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

Eeva Sliz, ·1,2,3 Marita Kalaoka, ·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

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.

Objective To assess genetic determinants of 16 circulating cytokines and cell adhesion molecules (inflammatory phenotypes) in Finns.

Methods Genome-wide associations of the inflammatory phenotypes were studied in Northern Finland Birth Cohort 1966 (N=5284). A subsequent meta-analysis was completed for 10 phenotypes available in a previous genome-wide association study, adding up to 13 577 individuals in the study.

Results We identified seven novel and six previously reported genetic associations (p<3.1×10−8). Three loci were associated with soluble vascular cell adhesion molecule-1 (sVCAM-1) level, one of which was the ABO locus that has previously been associated with soluble E-selectin (sE-selectin) and intercellular adhesion molecule-1 (sICAM-1) levels. Our findings suggest that the blood type B associates primarily with sVCAM-1 level, while the A1 subtype shows a robust effect on sE-selectin and sICAM-1 levels. The genotypes in the ABO locus associating with higher soluble adhesion molecule levels tend to associate with lower circulating cholesterol levels 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 in the ABO locus and that elevated soluble adhesion molecule levels per se may not increase risk for cardiovascular disease.

INTRODUCTION

It is currently established that inflammatory load may play a role in the aetiology of autoimmune and infectious diseases, but also in a broad range of other diseases, such as chronic cardiometabolic disorders, neurodegenerative diseases and cancer.

The risk for these diseases increases with age, and due to the world’s ageing population their prevalence is likely to expand. Moreover, these diseases often co-occur, which is likely due to shared inflammation-related pathophysiology.

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. Circulating cytokines, growth factors, chemokines and cell adhesion molecules (CAMs) (hereafter inflammatory phenotypes) are fundamental mediators of inflammatory responses. Genes encoding these molecules and their receptors play a crucial role in modulating the related functions. Previous studies have identified loci associating with levels of inflammatory phenotypes, but the understanding of the exact regulatory mechanisms is still incomplete.

To add insights into the genetic mechanisms contributing to the inflammatory load, we performed a genome-wide association study (GWAS) of 16 circulating inflammatory phenotypes in 5284 individuals from Northern Finland Birth Cohort 1966 (NFBC1966) and a subsequent meta-analysis of 10 phenotypes in three other Finnish population cohorts, 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 NFBC1966 comprises 96% of all births during 1966 in the two northermost 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. The data analysed in the present study are from the 31 years’ follow-up when clinical examinations and blood sampling were completed for altogether 6033 individuals, 5284 of whom had body mass index (BMI), inflammatory phenotypes and genotype data available (a maximum number of individuals per inflammatory marker of 5100). Genotyping of the samples was completed...
using 370 k Illumina HumanHap arrays (Illumina, California, USA), and subsequent imputation was performed based on the 1000 Genomes 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, California, USA) with Milliplex Human Chemokine/Cytokine and CVD/Cytokine kits (Cat# HCYTOMAG-60K-12 and Cat# SPR349; Millipore, St Charles, Missouri, USA) and Bio-Plex Manager Software V4.3 as previously described.15 The 16 inflammatory phenotypes studied in the NFBC1966 were interleukin (IL) 1-alpha, IL-1-beta (IL1β), IL4, IL6, IL8, IL17, IL1 receptor antagonist (IL1ra), interferon gamma-induced protein 10 (IP10), monocyte chemotractant protein 1 (MCP1), tumour necrosis factor alpha (TNFα), vascular endothelial growth factor (VEGF), plasminogen activator inhibitor 1, soluble CD40 ligand, 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.7 The study included up to 8293 Finnish individuals from the Cardiovascular Risk in Young Finns Study (YFS)12 and FINRISK (www.thl.fi/finsirk),13 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 are from 2019 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 5 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 670 k Illumina HumanHap arrays (Illumina) and imputed based on 1000 Genomes 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 V6.0 software (Bio-Rad Laboratories) as previously described.14

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. First, rank-based inverse transformation was applied to normalise the phenotypes. Preceding the analyses, linear regression models were fitted to adjust the transformed inflammatory phenotypes for age, sex, BMI and the 10 first genetic principal components to control for population stratification. The resulting residuals were again normalised with inverse transformation, and the adjusted and transformed residuals were used as phenotypes in the analyses.

