ARTICLE

Multivariate Genome-wide Association Analysis of a Cytokine Network Reveals Variants with Widespread Immune, Haematological, and Cardiometabolic Pleiotropy

Artika P. Nath,[1,2,3,*] Scott C. Ritchie,[1,2] Nastasiya F. Grinberg,[4] Howard Ho-Fung Tang,[1] Qin Qin Huang,[1,5] Shu Mei Teo,[1,2] Ari V. Ahola-Olli,[6,7] Peter Würtz,[8,9] Aki S. Havulinna,[7,10] Kristiina Santalahti,[11] Niina Pitkänen,[12] Terho Lehtimäki,[13,14] Mika Kähönen,[14,15] Leo-Pekka Lyytikäinen,[13,14] Emma Raitoharju,[13,14] Ilkka Seppälä,[13,14] Antti-Pekka Sarin,[7,10] Samuli Ripatti,[7,16,17] Aarno Palotie,[7,17,18,19,20] Markus Perola,[7,10] Jorma S. Viikari,[21,22] Sirpa Jalkanen,[11] Mikael Maksimow,[11] Marko Salmi,[23] Chris Wallace,[4,24] Olli T. Raitakari,[12,25] Veikko Salomaa,[11] Gad Abraham,[1,2,5] Johannes Kettunen,[11,26,27,28] and Michael Inouye[1,2,5,29,*]

Cytokines are essential regulatory components of the immune system, and their aberrant levels have been linked to many disease states. Despite increasing evidence that cytokines operate in concert, many of the physiological interactions between cytokines, and the shared genetic architecture that underlies them, remain unknown. Here, we aimed to identify and characterize genetic variants with pleiotropic effects on cytokines. Using three population-based cohorts (n = 9,263), we performed multivariate genome-wide association studies (GWAS) for a correlation network of 11 circulating cytokines, then combined our results in meta-analysis. We identified a total of eight loci significantly associated with the cytokine network, of which two (*PDGFRB* and *ABO*) had not been detected previously. In addition, conditional analyses revealed a further four secondary signals at three known cytokine loci. Integration, through the use of Bayesian colocalization analysis, of publicly available GWAS summary statistics with the cytokine network associations revealed shared causal variants between the eight cytokine loci and other traits; in particular, cytokine network variants at the *ABO, SERPINE2,* and *ZFPM2* loci showed pleiotropic effects on the production of immune-related proteins, on metabolic traits such as lipoprotein and lipid levels, on blood-cell-related traits such as platelet count, and on disease traits such as coronary artery disease and type 2 diabetes.

Introduction

Cytokines are signaling molecules secreted by cells, and they are central to multiple physiological functions, especially immune regulation.[1] Broadly speaking, cytokines include chemokines, which drive movement of cells, and growth factors, which drive cell growth and proliferation. Changes in circulating cytokine levels have been associated with infection,[2] autoimmune diseases,[3] and malignancies,[4] as well as atherosclerosis and cardiovascular disease.[5,6] The expression of cytokines can be strongly regulated by genetic variation,[7] and several studies have identified *cis*-acting genetic variants associated with circulating levels of certain cytokines and their receptors under various conditions.[8–10] These initial studies laid the foundation for genetic investigation of circulating cytokine levels at a scale and breadth that may improve our understanding of individual differences in immune response, inflammation, infection, and common disease susceptibility.

Despite cytokines operating in concert to facilitate immune regulation, genome-wide association studies (GWAS) have typically focused on individual cytokines.[11–18] The most extensive cytokine GWAS to date

[1]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Victoria 3004, Australia; [2]Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, United Kingdom; [3]Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria 3010, Australia; [4]Cambridge Institute of Therapeutic Immunology and Infectious Disease, Department of Medicine, University of Cambridge, Cambridge CB2 0AW, United Kingdom; [5]Department of Clinical Pathology, University of Melbourne, Parkville, Victoria 3010, Australia; [6]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; [7]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki 00014, Finland; [8]Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki 00014, Finland; [9]Nightingale Health Ltd., Helsinki 00300, Finland; [10]National Institute of Health and Welfare, Helsinki 00271, Finland; [11]Medicity Research Laboratory, Department of Medical Microbiology and Immunology, University of Turku, Turku 20520, Finland; [12]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland; [13]Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland; [14]Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland; [15]Department of Clinical Physiology, Tampere University Hospital, Tampere 33521, Finland; [16]Department of Public Health, University of Helsinki, Helsinki 00014, Finland; [17]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; [18]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA; [19]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; [20]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; [21]Department of Medicine, University of Turku, Turku 20520, Finland; [22]Division of Medicine, Turku University Hospital, Turku 20520, Finland; [23]Medicity Research Laboratory and Institute of Biomedicine, University of Turku, Turku 20520, Finland; [24]MRC Biostatistics Unit, Institute of Public Health, Cambridge CB2 0SR, United Kingdom; [25]The Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520, Finland; [26]Computational Medicine, Centre for Life Course Health Research, University of Oulu, Oulu 90014, Finland; [27]NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio 70211, Finland; [28]Biocenter Oulu, University of Oulu, Oulu 90014, Finland; [29]The Alan Turing Institute, London, United Kingdom

*Correspondence: artika.nath@baker.edu.au (A.P.N.), mi336@medschl.cam.ac.uk (M.I.)

https://doi.org/10.1016/j.ajhg.2019.10.001.

separately analyzed individual levels of 41 circulating cytokines in approximately 8,000 individuals, identifying 27 distinct loci each associated with at least one cytokine.[19] Others have identified loci influencing cytokine production in response to pathogens.[20,21] While these previous GWAS utilized a univariate framework, analyzing each cytokine separately, studies of related traits indicate that a multivariate framework can confer greater statistical power, for example by taking advantage of the tightly co-regulated nature of both pro- and anti-inflammatory cytokines.

Several methods for multivariate GWAS of correlated phenotypes have been developed.[22–27] Simulations have shown that multivariate analysis can result in increased power to detect genetic associations with small or pleiotropic effects across phenotypes.[22,28–30] These have largely been conducted on metabolic traits where they have demonstrated a boost in statistical power. For example, multivariate analysis of four lipid traits led to a 21% increase in independent genome-wide significant variants compared to univariate analysis.[23] Similar findings were shown for other metabolic traits.[24,31] Moreover, complex genotype-phenotype dependencies have been revealed when variants were jointly tested with lipoprotein traits.[32] Notably, a multivariate GWAS of networks of highly correlated serum metabolites was able to detect nearly twice the number of loci compared to univariate testing, with downstream tissue-specific transcriptional analyses showing that the top candidate genes from multivariate analysis were upregulated in atherosclerotic plaques.[31]

In this study, we focused on correlated immune traits by leveraging the correlation structure within a network of 11 cytokines to perform a multivariate genome-wide scan in 9,263 individuals from three population-based cohorts. We then investigated the colocalization of cytokine-associated variants with those regulating gene expression in numerous tissues and cell types, circulating protein and metabolite levels, hematological traits, and disease states. Finally, we highlighted and characterized variants as potential master regulators of the cytokine network, with pleiotropic effects on production of inflammatory proteins, immune cell function, lipoprotein and lipid levels, and cardiometabolic diseases.

Material and Methods

Study Populations

Approval for the study protocols for each cohort was obtained from their respective ethics committees, and all subjects enrolled in the study gave written informed consent.

The Cardiovascular Risk in Young Finns Study (YFS) is a longitudinal prospective cohort study that commenced in 1980, with follow-up studies carried out every three years. The purpose of this study was to monitor the risk factors of cardiovascular disease in children and adolescents from different regions of Finland. In the baseline study, which was conducted in five Finnish metropolitan areas (Turku, Helsinki, Kuopio, Tampere, and Oulu), a total of 3,596 children and adolescents were randomly selected from the national public register, the details of which were described by Raitakari, et al.[33] Follow-up studies have been carried out every three years, in 1983, 1986, 1989, 2001, 2007, and 2011. For this current study, we utilized data from 2,204 participants (aged 30–45 years) who responded to the 2007 follow-up study (YFS07). Of these, 2,018 individuals had matched cytokine and genotype data available. Ethics were approved by the Joint Commission on Ethics of the Turku University and the Turku University Central Hospital.

The FINRISK cohorts were part of a cross-sectional population-based survey; such studies have been carried out every five years since 1972 in order to evaluate the risk factors of chronic diseases in the Finnish population.[34] Each survey has recruited a representative random sample of 6,000–8,800 individuals, within the age group of 25–74 years, chosen from the national population information system. This study utilized samples from the 1997 (FINRISK97) and 2002 (FINRISK02) collections, which recruited individuals from five or six (for FINRISK02) major regional and metropolitan areas of Finland; the provinces of North Karelia, Northern Savo, Northern Ostrobothnia, Kainuu, and Lapland; the Turku and Loimaa region of southwestern Finland; and the Helsinki and Vantaa metropolitan area. In total, 8,444 (aged 24–74 years) and 8,798 (aged 51–74 years) individuals participated in the FINRISK97 and FINRISK02 studies, respectively. Importantly, each FINRISK survey is an independent cohort, each comprising a different set of participants. Ethics were approved by the coordinating ethical committee of the Helsinki and Uusimaa hospital district, Finland. For FINRSK97, cytokines profiles were measured for all participants where high-quality blood samples were still available. For FINRISK02, cytokine profiling was restricted to older participants (>50 years) due to budget constraints. Cytokine measurements and matched genotype data were available for a subset of 5,728 FINRISK97 participants and 2,027 FINRISK02 participants.

Blood Sample Collection

Blood samples and detailed information on various physical and clinical variables for the YFS and FINRISK cohorts were collected using similar protocols to those described previously.[33,34] Venous blood was collected following an overnight fast for the YFS cohort, while non-fasting blood was collected for FINRISK. Samples were centrifuged, and the resulting plasma and serum samples were aliquoted into separate tubes and stored at −70°C for later analyses.

Genotype Processing and Quality Control

Genotyping in YFS and FINRISK cohorts was performed on whole blood genomic DNA. For YFS07 (n = 2,442), a custom 670K Illumina BeadChip array was used for genotyping. For FINRISK97 (n = 5,798), the Human670-QuadCustom Illumina BeadChip platform was used for genotyping. For FINRISK02 (n = 5,988), the Human670-QuadCustom Illumina BeadChip (n = 2,447) and the Illumina Human CoreExome BeadChip (n = 3,541) were used for genotyping. The Illuminus clustering algorithm was used for genotype calling,[35] and quality control (QC) was performed using the Sanger genotyping QC pipeline. This included removal of SNPs and samples with >5% genotype missingness followed by removal of samples with gender discrepancies. Genotypes were then imputed with IMPUTE2[36] through the use of the 1000 Genomes Phase 1 version 3 as the reference panel followed by removal of SNPs with call rate < 95%, imputation "info" score < 0.4, minor allele frequency < 1%, and Hardy-Weinberg

equilibrium p value $< 5 \times 10^{-6}$. In instances where data were generated using different genotyping platforms, overlapping SNPs were merged using PLINK version 1.90 software.[37] A total of 6,664,959, 7,370,592, and 6,639,681 genotyped and imputed SNPs passed QC in YFS, FINRISK97, and FINRISK02, respectively. Cryptic relatedness was assessed using identity by descent (IBD) estimates, and in cases where the pi-hat relatedness was greater than 0.1, one of the two individuals was randomly removed (n = 44 for YFS, n = 291 for FINRISK97, and n = 39 for FINRISK02). Genetic principal components (PCs) were obtained through principal component analysis (PCA) using FlashPCA[38] on ~60,000 linkage disequilibrium (LD)-pruned SNPs. LD-based pruning was performed to remove SNPs that exceeded an r^2 threshold of 0.05 through the use of PLINK's –indep-pairwise command (SNP window = 100, SNPs shifted each time = 10, r^2 threshold = 0.05).

Measurement of Cytokines

Concentrations of cytokines, chemokines, and growth factors (hereafter referred to as cytokines) were measured in serum (YFS07), EDTA plasma (FINRISK97), and heparin plasma (FINRISK02) using multiplex fluorescent bead-based immunoassays (Bio-Rad). A total of 48 cytokines were measured in YFS07 (n = 2,200) and FINRSK02 (n = 2,775) using two complementary array systems: the Bio-Plex ProTM Human Cytokine 27-plex assay and Bio-Plex ProTM Human Cytokine 21-plex assay. For FINRISK97, 19 cytokines were assayed on the Human Cytokine 21-plex assay system. All assays were performed in accordance with the manufacturer's instructions, except that beads, detection antibodies, and streptavidin-phycoerythrin conjugate were used at half their recommended concentrations. Fluorescence intensity values determined using the Bio-Rad's Bio-Plex 200 array reader were converted to concentrations from the standard curve generated by the Bio-Plex™ Manager 6.0 software. For each cytokine, a standard curve was derived by fitting a five-parameter logistic regression model to the curve obtained from standards provided by the manufacturer. Cytokines with concentrations at the lower and upper asymptotes of the sigmoidal standard curve were set to the concentration corresponding to the fluorescent intensity 2% above or below the respective asymptotes.

Cytokine Data Filtering, Normalization, and Clustering

The analysis was limited to 18 cytokines (Table S1) assayed in all three cohorts. Although Interleukin 1 receptor, type I (IL-1Ra) was assayed in all three cohorts, it was excluded from the analyses due to its inconsistent Pearson correlation pattern with the other 18 cytokines across the three datasets.

