Genome-wide association study identifies 17 new loci influencing concentrations of circulating cytokines and growth factors

Ahola-Olli AV¹, Würtz P², Havulinna AS³, Aalto K⁴, Pitkänen N¹, Lehtimäki T⁵, Kähönen M⁶,
Lyytikäinen LP⁵, Raitoharju E⁵, Seppälä I⁵, Sarin AP⁷, Ripatti S⁸, Palotie A⁹, Perola M⁷, Viikari
JS¹⁰, Jalkanen S⁴, Maksimow M⁴, Salomaa V³, Salmi M⁴, Kettunen J^{*2,3, 11, 12}, Raitakari OT^{*1,13}

* = equal contribution

Affiliations: 1) Research Centre of Applied and Preventive Cardiovascular Medicine, University of
Turku, Turku, Finland.

2) Computational Medicine, Center for Life Course Health Research, University of
Oulu, Oulu, Finland

3) National Institute of Health and Welfare, Helsinki, Finland

4) Medicity Research Laboratory, Department of Medical Microbiology and
Immunology, University of Turku, Turku, Finland

5) The Department of Clinical Chemistry, Fimlab Laboratories, Pirkanmaa Hospital
District, School of Medicine, University of Tampere, Tampere, Finland

6) The Department of Clinical Physiology, Tampere University Hospital, Tampere,
Finland

7) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki,
Finland; National Institute for Health and Welfare, Helsinki, Finland

8) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki,
Finland; National Institute for Health and Welfare, Helsinki, Finland; The Department
of Public Health, University of Helsinki, Helsinki, Finland;

9) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA; Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA; Department of Neurology, Massachusetts General Hospital, Boston, MA

10) The Department of Medicine, University of Turku and Division of Medicine, Turku University Hospital, Turku, Finland

11) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland

12) Biocenter Oulu, University of Oulu, Finland

13) The Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

Conflicts of interests: None

Word Count: Main text (excluding abstract and methods) 2,912. Abstract 114 words.

Running title: Cytokine genetics

Corresponding author:

Ari Ahola-Olli, BM
Centre of Applied and Preventive Cardiovascular Medicine
University of Turku
avahol@utu.fi

Abstract

Circulating cytokines and growth factors are regulators of inflammation and have been implicated in autoimmune and metabolic diseases. In this genome-wide association study (GWAS) up to $n=8,293$ Finns we identified 27 loci with genome-wide association ($P\text{-value} < 1.2 \times 10^{-9}$) for one or more cytokines, including 17 unidentified in previous GWASes. Fifteen of the associated SNPs had expression quantitative trait loci in whole blood. We provide strong genetic instruments to clarify the causal roles of cytokine signaling and upstream inflammation in immune-related and other chronic diseases. We further link known autoimmune disease variants including Crohn's disease, multiple sclerosis and ulcerative colitis with new inflammatory markers, which elucidate the molecular mechanisms underpinning these diseases and suggest potential drug targets.

Genome-wide association studies (GWAS) have identified multiple loci for inflammatory diseases^{1,2}. However, the biochemical pathways underlying the link from locus to complex disease have often remained elusive. Cytokines and growth factors (hereafter cytokines) are inflammatory regulators, and therefore important intermediate phenotypes for inflammatory diseases.³ These intermediate phenotypes can be exploited in GWASes to elucidate the biochemical pathways underlying the link from locus to disease susceptibility. Cytokines have been implicated in the progression of inflammatory bowel diseases, multiple sclerosis, atherosclerosis and in various forms of cancer. Here, we studied the genetic basis for blood levels of multiple cytokines to gain insights to the molecular intermediates and causal pathways related to inflammatory diseases.

The genetic basis of inflammatory diseases can inform drug development to prioritise new pharmaceutical targets. For example, an individual who has inherited a cytokine signaling disrupting allele is analogous to a participant in randomized trial receiving active drug designed to inhibit signaling via the pertinent cytokine. Cytokines are attractive targets since they exert their effects via cell surface receptors that are readily druggable. Characterization of the genetic underpinnings of cytokine signaling pathways is important for drug development, because pharmaceutical targets with human genetic support have twice the probability for eventual regulatory approval compared to non-supported targets⁴. Furthermore, since inflammatory diseases often share common pathology, the indications of approved drugs can often be expanded⁵. For example, a tumor necrosis factor alpha (TNF- α) antibody adalimumab has FDA approval for rheumatoid arthritis, juvenile idiopathic arthritis, ankylosing spondylitis, psoriatic arthritis and Crohn's disease, although initially studied for rheumatoid arthritis⁶⁻⁸. Therefore, to gain insights into the molecular intermediates and causal pathways related to autoimmune and metabolic diseases, we studied the genetic basis for circulating levels of 41 cytokines. In addition to informing drug development, our results provide novel information on the genetic regulation of normal physiological variation of cytokines among healthy individuals.

Results

We performed GWAS for the circulating concentrations of 41 cytokines; listed in **Supplementary Table 1**. The GWAS meta-analysis included up to 8,293 Finnish individuals from three independent population cohorts: the Cardiovascular Risk in Young Finns Study (YFS), FINRISK1997 and FINRISK2002. Study cohort characteristics are reported in **Supplementary Table 1**. On average, the YFS participants are younger than the FINRISK1997 and FINRISK2002 participants (37 vs. 60 years). Correlations between the cytokine levels are shown in **Supplementary Figure 1**. Four cytokines (VEGF, IL12p70, IL13 and IL10) form a tightly correlated cluster. Most of the cytokines included in the study are positively correlated with each other.