Genome-wide association tests were performed using snpset V2.5.1 software.15 Allele effects were estimated using an additive model (–freqset 1), and the option to centre 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 were used to perform meta-analyses with METAL software (V2011-03-25)16 for the 10 phenotypes (IL1β, IL1ra, IL4, IL6, IL8, IL17, IP10, MCP1, TNFα and VEGF) available in the previous GWAS.7 Genomic control correction was enabled (GENOMICCONTROL ON) to account for population stratification and cryptic relatedness. To estimate the heterogeneity of effect sizes between NFBC1966 and the previous GWAS, calculation of heterogeneity statistics based on Cochrane’s Q-test was enabled (ANALYZE HETEROGENEITY).

Supplemental genome-wide tests in NFBC1966

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

Conditional analyses and variance explained

To assess whether the identified loci harbour multiple independent association signals, we conducted conditional analyses by further adjusting the models with the locus-specific lead variants. The association tests were repeated within a 2 Mb window around the lead SNP for the phenotypes studied in the NFBC1966 population only. For the meta-analysed phenotypes, we applied a method proposed by Yang et al17 that enables conditional analyses of GWAS summary statistics. NFBC1966 was used as a reference sample to estimate linkage disequilibrium (LD) corrections in these analyses. The proportion of variance explained was calculated using all independent variants using the following formula:

\[
\text{Variance explained} = (\beta \sqrt{2 \times MAF (1 - MAF)})^2
\]

Here \(\beta\) is the variant’s effect estimate on the inflammatory phenotype and MAF denotes minor allele frequency.

Complementary association tests on soluble adhesion molecule levels

Complementary association tests were conducted to better evaluate the molecular mechanism explaining the two potentially independent association signals with soluble CAM levels in the ABO locus. Here, linear models were repeated within a 2 Mb window and further adjusted for the ABO blood type or rs507666 genotype tagging the A1 subtype.18 In addition, we determined the effect estimates of ABO blood types and ABO blood types stratified by rs507666 genotype on sE-selectin, sICAM-1 and sVCAM-1 levels: the adjusted and transformed CAM concentrations were as outcomes in the linear models and ABO blood types as categorical variables (individuals with blood type A vs non-A, and so on). Corresponding models were fitted for the rs507666-stratified blood types (individuals with blood type A and rs507666 G/G vs others, and so on).

Shared genetic influences on inflammatory and cardiovascular phenotypes

As previous evidence suggests that elevated concentrations of circulating markers of inflammation increase the risk of cardiovascular diseases (CVD),19 20 we further evaluated how variants in the loci associating with inflammatory phenotypes may relate to other cardiovascular traits. We used the gwas-pw method developed by Pickrell et al21 that estimates whether a locus harbours a genetic variant influencing one of the two phenotypes compared (models 1 and 2), if the same variant influences both phenotypes (model 3), or if separate variants within a locus influence the two phenotypes (model 4). Using the gwas-pw and open-access data provided by CARDioGRAM,22 MEGASTROKE consortium23 and
Global Lipids Genetics Consortium,\textsuperscript{24} we evaluated the shared genetic determinants of circulating levels of low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol (TotC) and total triglycerides (TG), as well as risk of coronary artery disease (CAD), ischaemic stroke and the inflammatory phenotypes showing significant genetic associations in the present study. The genome breakpoint data set for genetic determinants of circulating levels of low-density lipoprotein-cholesterol, high-density lipoprotein cholesterol in the NFBC1966 population. In chr17 the lead variant rs8176746 with concentrations of sE-selectin and sVCAM-1 were highly significant when adjusted for the ABO blood type (p=3.40×10\textsuperscript{-12} and p=3.43×10\textsuperscript{-17}, respectively). Overall, these results suggest that the association of the rs8176746 with sVCAM-1 level is independent of the A1 subtype, while the association of the rs2519093 with sE-selectin and sICAM-1 levels is independent of the ABO blood type. Statistical significances were abolished when the rs8176746 association with sVCAM-1 was adjusted for ABO blood type and rs2519093 association with sE-selectin or sICAM-1 was adjusted for rs507666.

