

Genetic study of circulating cytokines offers insight into the determinants, cascades and effects of systemic inflammation

Ville Karhunen MSc^{1*}, Dipender Gill BMBCh PhD^{1*}, Rainer Malik PhD², Mark J. Ponsford BMBCh MSc³,
Ari Ahola-Olli MD PhD⁴⁻⁶, Areti Papadopoulou MSc⁷, Saranya Palaniswamy PhD^{1,8}, Shivaprakash
Jagalur Mutt MSc⁹, Sylvain Sebert PhD⁸, Minna Männikkö PhD¹⁰, Juha Auvinen MD PhD⁸, Juha
Veijola PhD¹¹, Karl-Heinz Herzig MD PhD^{9,12}, Markku Timonen MD PhD^{8,13}, Sirkka Keinänen-
Kiukaanniemi MD PhD^{8,13,14}, Martin Dichgans MD^{2,15,16}, Marko Salmi MD PhD¹⁷, Sirpa Jalkanen MD
PhD¹⁷, Terho Lehtimäki PhD¹⁸, Veikko Salomaa MD PhD¹⁹, Olli Raitakari MD PhD^{20,21}, Simon A Jones
PhD³, Konstantinos K. Tsilidis PhD^{1,22†}, Marjo-Riitta Järvelin MD PhD^{1,8,13,23-25†}, Abbas Dehghan MD
PhD^{1,23, 26†}

*These authors contributed equally and are joint first.

†These authors contributed equally and are joint last.

¹Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,
London, UK

²Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich,
Germany

³Division of Immunology, Infection, and Inflammation, Tenovus Institute, Cardiff University, Cardiff,
UK

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

⁴The Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁵Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

⁶Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

⁷William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London UK

⁸Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland

⁹Research Unit of Biomedicine, Medical Research Center (MRC), University of Oulu, University Hospital, Oulu, Finland

¹⁰Northern Finland Birth Cohorts, Infrastructure for Population Studies, Faculty of Medicine, University of Oulu, Oulu, Finland.

¹¹Department of Psychiatry, Research Unit of Clinical Neuroscience, University of Oulu & Department of Psychiatry, University Hospital of Oulu, Finland

¹²Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, Poznan, Poland

¹³Unit of Primary Care, Oulu University Hospital, Oulu, Finland

¹⁴Healthcare and Social Services of Selänne, Pyhäjärvi, Finland and city of Oulu, Oulu, Finland.

¹⁵Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.

¹⁶German Centre for Neurodegenerative Diseases (DZNE), Munich, Germany.

¹⁷MediCity and Institute of Biomedicine, University of Turku, Turku, Finland.

¹⁸Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center, Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.

¹⁹Finnish Institute for Health and Welfare, Helsinki, Finland.

²⁰Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland

²¹Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

²²Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece

²³MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK

²⁴Biocenter Oulu, University of Oulu, Oulu, Finland

²⁵Department of Life Sciences, College of Health and Life Sciences, Brunel University London, London, United Kingdom

²⁶UK Dementia Research Institute at Imperial College London, London, UK

Contributors

AD, DG, KKT, MRJ and VK designed the study. VK, DG, AAO, RM and AP performed the analysis. All authors interpreted the results. VK, DG and MJP drafted the manuscript. All authors critically revised the manuscript for intellectual content. All authors approved the submitted version and are accountable for the integrity of the work.

Sources of Funding

VK is funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant (721567). DG was supported by the Wellcome Trust 4i Programme (203928/Z/16/Z) and British Heart Foundation Centre of Research Excellence (RE/18/4/34215) at Imperial College London. MP is a funded by the Welsh Clinical Academic Training (WCAT) programme and is a participant in the NIH Graduate Partnership Program. VS was supported by the Finnish Foundation for Cardiovascular Research. SP is funded by the the European Union's Horizon 2020 Framework EDCMET (825762), iHealth-T2D (643774), Yrjö Jahnsson Foundation and Päivikki and Sakari Sohlberg Foundation. KKT was supported by Cancer Research UK (C18281/A29019). MRJ is funded by the European Union's Horizon 2020 programmes: EDCMET (825762), DynaHEALTH (633595), LifeCycle (), iHealth-T2D (643774), LongITools (874739) and Medical Research Council/Biotechnology and Biological Sciences Research Council (PREcisE, MR/M013138/1, MR/S03658X/1, under Nutrition & Epigenome, The Joint Programming Initiative a Healthy Diet for a Healthy Life (JPI HDHL/EU-H2020, project number 665)). AD is funded by the Wellcome Trust seed award (206046/Z/17/Z).

This project has also received funding from the European Union's Horizon 2020 research and innovation programme (666881), SVDs@target (to MD; 667375), CoSTREAM (to MD); the DFG as part of the Munich Cluster for Systems Neurology (SyNergy, EXC EXC 2145 SyNergy – ID 390857198), the CRC 1123 (B3; to MD) and project DI 722/13-1; the Corona Foundation (to MD); the LMUexcellent fond (to MD); the e:Med program (e:AtheroSysMed; to MD) and the FP7/2007-2103

European Union project CVgenes@target (grant agreement number Health-F2-2013-601456; to MD). NFBC1966 received core funding for data generation and curation from the Academy of Finland (285547 (EGEA)), University Hospital Oulu, Finland (65354, 75617), Biocenter Oulu, Finland. The NFBCs are also funded by EU-H2020 EUCAN Connect (824989).

The Young Finns Study has been financially supported by the Academy of Finland: grants 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 for TAXINOMISIS and 848146 for To Aition); European Research Council (742927 for MULTIEPIGEN project); and Tampere University Hospital Supporting Foundation. The original cytokine measurements were funded by Academy of Finland (141136).

Acknowledgements

We thank all cohort members and researchers who participated in the NFBC1966 31-year study. We also wish to acknowledge the work of the NFBC project centre. The authors acknowledge the contributors of the data used in this work: BCAC, CARDIoGRAMplusC4D, CHARGE, CKDGen, DIAGRAM, GIANT, GLGC, MEGASTROKE, PRACTICAL and UK Biobank.

Disclosures

DG is employed part-time by Novo Nordisk. VS has received honoraria from Novo Nordisk and Sanofi for consultations and travel support from Novo Nordisk. He also has ongoing research collaboration with Bayer Ltd. These are all unrelated to the present study. The remaining authors have no conflicts of interest to declare.

Abstract

Cytokines are the signalling molecules that underlie inflammatory processes. Here, we performed genome-wide association study (GWAS) analyses of 47 circulating cytokines in up to 13,365 individuals to identify protein quantitative trait loci (pQTL). Applying a novel approach, we incorporated pQTL and expression quantitative trait loci (eQTL) data of 10,361 tissue samples in 635 individuals to identify biologically plausible genetic instruments to proxy the effect of cytokines. Using Mendelian randomization analysis, we explored the causal determinants of inflammatory cytokines, investigated inflammatory cascades and evaluated their effects on 20 diseases. We show evidence of body mass index (BMI), smoking and systolic blood pressure (SBP) being associated with inflammation, and specifically BMI affecting levels of active PAI-1, HGF, MCP1, sE-Selectin, sICAM1, TRAIL, IL6 and CRP. Our analysis highlights a key role of VEGF in influencing the levels of eight other inflammatory cytokines. Finally, we report evidence of sICAM affecting waist circumference and risk of major depressive disorder, evidence for TRAIL affecting the risk of cardiovascular diseases, breast and prostate cancer, and evidence for MIG affecting the risk of stroke. Overall, our results offer insight into inflammatory mediators of BMI, smoking and SBP, pleiotropic effects of VEGF, and circulating cytokines that increase the risk of cancer, cardiovascular, metabolic and neuropsychiatric diseases. All the studied cytokines represent pharmacological targets and therefore offer opportunities for clinical translation in diseases with inflammatory components.

Introduction

Cytokines, chemokines, growth factors and interferons (hereafter cytokines) are circulating protein signalling molecules that underlie inflammatory processes(1). Their levels are influenced by metabolic traits, and in turn have implications for various diseases(2). Cytokines are already targeted clinically for the prevention and treatment of a range of autoimmune and inflammatory processes(3), with further promise for cardiovascular(4), cancer(5), and neuropsychiatric diseases(6). Elucidation of the mechanisms relating cytokines to disease risk is important for the prioritisation of agents for study in clinical trials.

Availability of genotyped populations in which circulating cytokines have been measured has allowed for an investigation into the genetic determinants of systemic cytokine levels(7, 8). Similarly, variants associated with expression of cytokine genes across different tissues have been described(9). These genetic variants that associate with circulating cytokine levels and gene expression may be used as instrumental variables in the Mendelian randomization (MR) paradigm to study potential causal effects of inflammatory cytokines on disease risk(10, 11). In MR, randomly allocated genetic variants are used as unconfounded proxies for the effects of the exposure. These genetically predicted exposures are then evaluated for their association with the outcome of interest. MR can help overcome the confounding and reverse causation that hamper causal inference in conventional observational research(10, 12), and adds to the triangulation of causal evidence from multiple sources. With an estimated median cost of \$985 million required to take a new medicine to market(13), genetic data offers an opportunity to prioritise target candidates while saving on costs(14).

