Published in final edited form as:

Hum Genet. 2016 April ; 135(4): 415–423. doi:10.1007/s00439-016-1643-0.

Blood Group Antigen Loci Demonstrate Multivariate Genetic Associations with Circulating Cellular Adhesion Protein Levels in the Multi-Ethnic Study of Atherosclerosis

Nicholas B. Larson^{1,§}, Paul A. Decker¹, Christina L. Wassel², James S. Pankow³, Weihong Tang³, Naomi Q. Hanson⁴, Michael Y. Tsai⁴, and Suzette J. Bielinski⁵

¹Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

²Department Pathology and Laboratory Medicine, College of Medicine, University of Vermont, Burlington, VT, USA

³Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

⁴Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA

⁵Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Abstract

The cellular adhesion pathway is critical in the pathophysiology of atherosclerosis, and genetic factors contributing to regulation of circulating levels of related proteins may be relevant to risk prediction of cardiovascular disease. In contrast to conducting separate genome-wide protein quantitative trait loci (pQTL) mapping analyses of each individual protein, joint genetic association analyses of multiple quantitative traits can leverage cross-trait co-variation and identify simultaneous regulatory effects on protein levels across the pathway. We conducted a multi-pQTL (mpQTL) analysis of 15 proteins related to cellular adhesion assayed on 2,313 participants from the Multi-Ethnic Study of Atherosclerosis (MESA). We applied the MQFAM multivariate association analysis method in PLINK on normalized protein level residuals derived from univariate linear regression, adjusting for age, sex, and principal components of ancestry. Race/ethnicity-stratified analyses identified nine genome-wide significant ($P < 5 \times 10^{-8}$) loci associated with co-variation of protein levels. Although the majority of these SNPs were in proximity to structural genes of the assayed proteins, we discovered multiple loci demonstrating co-association with the circulation of at least two proteins. Of these, two significant loci specific to non-Hispanic white participants, rs17074898 at *ALOX5AP* ($P = 1.78 \times 10^{-8}$) and rs7521237 at *KIAA1614* ($P = 2.2 \times 10^{-8}$), would not have met statistical significance using univariate analyses. Moreover, common patterns of multi-protein associations were discovered at the *ABO* locus across race/ethnicity. These results indicate the biological relevance of blood group antigens on

§Corresponding author: Nicholas B. Larson, Ph.D., Mayo Clinic, 200 First Street SW, Rochester, MN 55905, Larson.nicholas@mayo.edu, (507) – 293 – 1700 (phone), (507) – 284 – 1516 (fax).

The authors declare no conflicts of interest.

regulation of circulating cellular adhesion pathway proteins while also demonstrating race/ethnicity-specific co-regulatory effects.

Keywords

adhesion proteins; ABO; multivariate; genome-wide association study

INTRODUCTION

The cellular adhesion pathway is a promising avenue of research to understanding molecular mechanisms of inflammation that could lead to biomarkers for prediction, early detection, and prognosis of atherosclerotic diseases. Members of these adhesion families in combination with a host of other types of molecules, including actin, thrombin, chemokines, kinases, transcription factors, cytokines, apolipoproteins, fibrinogen, growth factors, and matrix metalloproteinases, comprise the cellular adhesion pathway and link the processes of hemostasis, thrombosis, and inflammation. Despite the fact that adhesion proteins have been shown to predict increased risk of incident CHD (Bielinski et al. 2015; Cesari et al. 2003; Demerath et al. 2001; Hwang et al. 1997), many potentially important proteins and the corresponding coding genes have yet to be fully studied. Single-protein genome-wide association studies (GWASs) have been conducted for only a few select adhesion proteins (Barbalic et al. 2010; Pare et al. 2008; Pare et al. 2011; Paterson et al. 2009; Reiner et al. 2013).

In GWAS where multiple traits (e.g., protein measurements) are available for association analysis, it is common practice to assess these traits in a separate, or univariate, manner, with post-hoc comparisons of associated loci across traits. However, recently developed multivariate analysis methods (Ferreira and Purcell 2009; O'Reilly et al. 2012; Stephens 2013) accommodate the simultaneous analysis of multiple potentially correlated traits. Multivariate GWAS methods have been shown to be as or more powerful than univariate GWAS approaches, taking advantage of cross-trait covariance and reducing multiple testing dimensionality (Galesloot et al. 2014). Simultaneous assessment of multiple protein measurements relevant to the adhesion pathway may provide additional context to genetic modulation of these proteins at a systemic level. Therefore, in this study, we present a multi-protein quantitative trait locus (mpQTL) association analysis to identify genetic variants associated with co-variation in circulating levels of 15 proteins related to cellular adhesion in the Multi-ethnic Study of Atherosclerosis (MESA).

METHODS

Study Participants

MESA is a multi-center study that enrolled African (AFA), Chinese (CHN), Hispanic (HIS), and non-Hispanic white (EUR) Americans between 2000–02 (Exam 1), with no history of CVD in order to investigate subclinical and clinical cardiovascular endpoints in a large and diverse population. Participating centers were located in Baltimore, MD; Chicago, IL; Forsyth County, NC; Los Angeles County, CA; northern Manhattan, NY; and Saint Paul,

MN. The MESA study has been described in detail elsewhere (Bild et al. 2002). In order to measure a large number of circulating adhesion proteins in a representative sample of the MESA population, a stratified random sample including 720 individuals for each of the four races/ethnicities represented in MESA was used (N = 2880). This subgroup was randomly generated from the parent MESA cohort including all Exam 1 participants who gave consent for DNA sample use, and was stratified by race/ethnicity in order to reflect the diversity of the main MESA study. MESA and its ancillary studies were approved by the Institutional Review Board at participating centers and all participants gave written informed consent.

