Genome-wide association study identifies seven novel loci associating with circulating cytokines and cell adhesion molecules in Finns

Eeva Sliz[1,2,3], Marita Kalaoja[1,2,3], Ari Ahola-Olli[4,5], Olli Raitakari[5,6], Markus Perola[7,8,9], Veikko Salomaa[7], Terho Lehtimäki[10], Toni Karhu[3,11], Heimo Viinamäki[12], Marko Salmi[13], Kristiina Santalahti[13], Sirpa Jalkanen[13], Jari Jokelainen[2,14], Sirkka Keinänen-Kiukaanniemi[2,14,15], Minna Männikkö[16], Karl-Heinz Herzig[3,11,17,18], Marjo-Riitta Järvelin[2,3,19,20], Sylvain Sebert[2,3,21,*], Johannes Kettunen[1,2,3,*]

*these authors contributed equally

1. Computational Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland
2. Center for Life Course Health Research, University of Oulu, Oulu, Finland
3. Biocenter Oulu, Oulu, Finland
4. Department of Internal Medicine, Satakunta Central Hospital, Pori, Finland.
5. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
6. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
7. National Institute for Health and Welfare, Helsinki, Finland
8. Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
9. University of Tartu, Estonian Genome Center, Tartu, Estonia
10. Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
11. Institute of Biomedicine, University of Oulu, Oulu, Finland
12. Department of Psychiatry, University of Eastern Finland, and Kuopio University Hospital, Kuopio, Finland
13. Medicity Research Laboratory and Institute of Biomedicine, University of Turku, Turku, Finland
14. Unit of General Practice, Oulu University Hospital, Oulu, Finland
15. Oulu Deaconess Institute/Diapolis Oy Research Unit, Oulu, Finland

16. Northern Finland Birth Cohorts, Faculty of Medicine, University of Oulu, Oulu, Finland
17. Medical Research Center (MRC), University of Oulu, and Oulu University Hospital, Oulu, Finland
18. Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, Poznan, Poland
19. Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, Imperial College London, London, UK
20. Unit of Primary Care, Oulu University Hospital, Oulu, Finland
21. Department of Genomics and Complex Diseases, School of Public Health, Imperial College London, UK

Corresponding author:

Johannes Kettunen

Address: Faculty of Medicine, Computational Medicine, Center For Life Course Health Research, Aapistie 5A, P.O.Box 5000, 90014 University of Oulu, Oulu, Finland

Phone: +358 50 562 9718

E-mail: johannes.kettunen@oulu.fi

Abbreviations: GWAS, genome-wide association study; IL1α, interleukin 1-alpha; IL1β, interleukin 1-beta; IL1ra, interleukin 1 receptor antagonist; IL4, interleukin 4; IL6, interleukin 6; IL8, interleukin 8; IL17, interleukin 17; IP10, interferon gamma-induced protein 10; MAF, minor allele frequency; MCP1, monocyte chemoattractant protein 1; NFBC1966, Northern Finland Birth Cohort 1966; PAI-1, plasminogen activator inhibitor 1; sCD40L, soluble CD40 ligand; sE-selectin, soluble E-selectin; sICAM-1, soluble intercellular cell adhesion molecule 1; sVCAM-1, soluble vascular cell adhesion molecule 1; TNFα, tumor necrosis factor alpha; VEGF, vascular endothelial growth factor

Abstract

Background: Inflammatory processes contribute to the pathophysiology of multiple chronic conditions. Genetic factors play a crucial role in modulating the inflammatory load, but the exact mechanisms are incompletely understood.

Methods: To add understanding to the molecular mechanisms in inflammation, we performed a genome-wide association study (GWAS) on 16 circulating cytokines and cell adhesion molecules (inflammatory phenotypes) in Northern Finland Birth Cohort 1966 (NFBC1966, N=5,284). A subsequent meta-analysis was completed for 10 phenotypes available in a GWAS of three other Finnish population cohorts adding up to 13,577 individuals in the study. Complementary association tests were performed to study the effect of the ABO blood types on soluble adhesion molecule levels.

Results: We identified seven novel and confirmed six previously reported loci associating with at least one of the studied inflammatory phenotypes ($p<3.1x10^{-9}$). We observed three loci associating with the concentration of soluble vascular cell adhesion molecule-1 (sVCAM-1), one of which is the *ABO* locus that has been previously associated with soluble E-selectin (sE-selectin) and intercellular adhesion molecule-1 (sICAM-1) levels. Results from the complementary analyses suggest that the blood type B associates primarily with the concentration of sVCAM-1 while the A1 subtype shows a robust effect on sE-selectin and sICAM-1 levels. Furthermore, the genotypes in the *ABO* locus associating with higher soluble adhesion molecule levels tend to associate with lower low-density lipoprotein cholesterol level and lower cardiovascular disease risk.

Conclusion: The present results extend the knowledge about genetic factors contributing to the inflammatory load. Our findings suggest that two distinct mechanisms contribute to the soluble adhesion molecule levels at the *ABO* locus. The negative correlation between the genetic effects on soluble adhesion molecule levels and cardiovascular traits in this locus further suggests that increased soluble adhesion molecule levels per se may not be a risk factor for cardiovascular disease.

Introduction

It is currently established that inflammatory load may play a role in the etiology of autoimmune and infectious diseases, but also in a broad range of other diseases, such as chronic cardio-metabolic disorders (1), neurodegenerative diseases (2), and cancer (3). The risk for these diseases increases with age (4), and due to the world's aging population (5), their prevalence is likely to expand. Moreover, these diseases often co-occur which is likely due to shared inflammation related pathophysiology (6).

Inflammation is the body's physiological response to harmful stimuli involving multiple molecular and cellular interactions attempting to restore disturbances in tissue or systemic homeostasis (7). Circulating cytokines, growth factors, chemokines, and cell adhesion molecules (hereafter inflammatory phenotypes) are fundamental mediators of inflammatory responses. Genes encoding these molecules and their receptors play a crucial role in mediating the related functions. Previous studies have identified loci associating with levels of inflammatory phenotypes (8–10), but the understanding of the exact regulatory mechanisms is still incomplete.

To add insights to the genetic mechanisms contributing to the inflammatory load, we performed a genome-wide association study (GWAS) of 16 circulating inflammatory phenotypes in 5,284 individuals from Northern Finland Birth Cohort 1966 (NFBC1966), and a subsequent meta-analysis of 10 phenotypes in three other Finnish population cohorts (8) adding up to a total of 13,577 individuals in the study. We report identification of seven novel and replication of six loci associating with levels of the circulating inflammatory markers.

Methods

Study populations, genotyping and inflammatory phenotype quantification

Northern Finland Birth Cohort 1966

The Northern Finland Birth Cohort 1966 (NFBC1966) was initiated to study factors affecting preterm birth, low birth weight, and subsequent morbidity and mortality (*www.oulu.fi/nfbc*). It comprises 96% of all births during 1966 in the two northernmost provinces in Finland; altogether 12,058 children were live-born into the cohort, and the follow-ups occurred at the ages of 1, 14, 31, and 46 years (11, 12). The data analyzed in the present study is from the 31-years follow-up when clinical examinations and blood sampling was completed for altogether 6,033 individuals, 5,284 of whom had body mass index, inflammatory phenotypes and genotype data available (a maximum number of individuals per inflammatory marker 5,100). Genotyping of the samples was completed using 370k Illumina HumanHap arrays (Illumina Inc., CA, USA) and subsequent imputation was performed based on the 1000 Genome reference panel. A total of 16 inflammatory phenotypes were quantified from overnight fasting plasma samples using Bio-Rad's Bio-Plex 200 system (Bio-Rad Laboratories Inc., CA, USA) with Milliplex human chemokine/cytokine and CVD/cytokine kits (Cat# HCYTOMAG-60K-12 and Cat# SPR349; Millipore, St Charles, MO, USA) and Bio-Plex Manager Software version 4.3 as previously described (13). The 16 inflammatory phenotypes studied in the NFBC1966 were interleukin (IL) 1-alpha (IL1α), IL1-beta (IL1β), IL4, IL6, IL8, IL17, IL1 receptor antagonist (IL1ra), interferon gamma-induced protein 10 (IP10), monocyte chemoattractant protein 1 (MCP1), tumor necrosis factor alpha (TNFα), vascular endothelial growth factor (VEGF), plasminogen activator inhibitor 1 (PAI-1), soluble CD40 ligand (sCD40L), soluble E-selectin (sE-selectin), soluble intercellular adhesion molecule 1 (sICAM-1) and soluble vascular cell adhesion molecule 1 (sVCAM-1).