In this study, we conducted the largest genome-wide association studies (GWAS) of 47 circulating cytokines (Table 1) in up to 13,365 individuals and performed MR analyses to investigate the effect of cardiometabolic traits on their levels. Further, incorporating gene expression GWAS data from 10,361 samples on 53 tissues in 635 individuals, we aimed to identify genetic instruments with α

priori high biological relevance to proxy the effect of varying circulating cytokine levels. We applied these instruments to explore inflammatory cascades and effects on cancer, cardiovascular, metabolic and neuropsychiatric outcomes.

Results

Figure 1 shows the study design. We conducted GWAS on the circulating levels of 47 cytokines (Manhattan and QQ-plots are presented in Supplementary Figures 1-47) and used the summary statistics for the subsequent MR analysis. First, we examined potential causal effects of cardiometabolic traits on circulating cytokine levels. We found positive associations ($P < 0.0011$, applying a Bonferroni correction for testing of multiple cytokines) between:

- (i) genetically predicted body-mass index (BMI) and circulating levels of C-reactive protein (CRP), hepatic growth factor (HGF), interleukin (IL) 16, tumour necrosis factor related apoptosis-inducing ligand (TRAIL), monocyte chemoattractant protein-1 (MCP1), active plasminogen activator inhibitor-1 (activePAI1), soluble E-selectin (sE-selectin) and soluble vascular cell adhesion molecule-1 (sICAM1),
- (ii) genetically predicted smoking and levels of CRP and sICAM1, and
- (iii) genetically predicted systolic blood pressure (SBP) and CRP (Figure 2, Supplementary Figure 48, Supplementary Table 1).

To examine the impact of circulating cytokine levels on each other and on disease outcomes, we selected genetic instruments for circulating cytokine levels using two different criteria. In the first approach, we selected instrument variants within 500 kb of their corresponding gene locus and associated with their circulating levels at $P < 1 \times 10^{-4}$, which we term *cis*-protein quantitative trait loci (*cis*-pQTL) variants (Table 1). In the second approach, we chose instrument variants within 500 kb of the corresponding gene locus and associated with both its gene expression aggregated across tissues at $P < 1 \times 10^{-4}$, and its circulating cytokine levels at $P < 0.05$, which we term *cis*-expression quantitative

trait loci (*cis*-eQTL) variants (Table 1). The *cis*-pQTL and *cis*-eQTL instrument variants were available for 22 and 10 cytokines, respectively, with both types of instruments available for nine cytokines (Table 1). F-statistics (as a measurement of instrument strength, based on circulating protein levels) for all instrument variants ranged from 15 to 928 for *cis*-pQTLs (Supplementary Table 2) and from 5 to 178 for *cis*-eQTLs (Supplementary Table 3).

Next, we investigated the potential causal effects of circulating cytokine levels on each other. There was MR evidence ($P < 0.0011$) for associations between higher genetically predicted circulating cytokine levels and both increased and decreased levels of other circulating cytokines when using the *cis*-pQTL or *cis*-eQTL instrument selection criteria (Figure 3, Supplementary Figure 49, Supplementary Table 4). Most associations were seen for genetically predicted vascular endothelial growth factor (VEGF) and IL18, which were associated with circulating levels of eight and four other cytokines, respectively (Figure 3).

We used the GWAS-pairwise colocalization method(15) to further investigate the putative causal effects of circulating cytokine levels on each other based on our identified MR associations. We investigated proportionality of associations at the exposure gene locus with both the exposure and outcome traits, and thus any evidence for colocalization would further support a causal relationship. The results showed evidence of colocalization (posterior probability for shared variant > 0.9) for the associations of VEGF with interferon gamma (IFN γ), IL10, IL12p70, IL13, IL5 and IL17 at the *VEGF* gene locus, and for the association between monokine induced by interferon-gamma (MIG) and interferon gamma-induced protein 10 (IP10) at the *MIG* gene locus (Figure 4), providing supporting evidence for causality between the corresponding cytokines.

We finally conducted MR to examine the effect of circulating cytokine levels on 20 disease traits using outcome summary statistics from largest available GWAS, including cancer, cardiovascular, metabolic and neuropsychiatric outcomes (Supplementary Table 5). All these traits are previously shown to be related to inflammation(2). Using the *cis*-pQTL instrument selection criteria and

applying a Bonferroni correction for 20 outcomes, there was evidence of an association ($P < 0.0025$) between higher genetically predicted macrophage inflammatory protein-1-beta (MIP1 β) levels and increased risk of stroke, between higher genetically predicted sICAM1 levels and both increased risk of major depressive disorder (MDD) and lower waist circumference, and between higher genetically predicted TRAIL levels and both increased risk of coronary artery disease (CAD) and decreased risk of breast and prostate cancer (Figure 5, Supplementary Figure 50, Supplementary Table 4). Using the *cis*-eQTL criteria, there was evidence of an association between higher genetically predicted IL1 α levels and lower BMI (Figure 5). For the 15 cytokines where there were three or more instrument variants available (Table 1), weighted median and MR-PRESSO sensitivity analyses produced consistent MR estimates to the main inverse-variance weighted (IVW) analysis (Methods, Supplementary Figure 51, Supplementary Table 4). We compared the results of the two types of instrumental variables for the effects of genetically predicted cytokine levels on outcome traits and found a positive correlation between the *cis*-pQTL main MR estimates and the *cis*-eQTL main MR estimates (Pearson's correlation coefficient $r = 0.45$, $P = 1.2 \times 10^{-7}$, Figure 6).

Discussion

We performed a comprehensive MR analysis into the determinants, cascades, and effects of circulating cytokine levels by exploiting both pQTL and eQTL data. Our updated GWAS analyses of 47 cytokines facilitated investigation into the effects of cardiometabolic traits on systemic inflammation, and identified associations of genetically predicted BMI, SBP and smoking on circulating cytokines. GWAS of circulating cytokines in up to 13,365 healthy individuals coupled with genome-wide eQTL data of 10,361 tissue samples identified robust genetic instruments, with plausible biological relevance, for circulating cytokines. We applied these instruments in MR to explore inflammatory cascades and cytokine effects on disease risk. Our results suggest pleiotropic inflammatory effects of VEGF, and various circulating cytokines and adhesion molecules that increase risk of cancer, cardiovascular, metabolic and neuropsychiatric diseases.

We replicate associations of smoking with elevation of CRP and sICAM1, and of SBP with elevation of CRP, representing validated markers of systemic inflammation and future cardiovascular risk(16).

Elevated BMI emerges as an exposure that predicts multiple indices of inflammation, including thrombosis (e.g. plasminogen activation inhibitor-1), metabolic disease and altered endothelial function or immune homeostasis (HGF, MCP1, TRAIL, sICAM1 and sE-selectin). Elevation of these mediators has been reported across human and murine studies of obesity, with fatty-acid excess and relative hypoxia emerging as drivers of the inflammatory stress response(17-19). Obesity is a leading preventable threat to global health in terms of premature cardiovascular and cancer morbidity and mortality(20), suggesting adipose tissue as a relevant setting to contextualise potential mechanisms.

To better understand cytokine regulatory networks, we examined the effects of genetically predicted circulating cytokines on levels of other cytokines, as illustrated in Figure 3. For the MR findings that could be supported by evidence of colocalization at the relevant gene locus, we provide additional support for a causal mechanism. This innovative strategy enables unravelling of inflammatory pathways and insight into mediating mechanisms, which may be of relevance where a circulating cytokine is not suitable for clinical intervention, but its upstream or downstream mediators could instead be targeted. This is illustrated by strategies targeting MCP1, highlighted in recent work as a potential target to lower large-artery stroke and cardioembolic stroke risk(21).

Since the MCP1/CCL2-CCR2 axis regulates monocyte emigration from the bone marrow, prolonged therapeutic blockade may cause adverse monocyte depletion(22), prompting the need for alternative intervention strategies to overcome the depletion(23). Our results point to MCP1 affecting eotaxin expression, which would be consistent with their shared role in promotion and maintenance of the T helper cell 2 (Th2)-polarised immune response, illustrated by the co-expression of their receptors (CCR2 and CCR3, respectively) across human eosinophils(24), innate lymphoid cell, and Th2 cell subsets(25). Targeting the eotaxin/CCR3 pathway may therefore represent a novel therapeutic avenue to reduce cardiovascular risk. This precision approach appears

convergent with recent mechanistic insights illuminating the contribution of eosinophils in the initiation of thrombotic events during cardiovascular disease(26).