Genetic Data

The assayed genome-wide genotype data consisted of four individual genotype panels on all MESA participants who consented for genetic studies: the Illumina Exome BeadChip (Huyghe et al. 2013), the Illumina Cardio-MetaboChip (Voight et al. 2012), the Illumina iSelect ITMAT/Broad/CARe (IBC) Chip (Keating et al. 2008) and the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara). Each of the panels had quality control measures individually performed on the genotype data prior to merging them together using Plink v1.07 (Purcell et al. 2007) under genome build NCBI build 37, and in total included 1,222,482 directly genotyped polymorphic autosomal positions across all participants. Evaluation of relatedness revealed multiple participants belonged to sibships; to eliminate family structure one individual from each observed sibship was retained. Population stratification was assessed using STRUCTURE (Pritchard et al. 2000) and EIGENSTRAT (Price et al. 2006) was used to obtain ancestry informative principal components (PCs) based upon all unrelated MESA participants. The total number of informative PCs for subsequent analyses was selected based upon statistical and graphical evaluation of the corresponding eigenvalues.

Protein Measurements

Fifteen adhesion proteins were measured in plasma or serum samples obtained between September 2002 and February 2004 (Exam 2). Of the 2880 individuals included in the random sample, plasma was available for 2574 participants and serum was available for 2441 participants. Among these, 34 individuals were excluded due to CVD events prior to Exam 2, 1 due to cognitive impairment, and 3 individuals were excluded due to inconsistencies between the field center where they were enrolled and their self-reported race/ethnicity, as ethnic groups were specific for each field center during recruitment. Plasma was collected using EDTA as an anticoagulant and stored in ice, whereas serum was obtained allowing blood samples to clot at room temperature for 40 minutes. Both samples were centrifuged at either 4°C at 2,000g x 15 minutes or 3,000g x 10 minutes for a total of 30,000 g-minutes and subsequently stored at -70°C. Quantitative sandwich enzyme-linked immunosorbent assays (ELISA) were used to measure soluble proteins. Chemokine ligand 21 (CCL21), hepatocyte growth factor (HGF), interleukin-2 soluble receptor α (IL-2sR α), pro-matrix metalloproteinase 1 (pro-MMP-1), matrix metalloproteinase 2 (MMP-2), secretory leukocyte protease inhibitor (SLPI), soluble E-Cadherin (sE-cadherin), soluble intercellular adhesion molecule 1 (sICAM-1), soluble L-selectin (sL-selectin), soluble vascular cell adhesion molecule 1 (sVCAM-1), and tissue inhibitor of metalloproteinase 2 (TIMP-2) were measured in serum, while soluble P-selectin (sP-selectin), regulated on

activation normal T cell expressed and secreted (RANTES), stromal derived factor 1 α (SDF-1 α) and transforming growth factor β 1 (TGF β 1) were measured in plasma. The inter-assay coefficient of variation and the minimum detectable level for each of the ELISA assays, as well as the specific type of assay used, are summarized in Table S1.

Statistical Analysis

All circulating protein measurements were initially log-transformed and linear regression was used to obtain protein-level residuals after adjusting for age, sex, and the first three ancestry-informative PCs. Residuals were then centered and scaled per protein to unit variance prior to analysis. Multivariate association analysis of single SNPs with the vector of 15 cellular adhesion protein residuals was conducted using the MQFAM(Ferreira and Purcell 2009) multivariate test of association implemented in PLINK(Purcell et al. 2007). This method is based upon canonical correlation analysis (CCA) and identifies the linear combination of protein measurements (i.e., canonical variate) optimally correlated with SNP minor allele dosage under an additive genetic model, producing an F-statistic and a vector of CCA variable (i.e., protein) loadings, ω_p , such that $-1 \leq \omega_p \leq 1$ for $p = 1, \dots, 15$. These loadings correspond to correlations between the protein measurement and the canonical variate, and can be interpreted as an indication of the relative importance of that given protein for the association. All polymorphic autosomal SNPs with an empirical minor allele frequency (MAF) $\geq 5\%$ and variant genotype call rate $\geq 90\%$ in the respective racial/ethnic strata were considered for analysis. The relative signs of canonical loadings are arbitrary, as CCA is based upon the solution of an eigenvalue equation. To determine directionality of effect, we used linear regression to estimate univariate associations with the protein corresponding to the largest loading magnitude. Since MQFAM can also accommodate missing phenotype data, participants were included in the analysis if at least 8 out of 15 ($>50\%$) protein levels were observed. To adjust for multiple testing, a single-SNP association was declared to be statistically significant if the nominal P-value was below a genome-wide significance threshold of $\alpha = 5e-08$. All analysis steps were stratified by race/ethnicity; to combine analysis p-values across race/ethnicity for purposes of trans-ethnic meta-analysis, we applied Stouffer's weighted Z method(Stouffer 1949), with weights based on relative number of analyzed samples per race/ethnicity. As a p-value aggregation approach, this analysis does not take into account consistency of protein-specific effects (direction and/or magnitude) across race/ethnicity.

