washington	98195	US
Anugu, Pramod	University of Mississippi		Jackson	Mississippi	38677	US
Applebaum-Bowden, Deborah	National Institutes of Health		Bethesda	Maryland	20892	US
Arking, Dan	Johns Hopkins University		Baltimore	Maryland	21218	US
Arnett, Donna K	University of Kentucky		Lexington	Kentucky	40506	US
Ashley-Koch, Allison	Duke University		Durham	North Carolina	27708	US
Aslibekyan, Stella	University of Alabama		Birmingham	Alabama	35487	US
Assimes, Tim	Stanford University		Stanford	California	94305	US
Auer, Paul	University of Wisconsin Milwaukee		Milwaukee	Wisconsin	53211	US
Avramopoulos, Dimitrios	Johns Hopkins University		Baltimore	Maryland	21218	US
Barnard, John	Cleveland Clinic		Cleveland	Ohio	44195	US
Barnes, Kathleen	University of Colorado at Denver		Denver	Colorado	80204	US
Barr, R. Graham	Columbia University		New York	New York	10027	US
Barron-Casella, Emily	Johns Hopkins University		Baltimore	Maryland	21218	US
Beaty, Terri	Johns Hopkins University		Baltimore	Maryland	21218	US
Becker, Diane	Johns Hopkins University		Baltimore	Maryland	21218	US
Becker, Lewis	Johns Hopkins University		Baltimore	Maryland	21218	US

Beer, Rebecca	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US
Begum, Ferdouse	Johns Hopkins University		Baltimore	Maryland	21218	US
Beitelshees, Amber	University of Maryland		Baltimore	Maryland	21201	US
Benjamin, Emelia	Boston University, Massachusetts General Hospital	Boston University School of Medicine	Boston	Massachusetts	02118	US
Bezerra, Marcos	Fundação de Hematologia e Hemoterapia de Pernambuco - Hemope		Recife		52011-000	BR
Bielak, Larry	University of Michigan		Ann Arbor	Michigan	48109	US
Bis, Joshua	University of Washington		Seattle	Washington	98195	US
Blackwell, Thomas	University of Michigan		Ann Arbor	Michigan	48109	US
Blangero, John	University of Texas Rio Grande Valley School of Medicine	Human Genetics	Brownsville	Texas	78520	US
Boerwinkle, Eric	University of Texas Health at Houston		Houston	Texas	77225	US
Bowden, Donald W.	Wake Forest Baptist Health	Department of Biochemistry	Winston-Salem	North Carolina	27157	US
Bowler, Russell	National Jewish Health	National Jewish Health	Denver	Colorado	80206	US
Brody, Jennifer	University of Washington		Seattle	Washington	98195	US
Broeckel, Ulrich	Medical College of Wisconsin		Milwaukee	Wisconsin	53226	US
Broome, Jai	University of Washington		Seattle	Washington	98195	US

Bunting, Karen	New York Genome Center		New York	New York	10013	US
Burchard, Esteban	University of California, San Francisco		San Francisco	California	94143	US
Buth, Erin	University of Washington	Biostatistics	Seattle	Washington	98195	US
Cade, Brian	Brigham & Women's Hospital	Brigham and Women's Hospital	Boston	Massachusetts	02115	US
Cardwell, Jonathan	University of Colorado at Denver		Denver	Colorado	80204	US
Carty, Cara	Women's Health Initiative		Seattle	Washington	98109	US
Casaburi, Richard	University of California, Los Angeles		Los Angeles	California	90095	US
Casella, James	Johns Hopkins University		Baltimore	Maryland	21218	US
Chaffin, Mark	Broad Institute		Cambridge	Massachusetts	02142	US
Chang, Christy	University of Maryland		Baltimore	Maryland	21201	US
Chasman, Daniel	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Chavan, Sameer	University of Colorado at Denver		Denver	Colorado	80204	US
Chen, Bo-Juen	New York Genome Center		New York	New York	10013	US
Chen, Wei-Min	University of Virginia		Charlottesville	Virginia	22903	US
Chen, Yii-Der Ida	Los Angeles Biomedical Research Institute		Charlottesville	Virginia	90502	US
Cho, Michael	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Choi, Seung Hoan	Broad Institute		Cambridge	Massachusetts	02142	US
Chuang, Lee-Ming	National Taiwan University	National Taiwan University Hospital	Taipei		10617	TW