GWAS identifies 27 loci associated with circulating cytokines

Using an additive genetic model adjusted for the ten first genetic principal components, age, sex and body mass index, we tested for univariate associations between 10.7 million genetic polymorphisms and 41 cytokine concentrations. Genotype imputation was performed based on reference haplotypes provided by the 1000 Genomes Project September 2013 release⁹. The meta-analyses identified 27 genome-wide significant loci ($P < 1.2 \times 10^{-9}$ accounting for 41 measures); 17 of which have not been associated with cytokines, C-reactive protein or white blood cell count in previous GWASes (**Table 1 and Figure 1**). With the aforementioned threshold for genome-wide significance and sample size, we had 84 % power to detect an association with effect size of 0.11 standard deviations per one additional copy of effect allele assuming minor allele frequency (MAF) of 50%. The potential candidate genes of the 27 genetic variants are also reported in Table 1. Results from the random effect meta-analysis for heterogenic single nucleotide polymorphisms (SNPs) (HetPval < 0.1) are reported in **Supplementary Table 2**. Effect estimates from the random effect models were coherent with effect estimates from the fixed effect models; however, for ten of

the heterogenic loci the P-value did not reach the genome-wide significance (1.2×10^{-9}) with the random effect model. Complete summary statistics from meta-analyses are available at (http://www.will_be_announced_later.com). Cytokine concentration means per allele in absolute concentration units for lead SNPs are reported in **Supplementary Table 3** and genomic lambda values in **Supplementary Table 4**. The lambda values ranged from 0.98 to 1.02 indicating that there was no overall inflation in the test statistics.

Novel loci with large effects included rs192989810 from 18q12.1 for TNF-related apoptosis inducing ligand (TRAIL) with $\beta=2.2$ SD at MAF of 1%. This SNP is located within *MEP1B*, and leukocytes from *MEP1B*^{-/-} mice have been shown to have impaired migration in response to monocyte chemotactic protein-1 (MCP1) and macrophage inflammatory protein-1 α (MIP1 α)¹⁰. Additional low-frequency variants with large effects included rs141053179 ($\beta=1.1$ SD) for interferon gamma-induced protein 10 (IP10), rs116924815 ($\beta=0.61$ SD) for stem cell growth factor beta (SCGFb), and rs4778636 ($\beta=-0.73$ SD) for interleukin-16 (IL16). Novel common variants (MAF>5%) with large effects (β up to -0.49) included rs2070074 and rs13412535 for Cutaneous T-cell attracting (CTACK) and platelet derived growth factor BB (PDGFbb). In addition to discovering new loci, our results confirmed associations for 10 genome-wide significant loci previously identified. These include two loci with interleukin-18 (IL18) concentration (2p22.3 and 11q23.1)^{11,12}. We also replicated two loci previously associated with vascular endothelial growth factor (VEGF; 6p21.1 and 9p24.2)¹³ and one additional locus previously associated with monocyte chemotactic protein-1 (MCP1; 1q23.2)¹⁴. Manhattan plots, QQ-plots and detailed plots from heterogeneity for each cytokine are provided in the **Supplementary note**. For 6p21.1 associated with vascular endothelial growth factor (VEGF), the association signal was narrow and hereby suggested only few causal SNP candidates (Supplementary note, page 55). After conditioning on rs6921438, the subsequent lead SNP was rs12214617 which is located 46kB from rs6921438. According to Ensembl, the secondary association signal rs12214617 is located on the promoter

flanking region in various cell types including monocytes and HepG2 cells, thus suggesting a potential role in gene regulation. A previous GWAS of VEGF levels identified the same lead SNP associated with VEGF concentration in 6p21.1¹³. Our results confirm this association but the secondary associations signal was different from ours, potentially due to different imputation reference. For some of the genetic loci associated with cytokines the association signal is much wider, which makes the identification of the potential functional variant more challenging. For instance, the signal spans approximately 600kB for IL18-associated 2p22.3. The association peak for PDGFbb in 2q36.1 is mainly formed by only two variants (rs13412535 and rs68066031) in strong linkage disequilibrium. However, after conditioning on rs13412535 the peak is lost. According to Ensembl, the rs13412535 is located in a transcription factor binding site and DNase peak indicating accessible chromatin. Thus, rs13412535 is likely affecting to transcription factor binding and gene expression of *SERPINE2*.

Four of the identified loci were associated with the concentration of more than one cytokine (1q23.2, 6p21.1, 3p22.1, and 10q21.3). The 1q23.2 harbors rs12075, which is genome-wide significant for three cytokines: eotaxin, MCP1 and growth-regulated oncogene- α (GRO α). The lead SNP is located within *DARC*, which encodes a receptor for multiple cytokines and is also a human erythrocyte blood group antigen^{15,16}. This suggests that the effect of rs12075 on the three cytokines is not directly mediated via any of the three pertinent cytokines, but rather through altering binding of these cytokines to the DARC receptor (**Fig. 2a**). Similarly, the effect of rs2228467 at 3p22.1 on eotaxin and MCP1 appears to be mediated via chemokine-binding protein 2 (*CCBP2*), rather than directly via eotaxin or MCP1 (**Fig. 2b**). In contrast, a variant at 6p21.1 (rs6921438) is associated with concentration of five cytokines (VEGF, IL12p70, IL7, IL10, and IL13). This locus contains a *VEGFA* gene encoding VEGF, which suggest that VEGF might regulate the concentration of other cytokines associated with rs6921438. To test this directional mediation, we performed Mendelian randomization analysis using a genetic score of rs6921438 and rs12214617 (variants at 6p21.1

independently associated with VEGF) as the instrumental variable¹⁷. The observational and causal effect estimates were matching each other, supporting the role of VEGF as an upstream regulator of IL12p70, IL7, IL10, and IL13 (**Fig. 2c-d**). Variants in 10q21.3 were associated with VEGF and IL12p70. The lead SNP within this locus is rs10761731, which has the smallest association test P-value for VEGF. The ratio of the IL12p70 effect size divided by the effect size of VEGF remains approximately constant in both loci (6p21.1 and 10q21.3) suggesting tight regulatory effect of VEGF on IL12p70 concentration.