RESULTS

After subject-level exclusions, a total of $N = 2,268$ participants ($N_{\text{AFA}} = 516$, $N_{\text{CHN}} = 570$, $N_{\text{EUR}} = 611$, $N_{\text{HIS}} = 571$) were eligible for mpQTL association analysis. Distribution summaries of the 15 proteins, separated by race/ethnicity, are presented in Table 1. A total of 14 out of 15 proteins exhibited significant mean differences in protein values by race/ethnicity when adjusting for age and sex; results were comparable with additional covariate adjustment of body mass index, alcohol use, and smoking status. In general, CHN participants tended to have lower circulation of cellular adhesion proteins than the other races/ethnicities, with the notable exception of TGF β -1. In contrast, EUR and HIS participants tended to have higher protein concentrations, while AFA participants

corresponded to the lowest levels of sE-cadherin and sVCAM-1. Graphical representations of the correlation structures of the residual protein measurements by race/ethnicity are presented in Figure S1, indicating strong positive correlations between protein pairs TIMP-2/MMP2 and RANTES/TGF- β 1 as well as comparable overall patterns across racial/ethnic groups.

Overall, 970,081 common (MAF $\geq 5\%$) SNPs were assessed in our stratified analyses, and the number, genetic location, and pattern of protein association for significantly associated SNPs varied by race/ethnicity. A total of nine distinct loci corresponded to significant associations in at least one race/ethnicity (Table 2), with 13 proteins exhibiting high-magnitude ($|\omega_p| > 0.25$) loadings among the top findings. Despite multiple strong associations identified in the other races/ethnicities, no significant findings were observed in African Americans. Most associated SNPs were in close physical proximity to one of the structural genes of the assayed proteins, correspondent to high-magnitude CCA protein loadings for that specific protein. Moreover, many protein loading patterns for loci significant across multiple racial/ethnic strata were similar in direction and magnitude. The complete set of significant and suggestive ($P < 1e-05$) stratified analysis results with corresponding protein loadings are presented in Tables S2–S5 and as Manhattan plots (Figures S2–S5).

There were a total of 850,009 SNPs that had association results for >1 race/ethnicity and were evaluated under our trans-ethnic meta-analysis (Table S6). The combined race/ethnicity analysis p-values are presented as a Manhattan plot in Figure 1, revealing seven distinct significant loci. Of these, only one (rs12722588 near *IL2RA*; $P = 7.66E-16$) was not previously identified in the stratified analysis. *IL2RA* encodes the alpha chain of IL2-sr protein complex. The significant meta-analysis SNP rs492602, although not in linkage disequilibrium (LD) with nearby CHN finding *FUT2* SNP rs1047781, tags SNP rs601338 in non-Asian populations ($r^2 > 0.99$). Of note, rs1047781 and rs601338 respectively represent population-specific functional SNPs that knockout *FUT2*-encoded protein function (Kelly et al. 1995).

Two associated loci were identified solely in non-Hispanic whites. SNP rs17074898 at the *ALOX5AP* locus ($P = 1.78E-08$) demonstrated high-magnitude loadings for multiple proteins, including sP-selectin, SDF-1 α , and TIMP-2, indicating a potential pathway-wide effect. SNP rs7521237 at the *KIAA1614* locus ($P = 2.2E-08$), corresponded to similar loading patterns. Protein loadings and univariate protein association analysis results using linear regression are presented in Figure 2 for both SNPs. Notably, each of these loci would not have been identified by single protein QTL analysis under the same genome-wide significance threshold (min. $P = 1.3E-05$).

Multiple significant SNPs were identified in proximity to the respective protein coding genes. For example, SNP rs6136 at the *SELP/SELL* locus was significantly associated with protein levels in EUR and marginally associated in HIS ($P = 2.84E-07$). Both findings demonstrated strong loadings for sP-selectin and, to a lesser degree, sL-selectin, however in HIS loading magnitudes were also high ($|\omega_p| > 0.25$) for four additional proteins (RANTES, SLPI, TGF- β -1, sICAM-1). Similar findings in our analyses were identified for MMP1 (*MMP1*), RANTES (*CCL5*), IL-2sr (*IL2RA*), and sICAM-1 (*ICAM1*). Investigation into

potential regulatory function using HaploReg(Ward and Kellis 2012) revealed all but two SNPs (rs10905876 and rs12722588 near *IL2RA*) corresponded to *cis*-acting expression QTLs (*cis*-eQTLs) for the proximal protein coding genes in relevant tissues (Table S7).

We observed multiple significant mpQTL SNP associations within *ABO* separately across CHN, EUR, and HIS, as well as in our combined meta-analysis. Moreover, these SNPs are known to collectively tag functional alleles that define the ABO blood group (i.e., A, B, and O). Effect heterogeneity was observed across *ABO* SNPs as well as by racial/ethnic strata, as indicated by the respective SNP protein loading vectors for CHN, EUR, and HIS (Figure 3). These differences were most pronounced for sE-cadherin and sICAM-1, with HIS results exhibiting reduced magnitude effects relative to CHN and EUR participants.

