An expanded analysis framework for multivariate GWAS connects inflammatory biomarkers to functional variants and disease

Sanni E. Ruotsalainen¹ · Juulia J. Partanen¹ · Anna Cichonska^{1,2,3} · Jake Lin¹ · Christian Benner¹ · Ida Surakka⁴ · FinnGen · Mary Pat Reeve¹ · Priit Palta^{1,5} · Marko Salmi^{6,7} · Sirpa Jalkanen^{6,7} · Ari Ahola-Olli^{1,8,9} · Aarno Palotie^{1,8,9} · Veikko Salomaa¹⁰ · Mark J. Daly^{1,8,9} · Matti Pirinen^{1,11,12} · Samuli Ripatti^{1,8,11} · Jukka Koskela^{1,8}

Received: 7 December 2019 / Revised: 2 August 2020 / Accepted: 4 September 2020 / Published online: 27 October 2020
© The Author(s), under exclusive licence to European Society of Human Genetics 2020

Abstract

Multivariate methods are known to increase the statistical power to detect associations in the case of shared genetic basis between phenotypes. They have, however, lacked essential analytic tools to follow-up and understand the biology underlying these associations. We developed a novel computational workflow for multivariate GWAS follow-up analyses, including fine-mapping and identification of the subset of traits driving associations (driver traits). Many follow-up tools require univariate regression coefficients which are lacking from multivariate results. Our method overcomes this problem by using Canonical Correlation Analysis to turn each multivariate association into its optimal univariate Linear Combination Phenotype (LCP). This enables an LCP-GWAS, which in turn generates the statistics required for follow-up analyses. We implemented our method on 12 highly correlated inflammatory biomarkers in a Finnish population-based study. Altogether, we identified 11 associations, four of which (*F5*, *ABO*, *C1orf140* and *PDGFRB*) were not detected by biomarker-specific analyses. Fine-mapping identified 19 signals within the 11 loci and driver trait analysis determined the traits contributing to the associations. A phenome-wide association study on the 19 representative variants from the signals in 176,899 individuals from the FinnGen study revealed 53 disease associations ($p < 1 \times 10^{-4}$). Several reported pQTLs in the 11 loci provided orthogonal evidence for the biologically relevant functions of the representative variants. Our novel multivariate analysis workflow provides a powerful addition to standard univariate GWAS analyses by enabling multivariate GWAS follow-up and thus promoting the advancement of powerful multivariate methods in genomics.

Introduction

Genome-wide association studies (GWAS) of biomarkers have been highly successful in identifying novel biological pathways and their impact on health and disease. Biomarkers

increase statistical power in GWAS, compared to disease diagnoses, due to their quantitative nature and lack of errors due to subjectivity, such as misclassification. Thus, biomarker GWAS have identified thousands of biomarker-associated loci and elucidated the mechanisms underlying numerous disease associations [1–3]. A recent study on 38 biomarkers in the UK Biobank (UKBB) identified over 1,800 independent genetic associations with causal roles in several diseases [4]. Proteomics and metabolomics integrated with genomics has also revealed causal molecular pathways connecting the genome to multiple diseases, e.g., autoimmune disorders and cardiovascular disease [5–8]. Although biomarkers are more closely related to pathophysiology, a single biomarker is usually an inaccurate estimator of complex disease due to phenotypic heterogeneity and individual variation. Therefore, combinations of biomarkers provide a more robust predictive molecular signature. Studies examining combinations of biomarkers are increasingly feasible given the availability of

These authors contributed equally: Sanni E. Ruotsalainen, Juulia J. Partanen

Members of FinnGen are listed below Acknowledgements.

Supplementary information The online version of this article (<https://doi.org/10.1038/s41431-020-00730-8>) contains supplementary material, which is available to authorized users.

✉ Samuli Ripatti
samuli.ripatti@helsinki.fi

✉ Jukka Koskela
jukka.t.koskela@helsinki.fi

Extended author information available on the last page of the article

Fig. 1 Study workflow. The novel LCP-GWAS method that enables follow-up analyses such as fine-mapping for multivariate GWAS is illustrated in the violet panel on the right.

biobank resources around the globe with deep phenotyping, i.e., precise and comprehensive data on phenotypic variation including quantitative measures such as biomarkers [9, 10].

Multivariate GWAS increases statistical power compared to univariate analysis, especially in the case of complex biological processes and correlated traits [8, 11, 12]. This leads to identifying multivariate associations that are otherwise missed by univariate analysis [8, 13]. Efficient software programs are available for performing multivariate GWAS such as metaCCA [14], yet multivariate analyses currently have shortcomings in interpreting the arising signals. Follow-up tools for fine-mapping causal variants within the associated loci are lacking and the subset of tested traits that drive the association signals have not been identified. These shortcomings are largely due to the lack of a multivariate counterpart to the univariate regression coefficients (beta estimates). Lack of these necessary follow-up tools has hindered the utilization of multivariate methods.

In this study, we developed a novel computational workflow for multivariate GWAS discovery and follow-up analyses including fine-mapping and identification of driver traits (Fig. 1). Our workflow includes (1) a customized version of the metaCCA software that overcomes the problem of missing beta estimates by turning each multivariate association into its optimal univariate Linear Combination Phenotype (LCP), enabling an LCP-GWAS, (2) fine-mapping, i.e., identifying putative causal variants underlying each association using

summary statistics from the LCP-GWAS and a multivariate extension to FINEMAP [15], and (3) determining the traits driving each multivariate association using a newly developed tool, MetaPhat [16] that efficiently decomposes the multivariate associations into a smaller set of underlying driver traits. Taken together, we present to our knowledge the first comprehensive framework to map multivariate associations into individual causal variants and a subset of driver traits. We demonstrate the potential of our workflow in a Finnish population-based cohort with 12 inflammatory biomarkers implicated in the pathogenesis of autoimmune disorders and cancer [17–19]. This set of highly correlated biomarkers is particularly advantageous for multivariate analysis as high correlation between traits increases the boost in statistical power achieved by multivariate methods. Using multivariate analysis, we identify additional hits compared to univariate analysis, totaling 11 independent associations. We follow them up in a phenome-wide association study (PheWAS) in the FinnGen study ($n = 176,899$) across 2367 disease endpoints and in the UKBB ($n = 408,910$) [10]. We discover multiple disease associations, as well as identify orthogonal evidence for the biological impact of the causal variants through several protein quantitative trait loci (pQTLs) within the multivariate loci.

Materials and methods

Study cohort and data

We studied 12 highly correlated inflammatory biomarkers in the population-based national FINRISK Study [20] collected in 1997 ($n = 6890$) (Table 1, Supplementary Fig. 1). The FINRISK Study is a large Finnish population survey of risk factors for chronic, non-communicable diseases, and it has been collected by independent random population sampling every five years beginning in 1972 with multiple recruiting waves. The 12 inflammatory biomarkers included five interleukins (IL-4, IL-6, IL-10, IL-12p70, IL-17), three growth factors (FGF2, PDGF-BB, VEGF-A), one colony-stimulating factor (G-CSF), one interferon (IFN- γ), one chemokine (SDF-1 α), and one tumor necrosis factor (TNF- β) (Table 1, Supplementary Fig. 1). Hierarchical clustering identified the cluster of 12 inflammatory biomarkers out of 66 quantitative traits of cardiometabolic or immunologic relevance (Supplementary Fig. 2, Supplementary Table 1, and Supplementary Methods). The 66 quantitative traits were measured as previously described [1, 20, 21].

Genotyping, imputation and quality control

Samples were genotyped using multiple different genotyping chips (Supplementary Table 2), for which pre-

Table 1 Characterization of the 12 inflammatory biomarker measurements.