Chung, Mina	Cleveland Clinic		Cleveland	Ohio	44195	US
Conomos, Matthew	University of Washington	Biostatistics	Seattle	Washington	98115	US
Cornell, Elaine	University of Vermont		Burlington	Vermont	05405	US
Correa, Adolfo	University of Mississippi	Medicine	Jackson	Mississippi	39216	US
Crandall, Carolyn	University of California, Los Angeles		Los Angeles	California	90095	US
Crapo, James	National Jewish Health		Denver	Colorado	80206	US
Cupples, L. Adrienne	Boston University		Boston	Massachusetts	02215	US
Curran, Joanne	University of Texas Rio Grande Valley School of Medicine		Brownsville	Texas	78520	US
Curtis, Jeffrey	University of Michigan		Ann Arbor	Michigan	48109	US
Custer, Brian	Vitalant Research Institute		San Francisco	California	94118	US
Damcott, Coleen	University of Maryland		Baltimore	Maryland	21201	US
Darbar, Dawood	University of Illinois at Chicago		Chicago	Illinois	60607	US
Das, Sayantan	University of Michigan		Ann Arbor	Michigan	48109	US
David, Sean	Stanford University		Stanford	California	94305	US
Davis, Colleen	University of Washington		Seattle	Washington	98195	US
Daya, Michelle	University of Colorado at Denver		Denver	Colorado	80204	US
de Andrade, Mariza	Mayo Clinic		Rochester	Minnesota	55905	US
DeBaun, Michael	Vanderbilt University		Nashville	Tennessee	37235	US
Deka, Ranjan	University of Cincinnati		Cincinnati	Ohio	45220	US
DeMeo, Dawn	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Devine, Scott	University of Maryland		Baltimore	Maryland	21201	US

Do, Ron	Icahn School of Medicine at Mount Sinai		New York	New York	10029	US
Duan, Qing	University of North Carolina		Chapel Hill	North Carolina	27599	US
Duggirala, Ravi	University of Texas Rio Grande Valley School of Medicine		Edinburg	Texas	78539	US
Durda, Jon Peter	University of Vermont		Burlington	Vermont	05405	US
Dutcher, Susan	Washington University in St Louis		St Louis	Missouri	63130	US
Eaton, Charles	Brown University		Providence	Rhode Island	02912	US
Ekunwe, Lynette	University of Mississippi		Jackson	Mississippi	38677	US
Ellinor, Patrick	Massachusetts General Hospital		Boston	Massachusetts	02114	US
Emery, Leslie	University of Washington		Seattle	Washington	98195	US
Farber, Charles	University of Virginia		Charlottesville	Virginia	22903	US
Farnam, Leanna	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Fingerlin, Tasha	National Jewish Health	Center for Genes, Environment and Health	Denver	Colorado	80206	US
Flickinger, Matthew	University of Michigan		Ann Arbor	Michigan	48109	US
Fornage, Myriam	University of Texas Health at Houston		Houston	Texas	77225	US
Franceschini, Nora	University of North Carolina		Chapel Hill	North Carolina	27599	US
Fu, Mao	University of Maryland		Baltimore	Maryland	21201	US
Fullerton, Stephanie M.	University of Washington		Seattle	Washington	98195	US
Fulton, Lucinda	Washington University in St Louis		St Louis	Missouri	63130	US
Gabriel, Stacey	Broad Institute		Cambridge	Massachusetts	02142	US