To assess whether the genetic loci identified harbor additional independent variants, we performed conditional meta-analyses for each cytokine-lead SNP pairs. The 17q12 locus harbored five variants independently associated with macrophage inflammatory protein-1 β (MIP1b) concentration (**Supplementary Table 5**). These five statistically independent variants together explained 34% of the variation in circulating MIP1b. The variance explained by the independent variants for each cytokine are listed in **Supplementary Table 6**. The SNPs independently associated with VEGF explain 15% of the variation in circulating VEGF. The *MEP1B* locus had a large influence on circulating TRAIL and the variants identified here explain 14% of observed variance in TRAIL concentration. In addition, GRO α was shown to be under a strong genetic control as the identified SNPs explained 10% of the circulating GRO α variance.

Expression analyses link three SNPs, cytokine mRNA and circulating cytokine concentrations

To further elucidate the molecular mechanisms underlying the genome-wide association peaks, we performed peripheral blood expression quantitative trait loci (eQTL) analyses for lead SNPs. Transcriptomic data were meta-analyzed for 2,177 participants from YFS and FINRISK2007. These analyses identified 26 significant eQTLs for 15 SNPs. The eQTLs for SNPs in Table 1 are presented in **Table 2**; all significant eQTLs, including SNPs not listed in Table 1, are reported in **Supplementary Table 7**. Three SNPs were associated with both the mRNA encoding the cytokine

and the pertinent circulating cytokine concentration. Allele rs4778636-A in *IL16* was associated with decreased concentration of circulating IL16 (Table 1), but increased concentration of *IL16* mRNA (**Fig. 3**). The Ensembl's Variant Effect Predictor suggests that the A-allele creates a splice site variant leading to nonsense-mediated decay of *IL16* mRNA¹⁸. Increased *IL16* mRNA concentration may therefore be a compensatory mechanism to account for missing circulating IL16. The genome-wide association of rs4778636 with circulating IL16 confirms results from a small candidate gene study¹⁹. The MIP1b-associated variant (rs113877493) is located approximately 200kB away from a cluster of cytokine-coding genes in 17q21. According to our results, the MIP1b-associated SNP is an eQTL for four of the genes (*CCL4L1*, *CCL4L2*, *CCL3L1*, *CCL3L3*) belonging to this cluster. The variant which was associated with concentration of PDGFbb and resided at a transcription factor binding site (rs13412535) was associated with expression of *SERPINE2*. Interestingly, rs13412535 is located within an intron of *SERPINE2* within the gene sequence instead of upstream from the gene, where transcription factors have been traditionally thought to bind.

Novel drug targets for celiac disease and Behcet's disease and wider indication for IL2ra antagonists

To bridge the connection between the cytokine traits and disease, we searched the GWAS catalogue for SNP-trait and SNP-disease associations within a 1Mb window from cytokine lead SNPs²⁰ (www.genome.gov/gwastudies). The search window approach was selected since many GWASes for the complex diseases have been conducted with HapMap2 and therefore missed variants present in the 1000 Genomes imputation panel²¹. The results for cytokine-linked SNPs previously associated with disease status are listed in **Table 3**; the corresponding associations for cytokine-linked SNPs previously associated with quantitative traits are listed in **Supplementary Table 8**. The circulating interleukin-2 receptor alpha subunit (IL2ra) increasing variant (rs12722489-C) has previously been shown to increase the risk for Crohn's disease and multiple sclerosis^{1,22}. The

206 variant rs943072-G here associated with increased concentrations of VEGF and IL12p70 has
 207 previously been linked to increased risk for ulcerative colitis²³. In addition, both cytokines have
 208 been shown to be increased in patients with an inflammatory bowel disease^{24,25}. These results link
 209 the VEGF cascade depicted in Figure 2C to inflammatory bowel disease. Variant rs7616215-T here
 210 associated with increased concentration of MIP1b has previously been shown to decrease the risk
 211 for Behcet's disease²⁶. In addition to disease associations, the cytokine loci identified here were
 212 linked to 15 different molecular traits. For example, rs2251746 located at 1q23.2 within *FCERIA*
 213 and previously linked to IgE concentration was here associated with circulating concentration of
 214 GROa and MCP1.

Discussion

This study identified 27 (17 novel) genetic loci contributing to circulating concentration of one or more cytokines in the general population. In three cases, the genomic analyses linked the genetic variant via mRNA to the circulating cytokine concentration, and therefore likely pinpointed the causal gene. Genetic evidence identify a potentially novel drug target for treatment of Behcet's disease and celiac disease, since variants associated with decreased concentration of MIP1b in this study have previously been linked with increased risk of these diseases^{26,27}. Furthermore, our results point to a novel indication for IL2ra-antagonist, since the IL2-ra variant identified in our study has previously been associated with increased risk of Crohn's disease. The IL2ra-targeting antibody daclizumab reduces annualized relapse rate in multiple sclerosis patients by 50% compared to placebo²⁸. Our results suggest that daclizumab might be beneficial for patients with Crohn's disease as well.

The association of one locus to multiple circulating cytokines might help to understand clustering of autoimmune diseases. For example, a person with rheumatoid arthritis has more than 160-fold risk for autoimmune thyroiditis compared to the general population⁵. A genetic variant in *VEGFA* was here associated with increased circulating VEGF and shown to have a causal cascade effect on IL7, IL10, IL13 and IL12p70 (Fig 2C-D). Axitinib, a VEGF-receptor inhibitor, has been shown to reduce IL12p70 production in cultured monocyte-derived dendritic cells²⁹, which support our findings. Furthermore, *VEGFA*-transcription follows the same temporal pattern with *IL7* transcription during wound healing, and it has been suggested that the secretion of these two cytokines might be connected by the same pathway³⁰. These results from prior functional studies support our genetic evidence in humans that VEGF causally regulates secretion of IL7, IL10, IL13 and IL12p70 under normal physiological conditions in healthy individuals. VEGF contributes to pathogenesis of asthma by stimulating angiogenesis, edema, inflammation and airway remodeling. These effects are partly mediated via VEGF-mediated upregulation of IL13³¹. Most importantly, by

linking the VEGF cascade with the risk for ulcerative colitis, our results suggest that drugs targeting VEGF might have potential in treatment of ulcerative colitis²³. Furthermore, variants located 530 kB from rs10761731 (10q21.3; i.e. also associated VEGF like 6p21.1) were associated with ulcerative colitis². The same region had suggestive association with age at diagnosis of inflammatory bowel disease in another immuno-chip-based study³². This is further substantiated by GWAS of inflammatory bowel diseases, where the identified loci were enriched in pathways related VEGF cascade cytokines (IL12 and IL10)².