Characteristic	<i>n</i>
Sample size	6,890
male/female	3370/3520 (49% male)
Mean ± SD	
Age (year)	48 ± 13.3
Basic fibroblast growth factor (FGF2)	26.7 ± 49.5
Granulocyte colony stimulating factor (G-CSF)	141.2 ± 173.9
Interferon gamma (IFN-γ)	89.3 ± 160.8
Interleukin-4 (IL-4)	4.2 ± 5.4
Interleukin-6 (IL-6)	11.8 ± 29.7
Interleukin-10 (IL-10)	2.3 ± 7.4
Interleukin-12 p70 (IL-12p70)	26.4 ± 114.1
Interleukin-17 (IL-17)	65.2 ± 72.0
Platelet derived growth factor BB (PDGF-BB)	1199.0 ± 1152.8
Stromal cell derived factor 1 alpha, CXCL12 (SDF-1α)	123.4 ± 176.2
Tumor necrosis factor beta (TNF-β)	26.6 ± 66.9
Vascular endothelial growth factor A (VEGF-A)	16.6 ± 27.5

The cytokine concentrations are pg/ml.

n sample size, *SD* standard deviation.

imputation quality control (QC), phasing and imputation were done in multiple chip-wise batches (Supplementary Methods). Imputation of the genotypes was done utilizing a Finnish population-specific reference panel of 3775 high-coverage whole-genome sequences. Genotype imputation was followed by an additional post-imputation sample QC (Supplementary Methods) and variant QC (imputation INFO > 0.8, minor allele frequency > 0.002 and Hardy–Weinberg equilibrium *p* value > 1×10^{-6}). A total of 26,717 samples and 11,329,225 variants passed this rigorous quality control. All variants are reported based on the human genome reference sequence GRCh38.

Univariate and multivariate GWAS

Univariate genome-wide association analyses for the biomarkers were performed using a linear mixed model implemented in Hail [22], adjusting for age, sex, genotyping chip, first ten principal components of genetic structure and the genetic relationship matrix (GRM) (Supplementary Methods). The GRM was estimated using 73K independent high-quality genotyped variants (Supplementary Methods). We performed multivariate GWAS on the biomarkers using metaCCA [14], software that performs multivariate analysis by implementing Canonical Correlation Analysis (CCA) for a set of univariate GWAS summary statistics.

The objective of CCA is to find the linear combination of the *p* predictor variables ($X_1, X_2, \dots, X_p$) that is maximally correlated with a linear combination of the *q* response variables ($Y_1, Y_2, \dots, Y_q$). If we denote the respective linear

combinations by

$$X^* = \mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \dots + a_px_p$$

and

$$Y^* = LCP = \mathbf{b}'\mathbf{y} = b_1y_1 + b_2y_2 + \dots + b_qy_q$$

then finding the linear combination of the predictor variables that are maximally correlated with the linear combination of the response variables corresponds to finding vectors *a* and *b* that maximize

$$r = \frac{(Xa)'(Yb)}{\|Xa\|\|Yb\|} = \frac{\mathbf{a}'\sum_{xy}\mathbf{b}}{\sqrt{\mathbf{a}'\sum_{xx}\mathbf{a}}\sqrt{\mathbf{b}'\sum_{yy}\mathbf{b}}}$$

where $\sum_{xx}$, $\sum_{yy}$ and $\sum_{xy}$ represent the variance-covariance matrices of the predictor variables, response variables and both of them together, respectively. The maximized correlation *r* is the *canonical correlation* between *X* and *Y*. Multivariate GWAS is a special case of CCA with multiple response variables *Y*, but only one explanatory variable *X*, the genotypes at the variant tested.

Novel multivariate LCP-GWAS method

To enable follow-up analyses of multivariate GWAS results, such as fine-mapping, we developed a novel method to produce linear combination phenotypes (LCP) at the single variant level by extending the functionality of metaCCA. The updated metaCCA is available online at: <https://github.com/acichonska/metaCCA>.

LCPs were constructed as the weighted sum of the trait residuals, where the weights ($\mathbf{b} = [b_1, b_2, \dots, b_q]$) were chosen to maximize the correlation between the resulting linear combination of traits and the genotypes at the variant. We determined association regions by adding 1 Mb to each variant reaching genome-wide significance (GWS; *p* value < 5×10^{-8}) in the multivariate analysis and joining overlapping regions. We constructed LCPs for the lead variant, i.e., the variant with the smallest *p* value, in each of these regions, as a univariate representation of the multivariate association in that region. Next, we performed chromosome-wide LCP-GWAS for the constructed LCPs in a similar manner as for each of the biomarkers.

Fine-mapping multivariate associations

We used FINEMAP [15, 23] on the LCP-GWAS summary statistics to identify causal variants underlying the multivariate associations. FINEMAP analyses were restricted to a ± 1 Mb region around the GWS variants from the LCP-GWAS.

We assessed variants in the top 95% credible sets, i.e., the sets of variants encompassing at least 95% of the

probability of being causal (causal probability) within each causal signal conditional on other causal signals in the genomic region. Within these sets we excluded those sets that did not clearly represent one signal, determined by low minimum linkage disequilibrium (LD, $r^2 < 0.1$). Among each of the credible sets, the variant with the highest causal probability was chosen to represent the set as the representative variant.

To validate the multivariate fine-mapping results, we also performed conventional stepwise conditional analysis for all fine-mapping regions using LCPs. We iteratively conditioned on the lead variant in the region until the smallest p value in the region exceeded 5×10^{-8} .

Identifying driver traits

We determined the traits driving the multivariate associations for the representative variants of the credible sets identified by fine-mapping using the MetaPhat software developed in-house [16]. MetaPhat determines the set of driver traits for each multivariate association by performing multivariate testing using metaCCA iteratively on subsets of the traits, excluding one trait at a time until a single trait remains. At each iteration, the trait to be excluded is the one whose exclusion leads to the highest p value for the remaining subset of traits. The driver traits are determined as a set of traits that have been removed when the multivariate p value becomes non-significant ($p > 5 \times 10^{-8}$). The interpretation is that the driver traits make the multivariate association significant.

Phenome-wide association testing in FinnGen and UKBB

We performed a PheWAS in the FinnGen study for the representative variants of the credible sets identified by multivariate fine-mapping. FinnGen (<https://www.finnngen.fi/en>) is a large biobank study that aims to genotype 500,000 Finns and combine this data with longitudinal registry data, including national hospital discharge, death, and medication reimbursement registries, using unique national personal identification numbers. FinnGen includes prospective epidemiological and disease-based cohorts as well as hospital biobank samples. A total of 176,899 samples from FinnGen Data Freeze 4 with 2444 disease endpoints were analyzed using Scalable and Accurate Implementation of Generalized mixed model (SAIGE), which uses saddlepoint approximation (SPA) to calibrate unbalanced case-control ratios [24]. Additional details and information on genotyping and imputation are provided in the Supplementary Material and contributors of FinnGen are listed in the Acknowledgements.

FinnGen disease associations with p values $< 1 \times 10^{-4}$ were considered significant. We tested the p value threshold by sampling 1000 allele frequency-matched sets of n variants, where n represents the number of representative variants, from 8.2 million non-coding variants and determining a null distribution of the number of FinnGen associations passing the p value threshold. We confirmed the validity of the p value threshold by comparing the observed number of FinnGen associations passing the p value threshold to the null distribution (Supplementary Fig. 3). We excluded disease endpoints within the ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th Revision) chapters XXI and XXII from PheWAS analyses, resulting in 2367 disease endpoints analyzed. To confirm whether the FinnGen disease associations of the representative variants share a common causal variant with the most significantly associated variant (i.e., variant with smallest p value in FinnGen) within the locus, and thus evaluate their importance for the disease associations, the FinnGen disease associations were conditioned on the most significantly associated variant within the locus (± 0.5 MB of the representative variant). Finally, we assessed replication of the disease associations in the UKBB, where associations with p values < 0.05 were considered replicated given that the direction of effects were coherent. Phecodes from the UKBB were mapped to ICD-10 diagnosis codes using the PheCode map 1.2 [25]. The NHGRI-EBI GWAS Catalog [26] was used for assessing the novelty of the observed genetic associations.

