

Genome-Wide Association Study of High-sensitivity C-Reactive Protein, D-dimer and Interleukin-6 Levels in Multi-Ethnic HIV+ Cohorts

Brad T Sherman, Xiaojun Hu, Kanal Singh, Lillian Haine, Adam W. RUPERT, James D Neaton, Jens D Lundgren, Tomozumi Imamichi, Weizhong Chang, H. Clifford Lane, for the ESPRIT, SMART and START Study Groups

Supplemental Materials

SUPPLEMENTAL MATERIAL AND METHODS:

Biomarker Measurements: Methods for biomarker measurement at baseline for each study have been published previously^[1-4]. Briefly, levels of IL-6, hsCRP, and D-dimer were measured at baseline, before randomization, using stored plasma for patients who provided written consent. For the ESPRIT and START trials, laboratory measurements were performed by SAIC-Frederick (Frederick, MD; currently Leidos Biomedical Research, Inc.). For SMART, measurements were performed at the Laboratory for Clinical Biochemistry Research at the University of Vermont (Burlington). IL-6 was measured by the same method at each laboratory (ELISA; R&D Systems). ESPRIT and START D-dimer levels were measured using the VIDAS system (BioMerieux) and SMART D-dimer levels were measured by ELISA on the Sta-R analyzer, Liatest D-DI (Diagnostic Stago, Parsippany, NJ). hsCRP levels were measured with an R&D Systems ELISA assay for ESPRIT, a NBTMI nephelometer, N Antiserum to Human CRP (Siemens Diagnostics, Deerfield, IL) for SMART and a multiplex ECL (Vascular Injury Panel 2, Meso Scale Diagnostics) for START. It has previously been demonstrated that the measurement of IL-6, hsCRP and D-dimer from different laboratories for twenty duplicate samples from ESPRIT and SMART compared well even though the assays used to measure hsCRP and D-dimer levels were different^[1, 3]. Although the START trial was not among the trials in these two papers, the method for the measurement of IL-6 and D-dimer was the same for

START and ESPRIT and was conducted at the same laboratory. The only difference for START was the measurement of hsCRP with ECL from Meso Scale Diagnostics (MSD). It has been shown previously that MSD ECL assays were comparable to the WHO reference ELISA^[5].

SNP Quality Control and SNP Imputation: SNPs with any of the following were kept for GWAS analysis: (a) recommended by Axiom Analysis Suite (Thermo Fisher Scientific), (b) $\geq 90\%$ reproducibility from the internal control Ref103, (c) had a better quality than others for biallelic and multiallelic SNPs at one genomic position. As a result, 592,533 out of 770,558 markers remained for imputation.

SNP imputation to the 1000 Genomes phase 3 reference^[5] was performed on the total subjects with the genipe pipeline^[7] using PLINK, SHAPEIT (version 2.5)^[8] for phasing and IMPUTE2 (version 2.3.2)^[9] for imputation. The imputed SNPs with a confidence score INFO ≤ 0.8 and duplicates were removed. A total of 12,642,786 markers were retained after imputation.

For each study, SNPs with any of the following were excluded: (a) genotype missing rate $> 5\%$, (b) minor allele frequencies $< 1\%$, (c) Hardy–Weinberg equilibrium $p < 1 \times 10^{-6}$.

ADMIXTURE (version 1.3.0)^[10] was used to assign ancestry. The software estimates the relative admixture proportions of a priori defined ancestral groups contributing to the genome of each individual. 2504 individuals in 1000 Genomes phase 3^[5] were used as a reference group to generate a supervised set of five ancestral groups: European (EUR), African (AFR), Ad Mixed American (AMR), South Asian and East Asian. Common autosomal SNPs between the Affymetrix Axiom chip and 1000 Genomes genomic data after pruning were used to estimate the admixture proportions. Individual ancestry was assigned to the group with the maximum proportion.

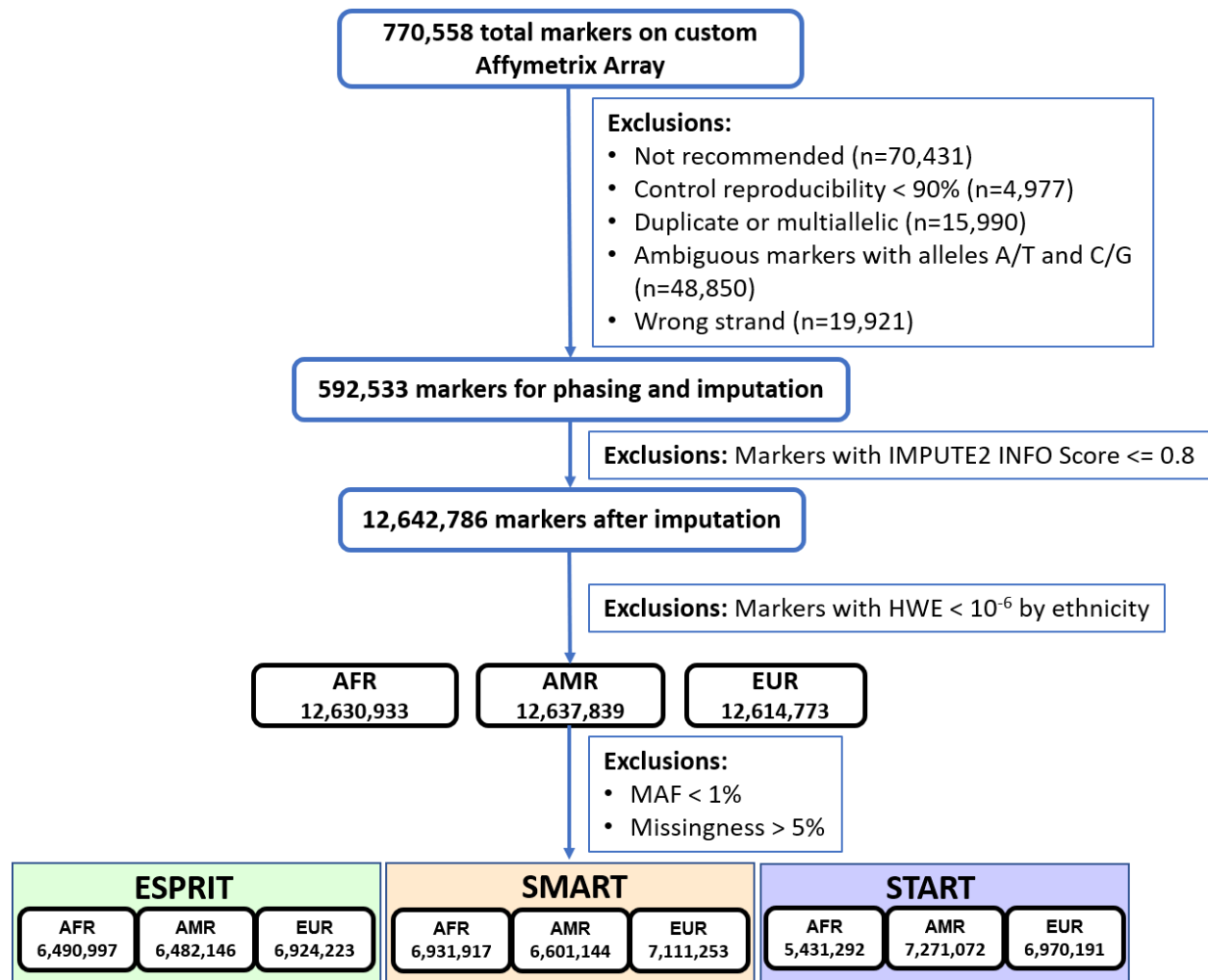
After the samples were assigned to their ancestry groups, as part of each cohort-based, ethnic specific association analysis, markers were further excluded based on MAF $< 1\%$ and missingness $> 5\%$.