Another result with direct implications for drug development was found for the TRAIL-associated *MEP1B* locus. Drugs targeting TRAIL-signaling include Conatumumab, which is currently in phase II trials for treatment of a variety of cancers (ClinicalTrials.gov Identifier: NCT01327612). The low-frequency variant here identified to have a major influence on circulating TRAIL levels could help to clarify the causality of TRAIL-signaling in the development of cancer and assess potential side-effects from TRAIL-lowering similarly as done for IL1 α / β and other targets³³.

Plasma IL16 levels have previously been associated to HIV progression³⁴, but the locus has not been linked to HIV-related traits in GWAS³⁵. This might be due to the small allele frequency in Europeans or possible recessive effects. However, our results offer a possibility to examine HIV progression in human knock-outs, since the frequency of rs4778636-AA homozygote is up to 25% in the Yoruba population of Nigeria³⁶. Circulating IL16 has been suggested to serve as a biomarker for impaired kidney transplant function, emphysema and interferon- β treatment efficacy in multiple sclerosis^{37–39}. If IL16 is used as a biomarker, the relatively high frequency of rs4778636-AA homozygotes in African and Asian ethnicities must be accounted for.

Although TNFa antibodies have been successfully used to treat various autoimmune diseases^{6,7}, no associations with cytokine concentrations were detected within the *TNF* locus (encoding TNF- α) in the MHC class III region at 6p21.33. The MHC region displays a unique pattern of linkage

disequilibrium: a subset of MHC haplotypes have moderate correlation with each other whereas variants within same haplotype block form a stronger correlation structure⁴⁰. This distinct linkage disequilibrium pattern may create complex epistasis to MHC locus, which might impede the association studies in this region. Furthermore, the cytokines were measured in conditions without stimulation. Measuring the cytokines concentrations after antigen stimulus might reveal additional loci contributing to circulating cytokine concentrations.

Variable degree of heterogeneity was found for of the 15 genetic loci identified. For most of the heterogenic loci, another SNP with genome-wide association but no heterogeneity could be found (see Supplemental note p. 111-124). The reason for occasional heterogeneity is likely differences in stability and accuracy of the cytokine experimental assay. In the case of the association between rs12075 and MCP1, differences in blood sample processing may also contribute: heparin treatment have been shown to release significant amounts of MCP1 from DARC, and thus affect the association signal caused by altered receptor binding properties⁴¹.

In conclusion, we identified 17 new loci contributing to the genetic regulation of circulating concentrations of cytokines. Improved understanding of the genetic basis of these inflammatory markers will help to clarify the causal roles of cytokine signaling and upstream inflammation in immune-related and other chronic diseases. By linking rs4778636 to undetectable concentration of IL16, these results enable studies of HIV and other inflammatory diseases in IL16 knock-out humans. In addition, we identified a potential novel drug target for treatment of Behcet's disease and celiac disease as well as indicated a possibility to expand indication of daclizumab from multiple sclerosis to Crohn's disease. These results provide the basis for further studies on the molecular regulation of the immune system in health and disease.

Methods

Study populations

The Cardiovascular Risk in Young Finns Study

The Cardiovascular Risk in Young Finns Study (YFS) is a multicenter follow-up study with randomly chosen subjects from Finnish cities of Helsinki, Kuopio, Oulu, Tampere, and Turku and their rural surroundings. The study began in 1980 when 3596 children and young adults participated to first cross-sectional survey. The follow-up visits have been conducted in 1983, 1986, 1989, 2001, 2007 and 2011. The present cross-sectional study includes 2,019 unrelated individuals who participated to 2007 follow-up and who had both cytokine measurements and genotype data available. In addition, gene expression data from 1,664 participants of the 2011 follow-up were analyzed for the present study. All participants gave written informed consent and the study was approved by local ethics committees⁴².

FINRISK

FINRISK surveys are population-based cross-sectional studies conducted every five years to monitor the levels of chronic disease risk factors in Finland. Each survey includes 25-74-year-old randomly chosen subjects from five geographical areas of Finland. The present study analyses cytokine data from participants of the 1997 and 2002 surveys. The study visit includes a clinical examination and semi-fasting blood sampling. For eQTL analyses, a peripheral blood sample was drawn to quantify mRNA expression profiles from a subset of 513 FINRISK2007 participants living in the Helsinki area⁴³.

Cytokine quantification

From YFS and FINRISK2002, total of 48 cytokines were measured by 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 software⁴⁴. The assays were performed according to manufacturer's instructions, except, that the amount of beads, detection antibodies and streptavidin-phycoerythrin conjugate were used with 50% lower concentrations than recommended. Only measures within the cytokine-specific detection range were included in the analyses. Cytokines with >90% of values missing were excluded (7 out of 48).

In FINRISK1997, a total of 17 cytokines overlapped with those measured in FINRISK2002 and YFS and were thus included in the GWAS. Genome-wide and cytokine data were available from up to 4608 FINRISK1997 participants and up to 1705 FINRISK2002 participants. The cytokine quantification was performed from EDTA plasma in FINRISK1997, from heparin plasma in FINRISK2002 and from serum in YFS.

Statistical analyses

Subjects whose cytokine concentration were below or above of laboratory analysis detection limits were omitted from the analyses. Cytokine distributions were first normalized with inverse transformation. The transformed phenotypes were then adjusted for age, sex, body mass index and ten first genetic principal components by calculating residuals of linear regression model. Subsequently, another inverse transformation was performed for model residuals to ensure normally distributed phenotypes. All effects sizes are hereby in SD-scaled units.