We also explored whether the fine-mapped representative variants or variants in LD with them ($r^2 > 0.6$) had previously been reported as pQTLs in studies by Suhre et al. [5], Sun et al. [6], Emilsson et al. [27] and Sasayama [28]. Regional overlap and architecture were visualized in Target Gene Notebook [29]. To validate the overlap of our pQTL findings, we performed Bayesian colocalization analysis using the COLOC package in R [30], within 200 kb from the representative variant, for all pQTL associations from data sets with full summary statistics available.

Results

Comparison of multivariate and univariate GWAS of 12 inflammatory biomarkers

We first tested for genome-wide associations of 12 highly correlated inflammatory biomarkers (Table 1, Supplementary Fig. 1) measured in 6890 FINRISK study participants using both multivariate and univariate methods. Pearson correlations between the biomarkers ranged from 0.64 to 0.93, with a mean of 0.80. Out of the 11,329,225 variants tested, 190 were significantly associated using both univariate and multivariate

Fig. 2 Power comparison between multivariate and univariate methods. Red and blue dots represent genetic variants reaching genome-wide significance only by the multivariate (metaCCA) or univariate method, respectively. Black dots reach the genome-wide significance threshold by both methods and gray dots do not by either method. Respective numbers are reported in the accompanying table.

analyses, 999 only by the multivariate analysis and two only by the univariate analysis using a Bonferroni-corrected p value threshold of $5 \times 10^{-8}/12$ (Fig. 2). A total of 1189 variants reached the significance threshold in the multivariate analysis compared to only 192 in the univariate analysis, reflecting a considerable increase in statistical power achieved by the multivariate analysis. When the univariate effect sizes were all in the same direction (e.g., *GP6* locus, all effects were positive), the gain in power was smaller compared to the situation where the effects were both positive and negative (e.g., *F5* locus). This is as expected, as all the 12 traits were positively correlated, and it is known that the gain in power in multivariate analyses is greatest when the correlation matrix and effect sizes differ from each other [31]. Despite the increase in power, the Type I error rate of the multivariate GWAS was preserved as the corresponding genomic inflation factor λ for all variants was 1.036, with no evidence of concerning genomic inflation due to Canonical Correlation Analysis. We also assessed the Type I error rate for three minor allele frequency (maf) bins ($\text{maf} < 0.01$, $0.01 < \text{maf} < 0.1$, and $\text{maf} > 0.1$) separately, with rare variants not showing noticeably more inflation than more common variants (Supplementary Fig. 4).

Within the 1189 genome-wide significant variants in the multivariate analysis, we identified 11 independently associated loci (Fig. 3 and Supplementary Fig. 5), four of which (*F5*, *C1orf140*, *PDGFRB* and *ABO*) were not detected by univariate analyses corrected for multiple testing (Fig. 3). The two variants that were significant only in the univariate analysis were both located in a locus (*JMJD1C*) that was found to be significant also by the multivariate analysis. Thus, no loci that were significant in the univariate analysis corrected for multiple testing went undetected by multivariate analysis. Eight of the 11 loci had previously been associated with at least one of the 12 biomarkers in the NHGRI-EBI GWAS catalog while three loci (*F5*, *C1orf140* and *PDGFRB*) were novel.

Comparing the multivariate and univariate lead variants in three loci significant in only one of the 12 univariate analyses (*CIQA*, *PCSK6*, and *VLDLR*), we noted that the multivariate and univariate lead variants were never the same. In the *CIQA* and *PCSK6* loci the lead variants from both analyses were in high LD (r^2 0.92 and 0.93, respectively), reflecting that the two methods were capturing the same association signal, while in the *VLDLR* locus LD between the lead variants was low ($r^2 = 0.27$). In the *CIQA* locus, an association with only $\text{TNF-}\beta$ of the 12 biomarkers was noted in the univariate results. The lead variant in the $\text{TNF-}\beta$ univariate GWAS was chr1:g.22720394C>T (rs78655189, $p = 2.2 \times 10^{-24}$), an intronic variant in the *EPHB2* gene. In contrast, the lead variant for the same locus in the multivariate analysis was chr1:g.22637683G>A (rs17887074, $p = 1.2 \times 10^{-73}$), a Finnish-enriched missense variant located in the *CIQA* gene. In the *PCSK6* locus both lead variants were intronic with similar multivariate p values (multivariate lead variant chr15:g.101451543G>T (rs11637184, $p = 2.4 \times 10^{-68}$), univariate PDGF-BB lead variant chr15:g.101446695T>A (rs11634270, $p = 1.3 \times 10^{-67}$)). In the *VLDLR* locus, where LD between the two lead variants was low, univariate fine-mapping of VEGF-A, the only associated biomarker, suggested that the common lead variant chr9:g.2692583C>G (rs2375981, allele frequency, AF = 47%) from the multivariate analysis was more likely causal than the lead variant chr9:g.2694711G>A (rs10967570, AF = 19%) from the VEGF-A univariate analysis (posterior probabilities 1.0 and 0.025, respectively).

Functional coding variants

GWAS hits are generally non-coding, although concentrated in regulatory regions [32], and enrichment of functional coding variants has been seen mainly only after fine-mapping e.g., in inflammatory bowel disease [33]. We, however, observed enrichment of functional coding variants in the multivariate GWAS hits already prior to fine-

Fig. 3 Manhattan plot of the multivariate GWAS results on 12 inflammatory biomarkers. Gene names colored in orange represent associations only detected by the multivariate method while black are

detected by both multivariate and univariate methods. 13 genome-wide significant nonsynonymous and splice-region variants are denoted with diamonds.

mapping. Considering all genome-wide significant variants in the multivariate GWAS, we found 13 nonsynonymous or splice-region variants with at least one such variant in five of the 11 multivariate loci (*C1QA*, *F5*, *SERPINE2*, *C6orf223*, and *GP6*) (Fig. 3). Out of the 13 variants, 11 were missense variants, one was a splice-region variant and one a frameshift variant. Only four missense variants at two loci were significantly associated in the univariate analyses. Two of the 11 missense variants led the multivariate association at their respective loci (chr1:g.22637683G>A (rs17887074) and chr19:g.55032292G>A (rs199588110), in the *C1QA* and *GP6* loci respectively) and were enriched (>1.5-fold) in Finns compared to non-Finnish, Swedish, Estonian Europeans (NFSEE) in the gnomAD genome reference database [34]. A total of six (46.2%) of the 13 variants were enriched in the Finnish population, highlighting the potential of utilizing isolated populations in GWAS.

We studied whether the multivariate genome-wide significant variants were enriched for missense, splice-region and frameshift variants compared to the 11.3 M variants analyzed. *P* values for enrichment were calculated using the χ^2 -test for the number of nonsynonymous and splice-region or missense variants within the genome-wide significant variants against the number of the corresponding subset of variants within all variants tested. The multivariate genome-wide significant variants were enriched for missense variants and missense, splice-region and frameshift variants (2.2-fold, $p = 0.015$, and 1.9-fold, $p = 8.8 \times 10^{-4}$, respectively).

Fine-mapping multivariate GWAS results

To identify the causal variants of the multivariate associations, we studied the likelihood of multiple variants contributing to the association signal in the 11 associated loci using FINEMAP [23]. Our novel multivariate LCP-GWAS method based on linear combinations calculated for each locus using multivariate metaCCA results enabled fine-mapping of the multivariate results. The number of credible sets varied from one to four for the multivariate associated loci (Supplementary Table 3), resulting in a total of 19 independent sets of variants considered putatively causal. All 183 variants within the 19 credible sets are available in Supplementary Table 3 and posterior probabilities for different numbers of causal signals for each locus are available in Supplementary Table 4.