Gan, Weiniu	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US
Gao, Yan	University of Mississippi		Jackson	Mississippi	38677	US
Gass, Margery	Fred Hutchinson Cancer Research Center		Seattle	Washington	98109	US
Gelb, Bruce	Icahn School of Medicine at Mount Sinai		New York	New York	10029	US
Geng, Xiaoqi (Priscilla)	University of Michigan		Ann Arbor	Michigan	48109	US
Germer, Soren	New York Genome Center		New York	New York	10013	US
Gignoux, Chris	Stanford University		Stanford	California	94305	US
Gladwin, Mark	University of Pittsburgh		Pittsburgh	Pennsylvania	15260	US
Glahn, David	Yale University		New Haven	Connecticut	06520	US
Gogarten, Stephanie	University of Washington		Seattle	Washington	98195	US
Gong, Da-Wei	University of Maryland		Baltimore	Maryland	21201	US
Goring, Harald	University of Texas Rio Grande Valley School of Medicine		San Antonio	Texas	78229	US
Gu, C. Charles	Washington University in St Louis		St Louis	Missouri	63130	US
Guan, Yue	University of Maryland		Baltimore	Maryland	21201	US
Guo, Xiuqing	Los Angeles Biomedical Research Institute		Los Angeles	California	90502	US
Haessler, Jeff	Fred Hutchinson Cancer Research Center, Women's Health Initiative		Seattle	Washington	98109	US

Hall, Michael	University of Mississippi		Jackson	Mississippi	38677	US
Harris, Daniel	University of Maryland		Baltimore	Maryland	21201	US
Hawley, Nicola	Yale University		New Haven	Connecticut	06520	US
He, Jiang	Tulane University		New Orleans	Louisiana	70118	US
Heavner, Ben	University of Washington	Biostatistics	Seattle	Washington	98195	US
Heckbert, Susan	University of Washington		Seattle	Washington	98195	US
Hernandez, Ryan	McGill University, University of California, San Francisco					CA
Herrington, David	Wake Forest Baptist Health		Winston-Salem	North Carolina	27157	US
Hersh, Craig	Brigham & Women's Hospital	Channing Division of Network Medicine	Boston	Massachusetts	02115	US
Hidalgo, Bertha	University of Alabama		Birmingham	Alabama	35487	US
Hixson, James	University of Texas Health at Houston		Houston	Texas	77225	US
Hokanson, John	University of Colorado at Denver		Denver	Colorado	80204	US
Hong, Elliott	University of Maryland		Baltimore	Maryland	21201	US
Hoth, Karin	University of Iowa		Iowa City	Iowa	52242	US
Hsiung, Chao (Agnes)	National Health Research Institute Taiwan	Institute of Population Health Sciences, NHRI	Miaoli County		350	TW
Huston, Haley	Blood Works Northwest		Seattle	Washington	98105	US
Hwu, Chii Min	Taichung Veterans General Hospital Taiwan		Taichung City		407	TW
Irvin, Marguerite Ryan	University of Alabama		Birmingham	Alabama	35487	US

Jackson, Rebecca	Ohio State University Wexner Medical Center	Internal Medicine, DIvision of Endocrinology, Diabetes and Metabolism	Columbus	Ohio	43210	US
Jain, Deepti	University of Washington		Seattle	Washington	98195	US
Jaquish, Cashell	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US
Jhun, Min A	University of Michigan		Ann Arbor	Michigan	48109	US
Johnsen, Jill	Blood Works Northwest, University of Washington		Seattle	Washington	98106	US
Johnson, Andrew	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US
Johnson, Craig	University of Washington		Seattle	Washington	98195	US
Johnston, Rich	Emory University		Atlanta	Georgia	30322	US
Jones, Kimberly	Johns Hopkins University		Baltimore	Maryland	21218	US
Kang, Hyun Min	University of Michigan	Biostatistics	Ann Arbor	Michigan	48109	US
Kaplan, Robert	Albert Einstein College of Medicine		New York	New York	10461	US
Kardia, Sharon	University of Michigan		Ann Arbor	Michigan	48109	US
Kathiresan, Sekar	Broad Institute		Cambridge	Massachusetts	02142	US
Kaufman, Laura	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Kelly, Shannon	Vitalant Research Institute		San Francisco	California	94118	US
Kenny, Eimear	Icahn School of Medicine at Mount Sinai		New York	New York	10029	US
Kessler, Michael	University of Maryland		Baltimore	Maryland	21201	US