Before normalization, cytokine data were subset to individuals with matched genotype data in YFS07 (n = 2,018), FINRISK97 (n = 5,728), and FINRISK02 (n = 2,775). We excluded individuals in YFS07 who reported febrile infection in the two weeks prior to blood sampling (n = 92). To identify extreme outlier samples, PCA was performed on the log2 transformed cytokine values through the use of the missMDA R package.[39] This method first imputed the missing cytokine values via a regularized iterative PCA algorithm implemented in the imputePCA function, then performed PCA. Three and two outlier samples were removed from FINRISK97 and FINRISK02, respectively. Based on IBD analysis described above, 44 (YFS07), 291 (FINRISK97), and 39 (FINRISK02) individuals were also removed. After filtering, a total of 1,843, 5,434, and 1,986 individuals passed QC in YFS07, FINRISK97, and FINRISK02, respectively, and these were used for downstream analysis.

Since all 18 cytokines displayed non-Gaussian distributions, we performed normalization of cytokine levels. For YFS07, the lower limit of detection (LOD) was available for each cytokine. Reported values that were below the LOD were indistinguishable from background noise signals or instrument error[40] and were excluded and treated as missing. For FINRISK97 and FINRISK02, the detection limits were not available; however, it was observed that these two datasets exhibited a bimodal distribution, with the leftmost peak below the expected LOD when compared to the YFS dataset. Individuals in the leftmost peak were therefore set to missing. The log2-transformed cytokine values were then normalized to follow standard Gaussian distributions (with mean of 0 and SD of 1) using rank-based inverse normal transformation (rntransform) as implemented in the GenABEL R package.[41] For each study group, residuals for all cytokines were calculated by regressing the normalized cytokine values on age, sex, BMI, lipid and blood pressure medication, pregnancy status (FINRISK97), and the first 10 genetic PCs through the use of a multiple linear regression model. Of note, information on pregnancy status was only available for the FINRISK97 cohort (n = 52; ~2% of the women). The FINRISK02 is an older cohort (aged 51-74 years), so we do not expect any pregnant women in this cohort. Density distribution plots were generated to confirm that the resulting cytokine residuals were still normally distributed (data not shown).

Detection of groups of correlated cytokines was done in FINRISK97, the cohort with the largest sample size. Pairwise Pearson correlation was performed among residuals of 18 cytokines. These cytokines were then subjected to hierarchical clustering, with one minus the absolute correlation coefficient used as the dissimilarity metric. We then defined a cytokine network—a group of 11 cytokines that were moderately to highly correlated ($r > 0.57$)—for subsequent use in the multivariate analysis.

Statistical Analysis

Univariate association analysis was carried out with linear regression in PLINK,[37] where the residuals of each cytokine were regressed on each SNP genotype. Summary statistics at each marker across three datasets were then combined in a meta-analysis using the METAL software program,[42] which implemented a weighted Z-score method. Since 11 hypothesis tests were performed for each SNP, genome-wide significance was formally set at p value $< 4.55 \times 10^{-9}$, i.e., dividing the standard genome-wide significance threshold (p value $< 5 \times 10^{-8}$) by 11.

Multivariate testing (MV) was performed under the canonical correlation framework implemented in PLINK (MV-PLINK),[22] which extracted the linear combination of traits most highly correlated with genotypes at a particular SNP. The test is based on Wilks' Lambda ($\lambda = 1-\rho^2$), where ρ is the canonical correlation coefficient between the SNP and the cytokine network. Corresponding p values were computed by transforming Wilks' Lambda into a statistic that approximates an *F* distribution, and the loadings for each cytokine represented their individual contributions toward the multivariate association result.[22] Since the multivariate beta-coefficients and standard errors were not calculated by MV-PLINK, the cohort-level multivariate p values were combined in a meta-analysis using the weighted Z-score method[43,44] implemented in the metap R package. In brief, the p values for each dataset were transformed into unsigned Z-scores and weighted by their respective sample sizes, and the sum of each of these

weighted Z-scores was then divided by the square root of the sum of squares of the sample size for each study. The combined weighted Z-scores obtained were then back-transformed into p values. Complete summary statistics from meta-analyses will be made available through the NHGRI-EBI GWAS Catalog.

To assess the inflation of the test statistics as a result of population structure, quantile-quantile (Q-Q) plots of observed-versus-expected $\log_{10}$ p values were generated from the multivariate analyses of the three datasets, both individually and meta-analyzed. Corresponding genomic inflation factor (λ) was calculated by taking the ratio of the median observed distribution of p values to the expected median.

To investigate the existence of additional independent signals within the significant multivariate loci, a conditional stepwise multivariate meta-analysis was performed within each locus. For each study cohort, the lead SNP at each locus (p value $< 5 \times 10^{-8}$), together with other covariates, was fitted in a linear regression model for each cytokine in the network. The resulting residuals were provided as an input for the multivariate test of the locus being assessed. The cohort-level conditional p values were then combined in a meta-analysis. The stepwise conditional analysis was repeated in the univariate model with the lead multivariate SNPs until no additional significant signal was identified.

Colocalization Analysis

Bayesian colocalization tests between cytokine-network-associated signals and the following trait- and disease-associated signals were performed using the COLOC R package.[45] For whole blood *cis* expression quantitative trait loci (eQTLs), we downloaded publicly available summary data from the eQTLGen Consortium portal. The eQTLGen Consortium analysis is the largest meta-analysis of blood eQTLs to date and comprises of 31,684 blood and peripheral blood mononuclear cell (PBMC) samples from a total of 37 datasets.[46] For immune cell *cis*-eQTLs, we either generated *cis*-eQTL summary data in resting B cells,[47] resting monocytes,[48] and stimulated monocytes with interferon-γ or lipopolysaccharide,[48] or obtained publicly available *cis*-eQTL summary data generated by the BLUEPRINT consortium in neutrophils and $CD4^+$ T cells.[57] For *cis*-eQTL mapping in B cells and monocytes (resting and stimulated), information on accessing the raw gene expression and genotype data, data pre-processing, and *cis*-eQTL analysis has been described in a previous study.[50] For protein QTLs (pQTLs), we used publicly available SomaLogic plasma protein GWAS summary statistics from the INTERVAL study.[17] A colocalization test was performed for loci where the cytokine-network-associated variants (within 200 kb from the lead SNP) were also influencing protein levels, either in *cis* (*cis*-pQTLs) or *trans* (*trans*-pQTLs), at pQTL p value $< 1 \times 10^{-6}$. For disease or complex trait associations, we compiled summary statistics of 185 diseases and quantitative traits from GWAS studies conducted in European ancestry individuals, which were accessed from the UK Biobank (Table S2), or downloaded from either ImmunoBase, the NHGRI-EBI GWAS Catalog, or LD Hub. Here, we only considered immune-related and cardiometabolic diseases. For each cytokine network locus, we only tested traits or diseases with the minimum association p value $< 1 \times 10^{-6}$ at this locus. COLOC requires either beta-coefficients and its variance, or p values, for each SNP, in addition to MAF and sample size. Since PLINK multivariate did not produce beta values and standard errors, we instead used meta-analyzed p values for the multivariate cytokine GWAS summary data. For each association pair assessed for colocalization, SNPs within 200 kb of the lead multivariate cytokine GWAS SNP were considered. COLOC (coloc.abf) was run with default parameters and priors. COLOC computed posterior probabilities for the following five hypotheses: PP0, no association with trait 1 (cytokine GWAS signal) or trait 2 (e.g., eQTL signal); PP1, association with trait 1 only (i.e., no association with trait 2); PP2, association with trait 2 only (i.e., no association with trait 1); PP3, association with trait 1 and trait 2 by two independent signals; and PP4, association with trait 1 and trait 2 by shared variants. In practice, evidence of colocalization was defined by PP3 + PP4 $\geq$ 0.99 and PP4/PP3 $\geq$ 5, a cut off previously suggested.[50] To further explore the possibility of colocalization with secondary multivariate cytokine-network- associated signals, we conducted colocalization analyses with conditional p values obtained from the stepwise conditional multivariate GWAS meta-analysis. A sensitivity analysis was further performed using the "sensitivity" function of the COLOC package. The sensitivity analysis takes the COLOC output and assesses whether the posterior inference is robust to the priors used in the colocalization analysis at a predefined rule, which was set to PP3 + PP4 $\geq$ 0.99 and PP4/PP3 $\geq$ 5 (threshold used as evidence for colocalization). The predefined decision rule determines the values of the posterior probabilities considered acceptable for the given priors. The sensitivity analysis demonstrated that the posterior for each colocalizing pair was robust to the priors chosen, in particular to the choice of p12 (1×10^{-5}); hence, all colocalization analysis was performed using default priors. The output from the sensitivity analysis was indicated as "pass" in all the colocalization results reported.

Multi-trait colocalization analysis was performed with the MOLOC tool[51] using default prior probabilities. This analysis was performed to assess whether the pairwise cytokine-to-disease and cytokine-to-molecular trait colocalizations, at the *ABO* locus, involved the same shared casual variant in a three-way colocalization analysis (e.g., CAD-to-cytokine network-to-protein). Only SNPs that were within 200 kb of the lead multivariate cytokine GWAS SNP and were common among all three datasets were assessed. We considered a posterior probability of associations (PPA) threshold of $\geq$80% as strong evidence that the disease, cytokine network, and complex trait (e.g., eQTL, proteins, metabolites, or blood cell traits) colocalized and shared a causal variant.

Results

Summary of Cohorts and Data

Our final dataset comprised a total of 9,267 individuals enrolled in three population-based studies, YFS07 (n = 1,843), FINRISK97 (n = 5,438), and FINRISK02 (n = 1,986), all of whom had available genome-wide genotype data and quantitative measurements of 18 cytokines (Table S1). Characteristics of the study cohorts are summarized in Table 1. Genotypes for the three datasets were imputed with IMPUTE2[36] using the 1000 Genomes Phase 1 version 3 of the reference panel. After QC, a total of 6,022,229 imputed and genotyped SNPs were available across all cohorts. Cytokine levels were measured in serum and plasma through the use of Bio-Plex ProTM Human Cytokine 27-plex and 21-plex assays, then subsequently normalized and adjusted for covariates, including age, sex, BMI, pregnancy status, blood-pressure-lowering medication,

Table 1. Summary of Descriptive Characteristics of the Three Study Cohorts

Characteristics	FINRISK97	FINRISK02	YFS07
Collection year	1997	2002	2007
Number of individuals with matched cytokine and genotype data	5,438	1,986	1,843
Number of males (%)	2,637 (48.5)	991(49.9)	841 (45.6)
Mean age in years (and range)	47.6 (24–74)	60.3(51–74)	37.7 (30–45)
BMI (kg/m^2); mean ± SD.	26.6 ± 4.6	28.1 ± 4.5	25.9 ± 4.6
Number of individuals on lipid lowering drugs (%)	174 (3.2)	284 (14.3)	40 (2.2)
Number of individuals on blood pressure treatment drugs (%)	698 (12.8)	512 (25.8)	127 (6.9)

Abbreviations: BMI, body mass index; YFS, Young Finns Study
The numbers beside the cohort names refer to the calendar year (collection year) in which the samples and clinical information were obtained from each cohort.

lipid-lowering medication, and population structure (see Material and Methods). An overview of the study is shown in Figure 1.

A Correlation Network of Circulating Cytokines

To characterize the correlation structure of circulating cytokines, we utilized the largest dataset available (FINRISK97) and the set of 18 cytokines overlapping all three cohorts. IL-18 was very weakly correlated with other cytokines (Figure 2A), while TRAIL, SCF, HGF, MCP-1, EOTAXIN, and MIP-1b showed moderate correlation with the others. A distinct set of 11 cytokines showed high correlation among themselves (median r = 0.75). In the smaller cohorts (YFS07 and FINRISK02), the cytokine correlation structure was similar but weaker (Figure S1), and the set of 11 cytokines also showed relatively high correlation (YFS07 median r = 0.42; FINRISK02 median r = 0.46). We used this set of 11 cytokines (denoted below as the cytokine network) for multivariate association analysis.

The cytokine network included both anti-inflammatory (IL-10, IL-4, IL-6) and pro-inflammatory (IL-12, IFN-γ, IL-17) cytokines as well as growth factors (FGF-basic, PDGF-BB, VEGF-A, G-CSF) and a chemokine (SDF-1a) involved in promoting leukocyte extravasation and wound healing.[52–54] These cytokines were all positively correlated, which is likely indicative of counter-regulatory (negative-feedback) mechanisms among pro-inflammatory and anti-inflammatory pathways, such as those of IFN-γ and IL-10.[55]

Multivariate Genome-Wide Association Analysis for Cytokine Loci

We performed a multivariate GWAS on the cytokine network in each cohort separately, then cohort-level results were combined using meta-analysis (see Material and Methods). Since one hypothesis test (corresponding to the cytokine network) was performed for each SNP, a genome-wide significance threshold of $p < 5 \times 10^{-8}$ was used. Minimal inflation was observed for the cohort-level and meta-analysis test statistics with lambda (λ) inflation ranging from 1.00–1.02 (Figure S2A–S2D). To directly compare the statistical power of multivariate to univariate GWAS, we first performed univariate analysis in each dataset by regressing each of the cytokines in the cytokine network individually on each SNP, and we then combined the results in a meta-analysis. To account for the 11 cytokines tested, the genome-wide significance threshold was set at $p < 4.55 \times 10^{-9}$. For comparison, we selected the smallest univariate meta-analysis p value for any cytokine at a given locus.