DISCUSSION

Using a large, multi-ethnic cohort, we observed substantial variation in protein levels by race/ethnicity, indicating potential genetic factors regulating cellular adhesion pathway activity. Although mean levels of the majority of circulating adhesion proteins (e.g., sP-selectin, sL-selectin, sICAM-1) were generally lowest amongst Chinese American participants, rates of subclinical atherosclerosis (defined by presence of coronary artery calcium) for CHN in MESA (59.2%) are second only to EUR participants (70.1%)(Bild et al. 2005). These results indicate that there may exist complex regulatory effects on circulating cellular adhesion protein profiles in atherogenesis beyond broad unidirectional dysregulation of the pathway. Multivariate methods for mpQTL mapping can identify pathway-wide genetic effects of protein regulation while leveraging cross-trait correlation to improve statistical power. Using a large, multi-ethnic cohort, we identified multiple genetic loci with strong multivariate associations with circulating levels of cellular adhesion proteins assayed in serum or plasma. Many of the associated SNPs identified were in close proximity to the structural genes of one of the measured proteins and corresponded to prior evidence of eQTL associations. For example, SNP rs12938, located in the 3' UTR of *SELL*, demonstrated high-magnitude loadings for sL-selectin across all race/ethnicities. Beyond structural genes for the measured proteins, some distal loci also demonstrated multivariate associations with adhesion proteins, such as *ABO*.

ABO blood type has been widely studied for risk association with atherosclerosis(Gong et al. 2014) and CVD(He et al. 2012; Medalie et al. 1971; Saha et al. 1973; von Beckerath et al. 2004; Zhang et al. 2012) and is most well-known to be strongly associated with circulation of von Willebrand Factor, a procoagulant with 22–30% reduced levels in Type O subjects relative to non-O individuals(Souto et al. 2000; Tirado et al. 2005). This glycoprotein is recognized to be directly modified by ABO glycosyltransferases, and Type O subjects have been shown to be at reduced risk for venous thromboembolism(Ohira et al. 2007; Sode et al. 2013; Wiggins et al. 2009; Zakai et al. 2014). Recent GWASs(Barbalic et al. 2010; Chen et al. 2015; Kiechl et al. 2011) have also identified multiple *ABO* SNP associations with circulating levels of a number of markers of endothelial function, including sP-selectin and sICAM-1. Mechanistically, it has been suggested that ABO glycosyltransferases may regulate endothelial markers through cleavage or proteolysis, since *ABO* associations with sP-selectin have not been replicated for platelet-bound

levels(Barbalic et al. 2010). Our mpQTL findings are consistent with these findings while also providing additional context of other protein effects across the cellular adhesion pathway. Specifically, SNPs perfectly tagging the A1 allele (e.g., rs507666) corresponded to lower sP-selectin and sE-cadherin levels across all races/ethnicities. Moreover, in AFA ($P = 3.4E-06$), we observed evidence of increased levels of RANTES and TGF β -1, while HIS participants exhibited increased circulation of SLPI. These results, like others(Chen et al. 2015; Kiechl et al. 2011), seemingly contradict the hypothesis that sP-selectin may mediate the association between ABO blood type and atherosclerotic CVD, as increased sP-selectin has been shown to correlated with higher CVD risk(Ridker et al. 2001). However, potential explanations have been proffered that accommodate this apparently paradoxical relationship(Kiechl et al. 2011). While underlying population differences in overall ABO allele frequencies may contribute to the significant differences we observed in cellular adhesion protein distributions by race/ethnicity, effect heterogeneity of *ABO* alleles across populations may also be a contributing factor. Further research is necessary to understand the molecular mechanisms underlying these racial/ethnic differences and their role in atherogenesis.

ABO and *FUT2* SNP associations with the canonical variates also indicated sizable loadings for sE-cadherin. E-cadherin plays a critical role in epithelial cell-cell adhesion, and its soluble form is a biomarker for inflammatory response. Lower circulation of sE-cadherin levels is also indicative of reduced cellular adhesion and increased cellular motility, and loss of E-cadherin expression is commonly observed in many malignant cancers(van Roy and Berx 2008). *FUT2* encodes the enzyme galactoside 2-alpha-L-fucosyltransferase 2 (FUT2), a Golgi stack membrane protein that contributes to the synthesis of an H antigen precursor and of relevance to the Lewis blood group system. Subjects with at least one functional *FUT2* allele are referred to phenotypically as secretors, characterized by the additional presence of ABH antigens in plasma other bodily fluids(Slomiany and Slomiany 1978). Our findings for both CHN and our meta-analysis demonstrate independent associations of population-specific functional *FUT2* knockout variants, lending credence to the non-secretor phenotype playing a large role in E-cadherin circulation. E-cadherin is one of many proteins targeted for glycosylation, and sugar remodeling is critical for E-cadherin functional regulation (Zhao et al. 2008). Moreover, core N-glycan fucosylation of E-cadherin by fucosyltransferase FUT8 has been shown to be critical to regulation of E-cadherin expression and function(Osumi et al. 2009). Oligosaccharide modification by either ABO glycosyltransferases or FUT2 may similarly alter E-cadherin and impact cellular adhesion.