To further evaluate the related molecular mechanisms, we determined the effect estimates of the ABO blood types and ABO blood types stratified by rs507666 genotype on soluble CAM levels. The blood type A showed negative associations with the levels of all the three CAMs and the effect was the most robust on the sE-selectin level (figure 2, left panel). However, major discrepancies in the effect directions were seen when the analyses were stratified by the rs507666 genotype (figure 2, right panel). Congruent with previous reports,\textsuperscript{18,26} the present results suggest that the A1 subtype/rs507666 influences sE-selectin or sICAM-1 levels. In contrast, the blood type B seems to attribute predominantly to sVCAM-1 level, while the A1 subtype/rs507666 shows only a modest effect on sVCAM-1.

**RESULTS**

The basic characteristics of the NFBC1966 study population are provided in table 1. Inflammatory phenotype distributions are tabulated in online supplementary table S1, and their correlation structure is shown in online supplementary figure S1. Using a threshold of p<3.1×10\textsuperscript{-15} for statistical significance (standard genome-wide significance level p<5×10\textsuperscript{-8} corrected for 16 phenotypes tested), we identified seven novel and six previously reported loci associating with one or more of the inflammatory phenotypes. The results are summarised in table 2 and combined Manhattan plots are shown in figure 1. Manhattan plots and Q-Q plots for each inflammatory phenotype are provided in online supplementary figure S2 A–Z. Genomic inflation factor values range between 0.99 and 1.02, suggesting no inflation in the test statistics (online supplementary table S2). Online supplementary table S3 lists the traits associated previously with the loci showing novel 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 (rs117238625 is in LD (r\textsuperscript{2}=1 in NFBC1966) with rs117468318 that locates in the 5′ untranslated region (UTR) region of HSP90B1 (heat shock protein 90 kDa beta member 1) and, according to RegulomeDB,\textsuperscript{28} is likely to affect transcription factor binding. The association signal in chr17 locates near ABCA8 (ATP binding cassette subfamily A member 8) encoding one of the ATP binding cassette transporters.

Variations in sialytransferase encoding genes show an effect on sE-selectin level

For sE-selectin level, we identified a novel association in 11q24.2 in the region of ST3GAL4 (ST3 beta-galactoside alpha-2,3-sialyltransferase 4). We identified a suggestive signal with sE-selectin level also in 3q12.1 near ST3GAL6 (ST3 beta-galactoside alpha-2,3-sialyltransferase 6), but the association was not significant after multiple correction (p=1.75×10\textsuperscript{-8}). Both of the sialyltransferase genes have been implicated in the production of functional E-selectin, P-selectin and L-selectin ligands in mice.\textsuperscript{29}

Two independent association signals on sICAM-1 level near ICAM1

We replicated the previously reported association for sICAM-1 level in 19p13.2 near ICAM1 (intracellular adhesion molecule 1).\textsuperscript{18,30} When the primary association test was conditioned for the lead variant rs117960796, another significant association was detected (rs74428614, p=1.14×10\textsuperscript{-16}) indicative of more than one independent variant contributing to sICAM-1 level in this locus.