We identified eight loci reaching genome-wide significance for the cytokine network (Figure 2B; Table 2). The strongest association was rs7767396 (meta-p value = 6.93×10^{-306}), a SNP located 172 kb downstream of vascular endothelial growth factor A (*VEGFA* [MIM: 192240]) (Figure S3A). The *VEGFA* locus was previously identified in GWAS for individual cytokine levels, including VEGF-A, IL-7, IL-10, IL-12, and IL-13.[14,19] Consistent with these earlier results, we found that VEGF-A, IL-10, and IL-12 were the top three cytokines based on their trait loadings (relative contribution of each cytokine to the multivariate association result) in each cohort and also significantly associated with this locus in the univariate scans (Figure S4A). Multivariate analysis also confirmed four other previously known associations,[14,16,19] including loci harboring *SERPINE2* (MIM: 177010) (rs6722871; meta-p value = 1.19×10^{-59}), *ZFPM2* (MIM: 603693) (rs6993770; meta-p value = 4.73×10^{-8}), *VLDLR* (MIM: 192977) (rs7030781; meta-p value = 3.78×10^{-13}), and *PCSK6* (MIM: 167405) (rs11639051; meta-p value = 1.93×10^{-58}) (Figure 2B; Table 2; Figure S3B–S3E). The cytokine with the highest loading at each of these loci was consistent with those previously identified in univariate analysis (Figure S4B–S4E).

The multivariate GWAS also detected novel cytokine associations not identified in any previous univariate tests of these cytokines. These were three loci with genic lead SNPs in the candidate genes *F5* (MIM: 612309), *PDGFRB* (MIM: 173410), and *ABO* (MIM: 110300). The lead variant at the *F5* locus (rs9332599; meta-p value = 7.17×10^{-12}) is located in intron 12 of *F5* (Figure S3F). At the platelet-derived growth factor receptor-beta (*PDGFRB*) locus, the lead variant rs2304058 (meta-p value = 4.06×10^{-9}) is

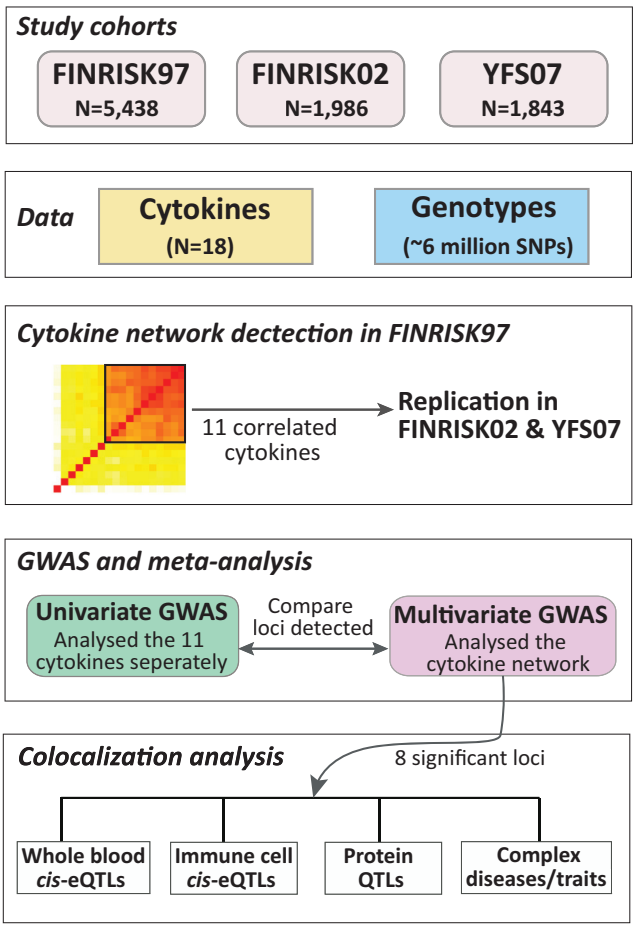

Figure 1. Overview of the Study Populations, Design, and the Analyses Conducted

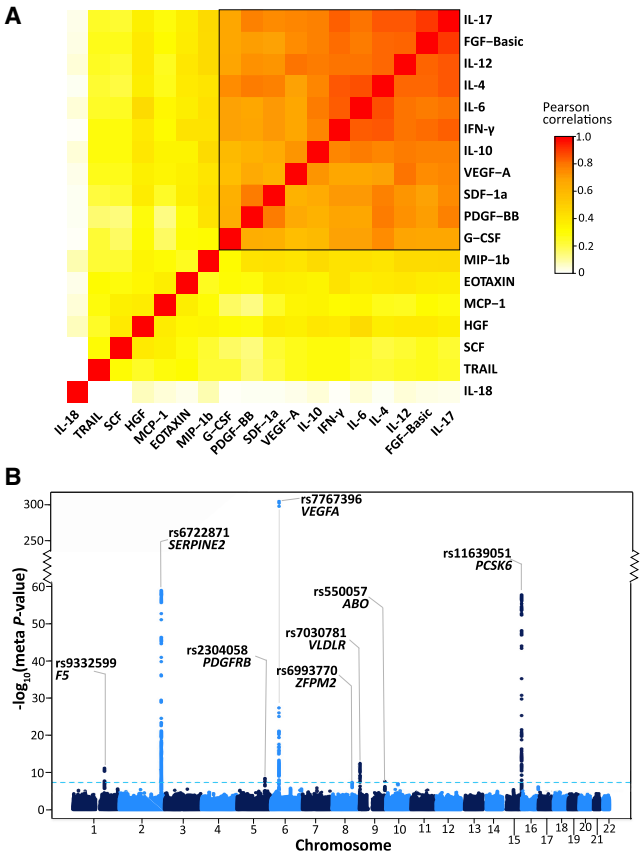

Figure 2. Multivariate GWA Analysis of a Network of 11 Correlated Cytokines in Three Finnish Cohorts

(A) Correlation heatmap of the 18 cytokines in the FINRISK97 cohort. Each cell presents the pairwise Pearson's correlation coefficient between the normalized cytokine residuals. The cytokines are ordered by hierarchical clustering, using 1 minus the absolute value of the correlations as the distance matrix. The color scale denotes the strength of the correlations, where red is a high positive correlation. The group of 11 tightly correlated cytokines (black box) was used for multivariate analysis.

(B) Manhattan plot for meta-analysis results from the multivariate GWAS of the cytokine network. The statistical strength of association (-$\log_{10}$ meta-p value; y axis) is plotted against all the SNPs ordered by chromosomal position (x axis). The sky-blue horizontal dashed line represents the genome-wide (meta-p value $< 5 \times 10^{-8}$) significance threshold. The lead SNP (lowest meta-p value) at each locus and the nearby genes are shown.

within intron 10 of *PDGFRB* (Figure S3G). At the *ABO* locus, the lead variant rs550057 (meta-p value = 2.75×10^{-8}) is within the first intron of *ABO* (Figure S3H); furthermore, rs550057 is located ~1.6 kb upstream of the erythroid cell specific enhancer, which contains a GATA-1 transcription factor binding site and has been shown to enhance the transcription of the *ABO* gene.[56]

To investigate the presence of multiple independently associated variants at each of the eight loci, we performed stepwise conditional multivariate meta-analysis. Three loci (*SERPINE2*, *VEGFA*, and *PCSK6*) exhibited evidence of multiple independent signals (Table S3). In addition to the lead variants (rs6722871, rs7767396, and rs11639051) at each of these three loci, we identified additional association signals (rs55864163, *SERPINE2*, meta-$p_{cond} = 9.03 \times 10^{-29}$; rs112215592, *SERPINE2*, meta-$p_{cond} = 2.10 \times 10^{-12}$; rs4714729, *VEGFA*, meta-$p_{cond} = 7.49 \times 10^{-10}$; and rs6598475, *PCSK6*, meta-$p_{cond} = 2.63 \times 10^{-17}$), which were independently associated with the cytokine network. We also performed conditional univariate analysis that adjusted for the lead multivariate SNPs, which were either the same lead univariate SNPs or in high LD ($r^2 = 0.99$). This univariate analysis also uncovered the same secondary signal at the *VEGFA* locus in association with VEGF-A cytokine levels (rs4714729; meta-$p_{cond} = 8.8 \times 10^{-13}$) (Table S3).

Colocalization of Cytokine Variants with *cis*-eQTLs in Whole Blood

To characterize the regulatory effects of the multivariate cytokine-associated loci, we queried the largest publicly available set of results for whole blood *cis*-eQTLs, a meta-analysis of 31,684 individuals, which was obtained from the eQTLGen Consortium database.[46] We found SNPs, lead or LD-proxy (r^2>0.5), at seven of the eight cytokine loci (*ABO*, *F5*, *PCSK6*, *PDGFRB*, *SERPINE2*, *VEGFA*, and *VLDLR*) with *cis*-regulatory effects (p value $< 1 \times 10^{-6}$) on gene expression (for a total of 17 unique genes) in blood

Table 2. Meta-Analyzed Results of Multivariate GWAS of Cytokine Network

Locus	Locus Region	Top SNP	Average MAF	Top Multivariate Meta-p value	Univariate Meta-p value (Top Cytokine)	Detection
F5	1q24.2	rs9332599	0.294	7.17×10^{-12}	9.21×10^{-3} (SDF1-a)	multivariate
SERPINE2	2q36.1	rs6722871	0.311	1.19×10^{-59}	3.55×10^{-18} (PDGF-BB)	both
PDGFRB	5q32	rs2304058	0.379	4.06×10^{-9}	1.52×10^{-5} (IL4)	multivariate
VEGFA	6p21.1	rs7767396	0.471	6.93×10^{-306}	3.10×10^{-201} (VEGF-A)	both
ZFPM2	8q23.1	rs6993770	0.221	4.73×10^{-8}	1.01×10^{-7} (IL12p70)	multivariate
ABO	9q34.2	rs550057	0.306	2.75×10^{-8}	4.9×10^{-3} (IL4)	multivariate
VLDLR	9p24.2	rs7030781	0.413	3.78×10^{-13}	6.78×10^{-14} (VEGF-A)	both
PCSK6	15q26.3	rs11639051	0.255	1.93×10^{-58}	1.19×10^{-26} (PDGF-BB)	both
JMJD1C	10q21.3	rs9787438	0.374	[a]1.30×10^{-7}	[a]8.96×10^{-12} (VEGF-A)	univariate

The table shows the meta-analysis p values for the top SNP (lowest p value) at each locus associated with the cytokine network in the multivariate analysis at genome-wide significance threshold ($p < 5 \times 10^{-8}$). The corresponding lowest meta-p value for the same top SNP in the univariate analysis with any single cytokine present in the cytokine network, given in brackets beside the meta-p value, was also reported.
[a]Instance where the top SNP at a locus crossed only the univariate significance threshold ($p < 4.55 \times 10^{-9}$), then the corresponding meta-p value for that SNP in the multivariate was also given. The univariate significance threshold was calculated from a Bonferroni correction for 11 cytokines tested ($p < 5 \times 10^{-8}/11$).

(Table S4). Using Bayesian colocalization analysis, we further demonstrated that associations at three of these loci colocalized with *cis*-eQTLs for *ABO*, *PCSK6,* and *SERPINE2* expression (Figure 3A–3C; Table S5). We did not observe colocalization with the secondary multivariate GWAS signals at the *PCSK6* and *SERPINE2* loci.

Colocalization of Cytokine Variants with Immune Cell-Specific *cis*-eQTLs

Next, we investigated the cell-type- or context-dependent regulatory effects of genetic variants associated with the cytokine network by interrogating previously published *cis*-eQTLs specific to resting B cells,[47] resting monocytes,[48] stimulated monocytes with interferon-γ or lipopolysaccharide,[48] resting neutrophils,[57] naive $CD4^+$ T cells[49,57] and $CD8^+$ T cells,[49] all isolated from healthy donors of European ancestry (Table S6). Three out of the eight cytokine network loci harbored *cis*-eQTLs (p value $< 1 \times 10^{-6}$) in at least one immune cell type, in either a stimulated or a non-stimulated state (Table S7). For example, SNPs at the *SERPINE2* locus were reported to have *cis*-eQTL effects across multiple immune cell types, including B cells, $CD4^+$, and $CD8^+$ T cells (Table S7).

Further, colocalization analysis showed that the cytokine network variants at *SERPINE2* had strong evidence of sharing a causal variant with *SERPINE2* *cis*-eQTLs in $CD4^+$ T cells and B cells, similar to the colocalization we observe in whole blood (Figure 3B; Table S8). Evidence of colocalization was not observed with the secondary multivariate GWAS signals at this locus.

Colocalization of Cytokine Variants with Plasma Protein QTLs

To investigate protein-level effects of cytokine network variants, we utilized plasma protein QTLs (pQTLs) from the INTERVAL study.[17] Colocalization analysis, considering only proteins with both *cis*- and *trans*-pQTLs, at the cytokine network loci, with association p value $< 1 \times 10^{-6}$, showed that all the eight cytokine network loci had strong evidence of shared causal variants with plasma levels of a total of 146 proteins (out of the 215 tested) (Table S9). Of these, the *ABO* and *ZFPM* cytokine network loci strongly colocalized with *trans*-pQTL signals for 55 (out of 81) and 87 (out of 98) proteins, respectively (Table 3; Table S9). Of these, 14 and 75 proteins shared the same causal lead pQTLs with the lead cytokine network variants at the *ABO* (rs550057) and *ZFPM2* (rs6993770) loci, respectively, suggesting these variants have extensive pleiotropic effects on multiple cytokines and proteins—which potentially have shared underlying pathophysiology and/or biology.