Among each of the 19 sets, the variant with the highest causal probability was chosen to represent the set as the representative variant (Table 2 and Supplementary Fig. 6). The 19 representative variants, included all except one (chr15g.101991748G>C, rs11637184 in the *PCSK6* locus) of the 11 lead variants from multivariate GWAS. Highlighting the importance of fine-mapping multivariate GWAS results, one of the four representative variants (chr15:g.101339772G>A, rs111482836) in the *PCSK6* locus was associated with disease in FinnGen, whereas the lead variant was not. Additionally, the 19 representative variants were further enriched for both missense variants and missense, splice-region and frameshift variants (37-

Table 2 Results of the 19 representative variants of the credible sets.

Variant ^a	Locus	AF ^b (FIN enrichment)	Multivariate <i>p</i> value	Minimum univariate <i>p</i> value (biomarker)	Driver traits ^c	Previous biomarker associations ^d	Fine-mapping posterior probability ^e	pQTL gene	pQTL <i>p</i> value	FinnGen disease associations ^f	FinnGen association statistics		Novel disease association ^g
											OR	<i>p</i> value	
chr1: g-22637683G>A (rs17887074)^h	<i>C1QA</i>	1.48% (4.64)	1.21E-73	1.70E-23 (TNF-β)	TNF-β	IFN-γ, IL-17, TNF-β	0.975	-	-	-	-	-	-
chr1: g-169515314T>G (rs3820060)ⁱ	<i>F5</i>	29.60%	6.15E-20	1.07E-3 (VEGF-A)	IL-4, IL-12p70	NOVEL	0.175	-	-	-	-	-	-
chr1: g-169505159C>T (rs61808983)^j	<i>F5</i>	4.01%	3.49E-06	4.03E-2 (VEGF-A)	-	NOVEL	0.533	EHBPI	7.1E-18	-	-	-	-
chr1: g-221344914A>G (rs151049317)	<i>C1orf140</i>	0.98%	1.79E-08	2.79E-2 (PDGF-BB)	PDGF-BB	NOVEL	1.000	-	-	-	-	-	-
chr2: g-224010157G>A (rs13412535)	<i>SERPINE2</i>	19.8%	< 1E-324	1.60E-37 (IL-10)	All 12 biomarkers	FGF2, IL-6, IL-10, IL-12p70, PDGF-BB	0.998	PDGF-BB	4.6E-13	Hypertrophic scar	1.34	7.5E-5	Yes
chr2: g-224036001del (rs58116674)	<i>SERPINE2</i>	71.5%	3.37E-78	6.18E-13 (IL-6)	PDGF-BB, SDF-1α, IL-4, IL-17, IL-6, IL-10, FGF2, TNF-β	FGF2, IL-6, IL-10, IL-12p70, PDGF-BB	0.067	-	-	-	-	-	-
chr2: g-224257750T>A (rs7578029)	<i>SERPINE2</i>	8.46%	1.02E-08	3.28E-4 (PDGF-BB)	PDGF-BB	FGF2, IL-6, IL-10, IL-12p70, PDGF-BB	0.046	-	-	Infections of the skin and subcutaneous tissue	1.12	9.7E-5	Yes
chr5: g-150128981C>G (rs2304058)	<i>PDGFRB</i>	37.7%	1.08E-15	2.46E-8 (SDF-1α)	SDF-1α	NOVEL	0.546	PDGFRB	2.3E-458	-	-	-	-
chr6: g-43957870G>A (rs6921438)	<i>C6orf223 / VEGFA</i>	48.3%	3.03E-296	3.38E-91 (VEGF-A)	VEGF-A, IL-12p70, IL-10	IL-10, IL-12p70, VEGF-A	1.000	VEGFA	7.8E-71	-	-	-	-
chr6: g-43974045A>G (rs4714726)	<i>C6orf223 / VEGFA</i>	45.5%	1.12E-11	6.95E-5 (VEGF-A)	VEGF-A	IL-10, IL-12p70, VEGF-A	0.055	-	-	-	-	-	-
chr9: g-2692583C>G (rs2375981)	<i>VLDLR / KCNV2</i>	48.0%	2.03E-17	1.29E-8 (VEGF-A)	VEGF-A, IL-12p70	IFN-γ, IL-10, IL-12p70, VEGF-A	0.748	-	-	-	-	-	-
chr9: g-2782458A>G (rs10122155)	<i>VLDLR / KCNV2</i>	43.3%	1.08E-04	5.43E-3 (VEGF-A)	-	IFN-γ, IL-10, IL-12p70, VEGF-A	0.162	-	-	-	-	-	-
chr9: g-133271182T>C (rs550057)	<i>ABO</i>	31.0%	8.49E-14	2.08E-5 (IL-4)	IL-4	FGF2	0.578	ALPI, CHST15, FAM177A1, JAG1	2.8E-19, 1.0E-30, 9.3E-19, 8.3E-14	Anemias, Other and unspecified anaemias, Other anaemias, Diseases of the blood and blood-forming organs, Visual field defects, Diseases of the eye and adnexa, Diseases of the ear and mastoid process	1.12, 1.10, 1.11, 1.06	4.7E-8, 4.9E-5, 2.6E-5, 2.9E-5	No
											1.24, 1.04	4.4E-5, 9.4E-6	Yes
											1.04	4.8E-5	Yes

Table 2 (continued)

Variant ^a	Locus	AF ^b (FIN enrichment)	Multivariate <i>p</i> value	Minimum univariate <i>p</i> value (biomarker)	Driver traits ^c	Previous biomarker associations ^d	Fine-mapping posterior probability ^e	pQTL gene	pQTL <i>p</i> value	FinnGen disease associations ^f	FinnGen association statistics		Novel disease association ^g
											OR	<i>p</i> value	
chr10: g.6328546C>A (rs7080386)	<i>JMJD1C</i>	38.1%	4.04E-08	1.86E-11 (VEGF-A)	VEGF-A	IFN- γ , IL-10, IL-12p70, VEGF-A	0.064	HB-EGF	1.60E-13	-	-	-	-
chr15: g.10133972G>A (rs111482836)	<i>PCSK6</i>	29.0%	3.42E-05	0.010 (PDGF- BB)	-	PDGF-BB	0.537	-	-	-	-	-	-
chr15: g.101345122G>A (rs12905972)	<i>PCSK6</i>	21.8%	0.035	0.027 (VEGF- A)	-	PDGF-BB	0.127	-	-	-	-	-	-
chr15: g.101446425G>T (rs6598475)	<i>PCSK6</i>	65.7%	1.27E-54	2.63E-8 (PDGF-BB)	PDGF-BB, SDF-1 α , IL- 4, IL-17	PDGF-BB	0.704	-	-	-	-	-	-
chr15: g.101452793C>A (rs11639051)	<i>PCSK6</i>	24.3%	2.71E-69	1.11E-13 (PDGF-BB)	PDGF-BB, SDF-1 α , IL- 4, IL-10	PDGF-BB	0.321	-	-	-	-	-	-
chr19: g.55032292G>A (rs199588110)^h	<i>GFP6</i>	0.33% (3.69)	8.54E-14	1.25E-17 (IL- 17)	All 12 biomarkers	G-CSF	1.000	-	-	Benign neoplasm of meninges	6.4	4.9E-5	Yes

^aBolded variants are lead variants. Genomic positions are mapped to Genome Assembly GRCh38.^bAF allele frequency, *FIN enrichment* AF in Finns compared to AF in non-Finnish, Swedish, Estonian Europeans (NFSEE) in the gnomAD genomes database; reported if it was at least 1.5-fold.^cDriver traits can only be determined for those variants with a genome-wide significant association in the multivariate analysis.^dPrevious associations with the 12 biomarkers were searched for in the NHGRI-EBI GWAS Catalog within a region encompassing ± 500 kb around the variant. An association was regarded novel if no associations with any of the 12 biomarkers had been reported in this region.^ePosterior probability of being causal, determined by FINEMAP.^fOnly associations that remain significant after conditioning are reported here. Closely related disease diagnoses are represented in a shared cell.^gNovelty of disease associations was assessed at gene-level.^hMissense variant.ⁱVariant was in high linkage disequilibrium ($r^2 = 0.997$) with a missense variant (chr1:g.169529737T>C, rs6030)^jVariant was in high linkage disequilibrium ($r^2 = 0.996$) with a missense variant within its credible set (chr1:g.169515529A>G, rs9332701) predicted deleterious by SIFT and probably damaging by PolyPhen.