Our approach reveals network complexity, with evidence of VEGF appearing a master regulator. This is consistent with an earlier report identifying VEGF as an upstream controller of IL12p70, IL7, IL10, and IL13⁷. We replicate and extend this analysis to reveal a wider range of cytokines within this cascade, including drivers of Th2-type responses (IL5 and IL13), TNF α , and IFN γ signalling (IFN γ , IL12) and immune resolution (IL10). With the exception of TNF α , the cytokines highlighted are notable for shared activation of the Jak-STAT signalling(27), a pathway which has evolved to sense and interpret changes in the tissue microenvironment and targetable by a range of therapeutics. Based on directionality, we infer VEGF blockade could lead to a reduction in multiple downstream molecules. Redundant angiogenic activity mediated by inflammatory and pro-angiogenic cytokines have recently been described in association with obesity, leading to treatment resistance to VEGF blockade in breast cancer(28). Thus, an overview of the cytokine regulatory network is, therefore, essential to support the discovery of common pathways and to explore cytokine redundancy. Our analysis highlights several putative cytokine networks, including the chemokines MIP1 β , IL1 α , and RANTES. This aligns with the shared potent pro-angiogenic effects of these molecules(29-31).

We also examined the effects of circulating cytokine levels on 20 disease outcomes. The results link circulating levels of TRAIL with CAD, and a protective effect on breast and prostate cancer risk. TRAIL was first described as a molecule capable of selectively inducing cancer cell apoptosis(32), but is now also recognised to be a potent inhibitor of VEGF-stimulated angiogenesis(33). Of therapeutic relevance, some triple-negative/basal-like breast cancer cells retain sensitivity to TRAIL as a single agent(34) whilst chemotherapeutic agents show promise in restoring sensitivity to TRAIL in hormone-refractory prostate cancer cells(35, 36). We saw evidence for associations between genetically predicted sICAM1 and increased risk of MDD, and lower waist circumference. Epidemiological studies have reported sICAM1 to be elevated in a range of neuropsychiatric disorders, but this association has been questioned due to confounders such as medication⁽³⁷⁾ and smoking⁽³⁸⁾. Our

study links genetically predicted smoking and BMI to increased sICAM1 levels, which in turn associate with elevated risk of MDD. Conversely, our MR association between genetically predicted MIP1 β levels and risk of stroke is consistent with the ability of MIP1 β to induce endothelial cell adhesion, and the formation of vascular plaques associated with stroke(39).

Furthermore, using eQTL instruments, our results suggest an effect of higher IL1 α levels on decreasing BMI. IL1 α is active following translation(40), and some *in vitro* evidence implies IL1 α attenuating differentiation of pre-adipocytes and suppressing lipid accumulation(41), suggesting biological relevance.

Thus, through analysis of genetically proxied cytokines, we suggest a pattern of pro- and anti-angiogenic and inflammatory effectors linking obesity to multi-morbidity. Given that all cytokines investigated in our study represent viable targets for pharmacological intervention³, these findings could have important clinical implications for the prevention of disease, and therefore warrant further investigation.

Our use of genetic instruments encompassed the cumulative lifelong effects of natural variation in circulating cytokine levels, thus making progress towards overcoming the confounding, reverse causation and measurement error biases that can cloud causal interpretation of associations identified in conventional observational research(42). This is particularly relevant given the complex interactions that underlie levels of circulating cytokines(43), and therefore similarly relates to our investigation of cytokine pathways and effects on disease risk. In our analyses, we aimed to maximise the validity of instruments selected to proxy levels of circulating cytokines by only considering variants that located at the corresponding gene locus. This strategy for selecting instruments is likely to both increase the biological relevance of the instrument, and decrease the possibility of horizontal pleiotropy(44). Our use of complementary pQTL and eQTL instrument selection criteria offered distinct but consistent findings and provided a time-efficient and cost-effective approach for identifying potential drug targets for the prevention of a broad range of

common and serious diseases. Furthermore, for associations between cytokines, we used colocalization analyses to strengthen the evidence for causal effects.

There are also limitations to our work. The MR analyses should not be directly extrapolated to infer the effect of a clinical intervention²⁵, as the instruments employed represent the cumulative effect of lifelong genetic predisposition, while a clinical intervention typically represents a discrete event at a particular time point(45). Our ability to assess causality between complex cytokine networks based on genetic methods remains limited by the availability of genetic instruments and the expression profiles of these molecules. Furthermore, the MR estimates may be biased by possible pleiotropic effects of the variants employed as instruments, where they influence the outcome under consideration through pathways independent of the cytokine being investigated. To minimise this risk, instruments for cytokines were selected from their corresponding gene loci to increase their specificity(44). Where possible, MR sensitivity analyses using methods that are more robust to the inclusion of pleiotropic variants were also incorporated(46), and provided consistent results. Another potential source of bias in the MR analyses relates to the use of weak instruments(47, 48). However, 29 out of 32 instruments included only variants with F-statistics > 10, the only exceptions being *cis*-eQTL instruments for IL1 α , MIF and sICAM1(49). Finally, our approach considered circulating cytokine levels and cytokine gene expression aggregated across tissues, rather than restricting to biologically relevant sites, thus potentially having impact on the generalizability of our findings.

In conclusion, we used large-scale genetic data relating to cytokine levels to assess causal links between cardiometabolic traits and systemic inflammation. By incorporating gene expression data, we identified valid and biologically plausible instruments for use in MR analyses and investigated effects of circulating cytokines on inflammatory pathways and disease risk. Our findings are supported by existing work and add insight into the role of distinct inflammatory processes in the aetiology of a broad range of common cardiovascular, malignant, metabolic and neuropsychiatric

diseases. Our work provides suggestive causal associations for further validation, after which efforts could be made towards identifying therapeutic opportunities in clinical practice.

Methods

Genome-wide association study cohorts

Northern Finland Birth Cohort 1966

Northern Finland Birth Cohort 1966 (NFBC1966) recruited pregnant women with expected date of delivery between 1st January and 31st December 1966(50). Overall, 12,055 mothers (over 96% of eligible women) were followed from pregnancy onwards, with 12,058 live-born children in the cohort. In the offspring 31-year data collection in 1997, all cohort members with known addresses in either the Northern Finland or Helsinki area were invited to a clinical examination(51). In total, data were received for 6033 participants, and DNA was successfully extracted for 5753 participants from fasted blood samples. Cytokines were quantified from overnight fasting plasma samples using Bio-Rad's Bio-Plex 200 system (Bio-Rad Laboratories, California, USA) with Milliplex Human Chemokine/Cytokine and CVD/Cytokine kits (Cat# HCYTOMAG-60K-12 and Cat# SPR349; Millipore, St Charles, Missouri, USA) and Bio-Plex Manager Software V.4.3 as previously described(52, 53). Genotyping was conducted using Illumina HumanCNV-370DUO Analysis BeadChip (Illumina, California, USA) and imputed using Haplotype Reference Consortium imputation reference panel.

The Cardiovascular Risk in Young Finns Study

The Cardiovascular Risk in Young Finns (YFS) is an ongoing follow-up study of 3,596 children and adolescents aged 3, 6, 9, 12, 15, or 18 years. The subjects were randomly chosen from five university cities and their rural surroundings using Finnish population register. The baseline survey was held in 1980 and subsequent follow-up visits involving all five centres have been arranged in 1983, 1986, 1989, 2001, 2007, 2011 and 2017. The latest follow-up included also children and parents of the original participants.

Genotyping have been performed using the blood samples drawn at 2001 follow-up visit.

Genotyping was performed with custom-build Illumina 670K array. The custom content replaced some poor performing SNPs on the Human610 BeadChip and added more CNV content after which the customized chip shared 562,643 SNPs with Illumina Human610 chip. Genotyping was performed for 2,556 samples. Prior to imputation, samples and probes with high missingness were excluded (MIND > 0.05 and GENO < 0.05). To exclude poorly functioning probes, we excluded SNPs deviating from Hardy-Weinberg equilibrium (HWE p-value < 1×10^{-6}). To exclude related samples, we used $\hat{\pi}$ cut-off of 0.20. The pair with greater missingness was removed. After the QC steps, the data set included 2,443 individuals and 546,674 probes. The imputation was performed with IMPUTE2 software by using 1000 Genomes Phase 3 release as reference panel. After imputation, poorly imputed and rare variants (INFO < 0.7 and MAC < 3) were removed.

Biorad's Bio-Plex Pro Human Cytokine 27-plex Assay and 21-plex Assay were used to quantify circulating concentrations of 48 cytokines from serum samples drawn at 2007 follow-up visit as previously described(54). Depending on the cytokine, imputed genotypes and cytokine concentrations were available for 116-2019 samples.

FINRISK

FINRISK surveys are population-based cross-sectional studies which began in 1972. A new sample is recruited every five years to monitor the health status of Finnish population. Subset of individual-level data from 1992-2012 surveys is available through THL Biobank. Cytokine quantification for FINRISK1997 and FINRISK2002 samples was performed similarly as in YFS, but quantification was done using EDTA plasma in FINRISK1997 and heparin plasma in FINRISK2002(55). In FINRISK1997, a custom 20-plex array was used in cytokine quantification. In FINRISK2002, only participants between 51 and 74 years were selected for the analysis. Imputation was performed using 1000 Genomes Phase 3 as reference panel. Poorly imputed variants (INFO < 0.7) and variants with low minor allele count (MAC < 3) were excluded. Depending on the cytokine, imputed genotypes and cytokine

concentrations were available for 3440-4613 samples from FINRISK1997 and 843-1705 samples from FINRISK2002.