The *ABO* locus colocalized with *trans*-pQTLs for several membrane proteins (B3GN2, endoglin, GOLM1, OX2G, and TPST2) and cell surface receptors (IL-3RA, LIFR, IGF-I R, and HGF receptors). *ABO* colocalization was also observed with *trans*-pQTLs for adhesion and immune-related molecules involved in leukocyte recruitment, cell adhesion, and transmigration, including sGP130, sICAM-1, sICAM-2, LIRB4, and P-selectin (Table 3; Table S9). At the *ZFPM2* locus, colocalization was seen with *trans*-pQTLs for proteins generally found in platelet granules (e.g., VEGF-A, PDGF-AA, PDGF-BB, PDGF-D, angiopoietin, and P-selectin). At the *SERPINE2* locus, we observed that, in addition to colocalizing with the *cis*-eQTL signal for *SERPINE2* expression, the cytokine-network-associated variants colocalized with the *cis*-pQTL variants for SERPINE2 protein levels (Table S9). Likewise, the *VEGFA* locus colocalized with a *cis*-pQTL for VEGF-A, and the *PDGFRB* locus with a *cis*-pQTL for PDGFRB.

Relationships of Cytokine Network Variants with Complex Traits and Diseases

Using the NHGRI GWAS Catalog,[58,59] we found that, across all eight cytokine network loci, 55 SNPs matched

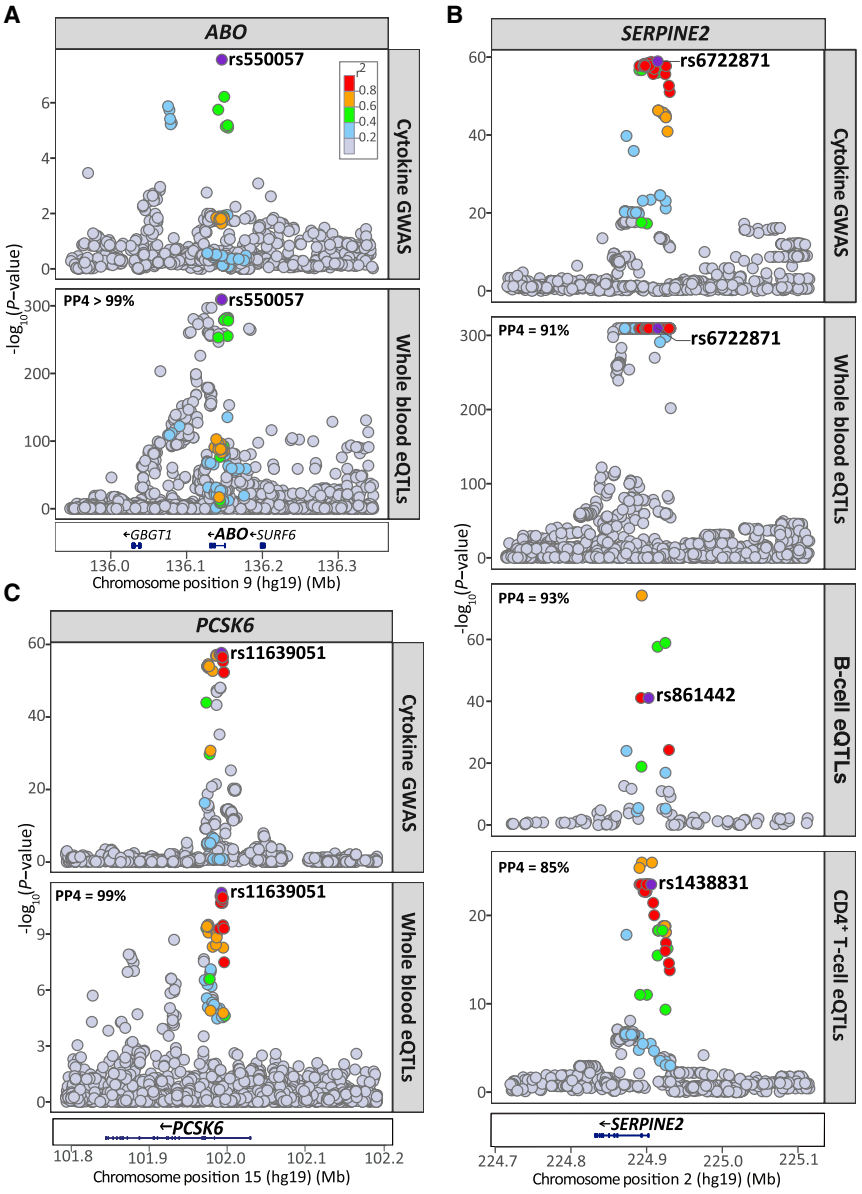

Figure 3. Regional Plots for the Cytokine Network Association and Whole Blood and Immune Cell *cis*-eQTL Association Signals at the *ABO*, *PCSK6*, and *SERPINE2* Locus

(A) The cytokine network GWAS signal (top) colocalizes with the whole blood *cis*-eQTLs signal for *ABO* (bottom) at the *ABO* locus on chromsome 9; (B) colocalizes with whole blood *cis*-eQTLs for *PCSK6* expression (bottom) at the *PCSK6* locus on chromosome 15; (C) colocalizes with the *cis*-eQTL signals for *SERPINE2* expression in whole blood (middle), B cells (middle), and $CD4^+$ T cells (bottom) at the *SERPINE2* locus on chromosome 2. For each plot, the circles represent the $-\log_{10}$ association p values (y axis) of SNPs plotted against their chromosomal position (x axis). The eQTL association plots show the lead cytokine network GWAS SNP tested in the colocalization analysis. The lead cytokine network GWAS SNP rs6722871 was not present in the B cell and $CD4^+$ T cell eQTL dataset, instead, the next top GWAS SNP present in each of the eQTL dataset (rs861442, B cell; rs1438831, $CD4^+$ T cell) is shown. For all regional plots, pairwise LD (r^2) in the region is colored with respect to the lead cytokine network GWAS SNP. LD was calculated from the 1000 Genomes European population.

SNPs previously associated with quantitative traits and diseases (Table S10). The lead cytokine network variant at *ZFPM2* (rs6993770) has previously been associated with various platelet traits, including platelet count, distribution width, plateletcrit (total platelet mass), and mean volume[17,60] (Table S10).

Next, GWAS summary statistics from a broad range of traits and diseases (Table S2), including hematopoietic traits, circulating metabolites, and immune- and cardiometabolic-related diseases, were compiled for colocalization analysis with the cytokine network loci. The two cytokine-network-associated loci, *ABO* and *ZFPM2*, exhibited strong evidence of colocalization for several traits and diseases. The *ZFPM2* locus colocalized not only with signals for several platelet trait associations, but also with other hematological trait-associated signals, including white blood cell counts and specifically neutrophil and basophil counts (Table 3; Table S11). The *ABO* locus showed colocalization with various QTLs for hematological traits, including red blood cell traits (haemoglobin concentration, red blood cell count, and hematocrit) and white blood cell counts, including granulocyte count and specifically eosinophil count (Table 3; Table S11). This is consistent with the *ABO* locus being identified as a pQTL for proteins involved in leukocyte activation as identified previously. Cytokine network variants at the *ABO* locus colocalized with those of intermediate-density, low-density, and very-low-density lipoprotein subclasses as well as glycosylated haemoglobin (HbA1c) (Table 3; Table S11), suggesting both inflammatory and metabolic effects. Notably, the same cytokine network variants at the *ABO* locus also strongly colocalized with signals associated with coronary artery disease (CAD), pulmonary embolism, ischemic stroke (MIM: 601367), and type 2 diabetes (T2D [MIM: 125853]) (Table 3, Table S11).

Multi-Trait Colocalization at the *ABO* Locus

Given its extensive pleiotropy and disease relevance, we performed three-way multi-trait colocalization (MOLOC)[51] at the *ABO* locus to assess shared genetic etiology between disease traits, the cytokine network, and

Table 3. Colocalization of Cytokine-Network-Associated Variants at the *ABO* and *ZFPM2* Loci with Those of Plasma Protein Levels, Quantitative Traits, and Disease Risk

Group/ Functions	Evidence	Names
ABO Locus (Chromosome 9)		
Diseases		
Cardiometabolic diseases	strong	pulmonary embolism, ischemic stroke, coronary artery disease, type 2 disease,
Cardiometabolic diseases	none	deep vein thrombosis
Blood Cell Traits		
Blood cell counts	Strong	white blood cell, granulocytes, basophils + eosinophils, basophils + neutrophils, eosinophils + neutrophils, eosinophils, neutrophils, hematocrit (%), haemoglobin, myeloid, red blood cells, platelet distribution width
Blood cell counts	Suggestive	basophils, reticulocytes
Blood cell counts	None	monocyte, platelet, plateletcrit (%), red cell distribution width
Metabolites		
IDL particle constituents	strong	Total cholesterol (IDL-C), free cholesterol (IDL-FC), total lipids (IDL-L), total particle concentration (IDL-P), phospholipids (IDL-PL), triglycerides (IDL-TG)
LDL subclass particle constituents	strong	For large particles: total cholesterol (L-LDL-C), cholesterol esters (L-LDL-CE), free cholesterol (L-LDL-FC), total lipids (L-LDL-L), total particle concentration (L-LDL-P), phospholipids (L-LDL-PL), For medium particles: total cholesterol (M-LDL-C), cholesterol esters (M-LDL-CE), total lipids (M-LDL-L), total particle concentration (M-LDL-P), phospholipids (M-LDL-PL) For small particles: total cholesterol (S-LDL-C), total lipids (S-LDL-L), total particle concentration (S-LDL-P)
VLDL subclass particle constituents	strong	For small particles: total cholesterol (S-VLDL-C), For extra-small particles: total lipids (XS-VLDL-L), phospholipids (XS-VLDL-PL)
Other	strong	HbA1c, Apolipoprotein B, total LDL cholesterol, total serum cholesterol
Proteins		
Chemokine activity	strong	FAM3B, FAM3D, MIP-5, TECK,
Chemokine activity	suggestive	CCL28
Chemokine receptors	strong	IL-3RA, HGF receptor, sGP130, VEGF-R2, VEGF-R3
Chemokine receptors	none	TCCR
Receptor function and/or signaling	strong	F177A, GP116, IGF-1R, IR, JAG1, MBL, PEAR1, PYY, SECTM1, SEMA6A, TLR4
Receptor function and/or signaling	suggestive	PLXB2
Receptor function and/or signaling	none	CD109, CD209, GFRAL, GPIV, LIF-R, Notch-1, PEAR1, sTIE1, sTIE2
Cell adhesion	strong	Cadherin-1, E-selectin, Endoglin, ICAM-4, ISLR2, Laminin, NCAM-L1, OX2G, P-selectin, sICAM-1, sICAM-2, sICAM-5
Cell adhesion	none	ADAM23, BCAM, Cadherin-5, Desmoglein-2, ESAM
Enzyme function	strong	B3GN2, B4GT1, B4GT2, Cathepsin S, CLIC5, DPEP2, FA20B, FUT10, GLCE, GNS, IAP, LPH, MA1A2, NDST1, QSOX2, ST4S6, TPST2, XXLT1
Enzyme function	none	ATS13, BGAT, CEL, CHSTB, DYR, MINP1, TLL1
Miscellaneous	strong	C1GLC, CASC4, GOLM1, KIN17, THSD1, TUFT1,
Miscellaneous	none	Factor VIII, OBP2B
ZFPM2 Locus (Chromosome 8)		
Blood Cell Traits		
Blood cell counts	strong	white blood cells, granulocytes, basophils + neutrophils, neutrophils + eosinophils, basophils, neutrophils, myeloid, platelets, plateletcrit (%), platelet distribution width, mean platelet volume
Proteins		
Cytokine/chemokine activity	strong	EDA, IL-7, PDGF-AA, PDGF-BB, PDGF-D, VEGF-A, NAP-2, RANTES, TARC
Immune response	strong	CLM2, COCH, CYTF, DB119

(*Continued on next page*)

Table 3. ***Continued***

Group/ Functions	Evidence	Names
Receptor function and/or signaling	strong	ANG-1, APP, BDNF, CD44, CGB2, CRIM1, Dkk-1, Dkk-4, EDAR, EPHB2, EPHB3, GI24, GRP, LIRB4, Mammaglobin-2, OBP2A, P2RX6, PAP1, PTPRD, RGS10, RGS3, RHOG, THA, MESD2
Receptor function and/or signaling	suggestive	Ephrin-A3
Receptor function and/or signaling	none	UNC5H4, sRAGE
Cell adhesion	strong	Galectin-7, KIRR2, MADCAM-1, MFGM, ON, P-Selectin, PCDG8, SCF, SPARCL1
Enzyme activity	strong	Arylsulfatase A, ASM3A, B4GT7, Cathepsin A, CHSTB, CPXM1, FUT8, GSTM1-1, INP5E, MMEL2, MYSM1, PAI-1, PDIA5, RIFK, SIRT5, SPTC1, UD2A1
Enzyme activity	none	PDE3A, ZFP91, LAML2, HECW1
Enzyme inhibitor	strong	SERPINE2, SPINK5, TICN3, WFD13
Transcription/translation	strong	APBB1, CENPW, HIF-1a, PAIP1
Transcription/translation	suggestive	ID2
Miscellaneous	strong	4EBP2, APLP2, ARL1, ASIC4, CA063, Coactosin-like protein, CQ089, DJB11, MPP7, NSG2, PROL1, RBM28, SATB1, SYT11, SYT17, TXNDC4
Miscellaneous	none	CNA2

Evidence: Evidence of colocalization, Strong = PP3+PP4 ≥ 0.99 and PP4/PP3 ≥ 5; Suggestive = PP3 + PP4 > 0.75 and PP4/PP3 > 3; None = association signal for the trait at the locus, but no evidence of colocalization
Refer to Table S9 for full descriptions of the proteins. The proteins have been grouped into broad functional categories using the Uniprot database.[85]

various molecular traits. Consistent with our pairwise COLOC analysis, the majority of these colocalizing proteins, lipids and lipoproteins, and blood cell traits also showed evidence of multi-trait colocalization (PPA ≥ 80%) with CAD-cytokine network, T2D-cytokine network, pulmonary embolism-cytokine network, and ischemic stroke-cytokine network pairs (Tables S12–S15). Overall, this suggested that the *ABO* locus contributes to the shared genetic architecture among several known cardiometabolic disease risk factors, which includes multiple inflammatory, haemostatic, and metabolic processes.