fold, $p = 1.3 \times 10^{-17}$, and 28-fold, $p = 1.4 \times 10^{-17}$, respectively) compared to multivariate genome-wide significant variants (2.2-fold, $p = 0.015$, and 1.9-fold, $p = 8.8 \times 10^{-4}$, respectively), as were the 183 variants in the credible sets (3.9-fold, $p = 0.050$, and 2.9-fold, $p = 0.050$, respectively). In one of the two credible sets in the *F5* locus a missense variant (chr1:g.169515529A>G, rs9332701), predicted deleterious by SIFT and probably damaging by PolyPhen, was found to be in high LD ($r^2 = 0.996$) with the representative non-coding variant chr1:g.169505159C>T (rs61808983) with a marginally smaller causal probability (46.1% vs. 53.3%). We assessed whether the causal probabilities changed in the credible set if the LCP was generated for the missense variant rs9332701 rather than the lead variant rs61808983. This had no notable effects on the causal probabilities (46.1% vs. 48.5%, 53.3% vs. 51.5% for rs9332701 and rs61808983, respectively).

To assess the possible bias toward the lead variants more generally, we constructed LCPs for all multivariate genome-wide significant variants in the *F5* locus ($n = 85$). For each of the variants, we compared the p value from LCP-GWAS in which the LCP was constructed for the *F5* lead variant to that in which the LCP was constructed for the variant itself (Supplementary Fig. 7). LCP-GWAS results indicated no significant bias toward the lead variant, and thus, no substantial bias in the fine-mapping results, even when the LD between the variants was only moderate. In addition, we assessed how the phenotype weights used to construct LCPs correlated among variants in the same locus, and also compared to them across loci. As expected, the phenotype weights were highly correlated for variants in high LD (e.g., in the same credible set or the same locus), but not across different loci (Supplementary Fig. 8).

Fine-mapping suggested at least as many causal signals as there were conditional rounds in stepwise conditional analysis ($n = 16$), thus verifying the results from FINEMAP. Further, 13 of the 19 (68.4%) representative variants were also conditioned on in the conditional analysis (Supplementary Table 5). The main benefit of fine-mapping is the probabilistic quantification of possible causal configurations that contain multiple variants. Such metrics are not available in standard implementations of stepwise conditional analysis.

Identifying driver traits

Next, we studied which traits were driving the multivariate associations in each of the 11 loci using metaPhat [16]. The number of driver traits for each of the 11 loci varied between one and all 12. The driver traits were very much in line with the univariate results; the most significantly associated biomarkers in the univariate GWAS were typically included among the driver traits (Table 2). In loci with multiple representative variants, driver traits for the variants

were generally subsets of the lead variant's driver traits, and a stronger multivariate association increased the number of driver traits. However, this relationship between multivariate p value and the number of driver traits did not hold across loci. Further, driver traits typically included all or some of the biomarkers that had previously been associated with the locus (Table 2).

Disease implications of the multivariate loci

Finally, we tested how the 19 representative variants in the 11 loci associated with disease risk among 2367 disease endpoints defined in FinnGen. Altogether, 53 disease associations were observed with seven representative variants. Two of these variants did not lead the multivariate associations at the 11 loci and thus would have gone unnoticed without fine-mapping.

To assess the relevance of the representative variants for their disease associations in FinnGen, the disease associations were conditioned on the variant with the strongest FinnGen disease association within the locus. In 13 of the 53 FinnGen disease associations with the representative variants, the representative variant or a variant in near perfect LD ($r^2 > 0.95$) led the association signal or remained significant after conditioning. We also tested the disease associations in the UKBB, where associations with p values < 0.05 were considered replicated given that the direction of effects were coherent (Supplementary Table 6).

In addition to disease associations, we explored whether the representative variants or variants in LD with them ($r^2 > 0.6$) had previously been reported as pQTLs. Several reported pQTLs [5, 6, 27, 28] in the 11 loci, most of which colocalized with the multivariate biomarker associations, provided evidence for the biologically relevant functions of the representative variants (Supplementary Table 7).

Here we further discuss results for the three multivariate loci with disease associations ($p < 1 \times 10^{-4}$) in FinnGen that remained significant after conditioning. The variants identified by multivariate testing for which the associations became insignificant after conditioning, were regarded unnecessary for the observed disease association. Full disease association results for the 11 loci are shown in Supplementary Table 8.

GP6 gene locus

Multivariate association and FinnGen disease associations

The Finnish enriched rare missense variant chr19:g.55032292G>A (rs199588110, AF = 0.33%, 3.7-fold enrichment), predicted deleterious by SIFT [35] and probably damaging by Polyphen [36], was suggested causal in the *GP6* locus. In FinnGen it led the association with benign

neoplasms of meninges (OR = 6.4, $p = 4.9 \times 10^{-5}$). The association was not replicated in the UKBB, although this may be due to impaired power as the AF of the Finnish enriched variant in the UKBB (0.036%) was roughly a tenth of its AF in FinnGen, and an inadequate match of the discovery and replication phenotypes, as UKBB phenotype definitions included all benign neoplasms of the brain and spinal cord and were not restricted to neoplasms of the meninges.

Driver traits

All 12 biomarkers were considered driver traits of the multivariate association. Cytokines, including many of the 12 biomarkers studied (e.g., IL-6, IL-4, PDGF-BB and VEGF-A), have been implicated in the autocrine regulation of meningioma cell proliferation and motility [37–40]. Further, higher expression levels of both PDGF-BB and VEGF occur in atypical and malignant meningiomas than in benign meningiomas [40, 41] and microvascular density regulated by VEGF has been linked with time to recurrence [42]. Several phase II clinical trials have tested therapies targeting VEGF and PDGF-BB signaling pathways as treatments for recurrent or progressive meningiomas [38] with promising results for two multifunctional tyrosine kinase inhibitors, sunitinib and PTK787/ZK 222584 that inhibit both VEGF and PDGF receptors [38, 43].

SERPINE2 gene locus

Multivariate association and FinnGen disease associations

The *SERPINE2* locus was the locus with the most significant association in the multivariate analysis ($p < 1 \times 10^{-324}$). Fine-mapping identified three independent association signals, represented by three representative variants (chr2:g.224010157G>A (rs13412535), chr2:g.224036001del (rs58116674), and chr2:g.224257750T>A (rs7578029)). One of them, the intronic lead variant rs13412535 from the multivariate analysis, increased the risk of hypertrophic scars (OR = 1.3, $p = 7.5 \times 10^{-5}$) and was in very high LD with the variant that led the disease association in FinnGen (chr2:g.224015781T>C, rs68066031, $r^2 = 0.99$). The association was not replicated in the UKBB, possibly due to differences in case ascertainment as the prevalence of hypertrophic scars was 6.5 times greater in FinnGen compared the UKBB (0.350% vs. 0.053%, respectively), and had not been previously reported at gene-level. Nonetheless, the variant in question had an association with another hypertrophic skin disorder, acquired keratoderma (OR = 1.5, $p = 0.02$) in the UKBB. Another representative variant, the intergenic variant rs7578029 increased the risk of infections of the skin and subcutaneous tissue (OR = 1.1, $p = 9.7 \times 10^{-5}$) and was in

very high LD with the variant that led the disease association in FinnGen (chr2:g.224261196C>T, rs13029443, $r^2 = 0.97$). The association did not replicate in the UKBB, which lacked a well-matching replication phenotype.