Principal Component Analysis (PCA): PCA was applied to account for population stratification. SNPs were pruned to exclude regions of high Linkage Disequilibrium (LD)^[11] and retain SNPs with r^2 between SNPs < 0.2. PCA was performed using GCTA^[12] with 185,491 SNP markers. After 10 eigenvectors, eigenvalues flatten out in the scree plot. Therefore, only the top 10 eigenvectors were analyzed. The first principal component (PC0) explains the most variance in the data.

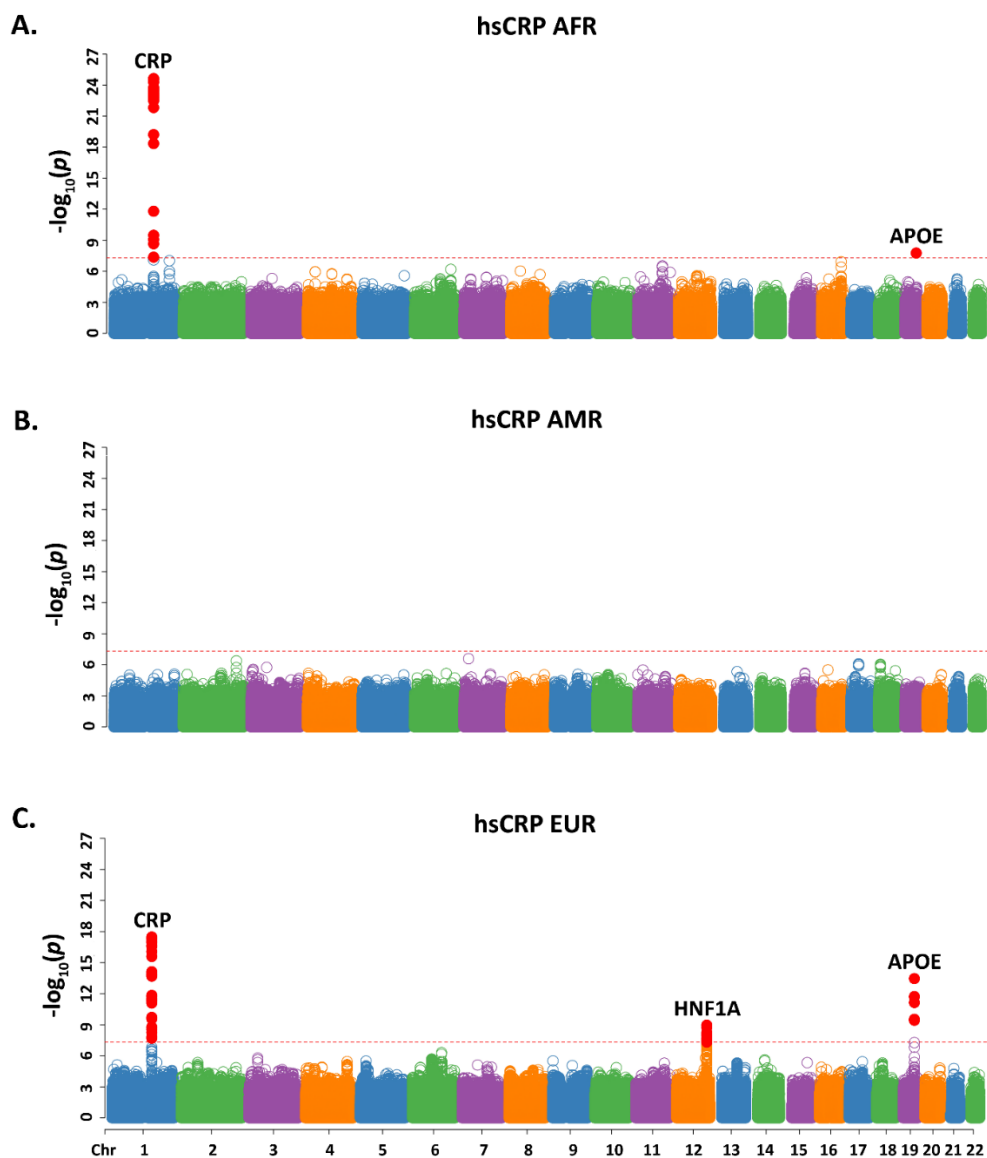
REFERENCES:

1. Nordell AD, McKenna M, Borges Á H, Duprez D, Neuhaus J, Neaton JD. **Severity of cardiovascular disease outcomes among patients with HIV is related to markers of inflammation and coagulation.** *J Am Heart Assoc* 2014; 3(3):e000844.
2. Borges AH, O'Connor JL, Phillips AN, Ronsholt FF, Pett S, Vjecha MJ, et al. **Factors Associated With Plasma IL-6 Levels During HIV Infection.** *J Infect Dis* 2015; 212(4):585-595.
3. Borges AH, O'Connor JL, Phillips AN, Baker JV, Vjecha MJ, Losso MH, et al. **Factors associated with D-dimer levels in HIV-infected individuals.** *PLoS One* 2014; 9(3):e90978.
4. Baker JV, Sharma S, Grund B, Rupert A, Metcalf JA, Schechter M, et al. **Systemic Inflammation, Coagulation, and Clinical Risk in the START Trial.** *Open Forum Infect Dis* 2017; 4(4).
5. The 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. **A global reference for human genetic variation.** *Nature* 2015; 526(7571):68-74.
6. Lemieux Perreault LP, Legault MA, Asselin G, Dube MP. **genipe: an automated genome-wide imputation pipeline with automatic reporting and statistical tools.** *Bioinformatics* 2016; 32(23):3661-3663.
7. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. **Fast and accurate genotype imputation in genome-wide association studies through pre-phasing.** *Nat Genet* 2012; 44(8):955-959.
8. Delaneau O, Zagury JF, Marchini J. **Improved whole-chromosome phasing for disease and population genetic studies.** *Nat Methods* 2013; 10(1):5-6.
9. Alexander DH, Novembre J, Lange K. **Fast model-based estimation of ancestry in unrelated individuals.** *Genome Res* 2009; 19(9):1655-1664.
10. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. **Data quality control in genetic case-control association studies.** *Nat Protoc* 2010; 5(9):1564-1573.
11. Yang J, Lee SH, Goddard ME, Visscher PM. **GCTA: a tool for genome-wide complex trait analysis.** *Am J Hum Genet* 2011; 88(1):76-82.