Discussion

In this study, we first identified a network of 11 correlated cytokines which are known to participate in a broad array of immune responses in circulation. These cytokines include those involved in the classical T_{H1} (IL-12, IFN-γ), T_{H2} (IL-4, IL-6, and IL-10), T_{H17} (IL-6, IL-17, and G-CSF), and T_{reg} (IL-10) responses[52,53] as well as the promotion of angiogenesis, tissue repair, and remodelling typically coinciding with inflammatory and post-inflammatory states (VEGF-A, FGF-basic, and PDGF-BB).[54] Although previous *in vitro* challenge studies[20,21] indicate antagonistic relationships among selected cytokines in the network, our analyses in >9,000 individuals are consistent with a previous study which utilized similar data,[19] showing that these 11 circulating cytokines are positively correlated in the general population. Therefore, at the population level, it is more likely that an equilibrium in circulating levels of disparate cytokines exists, possibly maintained by counter-regulatory mechanisms.

Our multivariate GWAS meta-analysis identified eight loci associated with the cytokine network, confirming six previously reported associations for circulating cytokine levels[14,16,19] as well as uncovering two additional signals (*PDGFRB* and *ABO)*, empirically demonstrating that jointly modeling correlated traits in a multivariate GWAS can increase statistical power to detect additional associations compared to the univariate approach. This contributes to the growing body of literature which shows, through both simulation and empirical analyses, that multivariate outperforms the univariate analysis, leading to the identification of novel pleiotropic loci.[22,28–30] On the other hand, we and others have also noted that in certain circumstances, the multivariate approach may suffer from power loss; for example, when the SNP influences nearly all the traits equally or the direction of genetic and cross-trait correlation is the same.[22,23,61]

Further, integrative genetic analyses revealed evidence for shared genetic influences between these loci, molecular QTLs, and complex trait and disease associations. This study identified several regions harboring cytokine-associated signals that colocalize with whole blood and/or immune cell-specific *cis*-eQTLs for a number of genes, including *SERPINE2, ABO*, and *PCSK6*, suggesting that these genes are possible candidates underlying the collective expression of cytokines in the cytokine network—or vice versa. Our findings also highlight the fact that the cytokine network associations at the pleiotropic loci, *ABO* and *ZFPM2,* overlap with signals associated with multiple traits, including cardiometabolic diseases, immune-related proteins, and platelet traits.

SERPINE2 encodes protease nexin-1, an inhibitor of serine proteases such as thrombin and plasmin, and is therefore implicated in coagulation, fibrinolysis, and tissue remodelling.[62] It shares similar functions with its better-known homolog *SERPINE1* (MIM: 173360), or

plasminogen activator inhibitor-1 (PAI-1), the elevation of which is associated with thrombosis and cardiovascular risk.[62] However, there is also evidence that *SERPINE2* has pleiotropic roles in immune and inflammatory regulation, roles that could be either dependent or independent of its function as a serine protease. It is expressed in many tissue types, and its expression can be induced by pro-inflammatory cytokines such as IL-1α.[63,64] Conversely, *SERPINE2* can itself influence inflammatory status: *SERPINE2* is a candidate susceptibility gene for chronic obstructive pulmonary disease, and *SERPINE2*-knockout mice exhibited extensive accumulation of lymphocytes in the lungs, through a mechanism linked to thrombin and NFκB activation.[64] We observed in our data that the cytokine network associations overlapped with the *SERPINE2* pQTL signal. Moreover, using immune cell-specific *cis*-eQTL data, we further demonstrated colocalization between the cytokine network and *SERPINE2* *cis*-eQTL signals specifically in $CD4^+$ T cells and B cells. This suggests that the association between *SERPINE2* and the cytokine network at this locus is at least partially driven by lymphocytic expression—consistent with *SERPINE2* itself influencing chemotaxis and recruitment of lymphocytes.[64] Our analyses demonstrate that the importance of *SERPINE2* in regulating immune and inflammatory processes is potentially greater than previously anticipated, and warrants further targeted research.

Like *SERPINE2*, the *ABO* locus has widespread pleiotropic effects. The most well-known function of *ABO* is its determination of blood group. The human *ABO* gene has three major alleles (A, B, and O) that determine ABO blood type. The A and B alleles encode for distinct "A" versus "B" glycosyltransferases that add specific sugar residues to a precursor molecule (H antigen) to form A versus B antigens, respectively.[65] The O allele results in a protein without glycosyltransferase activity.[65] The lead cytokine-associated variant rs550057 and its proxies in moderate LD (r^2 = 0.6; rs507666, rs687289) have been previously shown to determine the *ABO* allele,[66] but they have also been associated with circulating levels of inflammatory proteins such as sICAM-1, P-selectin, and ALP.[17,67,68] Our study showed that cytokine network associations at the *ABO* locus share colocalized signals with a host of other proteins and traits, including lipoproteins (IDL, LDL, and VLDL), proteins of immune function, immune cell subsets, and cardiometabolic diseases (Table 3); these results highlight the potential for shared molecular etiology among these traits. Our analyses highlight the potential genetic basis for numerous previous observations linking ABO blood group to an array of similar traits and phenotypes.[18,69–74] We also observed multi-trait colocalization among cardiometabolic diseases, cytokine network, and other features relating to multiple inflammatory (e.g., inflammatory proteins, cytokines, and cytokine receptors), haemostatic (blood cell traits), and metabolic processes (lipids and metabolites); this further strengthens the evidence for a shared causal variant. Altogether, these results suggest that certain genetic variants, e.g., at the *ABO* locus, influence the risk of cardiometabolic disease through a constellation of pleiotropic effects.

It could therefore be speculated, due to its involvement in multiple inflammatory, haemostatic, and metabolic processes, that the *ABO* gene influences the risk of cardiometabolic disease; however, our current understanding of the mechanisms behind this remains unclear. For instance, non-O blood groups have been associated with increased risk of cardiovascular disease, venous thromboembolism, stroke, and T2D.[70,75] However, the O blood group has itself been linked to elevated IL-10 and worse outcomes given existing coronary disease (risk of cardiovascular death, of recurrent myocardial infarction, and of all-cause mortality).[66] Other studies have suggested a role for von Willebrand factor (VWF), a coagulative factor which also expresses ABO antigens—in particular, the O phenotype is associated with lower VWF, which may explain reduced thrombotic and cardiovascular risk.[66,76] It has been suggested that the link between ABO blood group type and venous thromboembolism (VTE) is potentially driven by VWF and Factor VIII—non-O blood group individuals presented a higher risk of venous thromboembolism and had elevated levels of both VWF and Factor VIII.[77,78] Also relevant is the link between *ABO* and adhesion molecules such as E-selectin and sICAM-1 which are overexpressed in inflammatory states.[18,68,72,73] sICAM-1 is a known positive correlate with cardiovascular disease; however, it is the A blood group, not O, that is associated with reduced sICAM-1 levels, again complicating the picture.[72] Inferring the exact causal relationships among all these entities will require intricate follow-up experimental investigation, involving simultaneous examination of all key players. It is particularly unclear whether the link with cardiometabolic diseases may be due to its direct modification of H antigen, or of the glycosyltransferase activity of the encoded enzyme on other proteins, or some combination of both. In our study, formal causal inference (e.g., with Mendelian randomization) was not possible because the corresponding multivariate beta-coefficients and standard errors are not currently calculable and the locus itself has extensive pleiotropy.

The *ZFPM2* locus has been associated with platelet traits,[60] and our findings highlight its importance as a determinant of platelet and angiogenic cytokine activity. *ZFPM2* encodes a zinc finger cofactor that regulates the activity of GATA4, a transcription factor reported to play a critical role not only in heart development[79] but also in modulation of angiogenesis. In particular, GATA4 directly binds to the promoter of angiogenic factor *VEGFA* and regulates its expression,[80] and it has been shown that disruption of ZFPM2-GATA4 interaction alters the expression of *VEGFA* and other angiogenesis-related genes.[81] VEGF-A and PDGFR-BB, which are part of the cytokine network, have been found to be released via alpha granules of activated platelets, and serum VEGF-A levels correlate closely with blood platelet counts.[82–84] In our study, we show

that the cytokine-associated signal at the *ZFPM2* locus co-localized with GWAS signals for platelet traits (platelet count, platelet distribution width, and mean platelet volume) and platelet proteins (e.g., VEGF-A, PDGF-AA, PDGF-BB, PDGF-D, angiopoietin, and P-selectin). Our findings provide additional insights into the relationships between the *ZFPM2* locus, cytokines and various platelet-associated proteins, and their role in platelet biology. The lead cytokine network SNP rs6993770 has been reported to be a *trans*-eQTL in whole blood for gene products typically found in platelets and their receptors (e.g., CXCL5, GP9, MYL9, and VWF).[46] Collectively, these findings suggest that this locus regulates the number and/or cytokine activity of circulating platelets, and that this potentially occurs via interaction with *GATA4* (MIM: 600576) and regulation of *VEGFA*.

In conclusion, our study illustrates the utility of multivariate analysis of correlated immune traits and highlights potentially fruitful avenues of biological investigation for multivariate genetic signals. Our results highlight the fact that certain gene loci drive the expression of a cytokine network with immune, inflammatory, and tissue repair functions; and, simultaneously, these loci are implicated in the regulation of other haemostatic and metabolic functions, with relevance to human health and disease. This stresses the fact that the processes of inflammation, haemostasis, and repair often run concurrent with each other after injury, and that biological systems often feature ample redundancy and feedback loops within individual effectors.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.10.001.

Acknowledgments

This work was supported partially by the Victorian Government's OIS Program and by core funding from: the UK Medical Research Council (MR/L003120/1), the British Heart Foundation (RG/13/13/30194, RG/18/13/33946), and the NIHR (Cambridge BRC at the Cambridge University Hospitals NHS Foundation Trust). The YFS was supported by the Academy of Finland: 322098, 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 (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; EU Horizon 2020 (755320—TAXINOMISIS); ERC (742927—MULTIEPIGEN project); and Tampere University Hospital Supporting Foundation. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or DHSC. The authors acknowledge support from: A.P.N., Australian Postgraduate Award; M.I. and S.C.R., NIHR (Cambridge BRC at the Cambridge University Hospitals NHS Foundation Trust); G.A., NHMRC Early Career Fellowship (1090462); Q.Q.H., Melbourne Research Scholarship; H.H.T., NHMRC Postgraduate Scholarship; N.F.G. and C.W., Wellcome Trust (WT107881); C.W., Medical Research Council (MC_UU_00002/4); J.K., Sigrid Juselius Foundation, Academy of Finland (297338 and 307247) and Novo Nordisk Foundation (NNF17OC0026062); P.W., Novo Nordisk Foundation (15998) and Academy of Finland (312476 and 312477); T.L., Academy of Finland (322098); A.S.H., Academy of Finland (321356); and V.S., Finnish Foundation for Cardiovascular Research.

Declaration of Interests

Veikko Salomaa has consulted for Novo Nordisk and Sanofi and received honoraria from these companies. He also has ongoing research collaboration with Bayer Ltd. (All unrelated to the present study). The other authors declare no conflicts of interest.

Received: May 14, 2019
Accepted: September 30, 2019
Published: October 31, 2019

Web Resources

BLUEPRINT immune cell summary statistics, ftp://ftp.ebi.ac.uk/pub/databases/blueprint/blueprint_Epivar/
eQTLGen Consortium portal, http://www.eqtlgen.org/
GWAS Catalog, https://www.ebi.ac.uk/gwas/
ImmunoBase, https://www.immunobase.org/
LD Hub, http://ldsc.broadinstitute.org/ldhub/
OMIM, https://www.omim.org/
PLINK, https://www.cog-genomics.org/plink2
Summary statistics from the multivariate GWAS meta-analyses, https://www.ebi.ac.uk/gwas/downloads/summary-statistics

References

1. Dinarello, C.A. (2007). Historical insights into cytokines. Eur. J. Immunol. *37* (*Suppl 1*), S34–S45.
2. Vignali, D.A., and Kuchroo, V.K. (2012). IL-12 family cytokines: immunological playmakers. Nat. Immunol. *13*, 722–728.
3. O'Shea, J.J., Ma, A., and Lipsky, P. (2002). Cytokines and autoimmunity. Nat. Rev. Immunol. *2*, 37–45.
4. Dranoff, G. (2004). Cytokines in cancer pathogenesis and cancer therapy. Nat. Rev. Cancer *4*, 11–22.
5. Dranoff, H., Taleb, S., Mallat, Z., and Tedgui, A. (2011). Recent advances on the role of cytokines in atherosclerosis. Arterioscler Thromb. Vasc. Biol. *31*, 969–979.
6. Kaptoge, S., Seshasai, S.R., Gao, P., Freitag, D.F., Butterworth, A.S., Borglykke, A., Di Angelantonio, E., Gudnason, V., Rumley, A., Lowe, G.D., et al. (2014). Inflammatory cytokines and risk of coronary heart disease: new prospective study and updated meta-analysis. Eur. Heart J. *35*, 578–589.
7. de Craen, A.J., Posthuma, D., Remarque, E.J., van den Biggelaar, A.H., Westendorp, R.G., and Boomsma, D.I. (2005). Heritability estimates of innate immunity: an extended twin study. Genes Immun. *6*, 167–170.
8. Rafiq, S., Stevens, K., Hurst, A.J., Murray, A., Henley, W., Weedon, M.N., Bandinelli, S., Corsi, A.M., Guralnik, J.M., Ferruci,

L., et al. (2007). Common genetic variation in the gene encoding interleukin-1-receptor antagonist (IL-1RA) is associated with altered circulating IL-1RA levels. Genes Immun. *8*, 344–351.