The SNP rs17074898, significantly associated with adhesion proteins levels in EUR ($P = 1.7E-08$), corresponded to high-magnitude loadings across a large number proteins, particularly sP-selectin, TIMP-2, and SDF-1 α . This SNP is located approximately 24 kb upstream of the gene *ALOX5AP* and has been previously identified as a *cis*-eQTL for that gene in whole blood(Westra et al. 2013). *ALOX5AP* encodes the protein FLAP, the activating protein for 5-lipoxygenase (5-LOX), which in turn is necessary for leukotriene biosynthesis(Vickers 1995). Leukotrienes are inflammatory mediators with evidence of involvement in atherosclerosis(Capra et al. 2013; Haeggstrom and Funk 2011). The *ALOX5AP* locus itself has also been genetically associated with an increased risk of stroke

and/or myocardial infarction in multiple studies(Helgadottir et al. 2005; Helgadottir et al. 2004; Kajimoto et al. 2005), and FLAP is a therapeutic target for asthma and CVD(Evans et al. 2008). Our findings in EUR suggest FLAP dysregulation may have broad downstream regulatory effects on the cellular adhesion pathway and provides additional mechanistic information linking FLAP with CVD. The functional context of EUR-associated SNP rs7521237, displaying a similar protein association pattern to rs17074898, is less clear. An intronic variant within uncharacterized protein gene *KIAA1614*, rs7521237 has been previously reported to be a *cis*-eQTL of neighboring gene *STX6* (syntaxin-6) in brain tissues(Zou et al. 2012). Syntaxin-6 has been shown to play an important role in angiogenesis by regulating trafficking of VEGFR2 and $\alpha 5\beta 1$ integrin within endothelial cells(Jung et al. 2012), which in turn is regulated by cholesterol levels within the trans-Golgi network(Reverter et al. 2014). Although rs17074898 was also evaluated in HIS, this SNP was not genome-wide significant ($P = 0.05$). Similarly, SNP rs7521237 was evaluated for mpQTL association in AFA (MAF = 0.24) and found to be non-significant ($P = 0.95$). The population-specificity of these associations in non-Hispanic white participants may indicate effect modifying exposures and/or differences in underlying local linkage disequilibrium patterns, and validation in an independent cohort will be necessary.

There were no genome-wide significant mpQTL associations in African Americans despite relatively comparable sample size and protein measurement distributions to other racial/ethnic strata. Examination of AFA mpQTL association results for independent SNPs present in Table 2, significant in other races/ethnicities, revealed enrichment for lower p-values (Fisher's combined $P < 1e-16$) and comparable protein loading patterns for SNPs with $P_{AFA} < 0.05$. This indicates many of these effects may be present but at modest influences in African Americans, although the biological reasons for this are not immediately clear.

There are many notable strengths to this study, including the extensive amount of genotype data collected, the high quality and simultaneous assessment of the assayed proteins, and the well-studied multi-ethnic cohort. Multiple associations were also reproduced across race/ethnicity in our stratified analyses, providing internal replication of mpQTL loci for the cellular adhesion pathway. However, the numbers of participants in each race/ethnicity stratum, ranging from $N = 516$ to $N = 611$, are relatively low by commonly accepted GWAS standards and precluded evaluation of low-frequency/rare variation. Consequently, any modest effects from common variants and/or high impact rare variants will have gone undetected. Finally, additional independent replication of race/ethnicity-specific associations is necessary.

In summary, our mpQTL analysis of circulating levels of proteins in the cellular adhesion pathway revealed multiple significant associations of biological relevance while reproducing many previously identified univariate associations. Many findings indicated potential race/ethnicity-specific effects and/or allelic heterogeneity, notably blood antigen loci, and may be of particular interest to resolving population disparities in cardiovascular disease. Further research into the molecular mechanisms behind the modulation of adhesion pathway protein levels may improve our understanding cellular adhesion and cardiovascular risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Cardiometabochip genotyping data was supported in part by grants and contracts R01HL98077, N02-HL-64278, HL071205, UL1TR000124, DK063491, RD831697, and P50 ES015915. Funding for CARE IBC chip genotyping was provided by NHLBI Contract N01-HC-65226. Although the research described in this presentation has been funded in part by the United States Environmental Protection Agency through RD831697 to the University of Washington, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred. Funding for adhesion protein levels was provided by NHLBI by grant R01HL98077. 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 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-001079, UL1-TR-000040, and DK063491.

The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at <http://www.mesa-nhlbi.org>.