**Vascular endothelial growth factor**

In the meta-analyses, we identified a novel locus 4p16.2 with a large effect on VEGF (β=−2.38 SD). This locus harbours genes EVC (EvC ciliary complex subunit 1), EVC2 (EvC ciliary complex subunit 2) and STK32B (serine/threonine kinase 32B). In addition, we replicated two previously reported loci associating with significant when adjusted for the rs507666 indicative of the A1 subtype (p=4.98×10\textsuperscript{-15}). On the contrary, the associations of the rs2519093 with concentrations of sE-selectin and sVCAM-1 were highly significant when adjusted for the ABO blood type (p=3.40×10\textsuperscript{-12} and p=3.43×10\textsuperscript{-17}, respectively). Overall, these results suggest that the association of the rs8176746 with sVCAM-1 level is independent of the A1 subtype, while the association of the rs2519093 with sE-selectin and sICAM-1 levels is independent of the ABO blood type. Statistical significances were abolished when the rs8176746 association with sVCAM-1 was adjusted for ABO blood type and rs2519093 association with sE-selectin or sICAM-1 was adjusted for rs507666.

Table 1 Basic characteristics of the Northern Finland Birth Cohort 1966 study population

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of individuals</td>
<td>5284</td>
</tr>
<tr>
<td>Number of men (%)</td>
<td>2543 (48.1)</td>
</tr>
<tr>
<td>Age, years</td>
<td>31.1±0.4</td>
</tr>
<tr>
<td>Body mass index, kg/m\textsuperscript{2}</td>
<td>24.4±4.0</td>
</tr>
<tr>
<td>Glucose, mmol/L</td>
<td>5.1±0.7</td>
</tr>
<tr>
<td>Low-density lipoprotein-cholesterol, mmol/L</td>
<td>3.0±0.9</td>
</tr>
<tr>
<td>High-density lipoprotein-cholesterol, mmol/L</td>
<td>1.6±0.4</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>124.2±13.6</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>76.8±11.7</td>
</tr>
</tbody>
</table>

Values are means±SD.
### Table 2  Significant loci associating with the circulating inflammatory phenotypes