9. Interleukin 1 Genetics Consortium (2015). Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. Lancet Diabetes Endocrinol. *3*, 243–253.
10. Hollegaard, M.V., and Bidwell, J.L. (2006). Cytokine gene polymorphism in human disease: on-line databases, Supplement 3. Genes Immun. *7*, 269–276.
11. Larsen, M.H., Albrechtsen, A., Thørner, L.W., Werge, T., Hansen, T., Gether, U., Haastrup, E., and Ullum, H. (2013). Genome-wide association study of genetic variants in LPS-stimulated IL-6, IL-8, IL-10, IL-1ra and TNF-α cytokine response in a Danish cohort. PLoS ONE *8*, e66262.
12. Matteini, A.M., Li, J., Lange, E.M., Tanaka, T., Lange, L.A., Tracy, R.P., Wang, Y., Biggs, M.L., Arking, D.E., Fallin, M.D., et al. (2014). Novel gene variants predict serum levels of the cytokines IL-18 and IL-1ra in older adults. Cytokine *65*, 10–16.
13. Tekola Ayele, F., Doumatey, A., Huang, H., Zhou, J., Charles, B., Erdos, M., Adeleye, J., Balogun, W., Fasanmade, O., Johnson, T., et al. (2012). Genome-wide associated loci influencing interleukin (IL)-10, IL-1Ra, and IL-6 levels in African Americans. Immunogenetics *64*, 351–359.
14. Debette, S., Visvikis-Siest, S., Chen, M.H., Ndiaye, N.C., Song, C., Destefano, A., Safa, R., Azimi Nezhad, M., Sawyer, D., Marteau, J.B., et al. (2011). 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.
15. He, M., Cornelis, M.C., Kraft, P., van Dam, R.M., Sun, Q., Laurie, C.C., Mirel, D.B., Chasman, D.I., Ridker, P.M., Hunter, D.J., et al. (2010). Genome-wide association study identifies variants at the IL18-BCO2 locus associated with interleukin-18 levels. Arterioscler. Thromb. Vasc. Biol. *30*, 885–890.
16. Choi, S.H., Ruggiero, D., Sorice, R., Song, C., Nutile, T., Vernon Smith, A., Concas, M.P., Traglia, M., Barbieri, C., Ndiaye, N.C., et al. (2016). Six novel loci associated with circulating VEGF levels identified by a meta-analysis of genome-wide association studies. PLoS Genet. *12*, e1005874.
17. 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 *558*, 73–79.
18. Sliz, E., Kalaoja, M., Ahola-Olli, A., Raitakari, O., Perola, M., Salomaa, V., Lehtimäki, T., Karhu, T., Viinamäki, H., Salmi, M., et al. (2019). Genome-wide association study identifies seven novel loci associating with circulating cytokines and cell adhesion molecules in Finns. J. Med. Genet. *56*, 607–616.
19. 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. *100*, 40–50.
20. Li, Y., Oosting, M., Deelen, P., Ricaño-Ponce, I., Smeekens, S., Jaeger, M., Matzaraki, V., Swertz, M.A., Xavier, R.J., Franke, L., et al. (2016). Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. Nat. Med. *22*, 952–960.
21. Li, Y., Oosting, M., Smeekens, S.P., Jaeger, M., Aguirre-Gamboa, R., Le, K.T.T., Deelen, P., Ricaño-Ponce, I., Schoffelen, T., Jansen, A.F.M., et al. (2016). A functional genomics approach to understand variation in cytokine production in humans. Cell *167*, 1099–1110.e14.
22. Ferreira, M.A., and Purcell, S.M. (2009). A multivariate test of association. Bioinformatics *25*, 132–133.
23. O'Reilly, P.F., Hoggart, C.J., Pomyen, Y., Calboli, F.C.F., Elliott, P., Jarvelin, M.R., and Coin, L.J.M. (2012). MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS ONE *7*, e34861.
24. Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat. Methods *11*, 407–409.
25. Turley, P., Walters, R.K., Maghzian, O., Okbay, A., Lee, J.J., Fontana, M.A., Nguyen-Viet, T.A., Wedow, R., Zacher, M., Furlotte, N.A., et al.; 23andMe Research Team; and Social Science Genetic Association Consortium (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat. Genet. *50*, 229–237.
26. Cichonska, A., Rousu, J., Marttinen, P., Kangas, A.J., Soininen, P., Lehtimäki, T., Raitakari, O.T., Järvelin, M.R., Salomaa, V., Ala-Korpela, M., et al. (2016). metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. Bioinformatics *32*, 1981–1989.
27. Mägi, R., Suleimanov, Y.V., Clarke, G.M., Kaakinen, M., Fischer, K., Prokopenko, I., and Morris, A.P. (2017). SCOPA and META-SCOPA: software for the analysis and aggregation of genome-wide association studies of multiple correlated phenotypes. BMC Bioinformatics *18*, 25.
28. Yang, Q., and Wang, Y. (2012). Methods for analyzing multivariate phenotypes in genetic association studies. J. Probab. Stat. *2012*, 652569.
29. Kim, S., and Xing, E.P. (2009). Statistical estimation of correlated genome associations to a quantitative trait network. PLoS Genet. *5*, e1000587.
30. van der Sluis, S., Posthuma, D., and Dolan, C.V. (2013). TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. PLoS Genet. *9*, e1003235.
31. Inouye, M., Ripatti, S., Kettunen, J., Lyytikäinen, L.P., Oksala, N., Laurila, P.P., Kangas, A.J., Soininen, P., Savolainen, M.J., Viikari, J., et al. (2012). Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet. *8*, e1002907.
32. Marttinen, P., Pirinen, M., Sarin, A.P., Gillberg, J., Kettunen, J., Surakka, I., Kangas, A.J., Soininen, P., O'Reilly, P., Kaakinen, M., et al. (2014). Assessing multivariate gene-metabolome associations with rare variants using Bayesian reduced rank regression. Bioinformatics *30*, 2026–2034.
33. 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. *37*, 1220–1226.
34. Borodulin, K., Vartiainen, E., Peltonen, M., Jousilahti, P., Juolevi, A., Laatikainen, T., Männistö, S., Salomaa, V., Sundvall, J., and Puska, P. (2015). Forty-year trends in cardiovascular risk factors in Finland. Eur. J. Public Health *25*, 539–546.
35. Teo, Y.Y., Inouye, M., Small, K.S., Gwilliam, R., Deloukas, P., Kwiatkowski, D.P., and Clark, T.G. (2007). A genotype calling

algorithm for the Illumina BeadArray platform. Bioinformatics *23*, 2741–2746.
36. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
37. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.
38. Abraham, G., and Inouye, M. (2014). Fast principal component analysis of large-scale genome-wide data. PLoS ONE *9*, e93766.
39. Josse, J., and Husson, F. (2016). missMDA: A package for handling missing values in multivariate data analysis. J. Stat. Softw. *70*, 1–31.
40. Whitcomb, B.W., and Schisterman, E.F. (2008). Assays with lower detection limits: implications for epidemiological investigations. Paediatr. Perinat. Epidemiol. *22*, 597–602.
41. Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijn, C.M. (2007). GenABEL: an R library for genome-wide association analysis. Bioinformatics *23*, 1294–1296.
42. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
43. Whitlock, M.C. (2005). Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J. Evol. Biol. *18*, 1368–1373.
44. Zaykin, D.V. (2011). Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J. Evol. Biol. *24*, 1836–1841.
45. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.
46. Võsa, U., Claringbould, A., Westra, H.J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S., et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. bioRxiv *447367*. https://doi.org/10.1101/447367.
47. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O., and Knight, J.C. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. *44*, 502–510.
48. Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., and Knight, J.C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343*, 1246949.
49. Kasela, S., Kisand, K., Tserel, L., Kaleviste, E., Remm, A., Fischer, K., Esko, T., Westra, H.J., Fairfax, B.P., Makino, S., et al. (2017). Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. PLoS Genet. *13*, e1006643.
50. Guo, H., Fortune, M.D., Burren, O.S., Schofield, E., Todd, J.A., and Wallace, C. (2015). Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. Hum. Mol. Genet. *24*, 3305–3313.
51. Giambartolomei, C., Zhenli Liu, J., Zhang, W., Hauberg, M., Shi, H., Boocock, J., Pickrell, J., Jaffe, A.E., Pasaniuc, B., Roussos, P.; and CommonMind Consortium (2018). A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics *34*, 2538–2545.
52. Zhu, J., and Paul, W.E. (2010). Heterogeneity and plasticity of T helper cells. Cell Res. *20*, 4–12.
53. Dong, C. (2008). TH17 cells in development: an updated view of their molecular identity and genetic programming. Nat. Rev. Immunol. *8*, 337–348.
54. Barrientos, S., Brem, H., Stojadinovic, O., and Tomic-Canic, M. (2014). Clinical application of growth factors and cytokines in wound healing. Wound Repair Regen. *22*, 569–578.
55. Hu, X., and Ivashkiv, L.B. (2009). Cross-regulation of signaling pathways by interferon-gamma: implications for immune responses and autoimmune diseases. Immunity *31*, 539–550.
56. Sano, R., Nakajima, T., Takahashi, K., Kubo, R., Kominato, Y., Tsukada, J., Takeshita, H., Yasuda, T., Ito, K., Maruhashi, T., et al. (2012). Expression of ABO blood-group genes is dependent upon an erythroid cell-specific regulatory element that is deleted in persons with the B(m) phenotype. Blood *119*, 5301–5310.
57. Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell *167*, 1398–1414.e24.
58. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.
59. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. *45* (D1), D896–D901.
60. 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 *167*, 1415–1429.e19.
61. Galesloot, T.E., van Steen, K., Kiemeney, L.A., Janss, L.L., and Vermeulen, S.H. (2014). A comparison of multivariate genome-wide association methods. PLoS ONE *9*, e95923.
62. Bouton, M.C., Boulaftali, Y., Richard, B., Arocas, V., Michel, J.B., and Jandrot-Perrus, M. (2012). Emerging role of serpinE2/protease nexin-1 in hemostasis and vascular biology. Blood *119*, 2452–2457.
63. Santoro, A., Conde, J., Scotece, M., Abella, V., Lois, A., Lopez, V., Pino, J., Gomez, R., Gomez-Reino, J.J., and Gualillo, O. (2015). SERPINE2 inhibits IL-1α-induced MMP-13 expression in human chondrocytes: Involvement of ERK/NF-κB/AP-1 pathways. PLoS ONE *10*, e0135979.
64. Solleti, S.K., Srisuma, S., Bhattacharya, S., Rangel-Moreno, J., Bijli, K.M., Randall, T.D., Rahman, A., and Mariani, T.J. (2016). Serpine2 deficiency results in lung lymphocyte accumulation and bronchus-associated lymphoid tissue formation. FASEB J. *30*, 2615–2626.
65. Yamamoto, F., Cid, E., Yamamoto, M., Saitou, N., Bertranpetit, J., and Blancher, A. (2014). An integrative evolution theory of histo-blood group ABO and related genes. Sci. Rep. *4*, 6601.
66. Johansson, Å., Alfredsson, J., Eriksson, N., Wallentin, L., and Siegbahn, A. (2015). Genome-wide association study identifies that the ABO blood group system influences