References

- Barbalic M, et al. Large-scale genomic studies reveal central role of ABO in sP-selectin and sICAM-1 levels. *Hum Mol Genet.* 2010; 19:1863–1872.10.1093/hmg/ddq061 [PubMed: 20167578]
- Bielinski SJ, et al. P-selectin and subclinical and clinical atherosclerosis: the Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis.* 2015; 240:3–9.10.1016/j.atherosclerosis.2015.02.036 [PubMed: 25744700]
- Bild DE, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol.* 2002; 156:871–881. [PubMed: 12397006]
- Bild DE, et al. Ethnic differences in coronary calcification: the Multi-Ethnic Study of Atherosclerosis (MESA). *Circulation.* 2005; 111:1313–1320.10.1161/01.CIR.0000157730.94423.4B [PubMed: 15769774]
- Capra V, Back M, Barbieri SS, Camera M, Tremoli E, Rovati GE. Eicosanoids and Their Drugs in Cardiovascular Diseases: Focus on Atherosclerosis and Stroke. *Med Res Rev.* 2013; 33:364–438.10.1002/med.21251 [PubMed: 22434418]
- Cesari M, et al. Inflammatory markers and onset of cardiovascular events: results from the Health ABC study. *Circulation.* 2003; 108:2317–2322.10.1161/01.CIR.0000097109.90783.FC [PubMed: 14568895]
- Chen Y, et al. Association of ABO blood group with P-selectin levels in Chinese Han healthy volunteers. *Transfusion.* 201510.1111/trf.13212
- Demerath E, Towne B, Blangero J, Siervogel RM. The relationship of soluble ICAM-1, VCAM-1, P-selectin and E-selectin to cardiovascular disease risk factors in healthy men and women. *Ann Hum Biol.* 2001; 28:664–678. [PubMed: 11726042]
- Evans JF, Ferguson AD, Mosley RT, Hutchinson JH. What's all the FLAP about?: 5-lipoxygenase-activating protein inhibitors for inflammatory diseases. *Trends Pharmacol Sci.* 2008; 29:72–78.10.1016/j.tips.2007.11.006 [PubMed: 18187210]
- Ferreira MA, Purcell SM. A multivariate test of association. *Bioinformatics.* 2009; 25:132–133.10.1093/bioinformatics/btn563 [PubMed: 19019849]
- Galesloot TE, van Steen K, Kiemeneij LA, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. *PLoS One.* 2014; 9:e95923.10.1371/journal.pone.0095923 [PubMed: 24763738]
- Gong P, et al. Relation of ABO blood groups to the severity of coronary atherosclerosis: an Gensini score assessment. *Atherosclerosis.* 2014; 237:748–753.10.1016/j.atherosclerosis.2014.10.107 [PubMed: 25463115]

- Haeggstrom JZ, Funk CD. Lipoxygenase and Leukotriene Pathways: Biochemistry, Biology, and Roles in Disease. *Chem Rev.* 2011; 111:5866–5898.10.1021/cr200246d [PubMed: 21936577]
- He M, Wolpin B, Rexrode K, Manson JE, Rimm E, Hu FB, Qi L. ABO blood group and risk of coronary heart disease in two prospective cohort studies. *Arterioscler Thromb Vasc Biol.* 2012; 32:2314–2320.10.1161/ATVBAHA.112.248757 [PubMed: 22895671]
- Helgadottir A, et al. Association between the gene encoding 5-lipoxygenase-activating protein and stroke replicated in a Scottish population. *Am J Hum Genet.* 2005; 76:505–509.10.1086/428066 [PubMed: 15640973]
- Helgadottir A, et al. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet.* 2004; 36:233–239.10.1038/ng1311 [PubMed: 14770184]
- Huyghe JR, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet.* 2013; 45:197–201.10.1038/ng.2507 [PubMed: 23263489]
- Hwang SJ, Ballantyne CM, Sharrett AR, Smith LC, Davis CE, Gotto AM Jr, Boerwinkle E. Circulating adhesion molecules VCAM-1, ICAM-1, and E-selectin in carotid atherosclerosis and incident coronary heart disease cases: the Atherosclerosis Risk In Communities (ARIC) study. *Circulation.* 1997; 96:4219–4225. [PubMed: 9416885]
- Jung JJ, Inamdar SM, Tiwari A, Choudhury A. Regulation of intracellular membrane trafficking and cell dynamics by syntaxin-6. *Biosci Rep.* 2012; 32:383–391.10.1042/BSR20120006 [PubMed: 22489884]
- Kajimoto K, et al. Validation of the association between the gene encoding 5-lipoxygenase-activating protein and myocardial infarction in a Japanese population. *Circ J.* 2005; 69:1029–1034. [PubMed: 16127181]
- Keating BJ, et al. Concept, Design and Implementation of a Cardiovascular Gene-Centric 50 K SNP Array for Large-Scale Genomic Association Studies. *PLoS One.* 2008; 310.1371/Journal.Pone.0003583
- Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB. Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J Biol Chem.* 1995; 270:4640–4649. [PubMed: 7876235]
- Kiechl S, et al. Association of variation at the ABO locus with circulating levels of soluble intercellular adhesion molecule-1, soluble P-selectin, and soluble E-selectin: a meta-analysis. *Circ Cardiovasc Genet.* 2011; 4:681–686.10.1161/CIRCGENETICS.111.960682 [PubMed: 22010135]
- Medalie JH, et al. Blood groups, myocardial infarction and angina pectoris among 10,000 adult males. *N Engl J Med.* 1971; 285:1348–1353.10.1056/NEJM197112092852404 [PubMed: 5001056]
- O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FC, Elliott P, Jarvelin MR, Coin LJ. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One.* 2012; 7:e34861.10.1371/journal.pone.0034861 [PubMed: 22567092]
- Ohira T, et al. ABO blood group, other risk factors and incidence of venous thromboembolism: the Longitudinal Investigation of Thromboembolism Etiology (LITE). *J Thromb Haemost.* 2007; 5:1455–1461.10.1111/j.1538-7836.2007.02579.x [PubMed: 17425663]
- Osumi D, et al. Core fucosylation of E-cadherin enhances cell-cell adhesion in human colon carcinoma WiDr cells. *Cancer Sci.* 2009; 100:888–895.10.1111/j.1349-7006.2009.01125.x [PubMed: 19302290]
- Pare G, et al. Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. *PLoS Genet.* 2008; 4:e1000118.10.1371/journal.pgen.1000118 [PubMed: 18604267]
- Pare G, et al. Genome-wide association analysis of soluble ICAM-1 concentration reveals novel associations at the NFKB1K, PNPLA3, RELA, and SH2B3 loci. *PLoS Genet.* 2011; 7:e1001374.10.1371/journal.pgen.1001374 [PubMed: 21533024]
- Paterson AD, et al. Genome-wide association identifies the ABO blood group as a major locus associated with serum levels of soluble E-selectin. *Arterioscler Thromb Vasc Biol.* 2009; 29:1958–1967.10.1161/ATVBAHA.109.192971 [PubMed: 19729612]

- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–909.10.1038/ng1847 [PubMed: 16862161]
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155:945–959. [PubMed: 10835412]
- Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575.10.1086/519795 [PubMed: 17701901]
- Reiner AP, et al. Genome-wide and gene-centric analyses of circulating myeloperoxidase levels in the charge and care consortia. *Hum Mol Genet.* 2013; 22:3381–3393.10.1093/hmg/ddt189 [PubMed: 23620142]
- Reverter M, et al. Cholesterol Regulates Syntaxin 6 Trafficking at trans-Golgi Network Endosomal Boundaries. *Cell Rep.* 2014; 7:883–897.10.1016/j.celrep.2014.03.043 [PubMed: 24746815]
- Ridker PM, Buring JE, Rifai N. Soluble P-selectin and the risk of future cardiovascular events. *Circulation.* 2001; 103:491–495. [PubMed: 11157711]
- Saha N, Toh CC, Ghosh MB. Genetic association in myocardial infarction. Ethnicity; ABO, Rh, Le, Xg blood groups; G6PD deficiency; and abnormal haemoglobins. *J Med Genet.* 1973; 10:340–345. [PubMed: 4204387]
- Slomiany BL, Slomiany A. ABH-blood-group antigens and glycolipids of human saliva. *Eur J Biochem.* 1978; 85:249–254. [PubMed: 639818]
- Sode BF, Allin KH, Dahl M, Gynzelberg F, Nordestgaard BG. Risk of venous thromboembolism and myocardial infarction associated with factor V Leiden and prothrombin mutations and blood type. *Cmaj.* 2013; 185:E229–237.10.1503/cmaj.121636 [PubMed: 23382263]
- Souto JC, et al. Functional effects of the ABO locus polymorphism on plasma levels of von Willebrand factor, factor VIII, and activated partial thromboplastin time. *Arterioscl Throm Vas.* 2000; 20:2024–2028.
- Stephens M. A unified framework for association analysis with multiple related phenotypes. *PLoS One.* 2013; 8:e65245.10.1371/journal.pone.0065245 [PubMed: 23861737]
- Stouffer, SA. The American soldier. Studies in social psychology in World War II. Vol. 1–2. Princeton University Press; Princeton: 1949.
- Tirado I, et al. The ABO blood group genotype and factor VIII levels as independent risk factors for venous thromboembolism. *Thromb Haemost.* 2005; 93:468–474. [PubMed: 15735796]
- van Roy F, Berx G. The cell-cell adhesion molecule E-cadherin. *Cell Mol Life Sci.* 2008; 65:3756–3788.10.1007/s00018-008-8281-1 [PubMed: 18726070]
- Vickers PJ. 5-Lipoxygenase-activating protein (FLAP). *J Lipid Mediat Cell Signal.* 1995; 12:185–194. [PubMed: 8777565]
- Voight BF, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *Plos Genet.* 2012; 8:e1002793.10.1371/journal.pgen.1002793 [PubMed: 22876189]
- von Beckerath N, Koch W, Mehilli J, Gorchakova O, Braun S, Schomig A, Kastrati A. ABO locus O1 allele and risk of myocardial infarction. *Blood Coagul Fibrinolysis.* 2004; 15:61–67. [PubMed: 15166945]
- Westra HJ, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013; 45:1238–1243.10.1038/ng.2756 [PubMed: 24013639]
- Wiggins KL, et al. ABO genotype and risk of thrombotic events and hemorrhagic stroke. *J Thromb Haemost.* 2009; 7:263–269.10.1111/j.1538-7836.2008.03243.x [PubMed: 19036074]
- Zakai NA, Judd SE, Alexander K, McClure LA, Kissela BM, Howard G, Cushman M. ABO blood type and stroke risk: the REasons for Geographic And Racial Differences in Stroke Study. *J Thromb Haemost.* 2014; 12:564–570.10.1111/jth.12507 [PubMed: 24444093]
- Zhang H, Mooney CJ, Reilly MP. ABO Blood Groups and Cardiovascular Diseases. *Int J Vasc Med.* 2012; 2012:641917.10.1155/2012/641917 [PubMed: 23133757]
- Zhao Y, et al. Branched N-glycans regulate the biological functions of integrins and cadherins. *Febs J.* 2008; 275:1939–1948.10.1111/j.1742-4658.2008.06346.x [PubMed: 18384383]

Zou F, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *Plos Genet.* 2012; 8:e1002707.10.1371/journal.pgen.1002707 [PubMed: 22685416]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig 1.