GWAS summary statistics from three Finnish population cohorts

Meta-analyses were conducted for 10 phenotypes available in a previous GWAS (8). The study included up to 8,293 Finnish individuals from The Cardiovascular Risk in Young Finns Study (YFS) (14) and FINRISK (*www.thl.fi/finriski*) (15) adding up to 13,577 individuals studied in the present meta-analyses. Shortly, YFS is

a population-based follow-up study started in 1980 comprising randomly chosen individuals from Finnish cities Helsinki, Kuopio, Tampere, Oulu, and Turku. The YFS data included in the previous GWAS is from 2,019 individuals who participated in the follow-up in 2007 and who had both inflammatory phenotype and genotype data available. FINRISK is a Finnish population survey conducted every five years to monitor chronic diseases and their risk factors. The surveys use independent, random, and representative samples from different geographical areas of Finland. The data included in the present meta-analyses were from participants of the 1997 and 2002 surveys. Genotypes were obtained using 670k Illumina HumanHap arrays (Illumina Inc., CA, USA) and imputed based on 1000 Genome reference panel. Inflammatory markers were quantified using Bio-Rad's premixed Bio-Plex Pro Human Cytokine 27-plex Assay and 21-plex Assay, and Bio-Plex 200 reader with Bio-Plex 6.0 software (Bio-Rad Laboratories Inc., CA, USA) as previously described (16, 17). Samples were serum in YFS, EDTA plasma in FINRISK1997, and heparin plasma in FINRISK2002.

Statistical analyses

GWAS and meta-analysis

To allow meta-analysis between the present results and the previous GWAS, the data processing and analysis model were done according to Ahola-Olli *et al*. (8): Preceding the analyses, linear regression models were fitted to adjust the inflammatory phenotypes for age, sex, BMI, and ten first genetic principal components to control for population stratification. The resulting residuals were normalized with inverse-rank based transformation, and the adjusted and transformed residuals were used as phenotypes in the analyses. Genome-wide association tests were performed using snptest_v2.5.1 software (18, 19). Allele effects were estimated using an additive model (-frequentist 1) and the option to center and scale the phenotypes was disabled (-use_raw_phenotypes). The GWAS results were filtered by including markers with model fit info > 0.8 and minor allele count > 10. Filtered data was used to perform meta-analyses by METAL software (v.2011-03-25) (20) for the 10 phenotypes (IL1β, IL1ra, IL17, IL4, IL6, IL8, IP10, MCP1, TNFα, and VEGF) available in the previous GWAS (8). Genomic control correction was enabled to account for population stratification and cryptic relatedness.

140 *Supplemental genome-wide tests in NFBC1966*

141 Individuals showing symptoms of an acute infection were omitted from the supplemental genome-wide tests
142 performed in the NFBC1966 population. Here, individuals reported having fever at the time of the blood
143 sampling and individuals having C-reactive protein (CRP) level > 10 mg/l were excluded. Otherwise the
144 analysis models were as above.

145 *Conditional analyses and variance explained*

146 To assess whether the identified loci harbor multiple independent association signals, we conducted
147 conditional analyses by further adjusting the models with the locus-specific lead variants. The association
148 tests were repeated within a 2Mb window around the lead SNP for the phenotypes studied in the NFBC1966
149 population only. For the meta-analyzed phenotypes, we applied a method proposed by Yang *et al*. that
150 enables conditional analyses of GWAS summary statistics (21, 22). NFBC1966 was used as a reference sample
151 to estimate linkage disequilibrium (LD) corrections in these analyses. The proportion of variance explained
152 was calculated using all independent variants using the following formula:

153 $$Variance\ explained = (\beta\, x\, \sqrt{2\, x\, MAF(1 - MAF)})^2$$

154 Here β is the variant's effect estimate on the inflammatory phenotype and MAF denotes minor allele
155 frequency.

156 *Complementary association tests on soluble adhesion molecule levels*

157 Complementary association tests within a 2Mb window were conducted to better evaluate the effect of the
158 ABO blood type on the association with soluble adhesion molecule levels at the *ABO* locus. For sE-selectin,
159 sICAM-1 and sVCAM-1 levels, linear models were fitted by further adjusting for the ABO blood type or
160 rs507666 genotype tagging the A1 subtype (23).

161 In addition, we determined the effect estimates of ABO blood types and ABO blood types stratified by
162 rs507666 genotype on sE-selectin, sICAM-1 and sVCAM-1 levels in linear models. Here, adjusted and
163 transformed soluble adhesion molecule concentrations were as outcomes and ABO blood types as

164 categorical variables (blood type A versus non-A, etc.); corresponding models were fitted for the stratified
165 blood types (blood type A with rs507666 G/G versus others, etc.)

166 *Correlations of the genetic effects*

167 As previous evidence suggests that the elevated concentrations of circulating markers of inflammation
168 increase the risk of cardiovascular diseases (CVD) (24, 25), we further evaluated how variants at the loci
169 associating with inflammatory phenotypes may relate to other cardiovascular traits. We extracted SNP
170 effects on coronary artery disease (CAD) risk, stroke risk, and low-density lipoprotein cholesterol (LDL-C) or
171 high-density lipoprotein cholesterol (HDL-C) levels from open-access data provided by CARDIoGRAM (26),
172 Stroke Consortia (27), and a metabolomics GWAS (28). First, data were filtered to include only the SNPs
173 available in all the three data sets within a 1Mb window around each lead variant. Next, subsets of
174 representative SNPs at the each of the significant loci were extracted using clumping function in PLINK
175 1.90b4.1 (29). Here, NFBC1966 was used as the reference sample to construct LD structures and r^2=0.2 was
176 used as the LD threshold while other parameters were as by default. The subsets of SNPs were used to
177 determine the linear relationships of the genetic effects (Z-scores) on inflammatory phenotypes versus other
178 traits for each significant loci identified in the present GWAS.

179 ## Results

180 Basic characteristics of the NFBC1966 study population is provided in Table 1. Inflammatory phenotype
181 distributions are tabulated in Table S1 and their correlation structure is shown in Figure S1. Using a threshold
182 of $p<3.1\times10^{-9}$ for statistical significance (standard genome-wide significance level $p<5\times10^{-8}$ corrected for 16
183 phenotypes tested), we identified seven novel and six previously reported loci associating with one or more
184 of the inflammatory phenotypes. The results are summarized in Table 1 and combined Manhattan plots are
185 shown in Figure 1. Manhattan plots and Q-Q plots for each inflammatory phenotype are provided in the
186 supplement (Figure S2 A-Z). Genomic inflation factor values range between 0.99-1.02 suggesting no inflation
187 in the test statistics (Table S2). Table S3 lists traits associated previously with the loci showing novel
188 associations with inflammatory phenotypes in the present study.

Cell adhesion molecules

The ABO locus shows large effects on sE-selectin, sICAM-1, and sVCAM-1 levels

We observed a novel effect on sVCAM-1 concentration in 9q34.2 near *ABO* (ABO, alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase) in the NFBC1966 population. This locus showed a robust association also with sE-selectin and sICAM-1 concentrations as previously reported (23, 30, 31). Noteworthy, the lead variant for sE-selectin and sICAM-1 associations (rs2519093) was different from the lead variant for sVCAM-1 association (rs8176746). The former variant is in LD (r^2=1 in NFBC1966) with rs507666 tagging the ABO blood type A subtype A1 whereas the latter variant tags the blood type B (23).

To better evaluate if the ABO blood types constitute the molecular mechanism explaining the association between the *ABO* locus and soluble adhesion molecule levels, we completed supplementary association tests further adjusted for ABO blood type or rs507666 genotype indicative of the A1 subtype. The results of the supplementary tests suggested that the association of the rs8176746 with sVCAM-1 concentration is independent of the A1 subtype ($p=4.98\times10^{-15}$ for the rs507666-adjusted association). On the contrary, the associations of the rs2519093 with concentrations of sE-selectin and sVCAM-1 remained highly significant when adjusted for ABO blood type ($p=3.40\times10^{-123}$ and $p=3.43\times10^{-17}$, respectively). Statistical significances were abolished when rs8176746 association with sVCAM-1 was adjusted for ABO blood type and rs2519093 association with sE-selectin or sICAM-1 was adjusted for rs507666.

We further determined the effect estimates of the ABO blood types and ABO blood types stratified by rs507666 genotype on soluble adhesion molecule levels. The blood type A showed negative associations with the levels of all the three adhesion molecules and the effect was the most robust on the sE-selectin level (Figure 2, Panel 1). However, major discrepancies in the effect directions were seen when the analyses were stratified by the rs507666 genotype (Figure 2, Panel 2). Congruent with previous reports (23, 30), the present results suggest that the A1 subtype/rs507666 genotype influences sE-selectin or sVICAM-1 levels. In contrast,

213 the blood type B seems to attribute predominantly to sVCAM-1 level while the A1 subtype/rs507666 shows
214 only a modest effect on sVCAM-1.