Khan, Alyna	University of Washington		Seattle	Washington	98195	US
Kinney, Greg	University of Colorado at Denver		Denver	Colorado	80204	US
Konkle, Barbara	Blood Works Northwest		Seattle	Washington	98104	US
Kooperberg, Charles	Fred Hutchinson Cancer Research Center		Seattle	Washington	98109	US
Kramer, Holly	Loyola University	Public Health Sciences	Maywood	Illinois	60153	US
Krauter, Stephanie	University of Washington		Seattle	Washington	98195	US
Lange, Christoph	Harvard School of Public Health	Biostats	Boston	Massachusetts	02115	US
Lange, Ethan	University of Colorado at Denver		Denver	Colorado	80204	US
Lange, Leslie	University of Colorado at Denver		Denver	Colorado	80204	US
Laurie, Cathy	University of Washington		Seattle	Washington	98195	US
Laurie, Cecelia	University of Washington		Seattle	Washington	98195	US
LeBoff, Meryl	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Lee, Jiwon	Brigham & Women's Hospital					
Lee, Seunggeun Shawn	University of Michigan		Ann Arbor	Michigan	48109	US
Lee, Wen-Jane	Taichung Veterans General Hospital Taiwan		Taichung City		407	TW
LeFaive, Jonathon	University of Michigan		Ann Arbor	Michigan	48109	US
Levine, David	University of Washington		Seattle	Washington	98195	US

Levy, Dan	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US
Lewis, Joshua	University of Maryland		Baltimore	Maryland	21201	US
Li, Yun	University of North Carolina		Chapel Hill	North Carolina	27599	US
Lin, Honghuang	Boston University		Boston	Massachusetts	02215	US
Lin, Keng Han	University of Michigan		Ann Arbor	Michigan	48109	US
Lin, Xihong	Harvard School of Public Health					
Liu, Simin	Brown University, Women's Health Initiative	Epidemiology	Providence	Rhode Island	02912	US
Liu, Yongmei	Wake Forest Baptist Health		Winston-Salem	North Carolina	27157	US
Loos, Ruth	Icahn School of Medicine at Mount Sinai		New York	New York	10029	US
Lubitz, Steven	Massachusetts General Hospital		Boston	Massachusetts	02114	US
Lunetta, Kathryn	Boston University		Boston	Massachusetts	02215	US
Luo, James	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US
Mahaney, Michael	University of Texas Rio Grande Valley School of Medicine		Brownsville	Texas	78520	US
Make, Barry	Johns Hopkins University		Baltimore	Maryland	21218	US
Manichaikul, Ani	University of Virginia		Charlottesville	Virginia	22903	US
Manson, JoAnn	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Margolin, Lauren	Broad Institute		Cambridge	Massachusetts	02142	US

Martin, Lisa	George Washington University		Washington	District of Columbia	20052	US
Mathai, Susan	University of Colorado at Denver		Denver	Colorado	80204	US
Mathias, Rasika	Johns Hopkins University		Baltimore	Maryland	21218	US
McArdle, Patrick	University of Maryland		Baltimore	Maryland	21201	US
McDonald, Merry-Lynn	University of Alabama		Birmingham	Alabama	35487	US
McFarland, Sean	Harvard University		Cambridge	Massachusetts	02138	US
McGarvey, Stephen	Brown University		Providence	Rhode Island	02912	US
McHugh, Caitlin	University of Washington	Biostatistics	Seattle	Washington	98145	US
Mei, Hao	University of Mississippi		Jackson	Mississippi	38677	US
Meyers, Deborah A	University of Arizona		Tucson	Arizona	85721	US
Mikulla, Julie	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US
Min, Nancy	University of Mississippi		Jackson	Mississippi	38677	US
Minear, Mollie	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US
Minster, Ryan L	University of Pittsburgh		Pittsburgh	Pennsylvania	15260	US
Mitchell, Braxton D.	University of Maryland		Baltimore	Maryland	21201	US
Montasser, May E.	University of Maryland		Baltimore	Maryland	21201	US
Musani, Solomon	University of Mississippi	Medicine	Jackson	Mississippi	39213	US