Cytokine genome-wide association study

We conducted GWAS for 41 cytokines in FINRISK+YFS population and for 16 cytokines in NFBC1966 (Supplementary Table 6)(8). The data pre-processing was done in a similar manner to previous GWAS analyses(7, 8). Inverse-normal rank transformation was first applied to the traits, before regressing the transformed measures on age, sex and the first 10 genetic principal components. In contrast to the previous analyses(7, 8), we did not add BMI as a covariate, as this could potentially introduce collider bias into consequent MR analyses(56). The inverse-normal rank transformation was again applied to the residuals of this regression, and these transformed residual estimates were used as response variables in the GWAS. The GWAS was conducted in each study using an additive genetic model with SNPTEST2 software(57). The results for variants which showed poor imputation quality (model info < 0.7) or low minor allele frequency (MAF, < 0.05) were discarded. For the ten cytokines available in both NFBC1966 and FINRISK+YFS (Supplementary Table 6), the summary statistics were pooled by inverse variance weighted fixed-effects meta-analysis using Metal software(58).

Gene expression GWAS summary statistics were obtained from the GTEx project (release version 7), and related to 10,361 samples from a multi-ethnic group of 635 individuals(9). Results from 53 tissues were pooled using fixed-effects meta-analysis to produce cross-tissue estimates of association with gene expression(9).

Metabolic trait instrument selection

The metabolic traits considered were BMI, low-density lipoprotein cholesterol (LDL-C), lifetime smoking (referred to as smoking), systolic blood pressure (SBP) and type 2 diabetes mellitus (referred to as diabetes). Genetic association estimates for BMI were obtained from the GIANT Consortium GWAS meta-analysis of 806,834 individuals of Europeanancestry(59). Genetic

association estimates for fasting serum LDL-C were obtained from the Global Lipids Genetic Consortium GWAS of 188,577 individuals of European-ancestry(60). Genetic association estimates for smoking were obtained from a GWAS of 462,690 European-ancestry individuals, with a continuous measure of smoking created based on self-reported age at initiation, age at cessation and cigarettes smoked per day(61). Genetic association estimates for SBP were obtained from a GWAS of 318,417 White British individuals, with correction made for any self-reported anti-hypertensive medication use by adding 10mmHg(62). Genetic association estimates for liability to type 2 diabetes mellitus came from the DIAGRAM Consortium GWAS meta-analysis of 74,124 cases and 824,006 controls of European ancestry(63). Instruments for each trait were selected as single-nucleotide polymorphisms (SNPs) that associated with that trait at genome-wide significance ($P < 5 \times 10^{-8}$) and were in pair-wise linkage disequilibrium (LD) $r^2 < 0.001$. All clumping was performed using the TwoSampleMR package in R(64).

Cytokine instrument selection

Two distinct instrument selection criteria were used to identify variants to proxy the effect of circulating cytokine levels in MR analyses. Firstly, variants located within 500 kb of the gene locus corresponding to a cytokine under study that also related to circulating levels of that cytokine with association P -value $< 1 \times 10^{-4}$ were selected. These are referred to as *cis*-pQTL. Secondly, variants located within 500 kb of the gene locus corresponding to a cytokine under study that related to both expression of that gene at $P < 1 \times 10^{-4}$ and circulating levels of that cytokine at $P < 0.05$ were selected. These are referred to as *cis*-eQTL. The gene locations were extracted per human genome build 19 (released February 2009) using UCSC Genome Browser (accessed on 18th June 2019) (65). Only variants for which both exposure and outcome genetic association estimates were available for any given MR analysis were considered as potential instruments for that analysis. All variants were clumped to a pairwise linkage disequilibrium threshold of $r^2 < 0.1$ using the TwoSampleMR package in R(64). The F statistic was calculated as a measure of the strength of the instruments(66).

Outcome genetic association estimates

Disease outcomes falling into the categories of cardiovascular (CAD and stroke), cancer (breast and prostate), metabolic (BMI, fasting glucose, fasting insulin, type 2 diabetes mellitus, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol and triglycerides) and neuropsychiatric (attention-deficit hyperactivity disorder, autism spectrum disorder, bipolar disorder, major depressive disorder and schizophrenia) traits were investigated, based on the availability of large-scale GWAS summary data. In addition, the effects of circulating cytokine levels on C-reactive protein levels, an established biomarker of inflammation, were investigated. Trait definitions and details of the original GWAS analyses are provided in Supplementary Table 5.

Mendelian randomization analysis

MR analysis was performed using the two sets of instruments (*cis*-pQTL and *cis*-eQTL) to investigate effects of genetically predicted circulating cytokine levels on levels of other circulating cytokines (Supplementary Table 6) and the considered disease outcomes (Supplementary Table 5). The Pearson correlation coefficient was used to estimate the similarity between MR estimates obtained for the same exposure-disease outcome associations when using the two different instrument selection criteria (pQTL and eQTL). The ratio method was used to obtain MR estimates, with first order weights used to generate standard errors(67). Where more than one instrument variant was available for a given analysis, MR estimates obtained from different instruments were pooled using the multiplicative random-effects inverse-variance weighted (IVW) MR method. MR is prone to bias when the genetic variants used as instruments affect the outcome through some pathway that is independent of the exposure under consideration, a phenomenon termed pleiotropy(68). To explore this possibility, MR methods that make distinct assumptions on the inclusion of pleiotropic variants were performed in sensitivity analyses where three or more instrument variants were available. Specifically, the weighted median and MR-PRESSO methods were used(69, 70). Briefly, the weighted median approach orders the MR estimates produced by each instrument by their magnitude weighted for their precision and produces an overall MR estimate based on the median value, with

standard error estimated by bootstrapping(69). It is a consistent approach when more than half of the weight for the analysis is derived from valid instruments (69). The MR-PRESSO detects outliers using the squared residuals from the regression of the variant-outcome association estimates on the variant-exposure association estimates with the intercept fixed to zero, and repeats such regression-based MR analysis after excluding any identified outlier variants(70). To account for multiple testing, we applied a Bonferroni correction accounting for the number of outcomes for each analysis; thus, for the MR analysis on 20 disease outcomes and 47 cytokines, the thresholds for statistical significance were 0.0025 and 0.0011, respectively. The MR sensitivity analyses were only performed to explore the robustness of the main IVW analysis to potential pleiotropy and as such no statistical sensitivity threshold was applied for these.

Colocalization

We used GWAS-pairwise for colocalization to further evaluate the cytokine pairs for which there was MR evidence of causal effect in either direction(15). The colocalization analysis investigates whether the same variant within a genetic locus is associated with both exposure and outcome trait. We used the Z scores from the cytokine GWAS summary statistics. The number of variants included in each analysed chunk was set automatically by using an approximately linkage disequilibrium independent block file(15). Results with a posterior probability of association > 0.9 within the gene locus of the putative causal cytokine of each pair were deemed statistically significant. We only considered the gene region for the exposure cytokine, and thus any evidence for colocalization would imply supporting evidence for our MR results.

Data availability and ethical approval

All supporting data for this work are available within the article, its supplementary files and the citations provided. *The GWAS summary statistics generated in this work will be made publicly available upon the publication of the article.* All study cohorts used in this work had already obtained

relevant ethical approval and written participant consent. Statistical analysis was performed using R (version 3.6.0). The software code for the analyses are available from the authors upon request.