215 *HSP90B1 and ABCA8 loci associate with sVCAM-1 levels*

216 We identified two other novel loci for sVCAM-1 (12q23.3 and 17q24.2) in the NFBC1966 population. In chr12
217 the lead variant rs117238625 is in LD (r^2=1 in NFBC1966) with rs117468318 locating in the 5' UTR region of
218 *HSP90B1* (heat shock protein 90kDa beta member 1) and, according to RegulomeDB (32), is likely to affect
219 transcription factor binding providing evidence for a possible regulatory mechanism. Variants in this locus
220 have been previously associated with stem cell growth factor beta levels (8) and corneal structure (33). The
221 association signal in chr17 locates near *ABCA8* (ATP binding cassette subfamily A member 8) encoding one
222 of the ATP binding cassette transporters. Other studies have identified associations in this locus with HDL
223 and LDL cholesterol levels (34–37), breast cancer risk (38), heart's electrical cycle related traits (QT interval,
224 QRS duration) (39–41).

225 *Variations in sialyltransferase encoding genes show an effect on sE-selectin level*

226 For sE-selectin level, we identified a novel association in 11q24.2 in the region of *ST3GAL4* (ST3 beta-
227 galactoside alpha-2,3-sialyltransferase 4). Other studies have associated other variants in this region with
228 mean platelet volume and platelet count (42), LDL cholesterol levels (34, 36) or pleiotropic associations with
229 LDL cholesterol and C-reactive protein (43), blood protein levels (44) and liver enzyme levels (45). We
230 identified a suggestive signal with sE-selectin level also in 3q12.1 near *ST3GAL6* (ST3 beta-galactoside alpha-
231 2,3-sialyltransferase 6), but the association was not significant after multiple correction ($p=1.75x10^{-08}$). Both
232 of the sialyltransferase genes have been implicated in the production of functional E-, P-, and L-selectin
233 ligands in mice (46).

234 *Two independent association signals on sICAM-1 level near ICAM1*

235 We replicated the previously reported association for sICAM-1 level in 19p13.2 near *ICAM1* (intracellular
236 adhesion molecule 1) (23, 44). When the primary association test was conditioned for the lead variant

237 rs117960796, another significant association was detected (rs74428614, $p=1.14x10^{-16}$) indicative of more
238 than one independent variant contributing to sICAM-1 level in this locus.

239 ## Vascular endothelial growth factor

240 In the meta-analyses, we identified a novel locus 4p16.2 with a large effect on VEGF (β=-2.38 SD). This locus
241 harbours genes *EVC* (EvC ciliary complex subunit 1), *EVC2* (EvC ciliary complex subunit 2), and *STK32B*
242 (serine/threonine kinase 32B). Mutations in this locus have been associated previously with Celiac disease
243 (47), coronary heart disease (48), essential tremor (49), and Ellis van Creveld syndrome (50, 51). In addition,
244 we replicated two previously reported loci associating with VEGF levels in 6p21.1 near *VEGFA* (vascular
245 endothelial growth factor A) and in 9p24.2 near *VLDLR* (very-low-density lipoprotein receptor) (8).

246 ## Pro-inflammatory cytokines

247 ### *Locus near DLEU1 shows a large effect on TNFα*

248 We identified a novel variant with a large effect on TNFα levels (β=2.13 SD) in 13q14.3 near *DLEU1* and *DLEU7*
249 (deleted in lymphocytic leukemia 1 and 7) in the meta-analyses. Previous GWAS findings in this locus include
250 associations with reticulocyte related traits (42) and tooth development (52).

251 ### *The HLA locus shows a small effect on IL-1β*

252 A novel variant at 6p22.1 in the human leukocyte antigen locus associating with IL-1β level was identified in
253 the meta-analyses. In the conditional analyses, we observed two independent association signals at this locus
254 (Table 1, Figure S2 J). The same locus and the same lead variant rs6917603 showed also a suggestive effect
255 on IL4 level (Figure S2 L), but the meta-analysed result was not significant after multiple correction
256 ($p=5.56x10^{-09}$). Variants in this region have been associated previously with multiple immune system related
257 traits such as white blood cell count (42), but also with several other traits, including lipid metabolism
258 phenotypes (53), schizophrenia (54) and lung cancer (55).

Chemokines

We replicated previously reported loci near *CXCL10* (C-X-C motif chemokine ligand 10) and *ACKR1* (atypical chemokine receptor 1) associating with IP10 levels and with MCP1 levels, respectively (8).

Supplemental genome-wide tests in NFBC1966

Altogether 236 individuals having fever or CRP>10 mg/l were excluded from the supplemental genome-wide tests performed in the NFBC1966 population. The results of the supplementary analyses were congruent with the original findings (Table S4).

Comparisons of SNP effects on inflammatory markers versus other traits

Elevated circulating concentrations of inflammatory markers is a risk factor for cardiovascular diseases (24, 25). In order to add insights how genetic variants associating with inflammatory phenotypes contribute to other cardiovascular health-related traits, we compared the SNP effects (Z-scores) at each significant locus versus corresponding SNP effects obtained from open-access data (26–28). The list of SNPs used to determine the correlations of the SNP effects on inflammatory phenotypes versus other traits is provided in Table S5. At the *ABO* locus, we observed a negative linear relationship between the SNP effects on sE-selectin and sICAM-1 levels and CAD risk, stroke risk as well as LDL-C and HDL-C levels (Figure 3A-B, Figure S3 A-B). For sVCAM-1 at the same locus, the overall trend was similar, but correlations did not reach statistical significance due to a low number of representative SNPs obtained in clumping (Figure 3C, Figure S3 C). The results at the other loci are provided in Figures S3 D-I.

Discussion

The present study examines genetic determinants of 16 circulating inflammatory phenotypes in 5,284 individuals from Northern Finland with a subsequent meta-analysis of 10 phenotypes in three other Finnish populations adding up to a total of 13,577 participants. We report seven novel and replication of six previously published genetic associations.

We detected a novel association for sVCAM-1 concentration at the *ABO* locus. This locus showed robust associations also with sE-selectin and sICAM-1 concentrations congruent with previous studies (23, 30, 31). The present GWAS findings suggested two distinct association signals at the *ABO* locus for the sE-selectin and sICAM-1 levels versus sVCAM-1 level. The results of the supplementary tests supported the perception that genetic variants in this locus may regulate the circulating concentration of adhesion molecules by at least two different mechanisms. The first mechanism comprises the blood type A subtype A1, tagged by rs507666-A, that has a robust lowering effect on sE-selectin and sICAM-1 levels (23, 30). The second mechanism involves the blood type B that seems to have an increasing effect on sVCAM-1 level. Others have suggested that the lowering effect of the A1 subtype on sE-selectin and sICAM-1 could arise from increased glycosyltransferase activity that possibly modifies the shedding of the adhesion molecules from the endothelium and/or their clearance rate from circulation (23, 30, 56). To the best of our knowledge, the underlying mechanism explaining the association between the blood type B and the higher sVCAM-1 concentration remains unknown and warrants research. VCAM-1-mediated adhesion involves interaction with galectin-3, a protein that has a specificity for galactosides (57, 58). As the B antigen holds an additional galactose monomer compared with the A and O antigens, and galectins are known to recognize blood type antigens (59), it raises the speculation that the amount of unbound sVCAM-1 in the circulation could be influenced by a possible competitive binding of galectin-3 with sVCAM-1 and the B antigen.

To evaluate how variants in the *ABO* locus may relate to other health outcomes, we compared the correspondence of genetic effects on soluble adhesion molecule levels versus cardiovascular health-related traits. We observed that the genetic effects on adhesion molecule levels were inclined to show a negative correlation with the genetic effects on LDL-C and HDL-C levels as well as lower risk for CAD and stroke. This denotes that the genotypes at the *ABO* locus associating with increased levels of soluble adhesion molecules tended to associate with lower circulating cholesterol level as well as lower risk of cardiovascular outcomes. This was unexpected since according to previous evidence increased soluble adhesion molecule levels are linked with atherosclerosis progression and vascular outcomes (25, 60, 61). Possible explanations unravelling the negative correlation advocate that soluble adhesion molecules may compete with leukocyte adhesion to

the endothelial molecules or that enhanced ectodomain shedding may contribute to the reduced recruitment of leukocytes to the subendothelial space thereby promoting cardioprotective effects (62). Additionally, the observed negative relationship between the genetic effects on soluble adhesion molecule and LDL-C levels suggests that altered cholesterol metabolism could contribute to the CAD risk associated with the *ABO* locus; the genetic effects of the same SNPs on LDL-C versus CAD risk showed a positive correlation (Figure 3). Nevertheless, further studies are warranted to understand the exact mechanisms.