<table>
<thead>
<tr>
<th>Study</th>
<th>Marker</th>
<th>Locus</th>
<th>Chr:Position</th>
<th>Candidate gene</th>
<th>Nearest gene(s)</th>
<th>Annotation</th>
<th>dbsNP reference</th>
<th>INFO</th>
<th>EA</th>
<th>EAF</th>
<th>Beta</th>
<th>P value</th>
<th>HetPVal</th>
<th>Variance explained</th>
<th>Total variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFBC1966</td>
<td>sE-selectin</td>
<td>9q34.2</td>
<td>9:136 141 870</td>
<td>ABO</td>
<td>ABO</td>
<td>Intronic</td>
<td>rs2519093</td>
<td>0.994</td>
<td>T</td>
<td>0.188</td>
<td>−0.903</td>
<td>4.48e-305</td>
<td>NA</td>
<td>0.249</td>
<td>0.258</td>
</tr>
<tr>
<td>sICAM-1</td>
<td>sICAM-1</td>
<td>9q34.2</td>
<td>9:136 141 870</td>
<td>ABO</td>
<td>ABO</td>
<td>Intronic</td>
<td>rs2519093</td>
<td>0.994</td>
<td>T</td>
<td>0.188</td>
<td>−0.352</td>
<td>7.43e-48</td>
<td>NA</td>
<td>0.038</td>
<td>0.118</td>
</tr>
<tr>
<td>sVCAM-1</td>
<td>sVCAM-1</td>
<td>9q34.2</td>
<td>9:136 131 322</td>
<td>ABO</td>
<td>ABO</td>
<td>Intronic</td>
<td>rs8176746</td>
<td>1.000</td>
<td>T</td>
<td>0.129</td>
<td>0.256</td>
<td>5.06e-19</td>
<td>NA</td>
<td>0.015</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12q23.3</td>
<td>12:104 448 391</td>
<td>HSP90B1</td>
<td>HSP90B1</td>
<td>Intronic</td>
<td>rs117238625</td>
<td>0.981</td>
<td>A</td>
<td>0.023</td>
<td>0.510</td>
<td>2.90e-14</td>
<td>NA</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17q24.3</td>
<td>17:66 923 805</td>
<td>ABCA8</td>
<td>ABCA8</td>
<td>Intergenic</td>
<td>rs112001035</td>
<td>0.883</td>
<td>A</td>
<td>0.060</td>
<td>−0.324</td>
<td>1.04e-13</td>
<td>NA</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>Meta-analyses</td>
<td></td>
<td></td>
<td></td>
<td>HLA locus</td>
<td>HLA locus</td>
<td>Intronic</td>
<td>rs6917603</td>
<td>1.000</td>
<td>C</td>
<td>0.251</td>
<td>−0.163</td>
<td>1.76e-12</td>
<td>1.00</td>
<td>0.010</td>
<td>0.015</td>
</tr>
<tr>
<td>IL1β</td>
<td>6p22.1</td>
<td>6:30 017 071</td>
<td>HLA locus</td>
<td>HLA locus</td>
<td>HLA locus</td>
<td>Intronic</td>
<td>rs9261224</td>
<td>1.000</td>
<td>T</td>
<td>0.035</td>
<td>0.261</td>
<td>1.31e-09*</td>
<td>1.00</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>IP10</td>
<td>4q21.1</td>
<td>4:36 899 176</td>
<td>CXCL10</td>
<td>CXCL10</td>
<td>CXCL10</td>
<td>Intronic</td>
<td>rs192716315</td>
<td>0.851</td>
<td>C</td>
<td>0.003</td>
<td>1.513</td>
<td>2.71e-13</td>
<td>1.00</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>MCP1</td>
<td>1q23.2</td>
<td>1:159 175 354</td>
<td>ACKR1</td>
<td>ACKR1</td>
<td>ACKR1</td>
<td>Missense</td>
<td>rs12075</td>
<td>1.000</td>
<td>A</td>
<td>0.469</td>
<td>0.148</td>
<td>1.43e-33</td>
<td>1.51e-13</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>TNFx</td>
<td>13q14.3</td>
<td>13:51 141 997</td>
<td>DLXEU1</td>
<td>DLXEU1</td>
<td>DLXEU1</td>
<td>Intronic</td>
<td>rs1704575</td>
<td>0.803</td>
<td>G</td>
<td>0.002</td>
<td>2.131</td>
<td>2.71e-09</td>
<td>1.00</td>
<td>0.018</td>
<td>0.018</td>
</tr>
<tr>
<td>VEGF</td>
<td>4p16.2</td>
<td>4:53 636 073</td>
<td>STK32B</td>
<td>STK32B</td>
<td>STK32B</td>
<td>Intronic</td>
<td>rs186725382</td>
<td>0.875</td>
<td>A</td>
<td>0.001</td>
<td>−2.380</td>
<td>4.53e-10</td>
<td>1.00</td>
<td>0.011</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>6p21.1</td>
<td>6:43 927 050</td>
<td>VEGFA</td>
<td>VEGFA</td>
<td>VEGFA</td>
<td>Intergenic</td>
<td>rs7767396</td>
<td>1.000</td>
<td>A</td>
<td>0.523</td>
<td>0.284</td>
<td>8.33e-105</td>
<td>1.22e-09</td>
<td>0.040</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>9p24.2</td>
<td>9:2 866 273</td>
<td>VDLR</td>
<td>VDLR, KCNV2</td>
<td>VDLR, KCNV2</td>
<td>Intergenic</td>
<td>rs7030781</td>
<td>0.959</td>
<td>T</td>
<td>0.373</td>
<td>−0.099</td>
<td>1.57e-13</td>
<td>5.34e-04</td>
<td>0.005</td>
<td></td>
</tr>
</tbody>
</table>

Statistical significance is considered at $p<3.1\times10^{-9}$. Novel findings are highlighted with bold font. All positions correspond to human genome build 37.