References

1. Akdis M, Burgler S, Cramer R, Eiwegger T, Fujita H, Gomez E, et al. Interleukins, from 1 to 37, and interferon-gamma: receptors, functions, and roles in diseases. *J Allergy Clin Immunol*. 2011;127(3):701-21 e1-70.
2. Furman D, Campisi J, Verdin E, Carrera-Bastos P, Targ S, Franceschi C, et al. Chronic inflammation in the etiology of disease across the life span. *Nat Med*. 2019;25(12):1822-32.
3. Rider P, Carmi Y, Cohen I. Biologics for Targeting Inflammatory Cytokines, Clinical Uses, and Limitations. *Int J Cell Biol*. 2016;2016:9259646.
4. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med*. 2017;377(12):1119-31.
5. Lee S, Margolin K. Cytokines in cancer immunotherapy. *Cancers (Basel)*. 2011;3(4):3856-93.
6. Khairova RA, Machado-Vieira R, Du J, Manji HK. A potential role for pro-inflammatory cytokines in regulating synaptic plasticity in major depressive disorder. *Int J Neuropsychopharmacol*. 2009;12(4):561-78.
7. Ahola-Olli AV, Wurtz P, Havulinna AS, Aalto K, Pitkanen N, Lehtimaki T, et al. Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am J Hum Genet*. 2017;100(1):40-50.
8. Sliz E, Kalaoja M, Ahola-Olli A, Raitakari O, Perola M, Salomaa V, et al. Genome-wide association study identifies seven novel loci associating with circulating cytokines and cell adhesion molecules in Finns. *Journal of Medical Genetics*. 2019;56(9):607-16.
9. GTEx Consortium, Laboratory Data Analysis Coordinating Center Analysis Working Group, Statistical Methods Analysis Working Group, Enhancing GTEx Groups, NIH Common Fund BCSSN, Biospecimen Collection Source Site RPCI, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-13.
10. Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.
11. Georgakis MK, Gill D, Rannikmae K, Traylor M, Anderson CD, Lee JM, et al. Genetically Determined Levels of Circulating Cytokines and Risk of Stroke. *Circulation*. 2019;139(2):256-68.
12. Gill D, Georgakis MK, Koskeridis F, Jiang L, Feng Q, Wei WQ, et al. Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. *Circulation*. 2019;140(4):270-9.
13. Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA*. 2020;323(9):844-53.
14. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47(8):856-60.
15. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016;48(7):709-17.
16. Ridker PM, Hennekens CH, Roitman-Johnson B, Stampfer MJ, Allen J. Plasma concentration of soluble intercellular adhesion molecule 1 and risks of future myocardial infarction in apparently healthy men. *The Lancet*. 1998;351(9096):88-92.
17. Hosogai N, Fukuhara A, Oshima K, Miyata Y, Tanaka S, Segawa K, et al. Adipose Tissue Hypoxia in Obesity and Its Impact on Adipocytokine Dysregulation. *Diabetes*. 2007;56(4):901-11.
18. Dobrian AD, Lieb DC, Cole BK, Taylor-Fishwick DA, Chakrabarti SK, Nadler JL. Functional and pathological roles of the 12- and 15-lipoxygenases. *Prog Lipid Res*. 2011;50(1):115-31.
19. O'Rourke RW, White AE, Metcalf MD, Olivas AS, Mitra P, Larison WG, et al. Hypoxia-induced inflammatory cytokine secretion in human adipose tissue stromovascular cells. *Diabetologia*. 2011;54(6):1480-90.
20. Afshin A, Collaborators. TGO. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *New England Journal of Medicine*. 2017;377(1):13-27.

21. Georgakis MK, Gill D, Rannikmäe K, Traylor M, Anderson CD, Lee J-M, et al. Genetically Determined Levels of Circulating Cytokines and Risk of Stroke. *Circulation*. 2019;139(2):256-68.
22. Serbina NV, Pamer EG. Monocyte emigration from bone marrow during bacterial infection requires signals mediated by chemokine receptor CCR2. *Nature Immunology*. 2006;7(3):311-7.
23. Winter C, Silvestre-Roig C, Ortega-Gomez A, Lemnitzer P, Poelman H, Schumski A, et al. Chrono-pharmacological Targeting of the CCL2-CCR2 Axis Ameliorates Atherosclerosis. *Cell Metabolism*. 2018;28(1):175-82.e5.
24. Dunzendorfer S, Kaneider NC, Kaser A, Woell E, Frade JMR, Mellado M, et al. Functional expression of chemokine receptor 2 by normal human eosinophils. *Journal of Allergy and Clinical Immunology*. 2001;108(4):581-7.
25. Weston CA, Rana BMJ, Cousins DJ. Differential expression of functional chemokine receptors on human blood and lung group 2 innate lymphoid cells. *The Journal of allergy and clinical immunology*. 2019;143(1):410-3.e9.
26. Uderhardt S, Ackermann JA, Fillep T, Hammond VJ, Willeit J, Santer P, et al. Enzymatic lipid oxidation by eosinophils propagates coagulation, hemostasis, and thrombotic disease. *J Exp Med*. 2017;214(7):2121-38.
27. Seif F, Khoshmirsafa M, Aazami H, Mohsenzadegan M, Sedighi G, Bahar M. The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells. *Cell Commun Signal*. 2017;15(1):23-.
28. Incio J, Ligibel JA, McManus DT, Suboj P, Jung K, Kawaguchi K, et al. Obesity promotes resistance to anti-VEGF therapy in breast cancer by up-regulating IL-6 and potentially FGF-2. *Science Translational Medicine*. 2018;10(432):eaag0945.
29. Suffee N, Hlawaty H, Meddahi-Pelle A, Maillard L, Louedec L, Haddad O, et al. RANTES/CCL5-induced pro-angiogenic effects depend on CCR1, CCR5 and glycosaminoglycans. *Angiogenesis*. 2012;15(4):727-44.
30. Liao Y-Y, Tsai H-C, Chou P-Y, Wang S-W, Chen H-T, Lin Y-M, et al. CCL3 promotes angiogenesis by dysregulation of miR-374b/ VEGF-A axis in human osteosarcoma cells. *Oncotarget*. 2016;7(4):4310-25.
31. Fahey E, Doyle SL. IL-1 Family Cytokine Regulation of Vascular Permeability and Angiogenesis. *Front Immunol*. 2019;10:1426-.
32. de Miguel D, Lemke J, Anel A, Walczak H, Martinez-Lostao L. Onto better TRAILs for cancer treatment. *Cell Death & Differentiation*. 2016;23(5):733-47.
33. Cantarella G, Risuglia N, Dell'eva R, Lempereur L, Albini A, Pennisi G, et al. TRAIL inhibits angiogenesis stimulated by VEGF expression in human glioblastoma cells. *Br J Cancer*. 2006;94(10):1428-35.
34. Greer YE, Gilbert SF, Gril B, Narwal R, Peacock Brooks DL, Tice DA, et al. MEDI3039, a novel highly potent tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) receptor 2 agonist, causes regression of orthotopic tumors and inhibits outgrowth of metastatic triple-negative breast cancer. *Breast Cancer Research*. 2019;21(1):27.
35. Fujiwara J, Sowa, Y., Horinaka, M., Koyama, M., Wakada, M., Miki, T., Sakai, T. The anti-obesity drug orlistat promotes sensitivity to TRAIL by two different pathways in hormone-refractory prostate cancer cells. *International Journal of Oncology*. 2012;40:1483-91.
36. Vondálová Blanářová O, Šafaříková B, Herůdková J, Krkoška M, Tománková S, Kahounová Z, et al. Cisplatin or LA-12 enhance killing effects of TRAIL in prostate cancer cells through Bid-dependent stimulation of mitochondrial apoptotic pathway but not caspase-10. *PLOS ONE*. 2017;12(11):e0188584.
37. Müller N. The Role of Intercellular Adhesion Molecule-1 in the Pathogenesis of Psychiatric Disorders. *Frontiers in Pharmacology*. 2019;10(1251).
38. Delgado GE, Krämer BK, Siekmeier R, Yazdani B, März W, Leipe J, et al. Influence of smoking and smoking cessation on biomarkers of endothelial function and their association with mortality. *Atherosclerosis*. 2020;292:52-9.

39. Tatara Y, Ohishi M, Yamamoto K, Shiota A, Hayashi N, Iwamoto Y, et al. Macrophage inflammatory protein-1beta induced cell adhesion with increased intracellular reactive oxygen species. *J Mol Cell Cardiol.* 2009;47(1):104-11.
40. Di Paolo NC, Shayakhmetov DM. Interleukin 1 α and the inflammatory process. *Nature immunology.* 2016;17(8):906-13.
41. Um J-Y, Rim H-K, Kim S-J, Kim H-L, Hong S-H. Functional polymorphism of IL-1 alpha and its potential role in obesity in humans and mice. *PloS one.* 2011;6(12):e29524-e.
42. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol.* 2012;41(1):161-76.
43. Zhou X, Fragala MS, McElhaney JE, Kuchel GA. Conceptual and methodological issues relevant to cytokine and inflammatory marker measurements in clinical research. *Curr Opin Clin Nutr Metab Care.* 2010;13(5):541-7.
44. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol.* 2016;45(5):1600-16.
45. Gill D, Walker VM, Martin RM, Davies NM, Tzoulaki I. Comparison with randomized controlled trials as a strategy for evaluating instruments in Mendelian randomization. *Int J Epidemiol.* 2019.
46. Slob EAW, Burgess S. A Comparison Of Robust Mendelian Randomization Methods Using Summary Data. *bioRxiv.* 2019:577940.
47. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol.* 2011;40(3):755-64.
48. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol.* 2016;40(7):597-608.
49. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical methods in medical research.* 2012;21(3):223-42.
50. Rantakallio P. Groups at risk in low birth weight infants and perinatal mortality. *Acta paediatrica Scandinavica.* 1969;193:Suppl 193:1+.
51. Järvelin M-R, Sovio U, King V, Lauren L, Xu B, McCarthy MI, et al. Early Life Factors and Blood Pressure at Age 31 Years in the 1966 Northern Finland Birth Cohort. *Hypertension.* 2004;44(6):838-46.
52. Saukkonen T, Mutt SJ, Jokelainen J, Saukkonen AM, Raza GS, Karhu T, et al. Adipokines and inflammatory markers in elderly subjects with high risk of type 2 diabetes and cardiovascular disease. *Sci Rep.* 2018;8(1):12816.
53. Palaniswamy S, Gill D, De Silva NM, Lowry E, Jokelainen J, Karhu T, et al. Could vitamin D reduce obesity-associated inflammation? Observational and Mendelian randomization study. *The American Journal of Clinical Nutrition.* 2020;111(5):1036-47.
54. Santalahti K, Maksimow M, Airola A, Pahikkala T, Hutri-Kähönen N, Jalkanen S, et al. Circulating Cytokines Predict the Development of Insulin Resistance in a Prospective Finnish Population Cohort. *The Journal of Clinical Endocrinology & Metabolism.* 2016;101(9):3361-9.
55. Santalahti K, Havulinna A, Maksimow M, Zeller T, Blankenberg S, Vehtari A, et al. Plasma levels of hepatocyte growth factor and placental growth factor predict mortality in a general population: a prospective cohort study. *Journal of Internal Medicine.* 2017;282(4):340-52.
56. Day FR, Loh PR, Scott RA, Ong KK, Perry JR. A robust example of collider bias in a genetic association study. *Am J Hum Genet.* 2016;98(2):392-3.
57. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39(7):906-13.
58. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-1.

59. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet.* 2019;28(1):166-74.
60. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45(11):1274-83.
61. Wootton RE, Richmond RC, Stuijzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med.* 2019:1-9.
62. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. *BMJ.* 2019;365:l1855.
63. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* 2018;50(11):1505-13.
64. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife.* 2018;7.
65. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002;12(6):996-1006.
66. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res.* 2012;21(3):223-42.
67. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37(7):658-65.
68. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology.* 2017;28(1):30-42.
69. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol.* 2016;40(4):304-14.
70. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet.* 2018;50(5):693-8.

Figure 1. Schematic presentation of the study. The analyses conducted within this study are highlighted with bolded box edges.

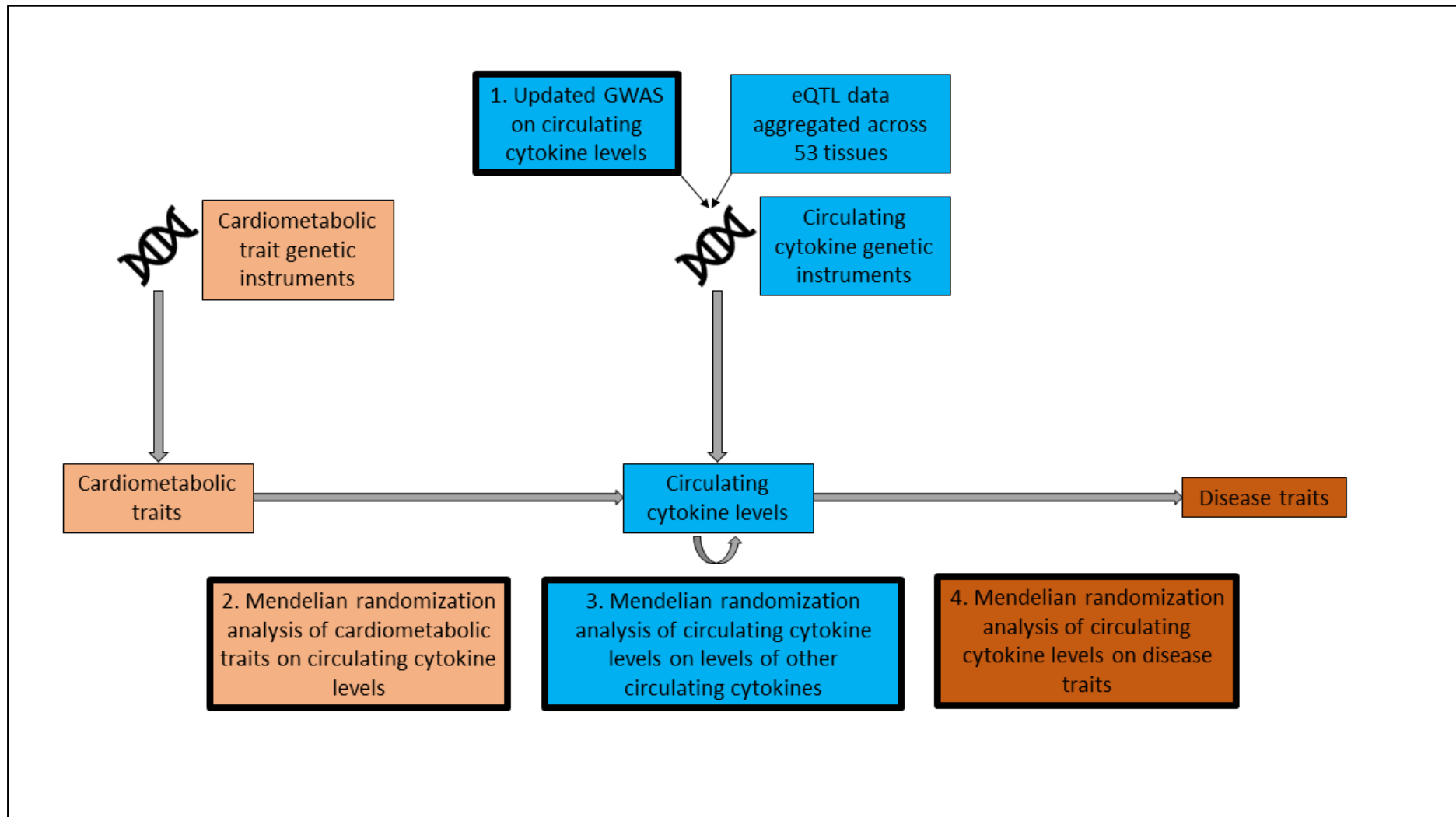


Figure 3. Mendelian randomization results of genetically predicted cytokine levels on levels of other circulating cytokines when considering *cis*- protein quantitative trait loci (solid lines) and *cis*-expression quantitative trait loci (dashed lines) instruments. The results are plotted only for effects with $P < 0.0011$ ($0.05/\text{number of cytokines}$). Red and blue lines indicate positive and negative associations respectively. The thickness of the line represents the absolute value of the effect size. The colours of the cytokines represent separate cytokine groups based on the results.