Another novel association with sVCAM-1 level was detected in chr12 near *HSP90B1* encoding heat shock protein gp96, a chaperone that is essential for assembly of 14 of 17 integrin pairs in the hematopoietic system (63). Integrin $\alpha4\beta1$, also known as very late antigen (VLA)-4, is an important ligand of VCAM-1 (64). The lead SNP of this locus is in LD with rs117468318 (r^2=1 in NFBC1966) that locates in the 5'UTR region of *HSP90B1* and, according to RegulomeDB (32), is likely to affect transcription factor binding suggesting a possible regulatory mechanism for the detected association. If altered transcription of *HSP90B1* had a downstream effect on integrin $\alpha4\beta1$ level, this could further modify the level of unbound sVCAM-1 in circulation.

The 3rd novel locus showing association with sVCAM-1 level was identified in chr17 near *ABCA8*. The lead SNP rs112001035 is an eQTL for *ABCA8* in multiple tissue types (65). ABCA8 has been shown to regulate levels of HDL-cholesterol with a mechanism that likely involves interaction with ABCA1 (66). If ABCA8 is involved in regulation of HDL level (66) and if plasma HDL levels contribute to VCAM-1 expression (67, 68), then altered expression of the *ABCA8* could influence circulating levels of sVCAM-1 via modulating HDL particle concentration. However, this hypothesis is not supported by the fact that the effect of the lead SNP on HDL particle concentration is negligible in a large metabolomics GWAS (β=-0.043 SD, p=0.049) (28). There is evidence suggesting that ABCA8 may be involved in sphingolipid metabolism (69) and it has been hypothesized that ABCA8 may be involved in the formation of specific membrane domains during ApoA-I lipidation (66). Thus, one could speculate that the association between the ABCA8 locus and sVCAM-1 level could be related to altered HDL composition rather than absolute particle concentration, which could contribute to altered endothelial homeostasis. However, more evidence is needed to draw conclusions.

We detected a novel effect of rs11220471 in chr11 near *ST3GAL4* on sE-selectin levels in the NFBC1966 population. *ST3GAL4* encodes a member of the glycosyltransferase 29 family of enzymes involved in protein glycosylation. In mice, St3Gal4 is needed for synthesis of functional selectin ligands (46) and it has been shown to participate to chemokine C-C motif ligand 5 (Ccl5)-dependent myeloid cell recruitment to inflamed endothelium (70). The altered levels or structure of selectin ligands due to variation in *ST3GAL4* could contribute to the levels of unbound sE-selectin in circulation, providing a biologically rational mechanism for the detected association.

In the meta-analyses, we detected a novel large effect locus for VEGF in chr4 (β=-2.38 SD) near genes *EVC* (EvC ciliary complex subunit 1), *EVC2* (EvC ciliary complex subunit 2) and *STK32B* (serine/threonine kinase 32B). Mutations in this locus have been associated previously with celiac disease (47), coronary artery disease (48), and Ellis-van Creveld syndrome, a rare recessive disorder characterized with congenital defects such as short ribs, postaxial polydactyly, growth retardation, and ectodermal and cardiac defects (50, 51). The expression level of *STK32B* has been associated with clinicopathological features of oral cavity cell carcinoma including peritumoral inflammatory infiltration, metastatic spread to the cervical lymph nodes, and tumour size (71). STK32B may play a role in the hedgehog signalling pathway, which has been implicated in metastasis and angiogenesis in cancer (71) and downregulated in celiac disease (72). The hedgehog signalling has shown to be involved in the regulation of VEGF expression during developmental angiogenesis in avian embryo (73). Thus, previous literature and our results advocate that STK32B may be involved in the regulation of VEGF levels possibly via hedgehog signalling-related mechanism.

The other novel findings obtained in meta-analysis include a large effect locus on TNFα level in chr13 (β=2.13 SD). The locus in 13q14.3 associating with TNFα locates near *DLEU1* and *DLEU7* (Deleted in Lymphocytic Leukemia 1 and 7). This region is recurrently deleted in tumours and hematopoietic malignancies (74, 75). *DLEU1* is a part of a transcriptionally coregulated gene cluster that modulates the activity of the NF-kB pathway (76) which is also modulated by TNFα (77). It is largely unknown how the DLEU1 and related DLEU2 regulate NF-kB activity (78); our result suggests that TNFα signalling might be involved in this mechanism.

At last, we identified a small effect locus in chr6 harbouring two independent association signals on IL-1β and showing suggestive association also on IL4 level. This association signal is in the region coding the human leukocyte antigen proteins, and further experimental evidence would be needed to identify the exact mechanism how the locus contributes to interleukin levels.

The strengths and limitations of our study should be considered. The sample size of the present study should provide adequate power for detecting genetic associations with circulating markers of systemic inflammation (9). The use of genetically isolated populations, such as inhabitants of Northern Finland, should further enhance the power for locus identification in GWAS settings (79). We were able to perform meta-analyses only for 10 out of the 16 inflammatory phenotypes analysed in the NFBC1966 population and, thus, a replication of the present findings in other populations would be helpful. The inter-assay coefficient of variability measures for sE-selectin and VEGF in particular are notably larger than 15% that is considered to be the limit for acceptable values (Table S1). However, to our consideration, all the findings identified in the present study locate on genome regions with biologically relevant genes. Furthermore, the extremely small p-values ($p=4.48x10^{-305}$ for sE-selectin at the *ABO* locus and $p=4.95x10^{-96}$ for VEGF at the *VEGFA* locus) and the replications of the previously reported loci speak for the data adequacy and adds confidence to the novel associations. Due to mismatches in genotype availability between the present results and open-access data sets or low number of SNPs obtained after exclusions, it was not possible to carry out meaningful comparisons of the genetic effects at all the significant loci.

The present results provide novel information on genetic mechanisms influencing levels of inflammatory phenotypes in circulation. The evident role of the *ABO* locus in the regulation of the soluble adhesion molecule levels in circulation likely encompasses at least two distinct mechanisms influencing sE-selectin, sICAM-1 and sVCAM-1 levels. Our findings provide evidence that increased soluble adhesion molecule concentrations per se may not be a risk factor for cardiovascular outcomes. In particular, genetic variation associating with increased sE-selectin or sICAM-1 levels at the *ABO* locus seem to contribute to lower CAD risk. Furthermore, genetic effects at the *ICAM1* locus providing a direct molecular link to sICAM-1

383 concentration do not correlate with the genetic effects on CAD risk nor stroke risk. Overall, the present study

384 extends the knowledge about the precise molecular pathways involved in inflammatory load.

References

1. Donath,M.Y. (2014) Targeting inflammation in the treatment of type 2 diabetes: time to start. *Nat. Rev. Drug Discov.*, **13**, 465–476.
2. Heneka,M.T., Kummer,M.P. and Latz,E. (2014) Innate immune activation in neurodegenerative disease. *Nat. Rev. Immunol.*, 10.1038/nri3705.
3. Coussens,L.M. and Werb,Z. (2002) Inflammation and cancer. *Nature*, 10.1038/nature01322.Inflammation.
4. Goldberg,E.L. and Dixit,V.D. (2015) Drivers of age-related inflammation and strategies for healthspan extension. *Immunol. Rev.*, 10.1111/imr.12295.
5. World Health Organization (2015) World report on ageing and health.
6. de Craen,A.J.M., Posthuma,D., Remarque,E.J., van den Biggelaar,A.H.J., Westendorp,R.G.J. and Boomsma,D.I. (2005) Heritability estimates of innate immunity: An extended twin study. *Genes Immun.*, 10.1038/sj.gene.6364162.
7. Medzhitov,R. (2008) Origin and physiological roles of inflammation. *Nature*, 10.1038/nature07201.
8. Ahola-Olli,A. V., Würtz,P., Havulinna,A.S., Aalto,K., Pitkänen,N., Lehtimäki,T., Kähönen,M., Lyytikäinen,L.P., Raitoharju,E., Seppälä,I., *et al.* (2017) Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am. J. Hum. Genet.*, 10.1016/j.ajhg.2016.11.007.
9. Sun,B.B., Maranville,J.C., Peters,J.E., Stacey,D., Staley,J.R., Blackshaw,J., Burgess,S., Jiang,T., Paige,E., Surendran,P., *et al.* (2018) Genomic atlas of the human plasma proteome. *Nature*, 10.1038/s41586-018-0175-2.
10. 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. *Am. J. Hum. Genet.*, **103**, 691–706.
11. Sabatti,C., Service,S.K., Hartikainen,A.-L., Pouta,A., Ripatti,S., Brodsky,J., Jones,C.G., Zaitlen,N.A., Varilo,T., Kaakinen,M., *et al.* (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.*, 10.1038/ng.271.
12. Järvelin,M.R., Sovio,U., King,V., Lauren,L., Xu,B., McCarthy,M.I., Hartikainen,A.L., Laitinen,J., Zitting,P., Rantakallio,P., *et al.* (2004) Early life factors and blood pressure at age 31 years in the 1966 Northern Finland birth cohort. *Hypertension*, 10.1161/01.HYP.0000148304.33869.ee.
13. Saukkonen,T., Mutt,S.J., Jokelainen,J., Saukkonen,A.M., Raza,G.S., Karhu,T., Härkönen,P., Eckel,J., Herzig,K.H., Rajala,U., *et al.* (2018) Adipokines and inflammatory markers in elderly subjects with high risk of type 2 diabetes and cardiovascular disease. *Sci. Rep.*, **8**, 1–8.
14. Raitakari,O.T., Juonala,M., Rönnemaa,T., Keltikangas-Järvinen,L., Räsänen,L., Pietikäinen,M., Hutri-Kähönen,N., Taittonen,L., Jokinen,E., Marniemi,J., *et al.* (2008) Cohort profile: The cardiovascular risk in young Finns study. *Int. J. Epidemiol.*, 10.1093/ije/dym225.
15. Borodulin,K., Tolonen,H., Jousilahti,P., Jula,A., Juolevi,A., Koskinen,S., Kuulasmaa,K., Laatikainen,T., Männistö,S., Peltonen,M., *et al.* (2017) Cohort Profile: The National FINRISK Study. *Int. J. Epidemiol.*, 10.1093/ije/dyx239.
16. Santalahti,K., Maksimow,M., Airola,A., Pahikkala,T., Hutri-Kähönen,N., Jalkanen,S., Raitakari,O.T. and Salmi,M. (2016) Circulating cytokines predict the development of insulin resistance in a prospective Finnish population cohort. *J. Clin. Endocrinol. Metab.*, 10.1210/jc.2016-2081.
17. Santalahti,K., Havulinna,A., Maksimow,M., Zeller,T., Blankenberg,S., Vehtari,A., Joensuu,H., Jalkanen,S., Salomaa,V. and Salmi,M. (2017) Plasma levels of hepatocyte growth factor and placental growth factor predict mortality in a general population: a prospective cohort study. *J. Intern. Med.*, 10.1111/joim.12648.
18. Marchini,J., Howie,B., Myers,S., McVean,G. and Donnelly,P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, 10.1038/ng2088.