Previous knowledge of gene function and driver traits

The *SERPINE2* gene encodes protease nexin-1, a protein in the serpin family of proteins that inhibits serine proteases, especially thrombin, and has therefore been implicated in coagulation and tissue remodeling [44]. The gene has been associated with chronic obstructive pulmonary disease and emphysema [45]. As previously reported, *SERPINE2* has been shown to inhibit extracellular matrix degradation [46] and overexpression of *SERPINE2* has been shown to contribute to pathological cardiac fibrosis in mice [47]. Additionally, serine protease inhibitor genes including *SERPINE2* have been noted to be heavily induced during wound healing [48]. According to GTEx the *SERPINE2* gene is most highly expressed in fibroblasts. Further, inflammation plays an important role in hypertrophic scar formation and cytokines including PDGF and VEGF are dysregulated in hypertrophic scars [49]. The lead variant had genome-wide significant associations with 11 of the 12 biomarkers and all 12 were regarded as driver traits of the association.

pQTLs

The lead variant (chr2:g.224010157G>A, rs13412535) is a pQTL impacting one of the driver traits, PDGF-BB levels (posterior probability of shared causal variant from colocalization analysis, PP = 5.06×10^{-5}), and an intronic variant chr2:g.224015781T>C (rs68066031) in high LD ($r^2 = 0.99$) with the lead variant is a pQTL for SERPINE2 [6, 27] (PP = 0.976). PDGF is considered essential in wound repair [50] and growth factors including PDGF are considered key players in the pathogenesis of hypertrophic scars [51]. PDGF enhances pathologic fibrosis in several tissues such as skin, lung, liver, and kidney by means of mitogenic and chemoattractant actions on the principal collagen-producing cell type, myofibroblasts, as well as stimulation of collagen production [52].

ABO gene locus

Multivariate association and FinnGen disease associations

An association with the *ABO* locus was only detected by multivariate analysis (minimum univariate $p = 2.1 \times 10^{-5}$ for the lead variant from multivariate analysis). Fine-mapping identified one association signal represented by the intronic lead variant chr9:g.133271182T>C (rs550057, aka rs879055593) from multivariate analysis ($p = 8.5 \times 10^{-14}$). It

was associated with 45 endpoints in FinnGen, such as endometriosis, heart failure, and statin usage. Most of these associations resulted from LD to other stronger regional associations, however, nine remained significant after conditioning on other lead variants within the *ABO* locus, including a risk-increasing effect on anemias, for which rs550057 led the genome-wide significant association signal ($p = 4.7 \times 10^{-8}$), visual field disturbances ($p < 6.5 \times 10^{-5}$), and diseases of the ear and mastoid process ($p = 4.8 \times 10^{-5}$). Replication of only two of the nine associations (other anemias and visual field defects) could be attempted in the UKBB due to poor phenotype matching and did not replicate; however, bearing relevance to the genome-wide significant finding in anemia, rs550057 led the association with red blood cell count in the UKBB ($p = 1.3 \times 10^{-212}$) [53].

Driver traits

IL-4 was the only driver trait of the multivariate association and has been implicated in the pathogenesis of many of the diseases associated with the locus. Aplastic anemia is considered to result primarily from immune-mediated bone marrow failure and an imbalance in Type I versus Type II T-cells that secrete IL-4 among other cytokines has been reported [54]. In endometriosis, IL-4 levels have been shown to be upregulated and induce the proliferation of endometriotic stromal cells [55, 56].

pQTLs

The lead variant chr9:g.133271182T>C (rs550057) is a pQTL impacting the levels of four proteins: ALPI (PP = 0.999), CHST15 (PP = 0.999), FAM177A1 (PP = 0.999), and JAG1 (PP = 0.995) [6]. Two of these proteins, carbohydrate sulfotransferase 15 (CHST15) and Jagged1 (JAG1), have been implicated in the pathogenesis of diseases associated with the locus. A small-interfering RNA targeting CHST15 improved myocardial function as well as reduced cardiac fibrosis, hypertrophy, and secretion of proinflammatory cytokines in rats with chronic heart failure [57]. Upregulation of JAG1 has been reported in the endometrium of patients with endometriosis compared to controls [58]. Alagille Syndrome mainly caused by mutations in the JAG1 gene, is accompanied by congenital heart defects and varying degrees of hypercholesterolemia [59].

Discussion

We developed a novel method for multivariate GWAS follow-up analyses and demonstrated the considerable boost in power provided by multivariate GWAS using 12 highly correlated inflammatory markers. In total, four out of 11

genome-wide significant loci were detected only by multivariate analysis when adjusting univariate GWAS for multiple testing. Multivariate analysis might also highlight more plausible candidates for causal variants than univariate analyses. For example, in the *CIQA* locus, the lead variant in the univariate GWAS of the driver trait TNF- β was an intronic variant in the *EPHB2* gene, whereas the lead variant for the locus in the multivariate analysis was a Finnish-enriched missense variant located in the *CIQA* gene which has been previously associated with immunologic diseases [60]. Our multivariate analysis may point toward a plausible mechanism underlying these associations via TNF- β levels.

Although both univariate and multivariate scans have previously been applied to these biomarkers [1, 61], these studies have suffered from the lack of essential follow-up analyses due to the absence of beta estimates in multivariate summary statistics. Our novel method enables two key follow-up analyses for multivariate GWAS: fine-mapping and trait prioritization. Our method solves the problem of missing effect sizes and standard errors required for fine-mapping by an extension of metaCCA followed by LCP-GWAS. This process allows for the transformation of CCA-based multivariate GWAS results into univariate summary statistics and thus extends the use of FINEMAP and other summary statistics-based tools to multivariate GWAS. Fine-mapping complex multivariate associations allows for assessing causality of the variants within the associated loci. This has not been previously feasible. We also further describe the multivariate associations by determining the traits driving the associations using MetaPhat. This workflow allows the identification of both the variants and traits underlying the multivariate associations.

Our study also elucidates the advantage of multivariate analysis combined with large biobank-based phenome-wide screening by discovering multiple novel disease associations. For example, in the *GP6* locus we observe a novel risk-increasing association between the Finnish enriched rare missense variant chr19:g.55032292G>A (rs199588110) and benign neoplasms of meninges. Altogether, a majority of the observed disease associations were for the *ABO* locus that was only detected by multivariate GWAS. All these associations, including a genome-wide significant association with anemia that replicated in the UKBB as an effect on red blood cell count, would have gone undetected had we used univariate GWAS. In addition to disease association discovery, our workflow promotes increasing insight into the pathophysiology underlying the associations by identifying the biomarkers driving the associations. Exploration of biological evidence including pQTLs, most of which colocalized with the multivariate biomarker associations, in the *GP6*, *SERPINE2*, and *ABO* loci orthogonally supports our evidence of causal variants and driver traits. For example, in the *SERPINE2* locus one of the three representative variants chr2:

g.224010157G>A (rs13412535) increased the risk of hypertrophic skin disorders in FinnGen and was a pQTL for PDGF-BB [6] that is considered a key player in the pathogenesis of hypertrophic scars [51], increasing evidence of the biologically relevant functions of this variant.

These methodological development and novel findings notwithstanding, our study has some limitations. First, our newly developed workflow for multivariate fine-mapping requires individual level genotype and phenotype data, problematic for some analysis settings. Additionally, the LCPs are optimized for the lead variants, potentially resulting in overestimation of the causal probability of these variants. We did not, however, see evidence of this in the *F5* locus where we constructed LCPs for each variant reaching genome-wide significance in the multivariate analysis and compared the *p* values from LCP-GWAS when the LCPs were constructed for either the lead variant or the variant itself. Due to the regionality of the LCP-GWAS, it should be noted that LCP-GWAS summary statistics cannot be used for genome-wide methods such as heritability estimation. We also acknowledge that the credible sets we chose for follow-up may not encompass all causal signals within the multivariate associations. The credible sets excluded due to low LD may arise from multiple signals included in the same set, resulting in small LD within the set. Further, some disease associations require replication and follow-up analyses.