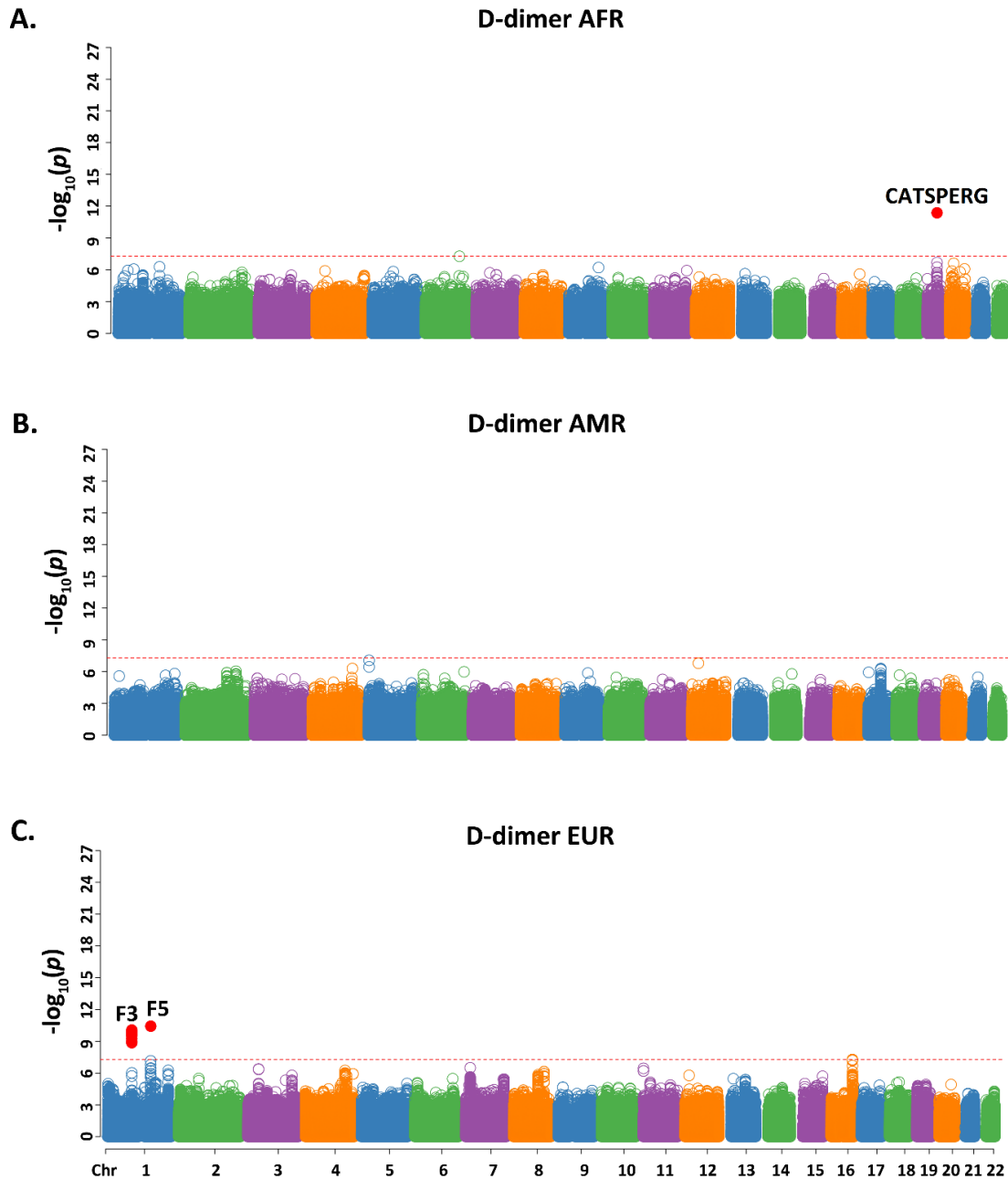
SUPPLEMENTAL FIGURES AND LEGENDS



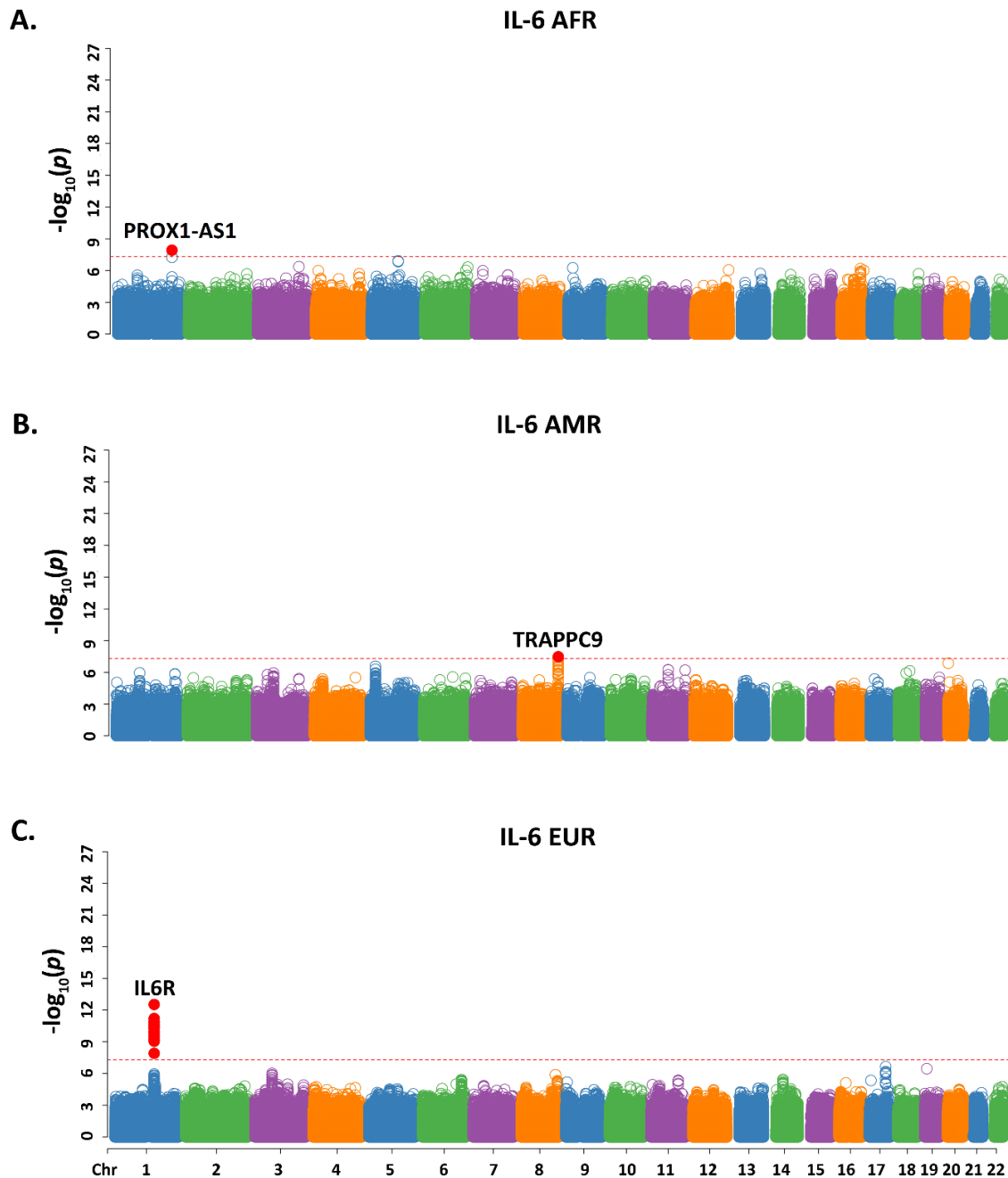
Supplemental Figure 1. SNP Imputation and Quality Control Workflow. **(1)** SNPs with any of the following were kept for GWAS analysis: (a) recommended by Axiom Analysis Suite (Thermo Fisher Scientific), (b) $\geq 90\%$ reproducibility from the internal control Ref103, (c) had a better quality than others for biallelic and multiallelic SNPs at one genomic position. For each study, SNPs with any of the following were excluded: (a) genotype missing rate $> 5\%$, (b) minor allele frequencies $< 1\%$, (c) Hardy–Weinberg equilibrium $p < 1 \times 10^{-6}$. **(2)** SNP imputation to the 1000 Genomes phase 3 reference^[5] was performed on the total subjects with the genipe pipeline^[7] using PLINK, SHAPEIT (version 2.5)^[8] for phasing and IMPUTE2 (version 2.3.2)^[9] for imputation. The imputed SNPs were QC with the confidence score INFO ≤ 0.8 and duplicates were removed. **(3)** After the samples were assigned to their ancestry groups, as part of each cohort-based, ethnic specific association analysis, markers were further excluded based on MAF $< 1\%$ and missingness $> 5\%$.



Supplemental Figure 2. Manhattan Plots of Ethnic Specific Meta-analyses Results for Genetic Associations with hsCRP. (A) hsCRP AFR results, (B) hsCRP AMR results, (C) hsCRP EUR results. Loci are labelled by the closest gene. Each point represents one SNP and is plotted by chromosomal location (x-axis) and $-\log_{10}(P)$ (y-axis). The dashed red line represents genome-wide significance ($P = 5 \times 10^{-8}$) and SNPs meeting this threshold are colored red. In contrast to the transeethnic results (**Figure 4. A**) which combined all participants from three ethnic groups shown here, there are generally less variants and loci reaching GWS for association with hsCRP level in each ethnic group due to a smaller sample size.



Supplemental Figure 3. Manhattan Plots of Ethnic Specific Meta-analyses Results for Genetic Associations with D-dimer. (A) D-dimer AFR results, (B) D-dimer AMR results, (C) D-dimer EUR results. Loci are labelled by the closest gene. Each point represents one SNP and is plotted by chromosomal location (x-axis) and $-\log_{10}(P)$ (y-axis). The dashed red line represents genome-wide significance ($P = 5 \times 10^{-8}$) and SNPs meeting this threshold are colored red. In contrast to the transethnic results (**Figure 4. B**) which combined all participants from three ethnic groups shown here, there are generally less variants and loci reaching GWS for association with D-dimer levels in each ethnic group due to a smaller sample size, except for the AFR-specific CATSPERG locus (**A**)



Supplemental Figure 4. Manhattan Plots of Ethnic Specific Meta-analyses Results for Genetic Associations with IL-6. (A) IL-6 AFR results, (B). IL-6 AMR results, (C) IL-6 EUR results. Loci are labelled by the closest gene. Each point represents one SNP and is plotted by chromosomal location (x-axis) and $-\log_{10}(P)$ (y-axis). The dashed red line represents genome-wide significance ($P = 5 \times 10^{-8}$) and SNPs meeting this threshold are colored red. In contrast to the transethnic results (**Figure 4. C**) which combined all participants from three ethnic groups shown here, there are generally less variants and loci reaching GWS for association with IL-6 levels in each ethnic group due to a smaller sample size, except for the AFR-specific locus PROX1-AS1 (A) and AMR-specific TRAPPC9 (B).