Mwasongwe, Stanford	University of Mississippi		Jackson	Mississippi	38677	US
Mychaleckyj, Josyf C	University of Virginia		Charlottesville	Virginia	22903	US
Nadkarni, Girish	Icahn School of Medicine at Mount Sinai		New York	New York	10029	US
Naik, Rakhi	Johns Hopkins University		Baltimore	Maryland	21218	US
Naseri, Take	Ministry of Health, Government of Samoa		Apia			WS
Natarajan, Pradeep	Broad Institute, Harvard University, Massachusetts General Hospital		Cambridge	Massachusetts	02138	US
Nekhai, Sergei	Howard University		Washington	District of Columbia	20059	US
Nelson, Sarah C.	University of Washington	Biostatistics	Seattle	Washington	98195	US
Nickerson, Deborah	University of Washington		Seattle	Washington	98195	US
North, Kari	University of North Carolina		Chapel Hill	North Carolina	27599	US
O'Connell, Jeff	University of Maryland		Balitmore	Maryland	21201	US
O'Connor, Tim	University of Maryland		Baltimore	Maryland	21201	US
Ochs-Balcom, Heather	University at Buffalo		Buffalo	New York	14260	US
Palmer, Nicholette	Wake Forest Baptist Health	Biochemistry	Winston-Salem	North Carolina	27157	US
Pankow, James	University of Minnesota		Minneapolis	Minnesota	55455	US
Papanicolaou, George	National Heart, Lung, and Blood Institute, National Institutes of Health		Bethesda	Maryland	20892	US

Parker, Margaret	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Parsa, Afshin	University of Maryland		Baltimore	Maryland	21201	US
Penchev, Sara	National Jewish Health		Denver	Colorado	80206	US
Peralta, Juan Manuel	University of Texas Rio Grande Valley School of Medicine		Edinburg	Texas	78539	US
Perez, Marco	Stanford University		Stanford	California	94305	US
Perry, James	University of Maryland		Baltimore	Maryland	21201	US
Peters, Ulrike	Fred Hutchinson Cancer Research Center, University of Washington		Seattle	Washington	98109	US
Peyser, Patricia	University of Michigan		Ann Arbor	Michigan	48109	US
Phillips, Lawrence S	Emory University		Atlanta	Georgia	30322	US
Phillips, Sam	University of Washington		Seattle	Washington	98195	US
Pollin, Toni	University of Maryland		Baltimore	Maryland	21201	US
Post, Wendy	Johns Hopkins University	Cardiology/Medicine	Baltimore	Maryland	21218	US
Powers Becker, Julia	University of Colorado at Denver	Medicine	Denver	Colorado	80204	US
Preethi Boorgula, Meher	University of Colorado at Denver		Denver	Colorado	80204	US
Preuss, Michael	Icahn School of Medicine at Mount Sinai		New York	New York	10029	US
Prokopenko, Dmitry	Harvard University		Cambridge	Massachusetts	02138	US
Psaty, Bruce	University of Washington		Seattle	Washington	98195	US
Qasba, Pankaj	National Heart, Lung, and Blood Institute,		Bethesda	Maryland	20892	US

	National Institutes of Health					
Qiao, Dandi	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Qin, Zhaohui	Emory University		Atlanta	Georgia	30322	US
Rafaels, Nicholas	University of Colorado at Denver		Denver	Colorado	80045	US
Raffield, Laura	University of North Carolina	Genetics	Chapel Hill	North Carolina	27599	US
Rao, D.C.	Washington University in St Louis		St Louis	Missouri	63130	US
Rasmussen-Torvik, Laura	Northwestern University		Chicago	Illinois	60208	US
Ratan, Aakrosh	University of Virginia		Charlottesville	Virginia	22903	US
Redline, Susan	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Reed, Robert	University of Maryland		Baltimore	Maryland	21201	US
Regan, Elizabeth	National Jewish Health		Denver	Colorado	80206	US
Reiner, Alex	Fred Hutchinson Cancer Research Center, University of Washington		Seattle	Washington	98109	US
Reupena, Muagututi‘a Sefuiva	Lutia I Puava Ae Mapu I Fagalele		Apia			WS
Rice, Ken	University of Washington		Seattle	Washington	98195	US
Rich, Stephen	University of Virginia		Charlottesville	Virginia	22903	US
Roden, Dan	Vanderbilt University	Medicine, Pharmacology, Biomedicla Informatics	Nashville	Tennessee	37235	US
Roselli, Carolina	Broad Institute		Cambridge	Massachusetts	02142	US
Rotter, Jerome	Los Angeles Biomedical Research Institute		Los Angeles	California	90502	US