Genome-wide association testing were performed using Snptest2 software version 2.5beta^{45,46}. Imputation inaccuracy was assessed with missing data likelihood score test. Allele effects were estimated from additive model (-frequentist 1). To prevent centering and scaling of phenotypes, use_raw_phenotypes option was enabled.

Meta-analyses were performed using METAL software (version 2011-03-25)⁴⁷. Prior to meta-analysis all markers with imputation info <0.7, model fit info <0.7 and minor allele count <10 were excluded. After filtering, 10.7 million markers were included in the meta-analysis. Each association test was weighted by sample size. Genomic control correction was used to account for population stratification and cryptic relatedness. Heterogeneity statistics based on Cochran's Q-test were calculated for all markers included in meta-analysis to estimate heterogeneity of effect sizes across the three cohorts. Statistical significance was set at $P < 1.2 \times 10^{-9}$, i.e. dividing the established genome-wide statistical significance ($P < 5 \times 10^{-8}$) by 41 (number of cytokines analyzed). Random effect meta-analyses were performed with *R* software. To assess whether cytokine-associated loci harbor multiple variants independently associated with cytokine concentration, the models were further adjusted with locus-specific lead SNP and the model was fitted against the concentrations of all included cytokines. All independent SNPs were used to calculate the total variance explained by the identified loci. The proportion of total variance explained by independent SNPs was calculated by following formula:

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

where β is the regression coefficient and MAF is minor allele frequency⁴⁸.

Instrumental variable analysis of the VEGF cascade effects were performed with two-stage least squares regression using *ivreg* from the R-package *AER*¹⁷. A weighted gene score based on the sum of cytokine-increasing alleles of the two independent variants associated with VEGF (rs6921438 and rs12214617) was used as the instrument. The estimates from the three individual cohorts were combined with inverse variance weighted meta-analyses. The observational associations were calculated as linear regression between the cytokines after adjusting for age, sex, body-mass index, and the ten first genetic principal components.

eQTL analyses

Gene expression data from peripheral blood leukocytes were available from 1,664 participants from the YFS 30-year follow-up survey in 2011. After preprocessing the expression levels were analyzed with an Illumina HumanHT-12 version 4 Expression BeadChip. Raw Illumina probe data was exported from Beadstudio and analysed in *R* using the *Bioconductor* packages. The HT-12 v4 BeadChip contains 47,231 expression probes. Probes expressed by <5% of the participants were excluded, resulting in 19,637 probes⁴⁹.

Gene expression data from peripheral blood leukocytes were analyzed for 518 individuals from FINRISK2007 with Illumina HumanHT-12 version 3 Expression BeadChips. Probes which were not autosomal, were complementary to cDNA from erythrocyte globin components and mapped to more than one genomic position were excluded. The filtering resulted to total of 35,419 probes included in the analyses⁴³.

Linear regression with probe as a dependent variable was used to test associations between cytokine-associated SNPs and transcripts. Age and sex were used as covariates. Genotype dosage was calculated for each included SNP with Qctool software⁵⁰. Probes with P-value <0.05 in both YFS and FINRISK07 were included in the eQTL meta-analyses. The eQTL results were combined by inverse variance weighted meta-analysis.

Database searches

To assess whether cytokine concentration could contribute to disease pathogenesis, we searched the GWAS catalogue for complex trait -associated SNPs within 1Mb window from the cytokine lead SNPs²⁰. The catalogue was downloaded on July 6, 2015 and catalogue SNPs with $P < 5 \times 10^{-8}$ were selected for further analyses. The statistical significance for SNP-cytokine association was inferred at $P < 1.2 \times 10^{-9}$.

Acknowledgements

The Young Finns Study has been financially supported by the Academy of Finland: grants 286284 (T.L.), 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001 for T.L.); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation of Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation (T.L.); Emil Aaltonen Foundation; and Yrjö Jahnsson Foundation. PW was supported by Finnish Diabetes Research Foundation, Novo Nordisk Foundation and Yrjö Jahnsson Foundation. VS was supported by the Academy of Finland, grant number 139635 and the Finnish Foundation for Cardiovascular Research.

We gratefully acknowledge Ville Aalto and Irina Lisinen for the expert technical assistance in the data management.

395 **Author contributions**

396 AVA-O, PW, JK and OTR wrote the manuscript. KA, SJ, MM and MS performed the cytokine
397 measurements. ER and IS acquired the expression data from YFS subjects and performed the
398 quality control. APS performed the genotype imputation in all involved studies. ASH, NP, TL, MK,
399 LPL, SR, AP, MP, JSV, VS gave critical comments regarding the manuscript. JK, OTR, TL, SJ and
400 VS supervised the research. OTR and VS organized the data collections.

401

References

1. Patsopoulos, N. A. *et al.* Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol.* **70**, 897–912 (2011).
2. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–24 (2012).
3. Akdis, M. *et al.* Interleukins, from 1 to 37, and interferon- γ : Receptors, functions, and roles in diseases. *J. Allergy Clin. Immunol.* **127**, 701–721 (2011).
4. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* (2015). doi:10.1038/ng.3314
5. Cooper, G. S., Bynum, M. L. K. & Somers, E. C. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J. Autoimmun.* **33**, 197–207 (2009).
6. Miller, A. V. & Ranatunga, S. K. M. Immunotherapies in Rheumatologic Disorders. *Med. Clin. North Am.* **96**, 475–496 (2012).
7. Ahluwalia, J. P. Immunotherapy in Inflammatory Bowel Disease. *Med. Clin. North Am.* **96**, 525–544 (2012).
8. Weinblatt, M. E. *et al.* Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. *Arthritis Rheum.* **48**, 35–45 (2003).
9. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–73 (2010).
10. Crisman, J. M., Zhang, B., Norman, L. P. & Bond, J. S. Deletion of the mouse meprin beta metalloprotease gene diminishes the ability of leukocytes to disseminate through extracellular matrix. *J. Immunol.* **172**, 4510–9 (2004).
11. Matteini, A. M. *et al.* Novel gene variants predict serum levels of the cytokines IL-18 and IL-1ra in older adults. *Cytokine* **65**, 10–16 (2014).
12. He, M. *et al.* Genome-Wide Association Study Identifies Variants at the IL18-BCO2 Locus Associated With Interleukin-18 Levels. *Arterioscler. Thromb. Vasc. Biol.* **30**, 885–890 (2010).