19. Marchini,J. and Howie,B. (2010) Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.*, 10.1038/nrg2796.
20. Willer,C.J., Li,Y. and Abecasis,G.R. (2010) METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 10.1093/bioinformatics/btq340.
21. Yang,J., Ferreira,T., Morris,A.P., Medland,S.E., Madden,P.A.F., Heath,A.C., Martin,N.G., Montgomery,G.W., Weedon,M.N., Loos,R.J., *et al.* (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*, 10.1038/ng.2213.
22. Yang,J., Lee,S.H., Goddard,M.E. and Visscher,P.M. (2011) GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, 10.1016/j.ajhg.2010.11.011.
23. Paré,G., Chasman,D.I., Kellogg,M., Zee,R.Y.L., Rifai,N., Badola,S., Miletich,J.P. and Ridker,P.M. (2008) 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.*, 10.1371/journal.pgen.1000118.
24. Ridker, Hennekens, Buring and Rifai (2000) C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *Ital. Heart J. Suppl.*, **1**, 1066–7.
25. Hwang,S.J., Ballantyne,C.M., Sharrett,A.R., Smith,L.C., Davis,C.E., Gotto,A.M. and Boerwinkle,E. (1997) 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*, 10.1161/01.cir.96.12.4219.
26. Nikpay,M., Goel,A., Won,H.H., Hall,L.M., Willenborg,C., Kanoni,S., Saleheen,D., Kyriakou,T., Nelson,C.P., CHopewell,J., *et al.* (2015) A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.*, 10.1038/ng.3396.
27. Traylor,M., Farrall,M., Holliday,E.G., Sudlow,C., Hopewell,J.C., Cheng,Y.C., Fornage,M., Ikram,M.A., Malik,R., Bevan,S., *et al.* (2012) Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): A meta-analysis of genome-wide association studies. *Lancet Neurol.*, 10.1016/S1474-4422(12)70234-X.
28. Kettunen,J., Demirkan,A., Würtz,P., Draisma,H.H.M., Haller,T., Rawal,R., Vaarhorst,A., Kangas,A.J., Lyytikäinen,L.P., Pirinen,M., *et al.* (2016) Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.*, 10.1038/ncomms11122.
29. Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M.A.R., Bender,D., Maller,J., Sklar,P., de Bakker,P.I.W., Daly,M.J., *et al.* (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.*, 10.1086/519795.
30. Karakas,M., Baumert,J., Kleber,M.E., Thorand,B., Dallmeier,D., Silbernagel,G., Grammer,T.B., Rottbauer,W., Meisinger,C., Illig,T., *et al.* (2012) A Variant In the Abo Gene Explains the Variation in Soluble E-Selectin Levels-Results from Dense Genotyping in Two Independent Populations. *PLoS One*, 10.1371/journal.pone.0051441.
31. Paterson,A.D., Lopes-Virella,M.F., Waggott,D., Boright,A.P., Hosseini,S.M., Carter,R.E., Shen,E., Mirea,L., Bharaj,B., Sun,L., *et al.* (2009) Genome-wide association identifies the ABO blood group as a major locus associated with serum levels of soluble E-selectin. *Arterioscler. Thromb. Vasc. Biol.*, 10.1161/ATVBAHA.109.192971.
32. Boyle,A.P., Hong,E.L., Hariharan,M., Cheng,Y., Schaub,M.A., Kasowski,M., Karczewski,K.J., Park,J., Hitz,B.C., Weng,S., *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, 10.1101/gr.137323.112.
33. Lu,Y., Vitart,V., Burdon,K.P., Khor,C.C., Bykhovskaya,Y., Mirshahi,A., Hewitt,A.W., Koehn,D., Hysi,P.G., Ramdas,W.D., *et al.* (2013) Genome-wide association analyses identify multiple loci associated with central corneal thickness and keratoconus. *Nat. Genet.*, 10.1038/ng.2506.
34. Global Lipids Genetics Consortium, Willer,C.J., Schmidt,E.M., Sengupta,S., Peloso,G.M., Gustafsson,S., Kanoni,S., Ganna,A., Chen,J., Buchkovich,M.L., *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, 10.1038/ng.2797.
35. Surakka,I., Horikoshi,M., Mägi,R., Sarin,A.P., Mahajan,A., Lagou,V., Marullo,L., Ferreira,T., Miraglio,B., Timonen,S., *et al.* (2015) The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.*,

10.1038/ng.3300.

36. Spracklen,C.N., Chen,P., Kim,Y.J., Wang,X., Cai,H., Li,S., Long,J., Wu,Y., Wang,Y.X., Takeuchi,F., *et al.* (2017) Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum. Mol. Genet.*, 10.1093/hmg/ddx062.
37. Teslovich,T.M., Musunuru,K., Smith,A. V., Edmondson,A.C., Stylianou,I.M., Koseki,M., Pirruccello,J.P., Ripatti,S., Chasman,D.I., Willer,C.J., *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 10.1038/nature09270.
38. Michailidou,K., Lindström,S., Dennis,J., Beesley,J., Hui,S., Kar,S., Lemaçon,A., Soucy,P., Glubb,D., Rostamianfar,A., *et al.* (2017) Association analysis identifies 65 new breast cancer risk loci. *Nature*, 10.1038/nature24284.
39. Arking,D.E., Pulit,S.L., Crotti,L., Van Der Harst,P., Munroe,P.B., Koopmann,T.T., Sotoodehnia,N., Rossin,E.J., Morley,M., Wang,X., *et al.* (2014) Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat. Genet.*, 10.1038/ng.3014.
40. van der Harst,P., van Setten,J., Verweij,N., Vogler,G., Franke,L., Maurano,M.T., Wang,X., Mateo Leach,I., Eijgelsheim,M., Sotoodehnia,N., *et al.* (2016) 52 Genetic Loci Influencing Myocardial Mass. *J. Am. Coll. Cardiol.*, 10.1016/j.jacc.2016.07.729.
41. Evans,D.S., Avery,C.L., Nalls,M.A., Li,G., Barnard,J., Smith,E.N., Tanaka,T., Butler,A.M., Buxbaum,S.G., Alonso,A., *et al.* (2016) Fine-mapping, novel loci identification, and SNP association transferability in a genome-wide association study of QRS duration in African Americans. *Hum. Mol. Genet.*, 10.1093/hmg/ddw284.
42. Astle,W.J., Elding,H., Jiang,T., Allen,D., Ruklisa,D., Mann,A.L., Mead,D., Bouman,H., Riveros-Mckay,F., Kostadima,M.A., *et al.* (2016) The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*, 10.1016/j.cell.2016.10.042.
43. Ligthart,S., Vaez,A., Hsu,Y.H., Stolk,R., Uitterlinden,A.G., Hofman,A., Alizadeh,B.Z., Franco,O.H. and Dehghan,A. (2016) Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. *BMC Genomics*, 10.1186/s12864-016-2712-4.
44. Suhre,K., Arnold,M., Bhagwat,A.M., Cotton,R.J., Engelke,R., Raffler,J., Sarwath,H., Thareja,G., Wahl,A., Delisle,R.K., *et al.* (2017) Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.*, 10.1038/ncomms14357.
45. Chambers,J.C., Zhang,W., Sehmi,J., Li,X., Wass,M.N., Van Der Harst,P., Holm,H., Sanna,S., Kavousi,M., Baumeister,S.E., *et al.* (2011) Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat. Genet.*, 10.1038/ng.970.
46. Yang,W.H., Nussbaum,C., Grewal,P.K., Marth,J.D. and Sperandio,M. (2012) Coordinated roles of ST3Gal-VI and ST3Gal-IV sialyltransferases in the synthesis of selectin ligands. *Blood*, 10.1182/blood-2012-04-424366.
47. Östensson,M., Montén,C., Bacelis,J., Gudjonsdottir,A.H., Adamovic,S., Ek,J., Ascher,H., Pollak,E., Arnell,H., Browaldh,L., *et al.* (2013) A Possible Mechanism behind Autoimmune Disorders Discovered By Genome-Wide Linkage and Association Analysis in Celiac Disease. *PLoS One*, 10.1371/journal.pone.0070174.
48. Slavin,T.P., Feng,T., Schnell,A., Zhu,X. and Elston,R.C. (2011) Two-marker association tests yield new disease associations for coronary artery disease and hypertension. *Hum. Genet.*, 10.1007/s00439-011-1009-6.
49. Müller,S.H., Girard,S.L., Hopfner,F., Merner,N.D., Bourassa,C. V., Lorenz,D., Clark,L.N., Tittmann,L., Soto-Ortolaza,A.I., Klebe,S., *et al.* (2016) Genome-wide association study in essential tremor identifies three new loci. *Brain*, 10.1093/brain/aww242.
50. Jayaraj,D., Maheswaran,T., Suresh,R. and Ganapathy,N. (2012) Ellis-van Creveld. *J. Pharm. Bioallied Sci.*, 10.4103/0975-7406.100257.
51. Temtamy,S.A., Aglan,M.S., Valencia,M., Cocchi,G., Pacheco,M., Ashour,A.M., Amr,K.S., Helmy,S.M.H., El-Gammal,M.A., Wright,M., *et al.* (2008) Long interspersed nuclear element-1 (LINE1)-mediated deletion of EVC, EVC2, C4orf6, and STK32B in ellis-van Creveld syndrome with borderline intelligence. *Hum.*