Ruczinski, Ingo	Johns Hopkins University		Baltimore	Maryland	21218	US
Russell, Pamela	University of Colorado at Denver		Denver	Colorado	80204	US
Ruuska, Sarah	Blood Works Northwest		Seattle	Washington	98107	US
Ryan, Kathleen	University of Maryland		Baltimore	Maryland	21201	US
Sabino, Ester Cerdeira	Universidade de Sao Paulo	Faculdade de Medicina	Sao Paulo		01310000	BR
Sakornsakolpat, Phuwanat	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Salimi, Shabnam	University of Maryland		Baltimore	Maryland	21201	US
Salzberg, Steven	Johns Hopkins University		Baltimore	Maryland	21218	US
Sandow, Kevin	Los Angeles Biomedical Research Institute	TGPS	Torrance	California	90502	US
Sankaran, Vijay	Harvard University		Cambridge	Massachusetts	02138	US
Scheller, Christopher	University of Michigan		Ann Arbor	Michigan	48109	US
Schmidt, Ellen	University of Michigan		Ann Arbor	Michigan	48109	US
Schwander, Karen	Washington University in St Louis		St Louis	Missouri	63130	US
Schwartz, David	University of Colorado at Denver		Denver	Colorado	80204	US
Sciurba, Frank	University of Pittsburgh		Pittsburgh	Pennsylvania	15260	US
Seidman, Christine	Harvard Medical School	Genetics	Boston	Massachusetts	02115	US
Seidman, Jonathan	Harvard Medical School					
Sheehan, Vivien	Baylor College of Medicine	Pediatrics	Houston	Texas	77030	US
Shetty, Amol	University of Maryland		Baltimore	Maryland	21201	US
Shetty, Aniket	University of Colorado at Denver		Denver	Colorado	80204	US

Sheu, Wayne Hui-Heng	Taichung Veterans General Hospital Taiwan		Taichung City		407	TW
Shoemaker, M. Benjamin	Vanderbilt University		Nashville	Tennessee	37235	US
Silver, Brian	UMass Memorial Medical Center		Worcester	Massachusetts	01655	US
Silverman, Edwin	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Smith, Jennifer	University of Michigan		Ann Arbor	Michigan	48109	US
Smith, Josh	University of Washington		Seattle	Washington	98195	US
Smith, Nicholas	University of Washington		Seattle	Washington	98195	US
Smith, Tanja	New York Genome Center		New York	New York	10013	US
Smoller, Sylvia	Albert Einstein College of Medicine		New York	New York	10461	US
Snively, Beverly	Wake Forest Baptist Health	Biostatistical Sciences	Winston-Salem	North Carolina	27157	US
Sofer, Tamar	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Sotoodehnia, Nona	University of Washington		Seattle	Washington	98195	US
Stilp, Adrienne	University of Washington		Seattle	Washington	98195	US
Streeten, Elizabeth	University of Maryland		Baltimore	Maryland	21201	US
Su, Jessica Lasky	Brigham & Women's Hospital					
Sung, Yun Ju	Washington University in St Louis		St Louis	Missouri	63130	US
Sylvia, Jody	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Szpiro, Adam	University of Washington		Seattle	Washington	98195	US

Sztalryd, Carole	University of Maryland		Baltimore	Maryland	21201	US
Taliun, Daniel	University of Michigan		Ann Arbor	Michigan	48109	US
Tang, Hua	Stanford University	Genetics	Stanford	California	94305	US
Taub, Margaret	Johns Hopkins University		Baltimore	Maryland	21218	US
Taylor, Kent D.	Los Angeles Biomedical Research Institute	Institute for Translational Genomics and Populations Sciences	Torrance	California	90502	US
Taylor, Simeon	University of Maryland		Baltimore	Maryland	21201	US
Telen, Marilyn	Duke University		Durham	North Carolina	27708	US
Thornton, Timothy A.	University of Washington		Seattle	Washington	98195	US
Tinker, Lesley	Women's Health Initiative		Seattle	Washington	98109	US
Tirschwell, David	University of Washington		Seattle	Washington	98195	US
Tiwari, Hemant	University of Alabama		Birmingham	Alabama	35487	US
Tracy, Russell	University of Vermont	Pathology & Laboratory Medicine	Burlington	Vermont	05405	US
Tsai, Michael	University of Minnesota		Minneapolis	Minnesota	55455	US
Vaidya, Dhananjay	Johns Hopkins University		Baltimore	Maryland	21218	US
VandeHaar, Peter	University of Michigan		Ann Arbor	Michigan	48109	US
Vasan, Ramachandran S.	Boston University		Boston	Massachusetts	02215	US
Vrieze, Scott	University of Colorado at Boulder, University of Minnesota		Boulder	Colorado	80309	US
Walker, Tarik	University of Colorado at Denver		Denver	Colorado	80204	US
Wallace, Robert	University of Iowa		Iowa City	Iowa	52242	US
Walts, Avram	University of Colorado at Denver		Denver	Colorado	80204	US