13. DeBette, S. *et al.* Identification of cis- and trans-Acting Genetic Variants Explaining Up to Half the Variation in Circulating Vascular Endothelial Growth Factor Levels. *Circ. Res.* **109**, 554–563 (2011).
14. Voruganti, V. S. *et al.* Genome-wide association replicates the association of Duffy antigen receptor for chemokines (DARC) polymorphisms with serum monocyte chemoattractant protein-1 (MCP-1) levels in Hispanic children. *Cytokine* **60**, 634–638 (2012).
15. Miller, L. H., Mason, S. J., Dvorak, J. A., McGinniss, M. H. & Rothman, I. K. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science* **189**, 561–3 (1975).
16. Horuk, R. *et al.* A receptor for the malarial parasite *Plasmodium vivax*: the erythrocyte chemokine receptor. *Science* **261**, 1182–4 (1993).
17. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–98 (2014).
18. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–70 (2010).
19. Kim, S. *et al.* Influence of genetic variation on plasma protein levels in older adults using a multi-analyte panel. *PLoS One* **8**, e70269 (2013).
20. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–6 (2014).
21. Huang, J., Ellinghaus, D., Franke, A., Howie, B. & Li, Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *Eur. J. Hum. Genet.* **20**, 801–5 (2012).
22. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–25 (2010).
23. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246–52 (2011).
24. Fuss, I. J. *et al.* Both IL-12p70 and IL-23 are synthesized during active Crohn's disease and are down-regulated by treatment with anti-IL-12 p40 monoclonal antibody. *Inflamm. Bowel Dis.* **12**, 9–15 (2006).

25. Kanazawa, S. *et al.* VEGF, basic-FGF, and TGF-beta in Crohn's disease and ulcerative colitis: a novel mechanism of chronic intestinal inflammation. *Am. J. Gastroenterol.* **96**, 822–8 (2001).
26. Kirino, Y. *et al.* Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. *Nat. Genet.* **45**, 202–7 (2013).
27. Garner, C. *et al.* Genome-wide association study of celiac disease in North America confirms FRMD4B as new celiac locus. *PLoS One* **9**, e101428 (2014).
28. Gold, R. *et al.* Daclizumab high-yield process in relapsing-remitting multiple sclerosis (SELECT): a randomised, double-blind, placebo-controlled trial. *Lancet* **381**, 2167–75 (2013).
29. Heine, A. *et al.* The VEGF-Receptor Inhibitor Axitinib Impairs Dendritic Cell Phenotype and Function. *PLoS One* **10**, e0128897 (2015).
30. Palagummi, S., Harbison, S. & Fleming, R. A time-course analysis of mRNA expression during injury healing in human dermal injuries. *Int. J. Legal Med.* **128**, 403–14 (2014).
31. Lee, C. G. *et al.* Vascular endothelial growth factor (VEGF) induces remodeling and enhances TH2-mediated sensitization and inflammation in the lung. *Nat. Med.* **10**, 1095–103 (2004).
32. Cleynen, I. *et al.* Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet (London, England)* (2015). doi:10.1016/S0140-6736(15)00465-1
33. Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. *lancet. Diabetes Endocrinol.* **3**, 243–53 (2015).
34. Baier, M., Werner, A., Bannert, N., Metzner, K. & Kurth, R. HIV suppression by interleukin-16. *Nature* **378**, 563 (1995).
35. Johnson, E. O. *et al.* Novel genetic locus implicated for HIV-1 acquisition with putative regulatory links to HIV replication and infectivity: a genome-wide association study. *PLoS One* **10**, e0118149 (2015).
36. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–11 (2001).

37. Alachkar, N. *et al.* Stem cell factor, interleukin-16, and interleukin-2 receptor alpha are predictive biomarkers for delayed and slow graft function. *Transplant. Proc.* **42**, 3399–405 (2010).
38. Bowler, R. P. *et al.* Integrative omics approach identifies interleukin-16 as a biomarker of emphysema. *OMICS* **17**, 619–26 (2013).
39. Nischwitz, S. *et al.* Interferon β -1a reduces increased interleukin-16 levels in multiple sclerosis patients. *Acta Neurol. Scand.* **130**, 46–52 (2014).
40. Blomhoff, A. *et al.* Linkage disequilibrium and haplotype blocks in the MHC vary in an HLA haplotype specific manner assessed mainly by DRB1*03 and DRB1*04 haplotypes. *Genes Immun.* **7**, 130–40 (2006).
41. Schnabel, R. B. *et al.* Duffy antigen receptor for chemokines (Darc) polymorphism regulates circulating concentrations of monocyte chemoattractant protein-1 and other inflammatory mediators. *Blood* **115**, 5289–99 (2010).
42. Raitakari, O. T. *et al.* Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.* **37**, 1220–6 (2008).
43. Inouye, M. *et al.* An immune response network associated with blood lipid levels. *PLoS Genet.* **6**, e1001113 (2010).
44. Ritchie, S. C. *et al.* The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection. *Cell Syst.* **1**, 293–301 (2015).
45. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–13 (2007).
46. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
47. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–1 (2010).
48. Park, J.-H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42**, 570–5 (2010).
49. Turpeinen, H. *et al.* A genome-wide expression quantitative trait loci analysis of proprotein convertase subtilisin/kexin enzymes identifies a novel regulatory gene variant for FURIN expression and blood pressure. *Hum. Genet.* **134**, 627–36 (2015).