Mutat., 10.1002/humu.20778.

52. Fatemifar,G., Hoggart,C.J., Paternoster,L., Kemp,J.P., Prokopenko,I., Horikoshi,M., Wright,V.J., Tobias,J.H., Richmond,S., Zhurov,A.I., *et al.* (2013) Genome-wide association study of primary tooth eruption identifies pleiotropic loci associated with height and craniofacial distances. *Hum. Mol. Genet.*, 10.1093/hmg/ddt231.
53. Kettunen,J., Tukiainen,T., Sarin,A.-P., Ortega-Alonso,A., Tikkanen,E., Lyytikäinen,L.-P., Kangas,A.J., Soininen,P., Würtz,P., Silander,K., *et al.* (2012) Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat. Genet.*, 10.1038/ng.1073.
54. Goes,F.S., Mcgrath,J., Avramopoulos,D., Wolyniec,P., Pirooznia,M., Ruczinski,I., Nestadt,G., Kenny,E.E., Vacic,V., Peters,I., *et al.* (2015) Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.*, 10.1002/ajmg.b.32349.
55. McKay,J.D., Hung,R.J., Han,Y., Zong,X., Carreras-Torres,R., Christiani,D.C., Caporaso,N.E., Johansson,M., Xiao,X., Li,Y., *et al.* (2017) Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.*, 10.1038/ng.3892.
56. Barbalic,M., Dupuis,J., Dehghan,A., Bis,J.C., Hoogeveen,R.C., Schnabel,R.B., Nambi,V., Bretler,M., Smith,N.L., Peters,A., *et al.* (2010) Large-scale genomic studies reveal central role of ABO in sP-selectin and sICAM-1 levels. *Hum. Mol. Genet.*, 10.1093/hmg/ddq061.
57. Scott,D.W. and Patel,R.P. (2013) Endothelial heterogeneity and adhesion molecules N-glycosylation: Implications in leukocyte trafficking in inflammation. *Glycobiology*, 10.1093/glycob/cwt014.
58. Rao,S.P., Wang,Z., Zuberi,R.I., Sikora,L., Bahaie,N.S., Zuraw,B.L., Liu,F.-T. and Sriramarao,P. (2007) Galectin-3 Functions as an Adhesion Molecule to Support Eosinophil Rolling and Adhesion under Conditions of Flow. *J. Immunol.*, 10.4049/jimmunol.179.11.7800.
59. Stowell,S.R., Arthur,C.M., Dias-Baruffi,M., Rodrigues,L.C., Gourdine,J.P., Heimburg-Molinaro,J., Ju,T., Molinaro,R.J., Rivera-Marrero,C., Xia,B., *et al.* (2010) Innate immune lectins kill bacteria expressing blood group antigen. *Nat. Med.*, 10.1038/nm.2103.
60. Pradhan,A.D., Rifai,N. and Ridker,P.M. (2002) Soluble intercellular adhesion molecule-1, soluble vascular adhesion molecule-1, and the development of symptomatic peripheral arterial disease in men. *Circulation*, 10.1161/01.CIR.0000025636.03561.EE.
61. Güray,U., Erbay, a R., Güray,Y., Yilmaz,M.B., Boyaci,A.A., Sasmaz,H., Korkmaz,S. and Kütük,E. (2004) Levels of soluble adhesion molecules in various clinical presentations of coronary atherosclerosis. *Int. J. Cardiol.*, 10.1016/j.ijcard.2003.07.014.
62. Kiechl,S., Pare,G., Barbalic,M., Qi,L., Dupuis,J., Dehghan,A., Bis,J.C., Laxton,R.C., Xiao,Q., Bonora,E., *et al.* (2011) 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.*, 10.1161/CIRCGENETICS.111.960682.
63. Staron,M., Yang,Y., Liu,B., Li,J., Shen,Y., Z????iga-Pfl??cker,J.C., Aguila,H.L., Goldschneider,I. and Li,Z. (2010) gp96, an endoplasmic reticulum master chaperone for integrins and Toll-like receptors, selectively regulates early T and B lymphopoiesis. *Blood*, 10.1182/blood-2009-07-233031.
64. Elices,M.J., Osborn,L., Takada,Y., Crouse,C., Luhowskyj,S., Hemler,M.E. and Lobb,R.R. (1990) VCAM-1 on activated endothelium interacts with the leukocyte integrin VLA-4 at a site distinct from the VLA-4/Fibronectin binding site. *Cell*, 10.1016/0092-8674(90)90661-W.
65. Lonsdale,J., Thomas,J., Salvatore,M., Phillips,R., Lo,E., Shad,S., Hasz,R., Walters,G., Garcia,F., Young,N., *et al.* (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, 10.1038/ng.2653.
66. Trigueros-Motos,L., Van Capelleveen,J.C., Torta,F., Castao,D., Zhang,L.H., Chai,E.C., Kang,M., Dimova,L.G., Schimmel,A.W.M., Tietjen,I., *et al.* (2017) ABCA8 regulates cholesterol efflux and high-density lipoprotein cholesterol levels. *Arterioscler. Thromb. Vasc. Biol.*, 10.1161/ATVBAHA.117.309574.
67. Dimayuga,P., Zhu,J., Oguchi,S., Chyu,K.Y., Xu,X.O.H., Yano,J., Shah,P.K., Nilsson,J. and Cercek,B. (1999) Reconstituted HDL containing human apolipoprotein A-1 reduces VCAM-1 expression and neointima formation following periadventitial cuff-induced carotid injury in apoE null mice. *Biochem. Biophys. Res. Commun.*, 10.1006/bbrc.1999.1278.