*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; HLA, human leukocyte antigen; HetPVal, p value of heterogeneity as estimated by Cochran’s Q-test; IL1β, interleukin 1-beta; INFO, imputation score in NFBC1966; IP10, interferon gamma-induced protein 10; MCP1, monocyte chemotactic protein 1; NA, not available; NFBC1966, Northern Finland Birth Cohort 1966; TNFα, tumour necrosis factor alpha; VEGF, vascular endothelial growth factor; dbsNP, single nucleotide polymorphism database; sE-selectin, soluble E-selectin; sICAM-1, soluble intercellular adhesion molecule -1; sVCAM-1, soluble vascular cell adhesion molecule-1.
Figure 1 The combined Manhattan plots for significant associations with inflammatory markers studied in (A) Northern Finland Birth Cohort 1966 and in (B) meta-analyses with three other Finnish population cohorts. Significance threshold \(p < 3.1 \times 10^{-9}\) derives from the standard \(p\) value limit for genome-wide significance \(p < 5 \times 10^{-8}\) 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. sE-selectin, soluble E-selectin; IL1b, interleukin 1-beta; IP10, interferon gamma-induced protein 10; MCP1, monocyte chemoattractant protein 1; sICAM-1, soluble intercellular adhesion molecule-1; sVCAM-1, soluble vascular cell adhesion molecule-1; TNFa, tumour necrosis factor alpha; VEGF, vascular endothelial growth factor.
**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.

**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 (online supplementary table S4).

**Comparisons of genetic effects on inflammatory phenotypes versus other traits**

Elevated circulating concentrations of inflammatory markers increase the risk for CVD.19 20 We used the gwas-pw method21 to evaluate the presence of shared genetic determinants between inflammatory phenotypes showing significant genetic association in the present study and other cardiovascular health-related traits (LDL-C, HDL-C, TOTC, TG, CAD risk, ischaemic stroke risk) obtained from open-access sources.22–24 Altogether 56 genomic regions showed robust statistical evidence for containing a genetic variant influencing one or more of the inflammatory phenotypes and at the same time one or more of the other traits studied (model 3 posterior probability greater than 0.99; online supplementary figure S3). The ABO locus was one of the loci harbouring variants influencing multiple traits. In this locus, we observed negative linear relationships between the SNP effects on sE-selectin and sICAM-1 levels and CAD risk, stroke risk, as well as LDL-C, HDL-C and TOTC levels (figure 3). The results in the other loci are provided in online supplementary figure S3.

**DISCUSSION**

The present study examines the genetic determinants of 16 circulating inflammatory phenotypes in 5284 individuals from Northern Finland with a subsequent meta-analysis of 10 phenotypes in 3 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 identified a novel association for sVCAM-1 concentration at the ABO locus. This locus was also associated with sE-selectin and sICAM-1 levels as observed previously.18–26 The present GWAS suggested two distinct association signals in the ABO locus for the sE-selectin and sICAM-1 levels versus sVCAM-1 level, and the supplementary tests provided further support for at least two mechanisms contributing to circulating concentrations of CAMs in this locus. The two mechanisms include the blood type A subtype A1, which has a robust lowering effect on sE-selectin and sICAM-1 levels,18–26 and the blood type B which seems to have an increasing effect on sVCAM-1 level. 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 CAMs from the endothelium and/or their clearance rate from circulation.19 26 The underlying mechanism explaining the association between the blood type B and 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.31 As the B antigen holds an additional adhesion involves interaction with galectin-3, a protein that has a specificity for galactosides.31 As the B antigen holds an additional
circulation could be influenced by a possible competitive binding of galectin-3 with sVCAM-1 and the B antigen.