interleukin-10 levels and the risk of clinical events in patients with acute coronary syndrome. PLoS ONE *10*, e0142518.
67. Masuda, M., Okuda, K., Ikeda, D.D., Hishigaki, H., and Fujiwara, T. (2015). Interaction of genetic markers associated with serum alkaline phosphatase levels in the Japanese population. Hum. Genome Var. *2*, 15019.
68. Yao, C., Chen, G., Song, C., Keefe, J., Mendelson, M., Huan, T., Sun, B.B., Laser, A., Maranville, J.C., Wu, H., et al. (2018). Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat. Commun. *9*, 3268.
69. Meo, S.A., Rouq, F.A., Suraya, F., and Zaidi, S.Z. (2016). Association of ABO and Rh blood groups with type 2 diabetes mellitus. Eur. Rev. Med. Pharmacol. Sci. *20*, 237–242.
70. Chen, Z., Yang, S.H., Xu, H., and Li, J.J. (2016). ABO blood group system and the coronary artery disease: an updated systematic review and meta-analysis. Sci. Rep. *6*, 23250.
71. Larson, N.B., Bell, E.J., Decker, P.A., Pike, M., Wassel, C.L., Tsai, M.Y., Pankow, J.S., Tang, W., Hanson, N.Q., Alexander, K., et al. (2016). ABO blood group associations with markers of endothelial dysfunction in the Multi-Ethnic Study of Atherosclerosis. Atherosclerosis *251*, 422–429.
72. Paré, G., Chasman, D.I., Kellogg, M., Zee, R.Y., 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. *4*, e1000118.
73. Qi, L., Cornelis, M.C., Kraft, P., Jensen, M., van Dam, R.M., Sun, Q., Girman, C.J., Laurie, C.C., Mirel, D.B., Hunter, D.J., et al. (2010). Genetic variants in ABO blood group region, plasma soluble E-selectin levels and risk of type 2 diabetes. Hum. Mol. Genet. *19*, 1856–1862.
74. 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. *8*, 14357.
75. Fagherazzi, G., Gusto, G., Clavel-Chapelon, F., Balkau, B., and Bonnet, F. (2015). ABO and Rhesus blood groups and risk of type 2 diabetes: evidence from the large E3N cohort study. Diabetologia *58*, 519–522.
76. O'Donnell, J., Boulton, F.E., Manning, R.A., and Laffan, M.A. (2002). Amount of H antigen expressed on circulating von Willebrand factor is modified by ABO blood group genotype and is a major determinant of plasma von Willebrand factor antigen levels. Arterioscler. Thromb. Vasc. Biol. *22*, 335–341.
77. Tirado, I., Mateo, J., Soria, J.M., Oliver, A., Martínez-Sánchez, E., Vallvé, C., Borrell, M., Urrutia, T., and Fontcuberta, J. (2005). The ABO blood group genotype and factor VIII levels as independent risk factors for venous thromboembolism. Thromb. Haemost. *93*, 468–474.
78. Schleef, M., Strobel, E., Dick, A., Frank, J., Schramm, W., and Spannagl, M. (2005). Relationship between ABO and Secretor genotype with plasma levels of factor VIII and von Willebrand factor in thrombosis patients and control individuals. Br. J. Haematol. *128*, 100–107.
79. Svensson, E.C., Tufts, R.L., Polk, C.E., and Leiden, J.M. (1999). Molecular cloning of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. Proc. Natl. Acad. Sci. USA *96*, 956–961.
80. Heineke, J., Auger-Messier, M., Xu, J., Oka, T., Sargent, M.A., York, A., Klevitsky, R., Vaikunth, S., Duncan, S.A., Aronow, B.J., et al. (2007). Cardiomyocyte GATA4 functions as a stress-responsive regulator of angiogenesis in the murine heart. J. Clin. Invest. *117*, 3198–3210.
81. Zhou, B., Ma, Q., Kong, S.W., Hu, Y., Campbell, P.H., McGowan, F.X., Ackerman, K.G., Wu, B., Zhou, B., Tevosian, S.G., and Pu, W.T. (2009). Fog2 is critical for cardiac function and maintenance of coronary vasculature in the adult mouse heart. J. Clin. Invest. *119*, 1462–1476.
82. Pal, E., Korva, M., Resman Rus, K., Kejžar, N., Bogovič, P., Strle, F., and Avšič-Županc, T. (2018). Relationship between circulating vascular endothelial growth factor and its soluble receptor in patients with hemorrhagic fever with renal syndrome. Emerg. Microbes Infect. *7*, 89.
83. Banks, R.E., Forbes, M.A., Kinsey, S.E., Stanley, A., Ingham, E., Walters, C., and Selby, P.J. (1998). Release of the angiogenic cytokine vascular endothelial growth factor (VEGF) from platelets: significance for VEGF measurements and cancer biology. Br. J. Cancer *77*, 956–964.
84. Graff, J., Klinkhardt, U., Schini-Kerth, V.B., Harder, S., Franz, N., Bassus, S., and Kirchmaier, C.M. (2002). Close relationship between the platelet activation marker CD62 and the granular release of platelet-derived growth factor. J. Pharmacol. Exp. Ther. *300*, 952–957.
85. UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. *43*, D204–D212.

The American Journal of Human Genetics, Volume *105*

Supplemental Data

Multivariate Genome-wide Association Analysis of a Cytokine Network Reveals Variants with Widespread Immune, Haematological, and Cardiometabolic Pleiotropy

Artika P. Nath, Scott C. Ritchie, Nastasiya F. Grinberg, Howard Ho-Fung Tang, Qin Qin Huang, Shu Mei Teo, Ari V. Ahola-Olli, Peter Würtz, Aki S. Havulinna, Kristiina Santalahti, Niina Pitkänen, Terho Lehtimäki, Mika Kähönen, Leo-Pekka Lyytikäinen, Emma Raitoharju, Ilkka Seppälä, Antti-Pekka Sarin, Samuli Ripatti, Aarno Palotie, Markus Perola, Jorma S. Viikari, Sirpa Jalkanen, Mikael Maksimow, Marko Salmi, Chris Wallace, Olli T. Raitakari, Veikko Salomaa, Gad Abraham, Johannes Kettunen, and Michael Inouye

Supplementary Text

Figures

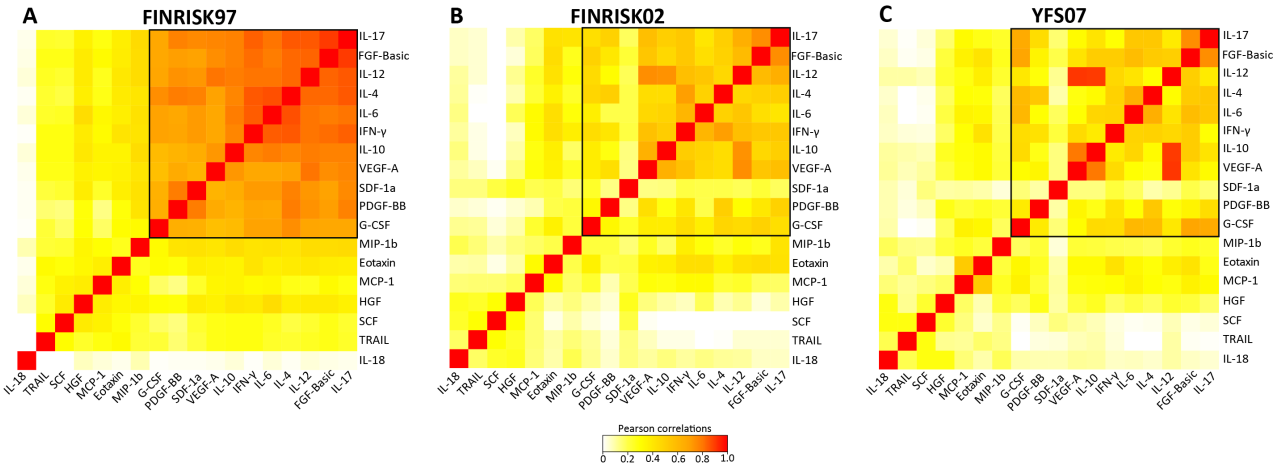

Figure S1: Comparison of cytokine-cytokine correlation in FINRISK07, FINRISK02, and YFS07.

The heatmaps show the correlations between the normalised cytokines residuals in the discovery dataset, **(A)** FINRISK97, and the replication datasets, **(B)** FINRISK02 and **(C)** YFS07. Each square represents the Pearson's correlation coefficient between the cytokines. The black box shows the correlation patterns among the 11 correlated cytokines (discovered using the FINRISK97) across the three datasets. The correlation matrix in FINRISK07 was subjected to hierarchical clustering using distance as 1 minus the absolute value of the correlations. The ordering of rows and columns in FINRISK02 and YFS07 was defined by the ordering in FINRISK07. The strength of the correlations is indicated by the colour on the scale, where red is a high positive correlation.

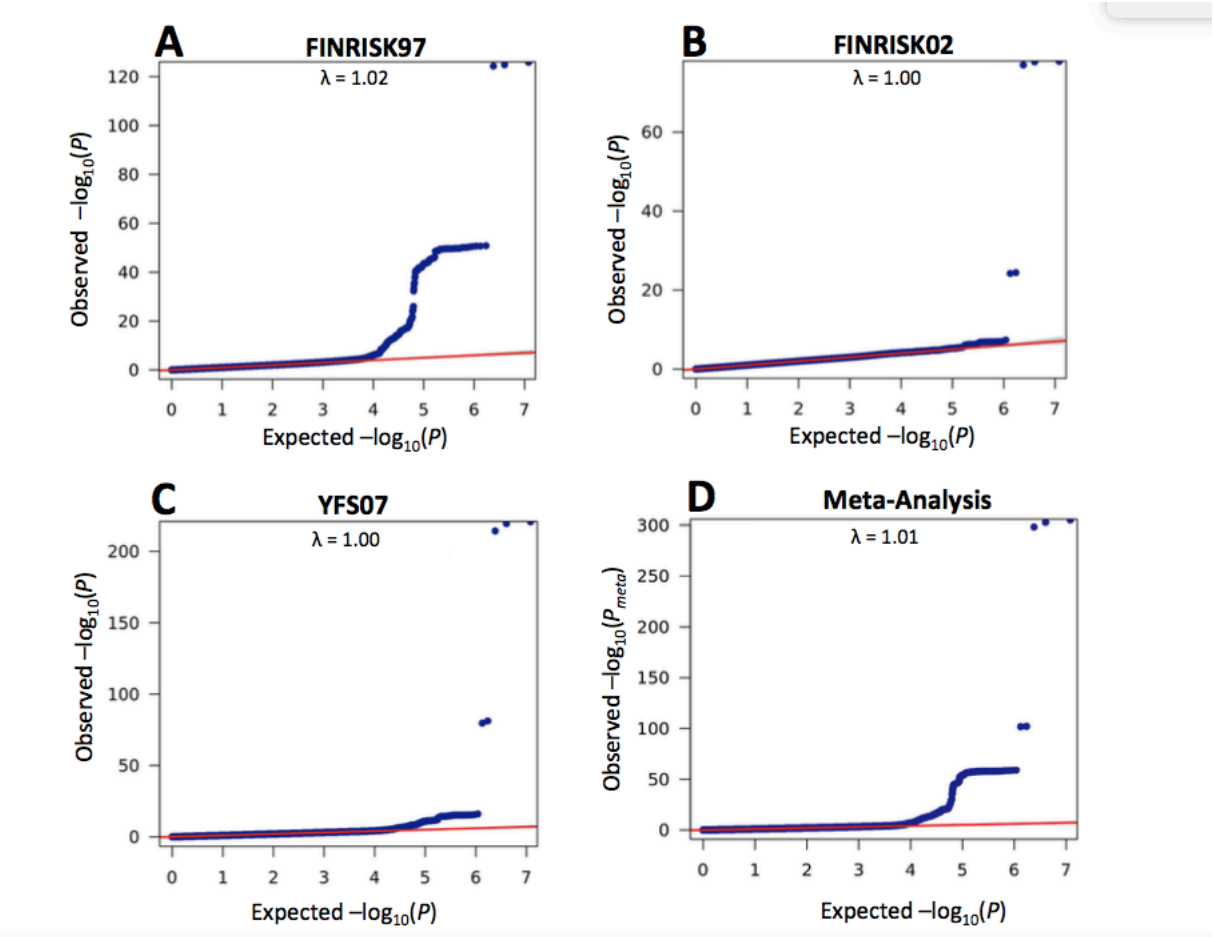

Figure S2: Quantile-quantile (Q-Q) plots resulting from the multivariate GWAS in the three cohorts separately and meta-analysis.

Q-Q plots of observed (*y*-axis) *vs.* expected *P*-values (*x*-axis) for each SNP from the multivariate genome-wide association in **(A)** FINRISK97, **(B)** FINRISK02, **(C)** YFS07, and **(D)** Meta-analysis of the three cohorts. The diagonal red line ($y=x$) indicates null hypothesis of no association. The inflation factor (λ) was between 1.0 – 1.02 suggesting that inflation from population substructure or other confounders was appropriately adjusted for.

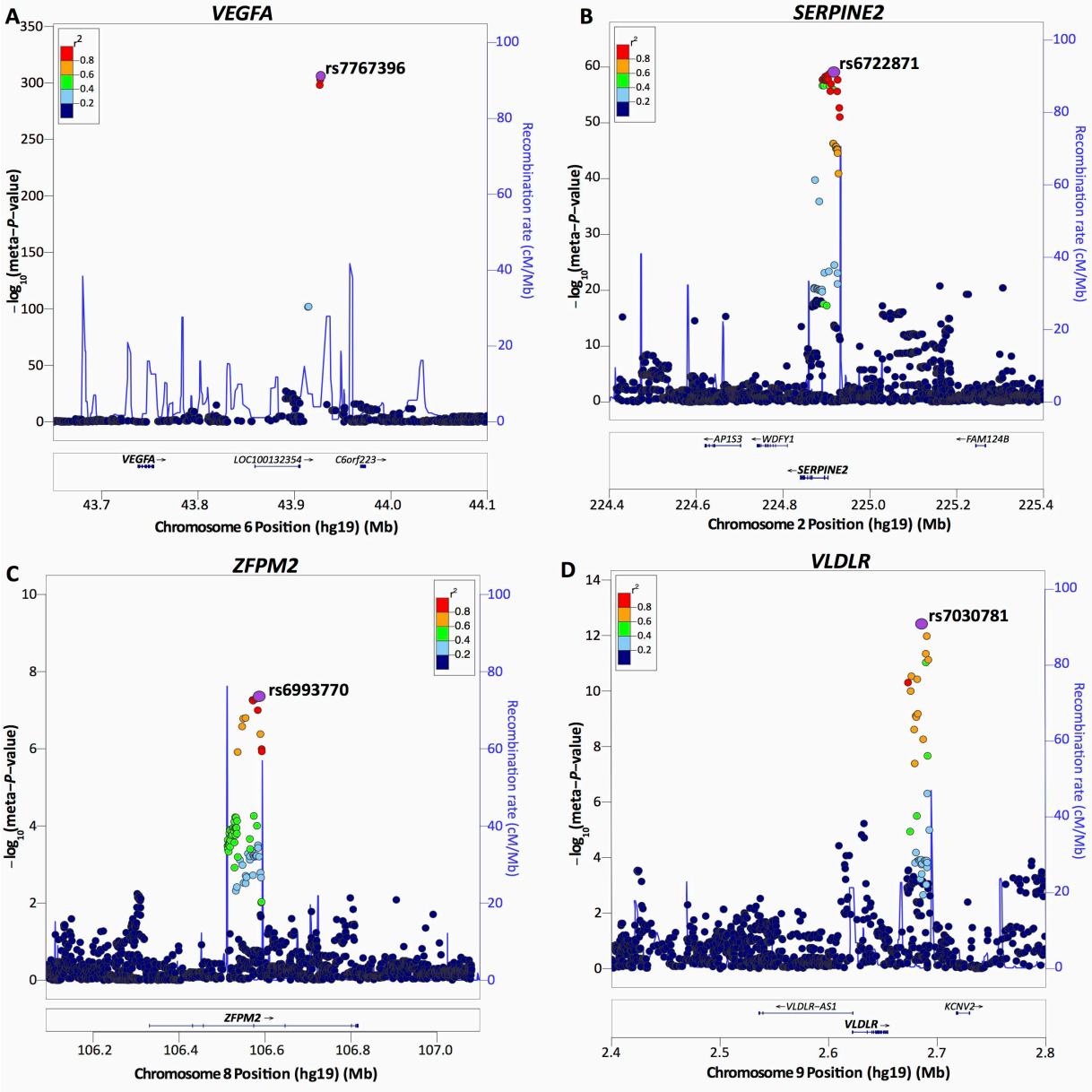

Figure S3: Regional association plots for each of the 8 loci associated with the cytokine network from the meta-analysed multivariate GWA analysis.