583 68. Cockerill,G.W., Rye,K.A., Gamble,J.R., Vadas,M.A. and Barter,P.J. (1995) High-Density Lipoproteins Inhibit
584 Cytokine-Induced Expression of Endothelial Cell Adhesion Molecules. *Arterioscler. Thromb. Vasc. Biol.*,
585 10.1161/01.ATV.15.11.1987.
586 69. Kim,W.S., Hsiao,J.-H.T., Bhatia,S., Glaros,E.N., Don,A.S., Tsuruoka,S., Shannon Weickert,C. and
587 Halliday,G.M. (2013) ABCA8 stimulates sphingomyelin production in oligodendrocytes. *Biochem. J.*,
588 10.1042/BJ20121764.
589 70. Döring,Y., Noels,H., Mandl,M., Kramp,B., Neideck,C., Lievens,D., Drechsler,M., Megens,R.T.A., Tilstam,P.
590 V., Langer,M., *et al.* (2014) Deficiency of the sialyltransferase St3Gal4 reduces Ccl5-mediated myeloid
591 cell recruitment and arrest. *Circ. Res.*, 10.1161/CIRCRESAHA.114.302426.
592 71. Parris,T.Z., Aziz,L., Kovács,A., Hajizadeh,S., Nemes,S., Semaan,M., Chen,C.Y., Karlsson,P. and Helou,K.
593 (2014) Clinical relevance of breast cancer-related genes as potential biomarkers for oral squamous cell
594 carcinoma. *BMC Cancer*, 10.1186/1471-2407-14-324.
595 72. Liang,R., Hinds,R., Abud,H.E. and Cheng,W. (2013) Hedgehog signalling is downregulated in celiac disease.
596 *Can. J. Gastroenterol.*
597 73. Moran,C.M., Myers,C.T., Lewis,C.M. and Krieg,P.A. (2012) Hedgehog regulates angiogenesis of
598 intersegmental vessels through the VEGF signaling pathway. *Dev. Dyn.*, 10.1002/dvdy.23795.
599 74. Wolf,S., Mertens,D., Schaffner,C., Korz,C., Dohner,H., Stilgenbauer,S. and Lichter,P. (2001) B-cell neoplasia
600 associated gene with multiple splicing (BCMS): the candidate B-CLL gene on 13q14 comprises more than
601 560 kb covering all critical regions. *Hum Mol Genet*.
602 75. Ouillette,P., Erba,H., Kujawski,L., Kaminski,M., Shedden,K. and Malek,S.N. (2008) Integrated genomic
603 profiling of chronic lymphocytic leukemia identifies subtypes of deletion 13q14. *Cancer Res.*,
604 10.1158/0008-5472.CAN-07-3105.
605 76. Garding,A., Bhattacharya,N., Claus,R., Ruppel,M., Tschuch,C., Filarsky,K., Idler,I., Zucknick,M., Caudron-
606 Herger,M., Oakes,C., *et al.* (2013) Epigenetic Upregulation of lncRNAs at 13q14.3 in Leukemia Is Linked
607 to the In Cis Downregulation of a Gene Cluster That Targets NF-kB. *PLoS Genet.*,
608 10.1371/journal.pgen.1003373.
609 77. Schütze,S., Wiegmann,K., Machleidt,T. and Krönke,M. (1995) TNF-induced activation of NF-kappa B.
610 *Immunobiology*.
611 78. Mao,X., Su,Z. and Mookhtiar,A.K. (2017) Long non-coding RNA: A versatile regulator of the nuclear factor-
612 κB signalling circuit. *Immunology*, 10.1111/imm.12698.
613 79. Hatzikotoulas,K., Gilly,A. and Zeggini,E. (2014) Using population isolates in genetic association studies.
614 *Brief. Funct. Genomics*, 10.1093/bfgp/elu022.
615
616

Acknowledgements

We gratefully acknowledge the contributions of the participants in the Northern Finland Birth Cohort 1966 study. We also thank all the field workers and laboratory personnel for their efforts. Data on coronary artery disease / myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG.

Funding

This work was supported by University of Oulu Graduate School [ES, MK], the Finnish Foundation for Cardiovascular Research [VS], Biocenter Oulu [SS], European Commission [DynaHEALTH – H2020 – 633595, SS], Academy of Finland [297338 and 307247, JK] and Novo Nordisk Foundation [NNF17OC0026062, JK]. NFBC1966 received financial support from University of Oulu Grant no. 65354, Oulu University Hospital Grant no. 2/97, 8/97, Ministry of Health and Social Affairs Grant no. 23/251/97, 160/97, 190/97, National Institute for Health and Welfare, Helsinki Grant no. 54121, Regional Institute of Occupational Health, Oulu, Finland Grant no. 50621, 54231. The Young Finns Study has been financially supported by the Academy of Finland: [grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi)]; the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals [grant X51001]; Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; and EU Horizon 2020 [grant 755320 for TAXINOMISIS]; and European Research Council [grant 742927 for MULTIEPIGEN project]; Tampere University Hospital Supporting Foundation.

Conflict of Interest

VS has participated in a conference trip sponsored by Novo Nordisk and received a honorarium from the same source for participating in and advisory board meeting. He also has ongoing research collaboration with Bayer Ltd.

642 # Tables

643 Table 1. Basic charcteristics of the NFBC1966 study population.

Characteristcs	
Total number of individuals	5284
Number of males (%)	2543 (48.1)
Age, years	31.1 ± 0.4
BMI, kg/m^2	24.4 ± 4.0
Glucose, mmol/l	5.1 ± 0.7
LDL-cholesterol, mmol/l	3.0 ± 0.9
HDL-cholesterol, mmol/l	1.6 ± 0.4
Systolic blood pressure, mmHg	124.2 ± 13.6
Diastolic blood pressure, mmHg	76.8 ± 11.7

Values are mean ± standard deviation.

644

645 Table 2. Significant loci associating with the circulating inflammatory phenotypes.

Study	Marker	Locus	Chr:Position	Candidate gene	Nearest gene(s)	Annotation	dbSNP refrerence	EA	EAF	Beta	P-value	Variance explained	Total variance explained
NFBC1966	sE-Selectin	9q34.2	9:136141870	*ABO*	*ABO*	Intronic	rs2519093	T	0.188	-0.903	4.48E-305	0.249	0.258
		11q24.2	**11:126266665**	***ST3GAL4***	***ST3GAL4***	**Intronic**	**rs11220471**	**G**	**0.212**	**-0.162**	**7.72E-12**	**0.009**	
	sICAM-1	9q34.2	9:136141870	*ABO*	*ABO*	Intronic	rs2519093	T	0.188	-0.352	7.43E-48	0.038	0.118
		19p13.2	19:10383403	*ICAM1*	*ICAM1*	Intronic	rs117960796	A	0.012	-1.669	8.03E-40	0.066	
		19p13.2	19:10497360	*ICAM1*	*CDC37*	Intergenic	rs74428614	A	0.163	0.226	1.14E-16 *	0.014	
	sVCAM-1	**9q34.2**	**9:136131322**	***ABO***	***ABO***	**Missense**	**rs8176746**	**T**	**0.129**	**0.256**	**5.06E-19**	**0.015**	0.038
		12q23.3	**12:104448391**	***HSP90B1***	***GLT8D2***	**Intronic**	**rs117238625**	**A**	**0.023**	**0.510**	**2.90E-14**	**0.012**	
		17q24.2	**17:66823805**	***ABCA8***	***ABCA8***	**Intergenic**	**rs112001035**	**A**	**0.060**	**-0.324**	**1.04E-13**	**0.012**	
meta-analyses	IL1β	**6p22.1**	**6:30017071**		***HLA* locus**	**Intronic**	**rs6917603**	**C**	**0.251**	**-0.163**	**1.76E-12**	**0.010**	0.015
		6p22.1	**6:30013887**		***HLA* locus**	**Intronic**	**rs9261224**	**T**	**0.035**	**0.261**	**1.31E-09 ***	**0.005**	
	IP10	4q21.1	4:76899176	*CXCL10*	*SAD1*	Intronic	rs192716315	C	0.003	1.513	2.71E-13	0.014	0.014
	MCP1	1q23.2	1:159175354	*ACKR1*	*ACKR1*	Missense	rs12075	G	0.561	0.125	1.90E-24	0.008	0.008
	TNFα	**13q14.3**	**13:51141997**	***DLEU1***	***DLEU1***	**Intronic**	**rs17074575**	**G**	**0.002**	**2.131**	**2.71E-09**	**0.018**	0.018
	VEGF	**4p16.2**	**4:5636073**	***STK32B***	***EVC2***	**Intronic**	**rs186725382**	**A**	**0.001**	**-2.380**	**4.53E-10**	**0.011**	0.052
		6p21.1	6:43927050	*VEGFA*	*C6orf223*	Intergenic	rs7767396	G	0.422	0.272	4.95E-96	0.036	
		9p24.2	9:2686273	*VLDLR*	*VLDLR, KCNV2*	Intergenic	rs7030781	T	0.373	0.099	1.57E-13	0.005	