To evaluate the shared genetic mechanisms, we compared the correspondence of genetic effects on inflammatory phenotypes versus cardiovascular health-related traits. We observed a negative relationship between the genetic effects on CAM levels and the genetic effects on LDL-C and HDL-C levels, as well as lower risk for CAD and ischemic stroke, in the ABO locus. This denotes that the genotypes in the ABO locus associating with higher levels of soluble CAMs tended to associate with lower circulating cholesterol levels as well as lower risk of cardiovascular outcomes. This was unexpected since previous evidence suggests that increased soluble CAM levels are linked with atherosclerosis progression and vascular outcomes. Possible explanations unravelling the negative correlation advocate that soluble CAMs may compete with leucocyte adhesion to the endothelial molecules or that enhanced ectodomain shedding may contribute to the reduced recruitment of leucocytes to the subendothelial space, thereby promoting cardioprotective effects. Our results suggesting a negative relationship between the genetic effects on soluble CAM and circulating cholesterol levels advocate 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 or TotC show positive correlation with CAD risk. Nevertheless, further studies are warranted to understand the exact mechanisms.

Another novel association with sVCAM-1 level was detected in chr12. The lead SNP of this locus is in LD with rs117468318 (r² = 1 in NFBC1966) that locates in the 5'UTR of HSP90B1 encoding heat shock protein gp96 and, according to RegulomeDB, is likely to affect transcription factor binding, suggesting a possible regulatory mechanism for the detected association. HSP90B1/gp96 is a chaperone that is essential for assembly of 14 of 17 integrin pairs in the haematopoietic system. Integrin α4β1 is an important ligand of VCAM-1; if altered transcription of HSP90B1 had a downstream effect on integrin α4β1 level, this could further modify the level of unbound sVCAM-1 in circulation.

The third novel locus showing association with sVCAM-1 level was identified in chr17 near ABCA8. The lead SNP rs112001035 is an expression quantitative trait locus (eQTL) for ABCA8 in multiple tissue types. If ABCA8 is involved in the regulation of HDL level via interaction with ABCA1 and if plasma HDL levels contribute to VCAM-1 expression, then altered expression of the ABCA8 could influence circulating levels of sVCAM-1 by 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 metabolomics GWAS (β = −0.043 SD, p = 0.049). There is evidence suggesting that ABCA8 may be involved in sphingolipid metabolism, and it has been hypothesised that ABCA8 may be involved in the formation of specific membrane domains during ApoA-I lipolysis. Thus, the association between the ABCA8 and sVCAM-1 level could be related to altered HDL composition possibly contributing to endothelial homeostasis rather than absolute particle concentration. 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. 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 STK32B. Mutations in this locus have been associated previously with coeliac disease, CA1 and Ellis-van Creveld syndrome. STK32B may play a role in the hedgehog signalling pathway, which has been implicated in metastasis and angiogenesis in cancer and downregulated in coeliac disease. The hedgehog signalling has shown to be involved in the regulation of VEGF expression during developmental angiogenesis in avian embryo. 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. This region is recurrently deleted in tumours and haematopoietic malignancies. DLEU1 is a part of a transcriptionally coregulated gene cluster that modulates the activity of the nuclear factor kappa B (NF-kB) pathway, which is also modulated by TNFα. It is largely unknown how the DLEU1 and related DLEU2 regulate NF-kB activity; 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 IL1β and showing suggestive association also on IL4 level. This association signal is in the region coding the human leucocyte antigen proteins, and
Complex traits