On the other hand, our study has many strengths. First, a prospective cohort study was used to assess deep phenotype data rarely available at large scale. Second, we are among the first to present phenome-wide results from FinnGen, a very large and well-phenotyped Finnish biobank study, and also make use of the UKBB in disease association follow-up, ensuring enough power for disease association detection. Finland has a public healthcare system and national health registries, which enable the vast and accurate phenotyping in FinnGen. Besides FinnGen, an additional advantage to performing the study in Finns is that deleterious variants are enriched in the Finnish population due to population history [21]. Furthermore, our reference panel for genotype imputation is from the same population as our discovery and follow-up data sets, which, as demonstrated also by others [62, 63], allows us to study variants that are enriched (and often unique) in the study-specific population.

In conclusion, we developed a novel workflow for multivariate GWAS discovery and follow-up analyses, including fine-mapping and identification of driver traits, and thus promote the advancement of powerful multivariate methods in genomic analyses. We demonstrate the benefit of applying this workflow by identifying novel associations and further describing previously reported associations with both biomarkers and diseases using a set of inflammatory

markers. We show that compared to univariate analyses, multivariate analysis of biomarker data combined with large biobank-based PheWAS reveals a considerably increased number of novel genetic associations with several diseases.

Acknowledgements We would like to thank Lea Urpa for proof-reading, and Sari Kivikko, Huei-Yi Shen, and Ulla Tuomainen for management assistance. We would like to thank all participants of the FINRISK, FinnGen and UKBB studies for their generous participation. The FINRISK data used for the research were obtained from THL Biobank. This research has been conducted using the UK Biobank Resource with application number 22627. This work was supported by the Academy of Finland Center of Excellence in Complex Disease Genetics [Grant No 312062 to SR, 312076 to MP, 312074 to AP, 312075 to MD]; Academy of Finland [Grant No 285380 to SR, 288509 to MP, 128650 to AP]; the Finnish Foundation for Cardiovascular Research [to SR, VS, and AP]; the Sigrid Jusélius Foundation [to SR, MP, and AP]; University of Helsinki HiLIFE Fellow grants 2017-2020 [to SR and MP]; Foundation and the Horizon 2020 Research and Innovation Programme [grant number 667301 (COSYN) to AP]; the Doctoral Programme in Population Health, University of Helsinki [to JJP and SER]; and The Finnish Medical Foundation [to JJP]. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and nine industry partners (AbbVie, AstraZeneca, Biogen, Celgene, Genentech, GSK, MSD, Pfizer and Sanofi). Following biobanks are acknowledged for collecting the FinnGen project samples: Auria Biobank (<https://www.auria.fi/biopankki/en>), THL Biobank (<https://thl.fi/fi/web/thl-biopankki>), Helsinki Biobank (<https://www.terveyskyla.fi/helsinginbiopankki/en>), Northern Finland Biobank Borealis (<https://www.ppsbp.fi/Tutkimus-ja-opetus/Biopankki>), Finnish Clinical Biobank Tampere (https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (<https://ita-suomenbiopankki.fi/>), Central Finland Biobank (<https://www.ksshp.fi/fi-FI/Potilaalle/Biopankki>), Finnish Red Cross Blood Service Biobank (<https://www.bloodservice.fi/Research%20Projects/biopanking>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

FinnGen Steering Committee: Aarno Palotie^{13,14}, Mark Daly^{13,14}

Pharmaceutical companies: Howard Jacob¹⁵, Athena Matakidou¹⁶, Heiko Runz¹⁷, Sally John¹⁷, Robert Plenge¹⁸, Mark McCarthy¹⁹, Julie Hunkapiller¹⁹, Meg Ehm²⁰, Dawn Waterworth²⁰, Caroline Fox²¹, Anders Malarstig²², Kathy Klinger²³, Kathy Call²³

University of Helsinki & Biobanks: Tomi Mäkelä²⁴, Jaakko Kaprio¹³, Petri Virolainen²⁵, Kari Pulkki²⁵, Terhi Kilpi²⁶, Markus Perola²⁶, Jukka Partanen²⁷, Anne Pitkäranta²⁸, Riitta Kaarteenaho²⁹, Seppo Vainio²⁹, Kimmo Savinainen³⁰, Veli-Matti Kosma³¹, Urho Kujala³²

Other Experts/ Non-Voting Members: Outi Tuovila³³, Minna Hendolin³³, Raimo Pakkanen³³

Scientific Committee Pharmaceutical companies: Jeff Waring¹⁵, Bridget Riley-Gillis¹⁵, Athena Matakidou¹⁶, Heiko Runz¹⁷, Jimmy Liu¹⁷, Shameek Biswas¹⁸, Julie Hunkapiller¹⁹, Dawn Waterworth²⁰, Meg Ehm²⁰, Dorothee Diogo²¹, Caroline Fox²¹, Anders Malarstig²², Catherine Marshall²², Xinli Hu²², Kathy Call²³, Kathy Klinger²³, Matthias Gossel²³

University of Helsinki & Biobanks: Samuli Ripatti^{13,14}, Johanna Schleutker²⁵, Markus Perola²⁶, Mikko Arvas²⁷, Olli Carpen²⁸, Reetta Hintala²⁹, Johannes Kettunen²⁹, Reijo Laaksonen³⁰, Arto Mannermaa³¹, Juha Paloneva³², Urho Kujala³²

Other Experts/ Non-Voting Members: Outi Tuovila³³, Minna Hendolin³³, Raimo Pakkanen³³

Clinical Groups Neurology Group: Hilkka Soininen³⁴, Valtteri Julkunen³⁴, Anne Remes³⁵, Reetta Kälviäinen³⁴, Mikko Hiltunen³⁴, Jukka Peltola³⁶, Pentti Tienari²⁸, Juha Rinne³⁷, Adam Ziemann¹⁵, Jeffrey Waring¹⁵, Sahar Esmaeili¹⁵, Nizar Smaoui¹⁵, Anne Lehtonen¹⁵, Susan Eaton¹⁷, Heiko Runz¹⁷, Sanni Lahdenperä¹⁷, Shameek Biswas¹⁸, John Michon¹⁹, Geoff Kerchner¹⁹, Julie Hunkapiller¹⁹, Natalie Bowers¹⁹, Edmond Teng¹⁹, John Eicher²¹, Vinay Mehta²¹, Padhraig Gormley²¹, Kari Linden²², Christopher Whelan²², Fanli Xu²⁰, David Pulford²⁰

Gastroenterology Group: Martti Färkkilä²⁸, Sampsa Pikkarainen²⁸, Airi Jussila³⁶, Timo Blomster³⁵, Mikko Kiviniemi³⁴, Markku Voutilainen³⁷, Bob Georgantas¹⁵, Graham Heap¹⁵, Jeffrey Waring¹⁵, Nizar Smaoui¹⁵, Fedik Rahimov¹⁵, Anne Lehtonen¹⁵, Keith Usiskin¹⁸, Joseph Maranville¹⁸, Tim Lu¹⁹, Natalie Bowers¹⁹, Danny Oh¹⁹, John Michon¹⁹, Vinay Mehta²¹, Kirsi Kalpala²², Melissa Miller²², Xinli Hu²², Linda McCarthy²⁰

Rheumatology Group: Kari Eklund²⁸, Antti Palomäki³⁷, Pia Isomäki³⁶, Laura Pirilä³⁷, Oili Kaipainen-Seppänen³⁴, Johanna Huhtakangas³⁵, Bob Georgantas¹⁵, Jeffrey Waring¹⁵, Fedik Rahimov¹⁵, Apinya Lertratanakul¹⁵, Nizar Smaoui¹⁵, Anne Lehtonen¹⁵, David Close¹⁶, Marla Hochfeld¹⁸, Natalie Bowers¹⁹, John Michon¹⁹, Dorothee Diogo²¹, Vinay Mehta²¹, Kirsi Kalpala²², Nan Bing²², Xinli Hu²², Jorge Esparza Gordillo²⁰, Nina Mars¹³