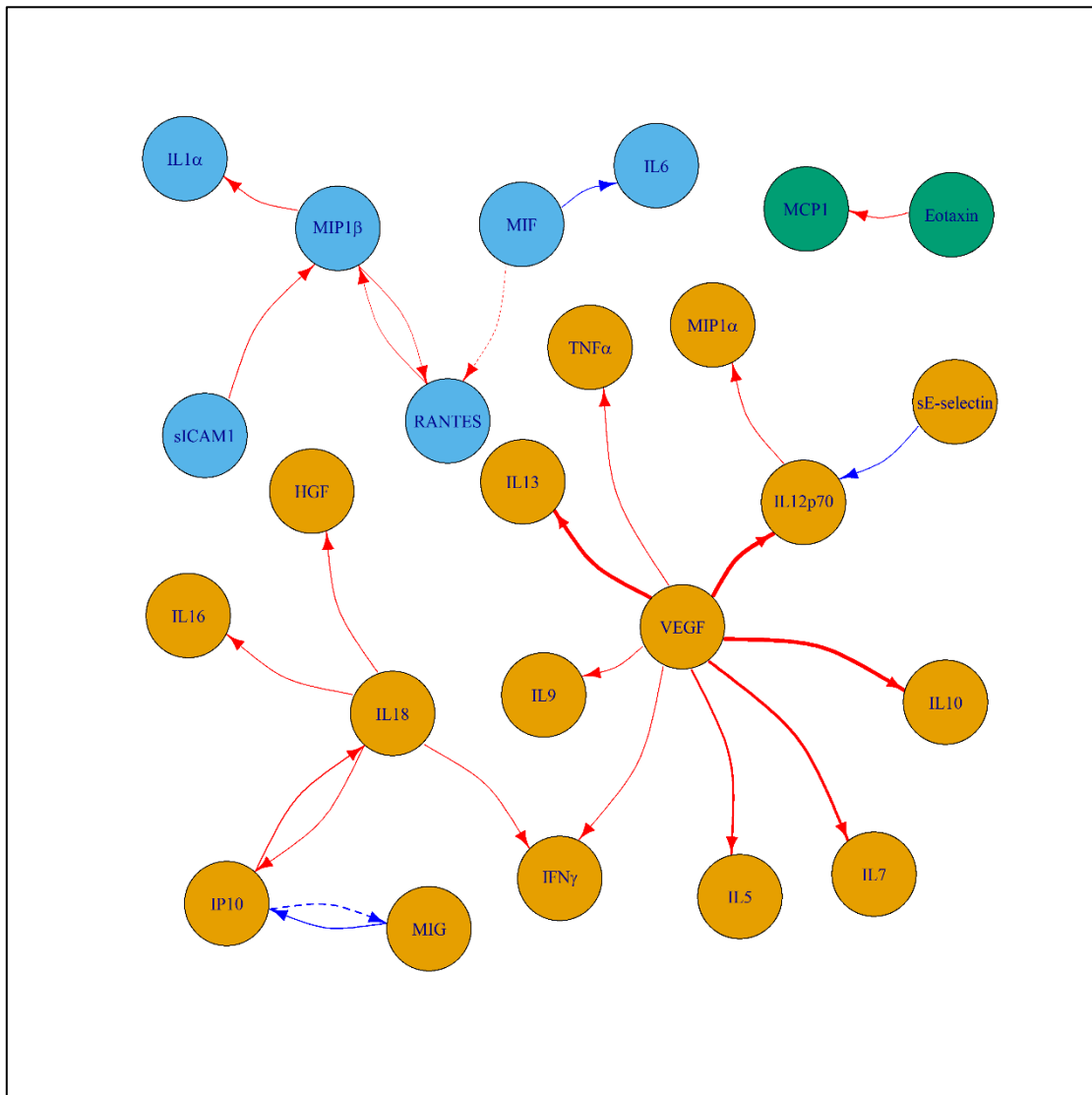


Figure 4. Colocalization for genome-wide association study summary statistics between selected cytokine pairs. Cytokine pairs that showed MR evidence for both causality ($P < 0.0011$ ($0.05/\text{number of cytokines}$)) and colocalization (posterior probability > 0.9) are plotted, within ± 500 kb of the gene locus of the exposure cytokine. The y-axis of $-\log_{10}(p)$ values is smoothed using generalized additive models.

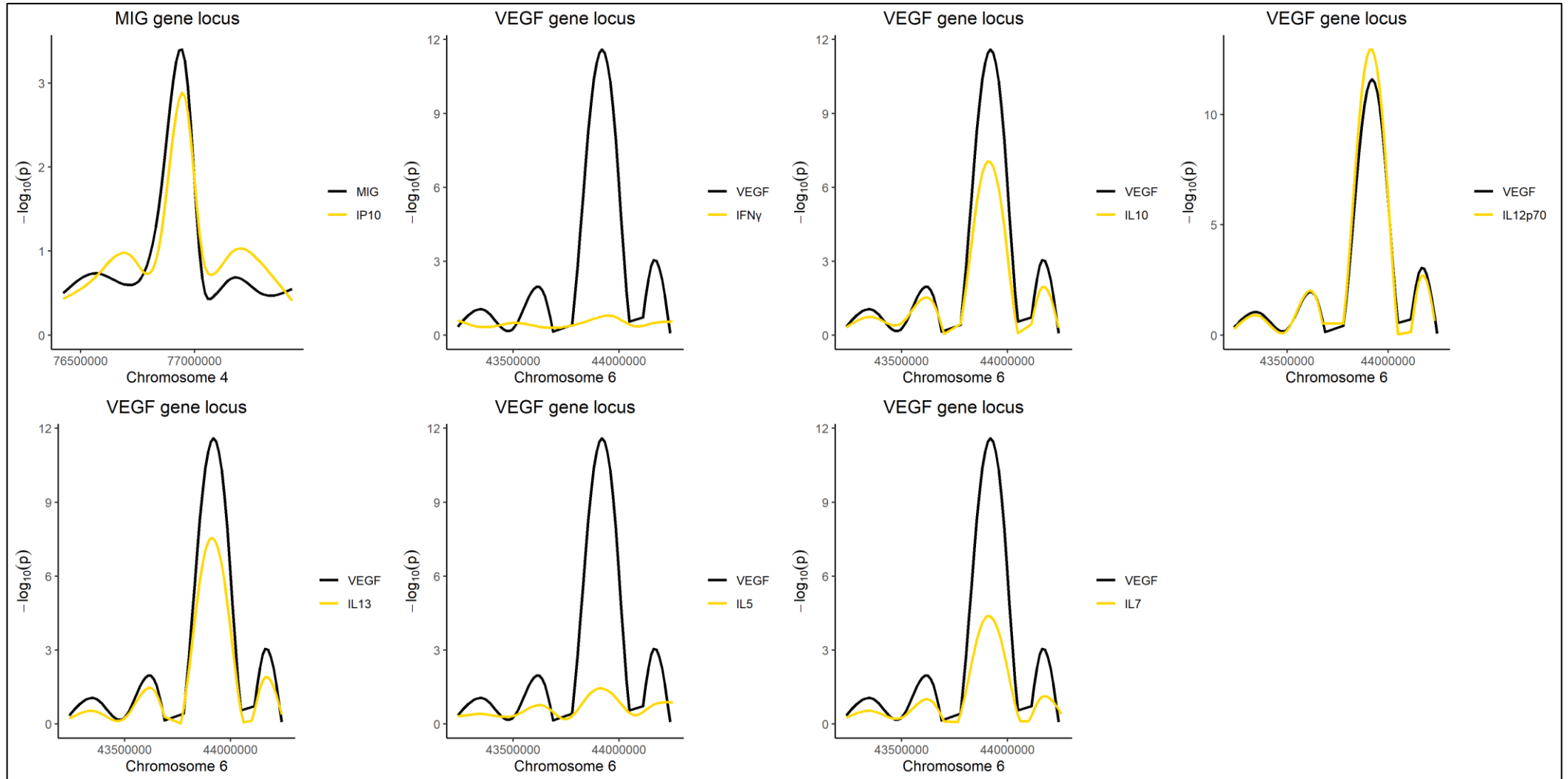


Figure 5. Mendelian randomization estimates for the effects of genetically predicted cytokine levels on disease outcomes when considering *cis*-pQTL (left – part A) and *cis*-eQTL (right – part B) instruments. After performing a Bonferroni correction for testing of multiple disease outcomes, associations with $P < 0.0025$ are denoted with an asterisk. pQTL: protein quantitative trait loci; eQTL: expression quantitative trait loci. ADHD: attention-deficit hyperactivity disorder; ASD: autism spectrum disorder; BPD: bipolar disease; MDD: major depressive disorder; Schz: schizophrenia; BMI: body-mass index; WC: waist circumference T2DM: type 2 diabetes mellitus; TC: total cholesterol; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; FI: fasting insulin; TG: triglycerides; CRP: C-reactive protein; CAD: coronary artery disease.

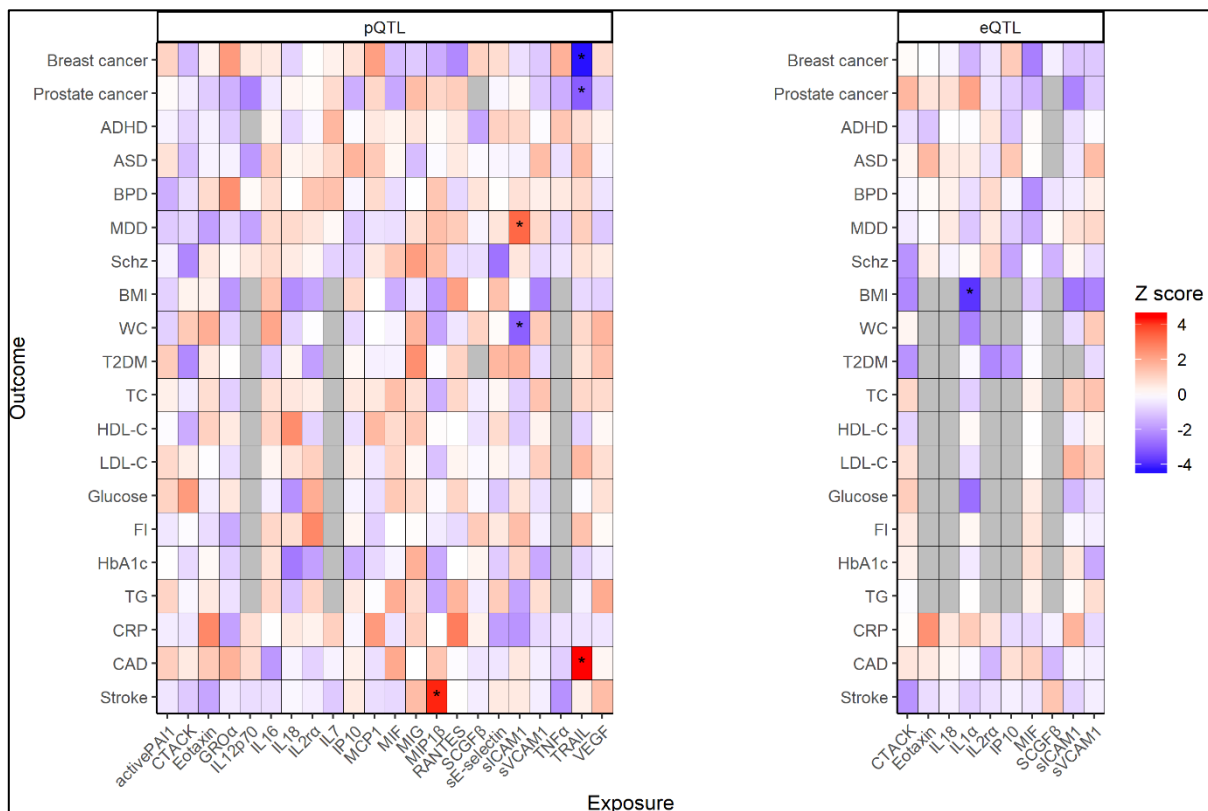


Figure 6. Scatter plot comparing Mendelian randomization (MR) estimates for the effects of genetically predicted cytokine levels on disease outcome traits when using the two different instrument selection criteria. The correlation between the MR estimates was quantified using Pearson's coefficient and a linear fit between effect estimates (blue line). x-axis: MR effect estimates using *cis*- protein quantitative trait loci instrument selection; y-axis: MR effect estimates using *cis*- expression quantitative trait loci instrument selection. *r*: correlation coefficient; *P*: p value.