further experimental evidence would be needed to identify the exact mechanism how the locus contributes to IL 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. The use of genetically isolated populations, such as inhabitants of Northern Finland, should further enhance the power for locus identification in Gwas settings. We were able to perform meta-analyses only for 10 out of the total of 16 inflammatory phenotypes, and the novel findings are largely based on NFBC1966 population only. Thus, replication of the present findings in other populations would be helpful. In particular, the associations of the novel rare, large-effect variants need to be interpreted with caution until the associations are validated in other populations. The interassay coefficient of variability measures for sE-selectin and VEGF in particular are notably larger than 15%, which is considered to be the limit for acceptable values (online supplementary table S1). However, to our consideration, all the findings identified in the present study locate on genome regions with biologically relevant genes. Furthermore, the replications of the previously reported loci speak for the data adequacy and add confidence to the novel associations. Finally, as we have not included functional experiments in this work, we are limited to previous literature when explaining the potential biological mechanism behind the identified associations.

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 CAM levels likely encompasses at least two distinct mechanisms influencing sE-selectin, sICAM-1 and sVCAM-1 levels. Our findings provide evidence that increased soluble CAM 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 seems to contribute to lower cardiovascular risk. Furthermore, genetic effects at the ICAM1 locus providing a direct molecular link to sICAM-1 concentration do not correlate with the genetic effects on CAD risk nor stroke risk. Overall, the present study extends the knowledge about the molecular pathways involved in inflammatory load.

Author affiliations

1Computational Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland
2Center for Life Course Health Research, University of Oulu, Oulu, Finland
3Biocenter Oulu, Oulu, Finland
4Department of Internal Medicine, Satakunta Central Hospital, Pori, Finland
5Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
6Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
7National Institute for Health and Welfare, Helsinki, Finland
8Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
9University of Tartu, Estonian Genome Center, Tartu, Estonia
10Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
11Institute of Biomedicine, University of Oulu, Oulu, Finland
12Department of Psychiatry, University of Eastern Finland, and Kuopio University Hospital, Kuopio, Finland
13Medicity Research Laboratory and Institute of Biomedicine, University of Turku, Turku, Finland
14Unit of General Practice, Oulu University Hospital, Oulu, Finland
15Oulu Deaconess Institute/Diapolis Oy Research Unit, Oulu, Finland
16Northern Finland Birth Cohorts, Faculty of Medicine, University of Oulu, Oulu, Finland
17Medical Research Center (MRC), University of Oulu, and Oulu University Hospital, Oulu, Finland
18Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, Poznan, Poland

REFERENCES

1Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, Imperial College London, London, UK
2Unit of Primary Care, Oulu University Hospital, Oulu, Finland
3Department of Genomics and Complex Diseases, School of Public Health, Imperial College, London, UK

Correction notice This article has been corrected since it was published Online First: values in table 2 have been amended.

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.

Contributors ES, SS and JK conceptualised the study. ES, MK and AA-O performed the statistical analyses. TK, K-HH, MS, KS and SJ performed the quantifications of the inflammatory markers. OR, MP, VS, TL, HV, MS, SJ, SK-K, MM, K-HH, M-RJ, SS and JK provided funding or other resources to conduct the study. SS and JK supervised the study. ES, SS and JK wrote the original manuscript. All authors contributed to revising the content and approved the final version.

Funding This work was supported by the University of Oulu Graduate School (ES, MK), the Finnish Foundation for Cardiovascular Research (VS), Biocenter Oulu (SS), Finnish Funding and Commission (DynAHEALTH - H2020 – 335595; SS), Academy of Finland (297338 and 307247); JK and Novo Nordisk Foundation (NNF1700C2006602; JK). NFBC1966 received financial support from University of Oulu (Grant No 65354), Oulu University Hospital (Grant No 2997 and 897), Ministry of Health and Social Affairs (Grant No 23/251/97, 16097 and 19097), National Institute for Health and Welfare, Helsinki (Grant No 54121), and Regional Institute of Occupational Health, Oulu, Finland (Grant No 50621 and 54231). The Young Finns Study has been financially supported by the Academy of Finland (grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendri) 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; Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS); European Research Council (grant 742927 for MULTIEPIGEN project); and Tampere University Hospital Supporting Foundation.

Competing interests 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.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.