A		rs3116635 chr1:159686493			
		A	G		
rs2794520 chr1:159678816	C	1545	659	2204	(0.729)
	T	813	5	818	(0.271)
		2358	664	3022	
		(0.78)	(0.22)		
B		rs11065394 chr12:121446922			
		A	G		
rs2393776 chr12:121424406	A	1290	606	1896	(0.627)
	G	1113	13	1126	(0.373)
		2403	619	3022	
		(0.795)	(0.205)		
C		rs4133213 chr1:154395212			
		A	C		
rs6427658 chr1:154400799	C	884	922	1806	(0.598)
	T	128	1088	1216	(0.402)
		1012	2010	3022	
		(0.335)	(0.665)		

Supplemental Figure 5. The relationships between variants of opposite effect size.

To assess the relationship between two different types of variants, we used LDpair in LDlink to obtain the distribution of different alleles of the two top significant variants. The LDpair we used is implemented in the NCBI website

(<https://ldlink.nci.nih.gov/?var1=rs2794520&var2=rs3091244&pop=CEU%2BTSI%2BFIN%2BGBR%2BIBS&tab=ldpair>). The data shown here were obtained on February 3, 2020.

(A) The variants of the second SAS in CRP locus associated with the level of hsCRP.

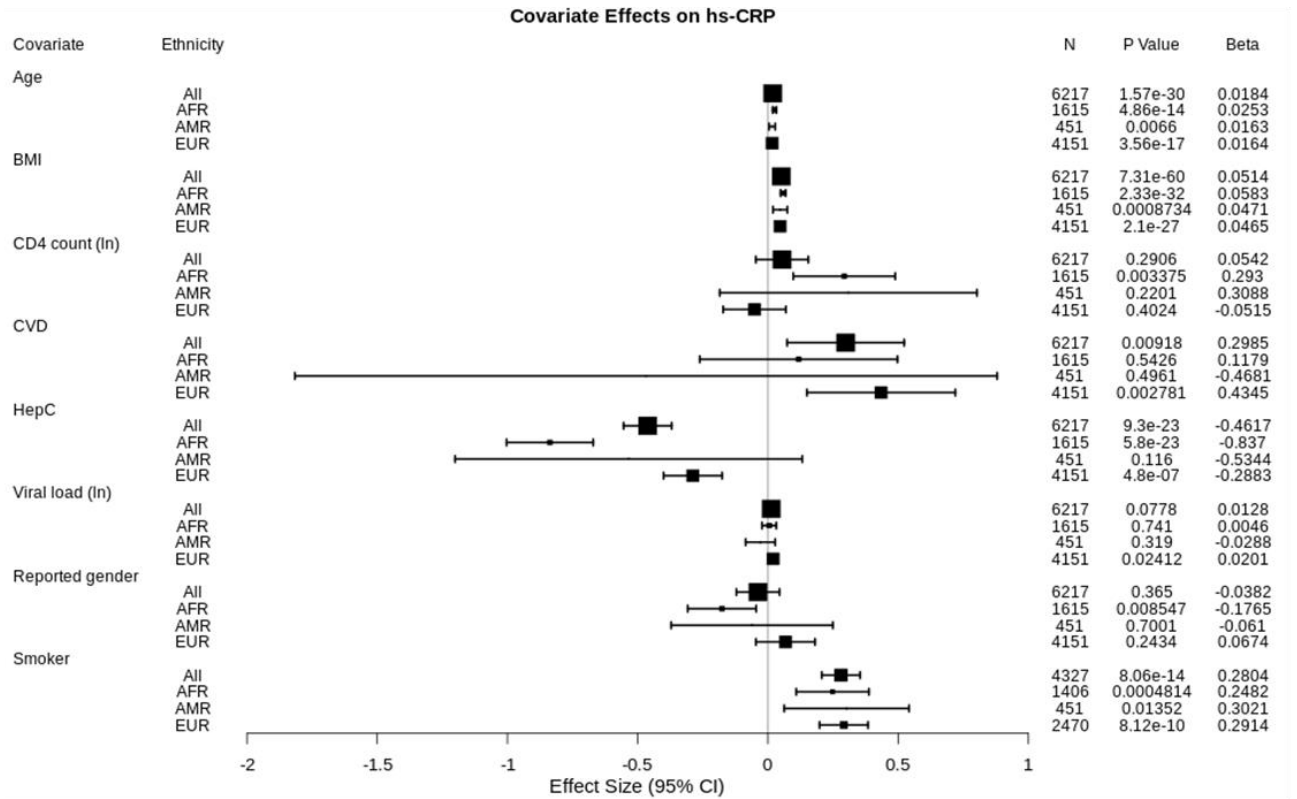
Most of the minor alleles (659 out of 664) of variant rs3116635 (G) from the positive effect size group commigrate with the major alleles of variant rs2794520 (C) from the negative effect size group within the AFR, AMR and EUR combined reference population.

Therefore, the two minor alleles have opposite effects.