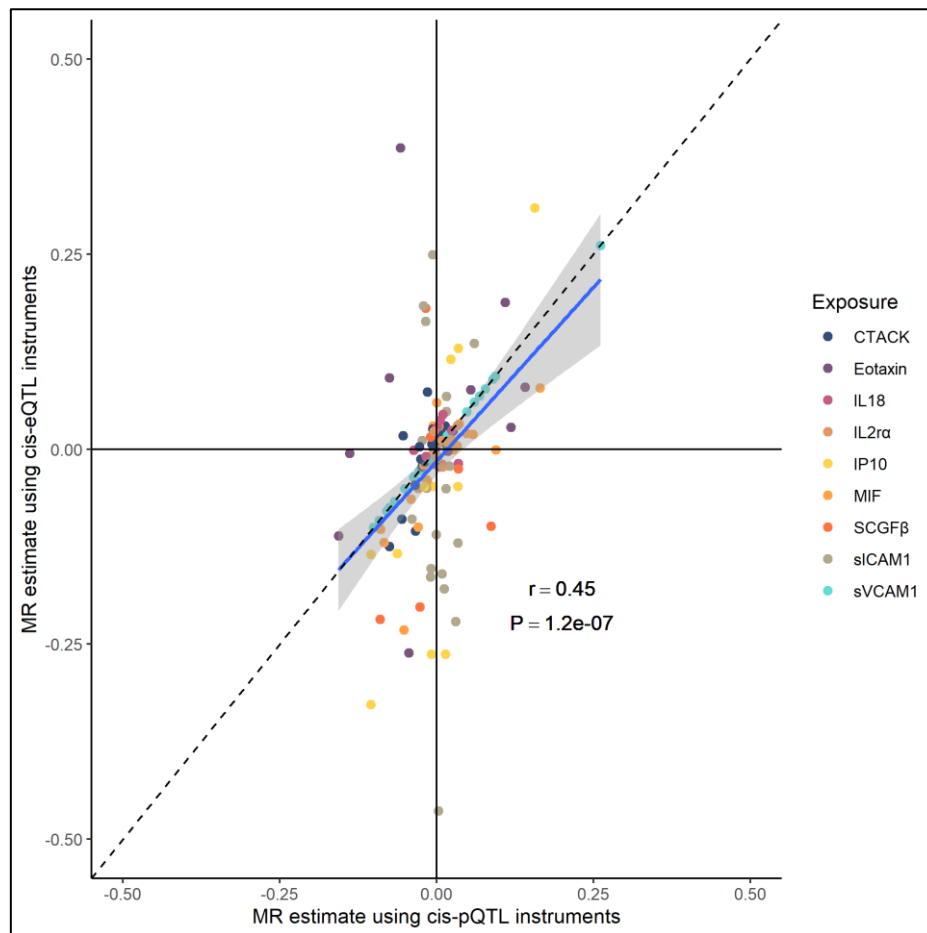


Table 1. Cytokines examined in this study. Chr: chromosome; GWAS: genome-wide association study; pQTL: protein quantitative trait loci; eQTL: expression quantitative trait loci. *Interleukin-12p70 has two coding genes.

Cytokine	Abbreviation	Gene	Chr	Start position (hg19)	End position (hg19)	GWAS sample size	Number of <i>cis</i> -pQTL instrument variants	Number of <i>cis</i> -eQTL instrument variants
Active plasminogen activator inhibitor-1	activePAI1	SERPINE1	7	100770370	100782547	5199	1	0
Beta nerve growth factor	β NGF	NGF	1	115828537	115880857	3531	0	0
Cutaneous T-cell attracting (CCL27)	CTACK	CCL27	9	34661893	34662689	3631	3	1
Eotaxin (CCL11)	Eotaxin	CCL11	17	32612687	32615199	8153	2	1
Basic fibroblast growth factor	FGFBasic	FGF2	4	123747863	123819390	7565	0	0
Granulocyte colony-stimulating factor	GCSF	CSF3	17	38171614	38174066	7904	0	0
Growth regulated oncogene-alpha (CXCL1)	GRO α	CXCL1	4	74735109	74737019	3505	5	0
Hepatocyte growth factor	HGF	HGF	7	81331444	81399452	8292	0	0
Interferon-gamma	IFN γ	IFNG	12	68548550	68553521	7701	0	0
Interleukin-10	IL10	IL10	1	206940948	206945839	7681	0	0
Interleukin-12p70	IL12p70	IL12A*	3	159706623	159713806	8270	1	0
		IL12B*	5	158741791	158757481		0	0
Interleukin-13	IL13	IL13	5	131993865	131996801	3557	0	0
Interleukin-16	IL16	IL16	15	81517640	81605104	3483	1	0
Interleukin-17	IL17	IL17A	6	52051185	52055436	12831	0	0
Interleukin-18	IL18	IL18	11	112013974	112034840	3636	6	1
Interleukin-1-alpha	IL1 α	IL1A	2	113531492	113542971	5014	0	1
Interleukin-1-beta	IL1 β	IL1B	2	113587337	113594356	8376	0	0
Interleukin-1 receptor antagonist	IL1ra	IL1RN	2	113885138	113891593	8595	0	0
Interleukin-2	IL2	IL2	4	123372626	123377650	3475	0	0

Interleukin-2 receptor, alpha subunit	IL2 α	IL2RA	10	6052657	6104333	3677	10	1
Interleukin-4	IL4	IL4	5	132009678	132018370	13183	0	0
Interleukin-5	IL5	IL5	5	131877136	131879214	3364	0	0
Interleukin-6	IL6	IL6	7	22766766	22771621	13252	0	0
Interleukin-7	IL7	IL7	8	79645007	79717758	3409	1	0
Interleukin-8 (CXCL8)	IL8	IL8	4	74606223	74609433	8597	0	0
Interleukin-9	IL9	IL9	5	135227935	135231516	3634	0	0
Interferon gamma-induced protein 10 (CXCL10)	IP10	CXCL10	4	76942269	76944689	8757	4	1
Monocyte chemotactic protein-1 (CCL2)	MCP1	CCL2	17	32582296	32584220	13365	1	0
Monocyte specific chemokine 3 (CCL7)	MCP3	CCL7	17	32597235	32599261	843	0	0
Macrophage colony-stimulating factor	MCSF	CSF1	1	110453233	110473616	840	0	0
Macrophage migration inhibitory factor (glycosylation-inhibiting factor)	MIF	MIF	22	24236565	24237409	3494	2	3
Monokine induced by interferon-gamma (CXCL9)	MIG	CXCL9	4	76922623	76928641	3685	2	0
Macrophage inflammatory protein-1a (CCL3)	MIP1 α	CCL3	17	34415603	34417506	3522	0	0
Macrophage inflammatory protein-1b (CCL4)	MIP1 β	CCL4	17	34431220	34433014	8243	22	0
Platelet derived growth factor BB	PDGFbb	PDGFB	22	39619685	39640957	8293	0	0
Regulated on Activation, Normal T Cell Expressed and Secreted (CCL5)	RANTES	CCL5	17	34198496	34207377	3421	1	0
Soluble CD40 ligand	sCD40L	CD40LG	X	135730336	135742549	5067	0	0
Stem cell factor	SCF	KITLG	12	88886570	88974250	8290	0	0
Stem cell growth factor beta	SCGF β	CLEC11A	19	51226605	51228981	3682	3	1
Stromal cell-derived factor-1 alpha (CXCL12)	SDF1 α	CXCL12	10	44872510	44880545	5998	0	0
soluble E-selectin	sE-selectin	SELE	1	169691781	169703220	5199	3	0
soluble intercellular adhesion molecule-1	sICAM1	ICAM1	19	10381517	10397291	5199	20	1
soluble vascular cell adhesion molecule-1	sVCAM1	VCAM1	1	101185196	101204601	5199	1	1
Tumor necrosis factor-alpha	TNF α	TNF	6	31543344	31546112	8522	2	0
Tumor necrosis factor-beta	TNF β	LTA	6	31539876	31542100	1559	0	0

TNF-related apoptosis inducing ligand	TRAIL	TNFSF10	3	172223298	172241297	8186	13	0
Vascular endothelial growth factor	VEGF	VEGFA	6	43737946	43754223	12155	26	0