569

570 50. Band, G. & Marchini, J. QCTOOL. *qctool* at
571 <<http://www.well.ox.ac.uk/~gav/qctool/#overview>>
572

Table 1 All loci associated with the concentration of at least one of the studied cytokines

Cytokine	Rs-id	Gene	Locus	Position	A1	A2	MAF	Info	Beta	SE	P-value	HetPVal
MCP1	rs12075	<i>DARC</i>	1q23.2	159175354	G	A	0.47	1.00	0.219	0.015	1.44×10^{-44}	8.76×10^{-78}
SCGFb	rs4656185	<i>F5</i>	1q24.2	169476326	G	A	0.29	1.00	0.205	0.026	1.16×10^{-15}	7.37×10^{-6}
IL18	rs385076	<i>NLRC4</i>	2p22.3	32489851	T	C	0.34	0.97	0.243	0.025	1.66×10^{-22}	0.07
PDGFbb	rs13412535	<i>SERPINE2</i>	2q36.1	224874874	G	A	0.19	0.87	0.335	0.021	2.46×10^{-55}	1.86×10^{-10}
Eotaxin	rs 2228467	<i>CCBP2</i>	3p22.1	42906116	T	C	0.07	1.00	0.416	0.029	2.27×10^{-46}	0.02
MIP1b	rs113010081	<i>CCRL2</i>	3p21.31	46457412	T	C	0.13	0.89	0.595	0.024	3.85×10^{-140}	0.33
IL17	rs1530455	<i>PDIA5</i>	3q21.1	122854899	T	C	0.36	0.95	-0.108	0.017	4.71×10^{-10}	0.81
TRAIL	esv2656942	<i>TNFSF10</i>	3q26.31	172274209	A	-	0.04	0.98	-0.431	0.041	1.74×10^{-25}	0.74
HGF	rs3748034	<i>HGFAC</i>	4p16.3	3446091	G	T	0.12	0.97	0.150	0.023	1.79×10^{-10}	0.50
GROa	rs508977	<i>CXCL1</i>	4q13.3	74762383	T	G	0.24	1.00	0.380	0.028	7.56×10^{-42}	0.03
IP10	rs141053179	<i>CXCL10</i>	4q21.1	77589911	C	G	0.01	0.99	1.103	0.140	2.81×10^{-15}	0.97
IL18	rs17229943	<i>OCN</i>	5q13.2	68682536	A	C	0.08	0.84	0.312	0.046	1.62×10^{-11}	0.98
VEGF	rs6921438	<i>VEGFA</i>	6p21.1	43925607	G	A	0.47	0.98	-0.490	0.018	2.09×10^{-171}	2.19×10^{-47}
HGF	rs5745687	<i>HGF</i>	7q21.11	81359051	C	T	0.04	1.00	-0.307	0.041	2.71×10^{-14}	0.29
VEGF	rs7030781	<i>VLDLR</i>	9p24.2	2686273	A	T	0.42	0.98	-0.137	0.017	2.57×10^{-15}	0.01
CTACK	rs2070074	<i>IL11RA</i>	9p13.3	34649442	A	G	0.11	1.00	-0.447	0.037	1.79×10^{-32}	5.55×10^{-5}
IL2ra	rs12722497	<i>IL2RA</i>	10p15.1	6095928	C	A	0.06	0.99	0.628	0.049	1.57×10^{-38}	0.18
VEGF	rs10761731	<i>JMJD1C</i>	10q21.3	65027610	A	T	0.36	0.99	0.119	0.017	1.01×10^{-11}	0.54
IL18	rs71478720	<i>IL18</i>	11q23.1	112009605	C	T	0.24	0.99	-0.267	0.028	3.71×10^{-22}	0.59
SCGFb	rs187503377	<i>STAB2</i>	12q23.3	104261835	C	T	0.02	0.93	0.965	0.097	1.34×10^{-23}	0.07
IL16	rs4778636	<i>IL16</i>	15q25.1	81591639	G	A	0.04	1.00	-0.727	0.063	1.11×10^{-30}	0.22
PDGFbb	rs4965869	<i>PCSK6</i>	15q26.3	101990320	C	T	0.26	1.00	0.184	0.018	5.66×10^{-24}	7.2×10^{-3}
MIP1b	rs113877493	<i>CCL4L1</i>	17q12	34812273	C	T	0.17	0.92	-0.612	0.022	1.62×10^{-173}	3.58×10^{-7}
TRAIL	rs192989810	<i>MEP1B</i>	18q12.1	29783353	C	T	0.01	0.96	2.221	0.078	8.40×10^{-175}	5.92×10^{-3}
SCGFb	rs116924815	<i>CLEC11A</i>	19q13.33	51230733	C	T	0.03	0.97	0.608	0.074	1.74×10^{-16}	0.03
MIF	rs2330634	<i>MIF</i>	22q11.23	24250795	C	G	0.37	0.99	-0.156	0.025	4.53×10^{-10}	0.11
CTACK	rs201003839	<i>PPARA</i>	22q13.31	46534944	A	ACGGGC	0.27	0.99	0.190	0.027	2.42×10^{-12}	1.36×10^{-3}

The same locus is occasionally associated with more than one cytokine but only the strongest trait-association is reported in the table. Allele 2 is the effect allele. Beta and SE are in SD-units. All positions refer to human genome build 37. Info column reports the imputation quality metric calculated by the Impute2 software. HetPVal reports the effect size heterogeneity statistics between the cohorts.