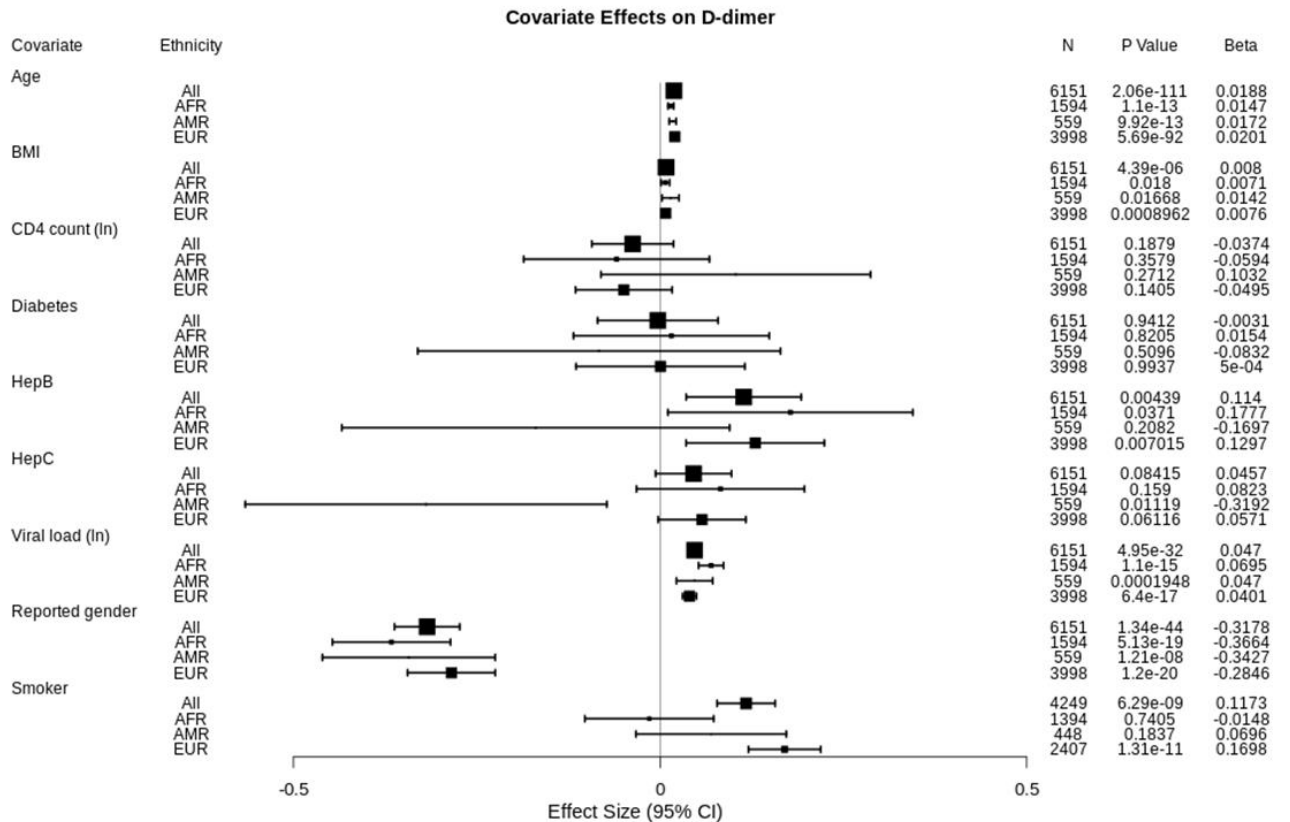
(B) The variants in HNF1A locus associated with the levels of hsCRP. Most of the minor alleles (606 out of 619) of variant rs11065394 (G) from the positive effect size group commigrate with the major alleles of variant rs2393776 (A) from the negative effect size group within the AFR, AMR and EUR combined reference population. Therefore, the two minor alleles have opposite effects.

(C) The variants in IL6R associated with IL-6 levels. Most of the minor alleles (884 out of 1,012) of variant rs4133213 (A) from the positive effect size group commigrate with the major alleles of variant rs6427658 (C) from the negative effect size group within the AFR, AMR and EUR combined reference population. Therefore, the two minor alleles have opposite effects.