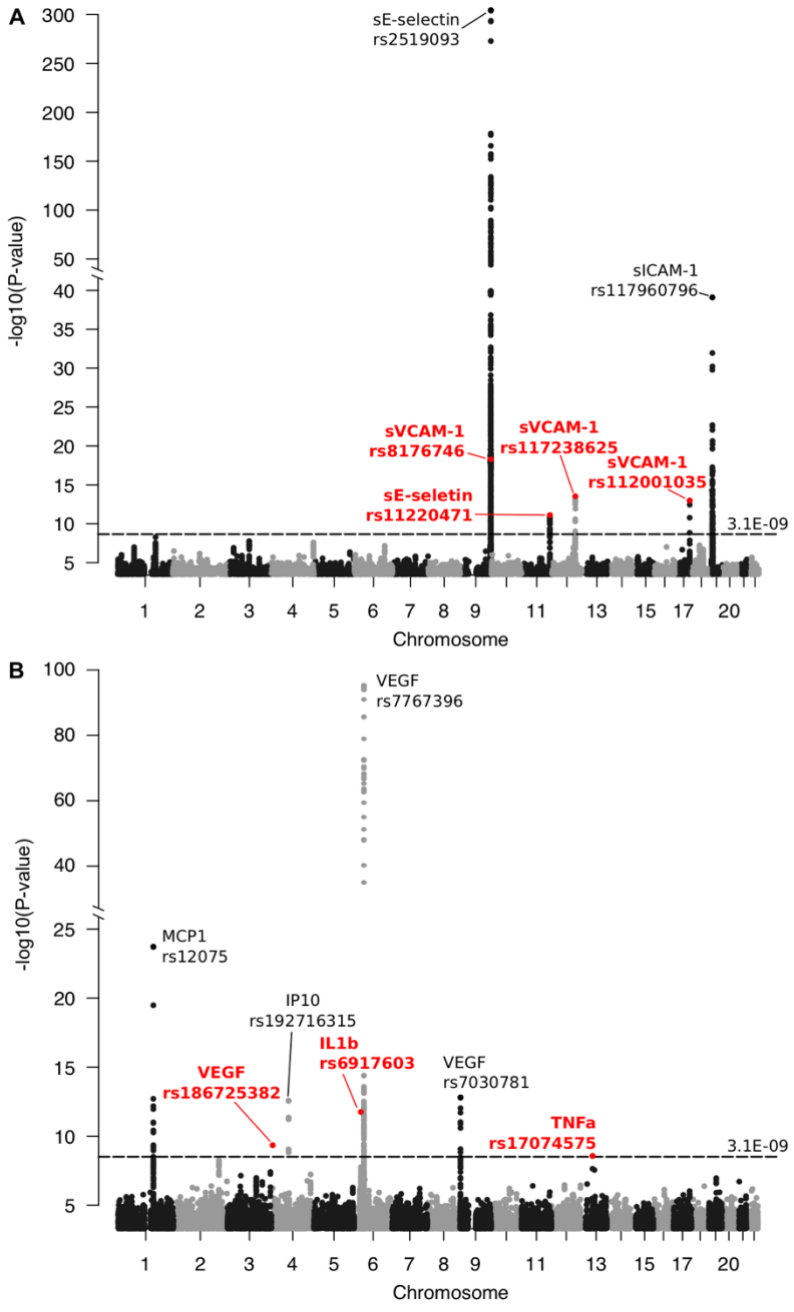
Statistical significance is considered at $p<3.1x10^{-9}$. Novel findings are highlighted with bold font. All positions correspond to human genome build 37. Asterisk (*) indicates associations that are significant after conditioning the analyses on the locus specific lead variant on the preceding row. EA, effect allele; EAF, effect allele frequency.

646

Figures

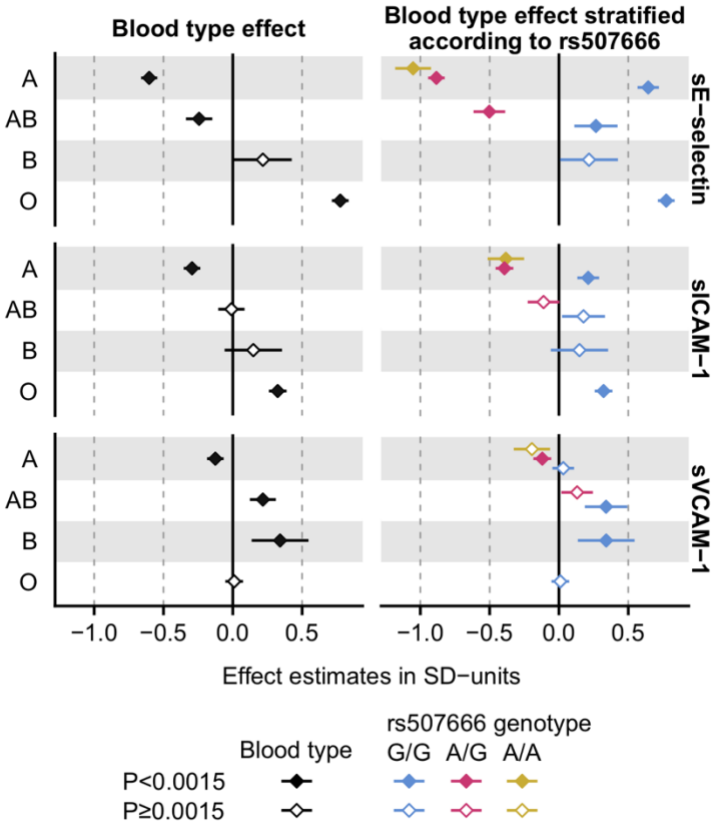
Figure 1. The combined Manhattan plots for significant associations with inflammatory markers studied in (A) NFBC1966 and in (B) meta-analyses with three other Finnish population cohorts.

Significance threshold $p<3.1x10^{-9}$ derives from the standard p-value limit for genome-wide significance $p<5x10^{-8}$ corrected for 16 markers examined in the present study. Novel association signals are highlighted with red font and replicated loci are marked with black font.



654 Figure 2. The effects of the ABO blood types and the A1 subtype on soluble adhesion molecule levels.

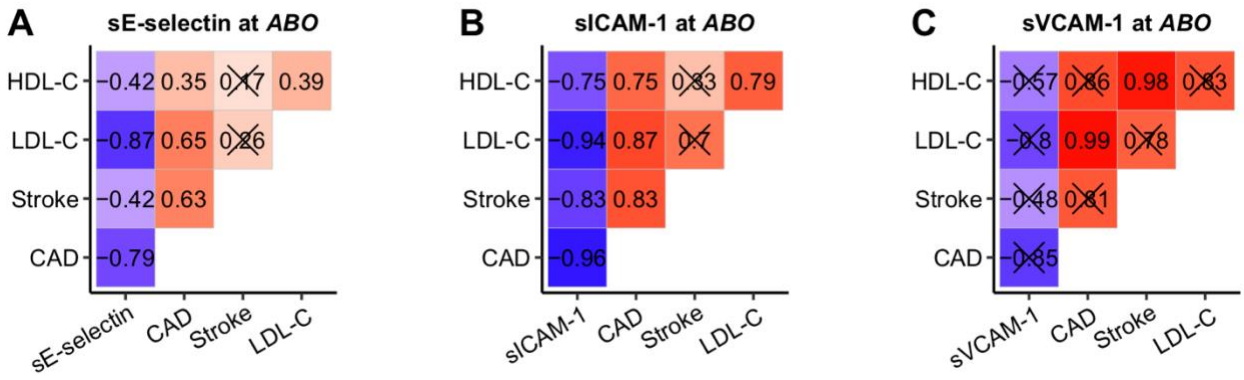
655 The effects of the ABO blood types on sE-selectin, sICAM-1 and sVCAM-1 levels were evaluated in linear
656 models, where adjusted (sex, age, BMI, and the ten first genetic principal components) and transformed
657 soluble adhesion molecule concentrations were used as outcomes and the ABO blood type served as
658 categorical variable (A versus non-A, etc.). Corresponding models were fitted for the ABO blood types
659 stratified by the rs507666-A allele count (0, 1 or 2), where the A allele tags the ABO subtype A1 having
660 enhanced glycosyltransferase activity (25). No individuals were found to have B or O blood type and one or
661 more copies of the rs507666-A allele and, thus, it was not possible to perform stratification within these
662 blood types.



663

664 Figure 3. The correlations of the SNP effects on soluble adhesion molecule levels versus other cardiovascular
665 health-related traits at the *ABO* locus.

666 The correspondence of the SNP effects on inflammatory phenotypes versus cardiovascular health-related
667 traits were determined for the SNPs associating with (A) sE-selectin, (B) sICAM-1, and (C) sVCAM-1 levels at
668 the *ABO* locus. SNP effects on coronary artery disease (CAD), stroke, and LDL-C or HDL-C were extracted from
669 CARDIoGRAM (26), Stroke Consortia (27), and a metabolomics GWAS (28) summary statistics, respectively.
670 The correlations (Pearson's *r*) of the genetic effects were estimated using subsets of representative SNPs
671 extracted from the summary statistics of the present GWAS using a clumping function in PLINK and r^2
672 threshold of 0.2. Prior to clumping, data were filtered to include only the SNPs available in all the three
673 datasets. Correlations with p≥0.05 are marked with a cross. The SNPs used for estimating the correlations
674 are listed in Table S5 and scatter plot representations as well as correlations at the other loci and other
675 inflammatory phenotypes are shown in Figure S3.



676