Wan, Emily	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Wang, Fei Fei	University of Washington		Seattle	Washington	98195	US
Wang, Heming	Brigham & Women's Hospital, Partners.org					
Watson, Karol	University of California, Los Angeles		Los Angeles	California	90095	US
Weeks, Daniel E.	University of Pittsburgh		Pittsburgh	Pennsylvania	15260	US
Weir, Bruce	University of Washington		Seattle	Washington	98195	US
Weiss, Scott	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Weng, Lu-Chen	Massachusetts General Hospital		Boston	Massachusetts	02114	US
Willer, Cristen	University of Michigan	Internal Medicine	Ann Arbor	Michigan	48109	US
Williams, Kayleen	University of Washington		Seattle	Washington	98195	US
Williams, L. Keoki	Henry Ford Health System		Detroit	Michigan	48202	US
Wilson, Carla	Brigham & Women's Hospital		Boston	Massachusetts	02115	US
Wilson, James	University of Mississippi	Department of Physiology and Biophysics	Jackson	Mississippi	39216	US
Wong, Quenna	University of Washington		Seattle	Washington	98195	US
Xu, Huichun	University of Maryland		Baltimore	Maryland	21201	US
Yanek, Lisa	Johns Hopkins University		Baltimore	Maryland	21218	US
Yang, Ivana	University of Colorado at Denver		Denver	Colorado	80204	US
Yang, Rongze	University of Maryland		Baltimore	Maryland	21201	US
Zaghloul, Norann	University of Maryland		Baltimore	Maryland	21201	US
Zekavat, Maryam	Broad Institute		Cambridge	Massachusetts	02142	US

Zhang, Yingze	University of Pittsburgh	Medicine	Pittsburgh	Pennsylvania	15260	US
Zhao, Snow Xueyan	National Jewish Health		Denver	Colorado	80206	US
Zhao, Wei	University of Michigan		Ann Arbor	Michigan	48109	US
Zhi, Degui	University of Texas Health at Houston		Houston	Texas	77225	US
Zhou, Xiang	University of Michigan		Ann Arbor	Michigan	48109	US
Zhu, Xiaofeng	Case Western Reserve University	Department of Population and Quantitative Health Sciences	Cleveland	Ohio	44106	US
Zody, Michael	New York Genome Center		New York	New York	10013	US
Zoellner, Sebastian	University of Michigan		Ann Arbor	Michigan	48109	US

TOPMed Inflammation Working Group

Stella Aslibekyan
Paul Auer
David Beame
Ferdouse Begum
Emelia Benjamin
Joshua Bis
Michael Bowers
Russell Bowler
Erin Buth
Brian Cade
Adolfo Correa
Jeffrey Curtis
Margaret Mengmeng Du
Josee Dupuis
Jon Peter Durda
Nauder Faraday
Ervin Fox
Sheila Gaynor

Rebecca Jackson
Min A Jhun
Anne Justice
Brian Kral
Martin Larson
David Levine
Honghuang Lin
Xihong Lin
Merry-Lynn McDonald
John McLenithan
Michael Mendelson
Julie Mikulla
Nancy Min
Joanne Murabito
Nels Olson
Nathan Pankratz
Ulrike Peters
Linda Polfus
Bruce Psaty
Jenn Purnell
Laura Raffield
Deepa Rastogi
Elizabeth Regan
Alex Reiner
Annabelle Rodriguez
Ester Cerdeira Sabino
Shabnam Salimi
Cassie Spracklen
Ryan Sun
Russell Tracy
Dhananjay Vaidya
Biqi Becky Wang
Fei Fei Wang
Kate Wehr
James Wilson
Lue Ping Zhao