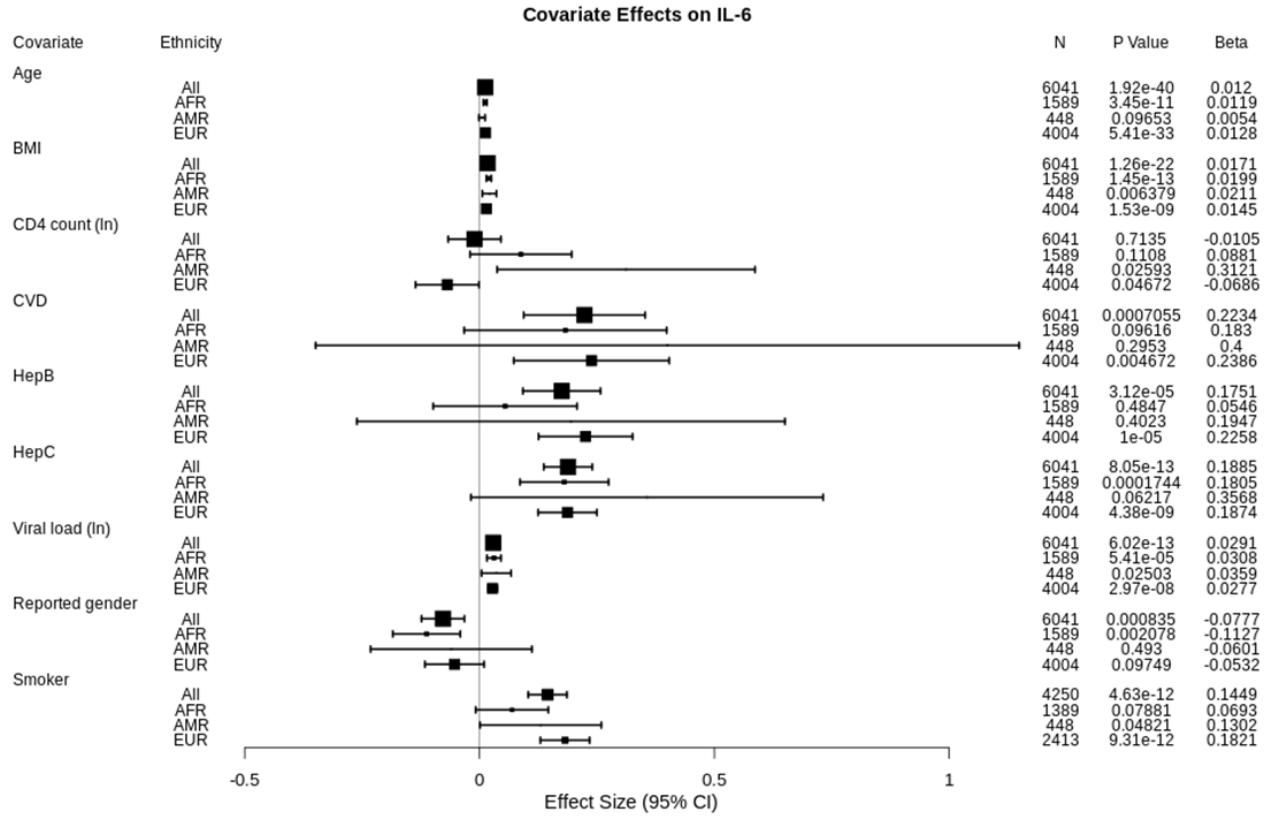
A



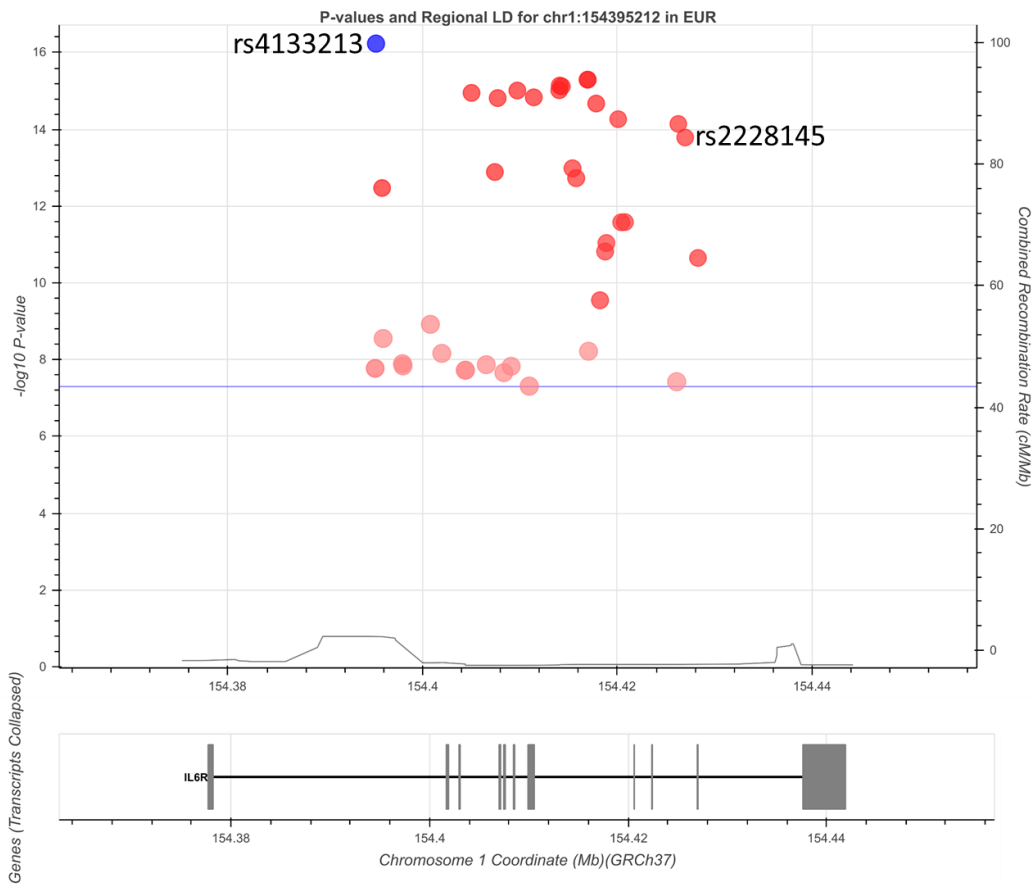
B



C



Supplemental Figure 6. The effect of the confounding factors on the levels of hsCRP (A), D-dimer (B) and IL-6 (C). N: number of patients have the measurement or record; *p* value: the evidence against a null hypothesis; Beta: the effect of the confounding factor on the biomarker.



Supplemental Figure 7. Diagram of p-value and Regional LD for chr1:154,395,212 in EUR. Horizontal axis represents genomic coordinates, Left vertical axis represents minus log p value and right vertical axis represents combined recombination rate (cM/Mb). The lower panel is aligned to the IL-6R structure. The blue dot is the lead variant, rs4133213, and the density of red color of other red dots shows the linkage disequilibrium(LD) of these variants with the lead variant. The published variant, rs2228145, is in close LD with rs4133213.

SUPPLEMENTAL TABLE

Supplemental Table 1. Baseline Characteristics of ESPRIT, SMART and START Patients.

Characteristic	ESPRIT	SMART	START
Cleaned samples	2,891	2,283	2,546
Mean age (SD)	40.9 (8.92)	44.5 (9.26)	37.3 (10.41)
Male (%)	2377 (82.22)	1681 (73.63)	2035 (79.93)
On ART (%)	2891 (100)	1815 (79.50)	0 (0)
*Race: white (%)	2178 (75.34)	948 (41.52)	1404 (55.15)
*Race: black (%)	269 (9.30)	876 (38.37)	578 (22.66)
*Race: hispanic (%)	NA	396 (17.35)	498 (19.56)
*Race: asian (%)	339 (11.73)	32 (1.40)	26 (1.02)
*Race: other (%)	105 (3.63)	31 (1.36)	41 (1.61)
Hepatitis B (%)	175 (6.81)	66 (2.90)	55 (2.24)
Hepatitis C (%)	374 (15.20)	388 (17.00)	109 (4.38)
Smoker (%)	NA	937 (41.04)	990 (38.88)
Myocardial infarction (%)	13 (0.45)	35 (1.53)	11 (0.43)
Median Viral load copies/ml (IQR)	50 (50-400)	400 (50-2784)	14,833 (3,507-46,000)
Median CD4 count cells/mm ³ (IQR)	451 (368-582)	572 (455-773)	651 (585-758)
Median BMI (IQR)	23.71 (21.78-25.88)	25.58 (22.94-28.98)	24.89 (22.50-28.31)
Samples with D-dimer measurement	2,778	2,207	2,417
Samples with hsCRP measurement	2,779	2,215	2,438
Samples with IL-6 measurement	2,778	2,188	2,436
Median D-dimer, ug/ml (IQR)	0.26 (0.18-0.37)	0.22 (0.13-0.40)	0.31 (0.22-0.47)
Median hsCRP ug/ml (IQR)	1.49 (0.69-3.21)	1.79 (0.76-4.51)	1.82 (0.77-4.15)
Median IL-6 pg/ml (IQR)	1.90 (1.30-2.80)	1.91 (1.19-3.23)	1.47 (1.02-2.21)

* Reported race

Supplemental Table 2. Loci Associated with Levels of hsCRP, D-dimer and IL-6

Biomarker	Associated Locus	Associated Variants	Ethnic Group	Pathway and Mechanism Relevant to the Biomarker Level	References
hsCRP	CRP	77	transethnic	Regulation of CRP expression	1 - 3
	HNF1A	34/6	transethnic/EUR	Regulation of CRP expression	4-6
	APOE	11	transethnic	Complexing with activated C1q	7
D-dimer	F3	17	transethnic	Complement and Coagulation Cascades	8, 9
	F5	1	transethnic	Complement and Coagulation Cascades	8, 10
	FGB	1	transethnic	Complement and Coagulation Cascades	11
	GCNT1	1	transethnic	Glycosyltransferase	12 -14
	CATSPERG	1	AFR	Near PSMD8 which is a subunit of proteasome	15
IL-6	IL-6R	40	transethnic	Receptor of IL-6 for classic and trans-signaling pathway	16, 17
	PROX1-AS1	1	AFR	Near PROX1 which is a homeodomain transcription factor; regulated by NF-kB pathway	18, 19
	TRAPPC9	1	AMR	Enhance NF-kB pathway	20-23