R. Manhattan plot of meta-analysis multivariate GWAS p-values derived using Stouffer's combined Z approach. Significant and suggestive loci are labeled by nearby protein coding genes likely relevant to the assayed proteins.

Fig 2.

R. Barplot depicting MQFAM protein loadings for EUR-associated mpQTL SNPs rs17074898 (*ALOX5AP*) and rs7521237 (*KIAA1614*). Statistical significance of each univariate protein SNP association is indicated next to each bar and was determined by a Wald test using linear regression on log-transformed protein levels, adjusting for age, sex, and ancestry informative PCs.

Fig 3.

R. Heatmaps of MQFAM protein loadings for all *ABO* SNPs that were designated as genome-wide significant in at least one race/ethnicity. Results are presented for Chinese, non-Hispanic White, and Hispanic Americans participants, and rows (SNPs) are ordered by physical position within the *ABO* gene. Linkage disequilibrium (R^2) values computed from EUR SNP data.

Table 1
Distribution summary of 15 analyzed cellular adhesion proteins by race/ethnicity.

Protein	Mean (SD) Protein Levels by Race/Ethnicity				P*
	AFA N = 516	CHN N = 570	EUR N = 611	HIS N = 571	
CCL-21 (pg/mL)	804 (277)	779 (281)	843 (280)	816 (283)	<1e-05
E-cadherin (ng/mL)	208 (56)	227 (64)	216 (59)	247 (75)	<1e-05
HGF (pg/mL)	979 (244)	868 (202)	1003 (253)	1101 (261)	<1e-05
ICAM-1 (ng/mL)	248 (83)	228 (76)	251 (77)	281 (96)	<1e-05
IL-2sr (pg/mL)	642 (263)	591 (229)	838 (348)	774 (289)	<1e-05
L-selectin (ng/mL)	862 (193)	836 (175)	956 (215)	905 (190)	<1e-05
MMP-1 (ng/mL)	6.0 (4.7)	5.5 (3.6)	5.4 (3.9)	5.5 (4.3)	0.16
MMP-2 (ng/mL)	199 (34)	190 (30)	196 (30)	198 (31)	<1e-05
P-selectin (ng/mL)	28.3 (9.0)	27.9 (8.6)	28.5 (9.4)	31.9 (9.8)	<1e-05
RANTES (pg/mL)	4383 (4684)	4405 (3352)	4191 (3987)	4322 (4319)	0.0020
SDF-1α (pg/mL)	1921 (463)	2004 (436)	1908 (434)	1942 (448)	2.6E-04
SLPI (pg/mL)	47362 (12124)	45030 (10298)	45517 (10085)	47518 (10722)	<1e-05
TGFβ-1 (pg/mL)	3924 (2559)	4386 (2244)	3714 (2171)	3889 (2265)	<1e-05
TIMP-2 (ng/mL)	82.1 (13.2)	77.1 (11.0)	82.9 (12.9)	82.0 (11.7)	<1e-05
VCAM-1 (ng/mL)	662 (213)	743 (229)	785 (224)	757 (237)	<1e-05

* P-value from likelihood ratio test for significance of race/ethnicity terms in linear regression of log-transformed protein values, adjusting for age and sex

Significant race/ethnicity-stratified multivariate genetic associations with circulating cellular adhesion protein levels, by locus (i.e., nearest gene(s)), along with top significant SNP and corresponding p-value. For each association, all proteins with loadings $|w_p| > 0.25$ are listed.

Table 2

Locus	Chr	Race/Ethnicity	Top SNP	P	Protein(s)
<i>SELP/SELL</i>	1	CHN	rs2223286	4.8E-08	E-cadherin L-selectin
		EUR	rs12938	2.8E-17	L-selectin
		EUR	rs6136	5.7E-17	P-selectin
<i>KIAA1614</i>	1	EUR	rs7521237	2.2E-08	P-selectin SDF1- α TIMP-2
<i>ABO</i>	9	CHN	rs651007	1.1E-21	E-cadherin ICAM-1 P-selectin
		EUR	rs651007	2.5E-17	E-cadherin P-selectin
		HIS	rs635634	5.1E-17	E-cadherin P-selectin SLPI
<i>IL2RA</i>	10	HIS	rs10905876	3.6E-12	CCL21 IL2-sr
<i>MMP1</i>	11	CHN	rs529381	8.0E-21	MMP-1
		EUR	rs484915	2.1E-14	MMP-1
		HIS	rs479095	1.2E-11	MMP-1 MMP-2 TIMP-2
<i>ALOX5AP</i>	13	EUR	rs17074898	1.7E-08	E-cadherin HGF P-selectin SDF1- α TIMP-2
<i>CCL5</i>	17	CHN	rs3760327	1.1E-10	RANTES
<i>ICAM1</i>	19	CHN	rs3093030	1.7E-09	ICAM-1
<i>FUT2</i>	19	CHN	rs1047781	4.3E-12	E-cadherin