Table 2 Results from eQTL analyses for locus specific lead SNPs

SNP locus	SNP	Probe	Probe locus	Gene	Other allele	Effect allele	Beta	SE	P-value	HetPval
1q23.2	rs12075	ILMN_1723684	1q23.2	<i>DARC</i>	G	A	0.111	0.017	6.95×10^{-11}	1.00
2q36.1	rs13412535	ILMN_1655595	2q36.1	<i>SERPINE2</i>	G	A	-0.309	0.018	9.65×10^{-63}	1.00
3p21.31	rs113010081	ILMN_1763322	3p21.31	<i>CCR3</i>	T	C	0.372	0.031	4.80×10^{-33}	0.87
4q13.3	rs508977	ILMN_1745522	4q13.3	<i>PF4VI</i>	T	G	0.257	0.031	7.54×10^{-17}	0.66
5q13.2	rs17229943	ILMN_1679620	5q13.2	<i>LOC728519</i>	A	C	0.342	0.036	6.72×10^{-22}	0.35
5q13.2	rs17229943	ILMN_1760189	5q13.2	<i>NAIP</i>	A	C	0.256	0.031	3.05×10^{-16}	0.36
5q13.2	rs17229943	ILMN_2260082	5q13.2	<i>NAIP</i>	A	C	0.228	0.032	8.09×10^{-13}	0.79
5q13.2	rs17229943	ILMN_1767377	5q13.2	<i>LOC153561</i>	A	C	0.226	0.037	6.88×10^{-10}	0.25
9p13.3	rs2070074	ILMN_1664912	9p13.3	<i>IL11RA*</i>	A	G	0.192	0.018	2.07×10^{-26}	0.36
9p13.3	rs2070074	ILMN_1720024	9p13.3	<i>IL11RA*</i>	A	G	0.081	0.013	8.30×10^{-10}	1.00
9p13.3	rs2070074	ILMN_1657475	9p13.3	<i>GALT</i>	A	G	-0.155	0.016	2.70×10^{-21}	0.90
15q25.1	rs4778636	ILMN_2290628	15q25.1	<i>IL16*</i>	G	A	0.636	0.023	2.70×10^{-161}	1.00
17q12	rs113877493	ILMN_2100209	17q12	<i>CCL4L1</i>	C	T	-0.389	0.020	2.14×10^{-87}	1.00
17q12	rs113877493	ILMN_2105573	17q12	<i>CCL3L3</i>	C	T	-0.391	0.039	4.51×10^{-23}	0.47
17q12	rs113877493	ILMN_2218856	17q12	<i>CCL3L1</i>	C	T	-0.180	0.021	1.72×10^{-18}	0.70
17q12	rs113877493	ILMN_1716276	17q12	<i>CCL4L2</i>	C	T	-0.164	0.021	1.27×10^{-14}	0.15
19q13.33	rs116924815	ILMN_1807359	19q13.33	<i>CLEC11A*</i>	C	T	0.450	0.023	1.99×10^{-85}	1.00
22q11.23	rs2330634	ILMN_1690982	22q11.23	<i>DDT</i>	C	G	0.196	0.006	4.75×10^{-211}	0.77

Probe IDs are from Illumina HumanHT-12 v.4 Expression BeadChip. * = The SNP is associated with cytokine concentration and with mRNA coding the cytokine. CTACK is also known as IL11RA and stem cell growth factor beta (SCGFb) as CLEC11A.

Table 3 SNPs previously associated with disease and here found to be associated with cytokine concentrations.

Locus	Rs-id	Disease	Cytokine	Effect allele / Disease		EAF	Beta	StdErr	P-value
				Other allele	risk increasing allele				
3p21.31	rs7616215	Behcet's disease	MIP1b	T	C	0.40	-0.142	0.016	4.8×10^{-19}
3p21.31	rs13096142	Celiac disease	MIP1b	C	T	0.29	-0.163	0.017	4.8×10^{-21}
6p21.1	rs943072	Ulcerative colitis	IL12p70	T	G	0.11	0.155	0.025	3.3×10^{-10}
6p21.1	rs943072	Ulcerative colitis	VEGF	T	G	0.11	0.176	0.027	2.0×10^{-11}
10p15.1	rs12722489	Crohn's disease & Multiple sclerosis	IL2ra	T	C	0.87	0.325	0.034	4.1×10^{-21}
10p15.1	rs12722515	Inflammatory bowel disease	IL2ra	A	C	0.87	0.328	0.034	1.6×10^{-21}

Each SNP reported in the table have been previously associated with corresponding disease or trait. P-values denote the SNP-cytokine association. EAF = Effect allele frequency.

Figure 1 Combined Manhattan plot of meta-analysis results from the GWAS of 41 circulating cytokines. Loci not previously associated with cytokine concentration in GWAS are bolded. The horizontal dashed line indicates the genome-wide significance threshold ($P < 1.2 \times 10^{-9}$) accounting for the number of cytokines tested in the study.

Figure 2 SNPs genome-widely associated with more than one cytokine. Variant rs12075-A is associated with increased concentrations of eotaxin, monocyte chemotactic protein-1 (MCP1) and growth regulated oncogene- α (GRO α). In addition, rs12075-A is associated with increased *DARC* mRNA. Since DARC is a cell surface receptor, the cluster effect on the three cytokines is unlikely to be mediated (in a cascade) through any of the three cytokines associated with rs12075 (**Panel a**). Similarly, effect of rs2228467 is mediated through a cell surface receptor (*CCBP2*) (**Panel b**). Variant rs6921438 in 6p21.1 is located near *VEGFA*. This suggests that VEGF is causally regulating the concentrations of IL7, IL10, IL12p70, and IL13 (**Panel c**). Instrumental variable analyses indicate that the inter-correlations between the cytokines match the causal effect estimates; this suggests a causal role of VEGF in mediating a cascade effect on IL7, IL10, IL12p70, and IL13 (**Panel d**).

Figure 3 Boxplots of interleukin-16 (IL16) cytokine and *IL16* mRNA concentrations. The mRNA and circulating cytokine concentration are depicted by the dark grey and white boxes, respectively. Variant rs4778636 is located within *IL16* gene. The asterisk indicates undetectable concentration of circulating IL16. The A-allele of rs4778636 leads to non-sense mediated decay of *IL16* mRNA and thus circulating concentration of IL16 is undetectable in rs4778636-AA homozygotes.

Supplementary Figure 1 Pearson correlations between the circulating concentrations of different cytokines. The correlations are based on inverse transformed values and are adjusted for age, sex, ten first genetic principal components, and body mass index.

Figure 1

Figure 2

a rs12075 within *DARC* in 1q23.2

b rs2228467 within *CCBP2* in 3p22.1

c rs6921438 near *VEGFA* in 6p21.1

d Instrumental variable analysis for VEGF cascade

Figure 3

Concentration of IL16 cytokine and *IL16* transcript by rs4778636 allele