REFERENCES:

1. Miller DT, Zee RY, Suk Danik J, Kozlowski P, Chasman DI, Lazarus R, et al. **Association of common CRP gene variants with CRP levels and cardiovascular events.** *Ann Hum Genet* 2005; 69(Pt 6):623-638.
2. Szalai AJ, Wu J, Lange EM, McCrory MA, Langefeld CD, Williams A, et al. **Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level.** *J Mol Med (Berl)* 2005; 83(6):440-447.
3. Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, et al. **Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels.** *Am J Hum Genet* 2005; 77(1):64-77.
4. Armendariz AD, Krauss RM. **Hepatic nuclear factor 1-alpha: inflammation, genetics, and atherosclerosis.** *Curr Opin Lipidol* 2009; 20(2):106-111.
5. Reiner AP, Gross MD, Carlson CS, Bielinski SJ, Lange LA, Fornage M, et al. **Common coding variants of the HNF1A gene are associated with multiple cardiovascular risk phenotypes in community-based samples of younger and older European-American adults: the Coronary Artery Risk Development in Young Adults Study and The Cardiovascular Health Study.** *Circ Cardiovasc Genet* 2009; 2(3):244-254.
6. Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, et al. **Control of pancreas and liver gene expression by HNF transcription factors.** *Science* 2004; 303(5662):1378-1381.
7. Yin C, Ackermann S, Ma Z, Mohanta SK, Zhang C, Li Y, et al. **ApoE attenuates unresolvable inflammation by complex formation with activated C1q.** *Nat Med* 2019; 25(3):496-506.
8. Smith NL, Huffman JE, Strachan DP, Huang J, Dehghan A, Trompet S, et al. **Genetic predictors of fibrin D-dimer levels in healthy adults.** *Circulation* 2011; 123(17):1864-1872.
9. Raffield LM, Zakai NA, Duan Q, Laurie C, Smith JD, Irvin MR, et al. **D-Dimer in African Americans: Whole Genome Sequence Analysis and Relationship to Cardiovascular Disease Risk in the Jackson Heart Study.** *Arterioscler Thromb Vasc Biol* 2017; 37(11):2220-2227.
10. Zaleski AL, Pescatello LS, Thompson PD, Taylor BA. **Protective effect of compression socks in a marathon runner with a genetic predisposition to thrombophilia due to Factor V Leiden.** *Phys Sportsmed* 2015; 43(3):324-327.

11. de Moerloose P, Casini A, Neerman-Arbez M. **Congenital fibrinogen disorders: an update.** *Semin Thromb Hemost* 2013; 39(6):585-595.
12. Fonseca KL, Maceiras AR, Matos R, Simoes-Costa L, Sousa J, Cá B, et al. **Deficiency in the glycosyltransferase *Gcnt1* increases susceptibility to tuberculosis through a mechanism involving neutrophils.** *Mucosal Immunol* 2020.
13. Perkey E, Maurice De Sousa D, Carrington L, Chung J, Dils A, Granadier D, et al. **GCNT1-Mediated O-Glycosylation of the Sialomucin CD43 Is a Sensitive Indicator of Notch Signaling in Activated T Cells.** *J Immunol* 2020; 204(6):1674-1688.
14. Kojima Y, Yoneyama T, Hatakeyama S, Mikami J, Sato T, Mori K, et al. **Detection of Core2 beta-1,6-N-Acetylglucosaminyltransferase in Post-Digital Rectal Examination Urine Is a Reliable Indicator for Extracapsular Extension of Prostate Cancer.** *PLoS One* 2015; 10(9):e0138520.
15. Driscoll J, Finley D. **A controlled breakdown: antigen processing and the turnover of viral proteins.** *Cell* 1992; 68(5):823-825.
16. Uciechowski P, Dempke WCM. **Interleukin-6: A Masterplayer in the Cytokine Network.** *Oncology* 2020; 98(3):131-137.
17. Jiang CQ, Lam TH, Liu B, Lin JM, Yue XJ, Jin YL, et al. **Interleukin-6 receptor gene polymorphism modulates interleukin-6 levels and the metabolic syndrome: GBCS-CVD.** *Obesity (Silver Spring)* 2010; 18(10):1969-1974.
18. Oliver G, Sosa-Pineda B, Geisendorf S, Spana EP, Doe CQ, Gruss P. **Prox 1, a prospero-related homeobox gene expressed during mouse development.** *Mech Dev* 1993; 44(1):3-16.
19. Flister MJ, Wilber A, Hall KL, Iwata C, Miyazono K, Nisato RE, et al. **Inflammation induces lymphangiogenesis through up-regulation of VEGFR-3 mediated by NF-kappaB and Prox1.** *Blood* 2010; 115(2):418-429.
20. Hu WH, Pendergast JS, Mo XM, Brambilla R, Bracchi-Ricard V, Li F, et al. **NIBP, a novel NIK and IKK(beta)-binding protein that enhances NF-(kappa)B activation.** *J Biol Chem* 2005; 280(32):29233-29241.
21. Qin M, Zhang J, Xu C, Peng P, Tan L, Liu S, et al. **Knockdown of NIK and IKKβ-Binding Protein (NIBP) Reduces Colorectal Cancer Metastasis through Down-Regulation of the Canonical NF-κB Signaling Pathway and Suppression of MAPK Signaling Mediated through ERK and JNK.** *PLoS One* 2017; 12(1):e0170595.
22. Zhang Y, Liu S, Wang H, Yang W, Li F, Yang F, et al. **Elevated NIBP/TRAPPC9 mediates tumorigenesis of cancer cells through NFκB signaling.** *Oncotarget* 2015; 6(8):6160-6178.
23. Wattacheril J, Lavine JE, Chalasani NP, Guo X, Kwon S, Schwimmer J, et al. **Genome-Wide Associations Related to Hepatic Histology in Nonalcoholic Fatty Liver Disease in Hispanic Boys.** *J Pediatr* 2017; 190:100-107.e102.

SUPPLEMENTAL DATA

Supplemental Data 1: GWAS meta-analysis results for the variants associated with the hsCRP levels.

Supplemental Data 2: GWAS meta-analysis results for the variants associated with the D-dimer levels.

Supplemental Data 3: GWAS meta-analysis results for the variants associated with the IL-6 levels.