(A) *VEGFA* locus, rs7767396 is an intergenic SNP located 172.83kb downstream of vascular endothelial growth factor A (*VEGFA*) gene on chromosome 6p21.1. **(B) *SERPINE2* locus**, rs6722871 lies 10.9kb upstream of *SERPINE2* on chromosome 2q36.1. **(C) *ZFPM2* locus**, rs6993770 lies within intron 4 of the zinc finger protein multitype 2 (*ZFPM2*) gene on chromosome 8q23.1. **(D) *VLDLR* locus**, rs7030781 is situated ~31.8kb away from the very low-density lipoprotein receptor (*VLDLR*) gene on chromosome 9p24.2. For each plot, the circles represent the -$\log_{10}$ meta-analysed *P*-values (y-axis) of SNPs plotted against their chromosomal position (x-axis). The lead SNP in each plot is denoted by a purple circle, and its pairwise LD (r^2) strength with other SNPs in the region, estimated from the "1000 genomes Mar 2012 EUR" population, is indicated by color. The blue lines indicate the recombination rates. The plots were generated using the LocusZoom online tool (http://locuszoom.sph.umich.edu/locuszoom/).

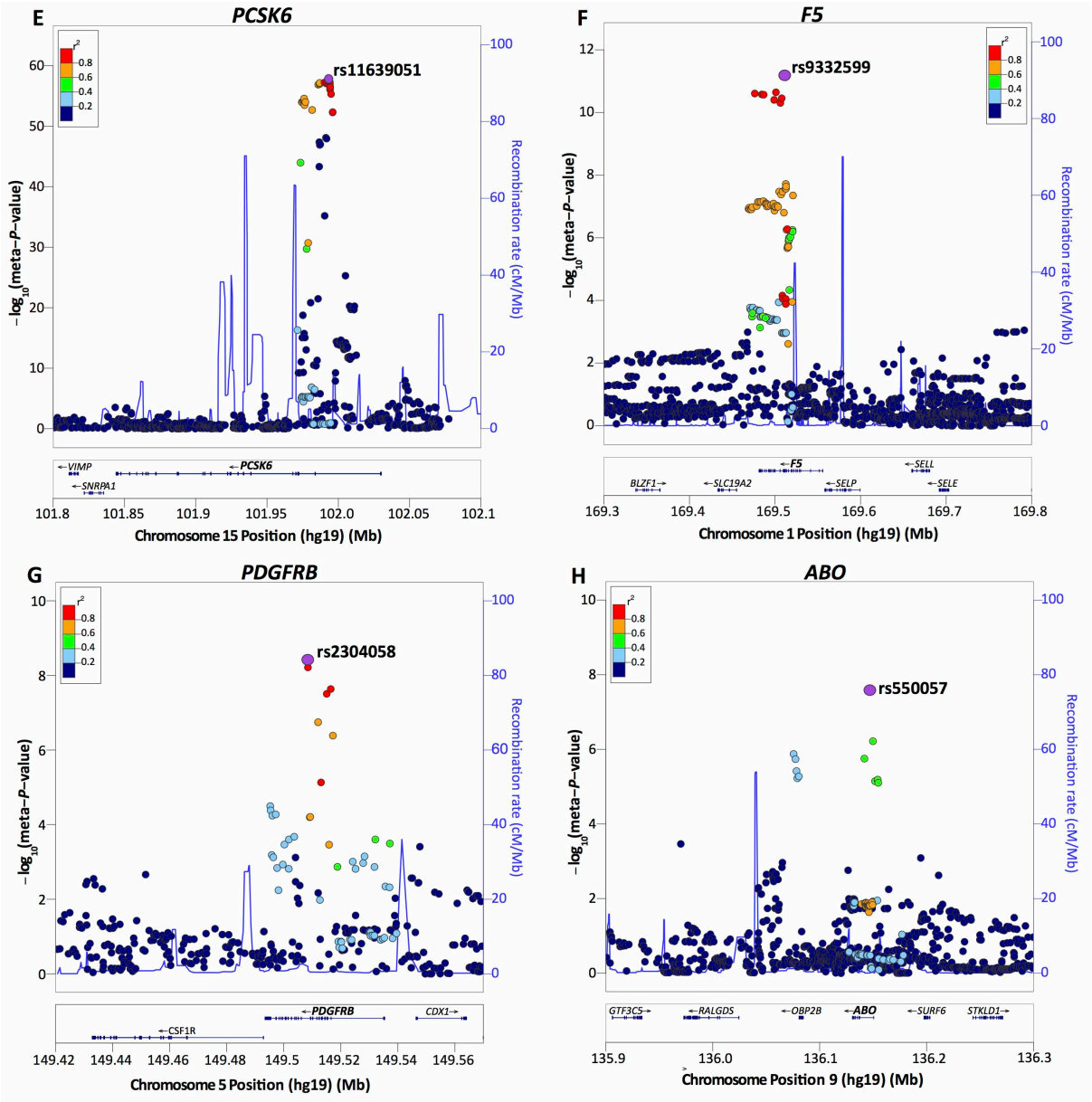

Figure S3: Regional association plots for each of the 8 loci associated with the cytokine network from the meta-analysed multivariate GWA analysis

(E) *PCSK6* locus, rs11639051 is located in the second intron of *PCSK6* (proprotein convertase subtilisin/kexin type 6) on chromosome 15q26.3. **(F) *F5* locus**, rs9332599 is located within intron twelve of factor V (*F5*) gene on chromosome 1q24.2. **(G) *PDGFRB* locus**, rs2304058 lies within the tenth intron of the platelet-derived growth factor receptor-beta (*PDGFRB*) gene on chromosome 5q32. **(H) *ABO* locus**, rs550057 is located within the first intron of *ABO* gene on chromosome 9q34.2.

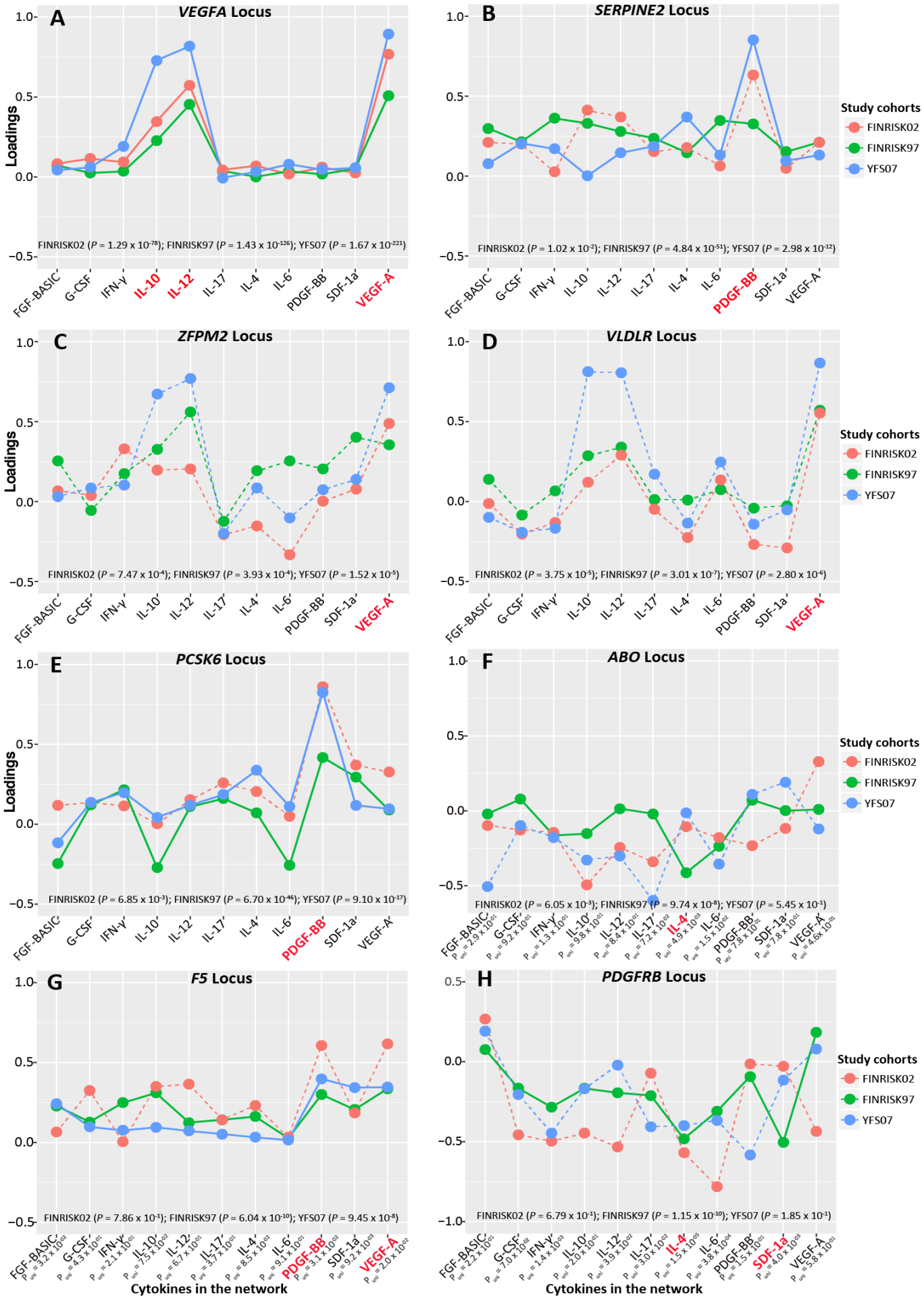
A
VEGFA Locus
FINRISK02 (P = 1.29 x 10-78); FINRISK97 (P = 1.43 x 10-126); YFS07 (P = 1.67 x 10-221)
B
SERPINE2 Locus
FINRISK02 (P = 1.02 x 10-2); FINRISK97 (P = 4.84 x 10-51); YFS07 (P = 2.98 x 10-12)
C
ZFPM2 Locus
FINRISK02 (P = 7.47 x 10-4); FINRISK97 (P = 3.93 x 10-4); YFS07 (P = 1.52 x 10-5)
D
VLDLR Locus
FINRISK02 (P = 3.75 x 10-5); FINRISK97 (P = 3.01 x 10-7); YFS07 (P = 2.80 x 10-6)
E
PCSK6 Locus
FINRISK02 (P = 6.85 x 10-3); FINRISK97 (P = 6.70 x 10-46); YFS07 (P = 9.10 x 10-17)
F
ABO Locus
FINRISK02 (P = 6.05 x 10-3); FINRISK97 (P = 9.74 x 10-8); YFS07 (P = 5.45 x 10-1)
G
F5 Locus
FINRISK02 (P = 7.86 x 10-1); FINRISK97 (P = 6.04 x 10-10); YFS07 (P = 9.45 x 10-8)
H
PDGFRB Locus
FINRISK02 (P = 6.79 x 10-1); FINRISK97 (P = 1.15 x 10-10); YFS07 (P = 1.85 x 10-1)
Loadings
Cytokines in the network
FGF-BASIC
G-CSF
IFN-γ
IL-10
IL-12
IL-17
IL-4
IL-6
PDGF-BB
SDF-1a
VEGF-A
Study cohorts
FINRISK02
FINRISK97
YFS07

Figure S4: Loadings, the contribution of each cytokine in the cytokine network to the multivariate association results with the lead SNPs at the *VEGFA*, *SERPINE2*, *ZFPM2*, *VLDLR* and *PCSK* locus in each study cohort.

The trait loadings are output results from MV-PLINK – the sign of the loadings for each cytokine in each cohort indicates whether the genetic variant influences different cytokines in the same or opposite effect direction. The dotted lines mean that the locus did not achieve 1×10^{-5} in a particular cohort. The top cytokine(s) for each locus is highlighted in red. The univariate meta-analysed *P*-values, for each cytokine, for the new loci (*ABO*, *PDGFRB* and *F5*) identified in the multivariate analysis are also provided.