Pulmonology Group: Tarja Laitinen³⁶, Margit Pelkonen³⁴, Paula Kauppi²⁸, Hannu Kankaanranta³⁶, Terttu Harju³⁵, Nizar Smaoui¹⁵, David Close¹⁶, Steven Greenberg¹⁸, Hubert Chen¹⁹, Natalie Bowers¹⁹, John Michon¹⁹, Vinay Mehta²¹, Jo Betts²⁰, Soumitra Ghosh²⁰

Cardiometabolic Diseases Group: Veikko Salomaa³⁸, Teemu Niiranen³⁸, Markus Juonala³⁷, Kaj Metsärinne³⁷, Mika Kähönen³⁶, Juhani Junttila³⁵, Markku Laakso³⁴, Jussi Pihlajamäki³⁴, Juha Sinisalo²⁸, Marja-Riitta Taskinen²⁸, Tiinamaija Tuomi²⁸, Jari Laukkanen³⁹, Ben Challis¹⁶, Andrew Peterson¹⁹, Julie Hunkapiller¹⁹, Natalie Bowers¹⁹, John Michon¹⁹, Dorothee Diogo²¹, Audrey Chu²¹, Vinay Mehta²¹, Jaakko Parkkinen²², Melissa Miller²², Anthony Muslin²³, Dawn Waterworth²⁰

Oncology Group: Heikki Joensuu²⁸, Tuomo Meretoja²⁸, Olli Carpen²⁸, Lauri Aaltonen²⁸, Annika Auranen³⁶, Peeter Karihtala³⁵, Saira Kauppi³⁵, Päivi Auvinen³⁴, Klaus Elenius³⁷, Relja Popovic¹⁵, Jeffrey Waring¹⁵, Bridget Riley-Gillis¹⁵, Anne Lehtonen¹⁵, Athena Matakidou¹⁶, Jennifer Schutzman¹⁹, Julie Hunkapiller¹⁹, Natalie Bowers¹⁹, John Michon¹⁹, Vinay Mehta²¹, Andrey Loboda²¹, Aparna Chhibber²¹, Heli Lehtonen²², Stefan McDonough²², Marika Crohns²³, Diptee Kulkarni²⁰

Ophthalmology Group: Kai Kaarniranta³⁴, Joni Turunen²⁸, Terhi Ollila²⁸, Sanna Seitsonen²⁸, Hannu Uusitalo³⁶, Vesa Aaltonen³⁷, Hannele Uusitalo-Järvinen³⁶, Marja Luodonpää³⁵, Nina Hautala³⁵, Heiko Runz¹⁷, Erich Strauss¹⁹, Natalie Bowers¹⁹, Hao Chen¹⁹, John Michon¹⁹, Anna Podgornaia²¹, Vinay Mehta²¹, Dorothee Diogo²¹, Joshua Hoffman²⁰

Dermatology Group: Kaisa Tasanen³⁵, Laura Huilaja³⁵, Katariina Hannula-Jouppi²⁸, Teea Salmi³⁶, Sirkku Peltonen³⁷, Leena Koulou³⁷, Ilkka Harvima³⁴, Kirsi Kalpala²², Ying Wu²², David Choy¹⁹, John Michon¹⁹, Nizar Smaoui¹⁵, Fedik Rahimov¹⁵, Anne Lehtonen¹⁵, Dawn Waterworth²⁰

FinnGen Teams: Administration Team Administration Team: Anu Jalanko¹³, Risto Kajanne¹³, Ulrike Lyhs¹³

Communication: Mari Kaunisto¹³

Analysis Team: Justin Wade Davis¹⁵, Bridget Riley-Gillis¹⁵, Danjuma Quarless¹⁵, Slavé Petrovski¹⁶, Jimmy Liu¹⁷, Chia-Yen Chen¹⁷, Paola Bronson¹⁷, Robert Yang¹⁸, Joseph Maranville¹⁸, Shameek Biswas¹⁸, Diana Chang¹⁹, Julie Hunkapiller¹⁹, Tushar Bhangale¹⁹, Natalie Bowers¹⁹, Dorothee Diogo²¹, Emily Holzinger²¹, Padhraig Gormley²¹, Xulong Wang²¹, Xing Chen²², Åsa Hedman²²,

Kirsi Auro²⁰, Clarence Wang²³, Ethan Xu²³, Franck Auge²³, Clement Chatelain²³, Mitja Kurki^{13,14}, Samuli Ripatti^{13,14}, Mark Daly^{13,14}, Juha Karjalainen^{13,14}, Aki Havulinna¹³, Anu Jalanko¹³, Kimmo Palin⁴⁰, Priit Palta¹³, Pietro Della Briotta Parolo¹³, Wei Zhou¹³, Susanna Lemmelä¹³, Manuel Rivas⁴¹, Jarmo Harju¹³, Aarno Palotie^{13,14}, Arto Lehisto¹³, Andrea Ganna¹³, Vincent Llorens¹³, Antti Karlsson²⁵, Kati Kristiansson²⁶, Mikko Arvas²⁷, Kati Hyvärinen²⁷, Jarmo Ritari²⁷, Tiina Wahlfors²⁷, Miika Koskinen²⁸, Olli Carpen²⁸, Johannes Kettunen²⁹, Katri Pylkäs²⁹, Marita Kalaoja²⁹, Minna Karjalainen²⁹, Tuomo Mantere²⁹, Eeva Kangasniemi³⁰, Sami Heikkinen³¹, Arto Mannermaa³¹, Eija Laakkonen³², Juha Kononen³²

Sample Collection Coordination: Anu Loukola²⁸

Sample Logistics: Päivi Laiho²⁶, Tuuli Sistonen²⁶, Essi Kaiharju²⁶, Markku Laukkanen²⁶, Elina Järvensivu²⁶, Sini Lähteenmäki²⁶, Lotta Männikkö²⁶, Regis Wong²⁶

Registry Data Operations: Kati Kristiansson²⁶, Hannele Mattsson²⁶, Susanna Lemmelä¹³, Tero Hiekkalinna²⁶, Manuel González Jiménez²⁶ Genotyping Kati Donner¹³

Sequencing Informatics: Priit Palta¹³, Kalle Pärn¹³, Javier Nunez-Fontana¹³

Data Management and IT Infrastructure: Jarmo Harju¹³, Elina Kilpeläinen¹³, Timo P. Sipilä¹³, Georg Brein¹³, Alexander Dada¹³, Ghazal Awaisa¹³, Anastasia Shcherban¹³, Tuomas Sipilä¹³

Clinical Endpoint Development: Hannele Laivuori¹³, Aki Havulinna¹³, Susanna Lemmelä¹³, Tuomo Kiiskinen¹³ Trajectory Team Tarja Laitinen³⁶, Harri Siirtola⁴², Javier Gracia Tabuenca⁴²

Biobank Directors: Lila Kallio⁴³, Sirpa Soini⁴⁴, Jukka Partanen⁴⁵, Kimmo Pitkanen⁴⁶, Seppo Vainio⁴⁷, Kimmo Savinainen⁴⁸, Veli-Matti Kosma⁴⁹, Teijo Kuopio⁵⁰

¹³Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland; ¹⁴Broad Institute, Cambridge, MA, USA; ¹⁵Abbvie, Chicago, IL, USA; ¹⁶Astra Zeneca, Cambridge, UK; ¹⁷Biogen, Cambridge, MA, USA; ¹⁸Celgene, Summit, NJ, USA; ¹⁹Genentech, San Francisco, CA, USA; ²⁰GlaxoSmithKline, Brentford, UK; ²¹Merck, Kenilworth, NJ, USA; ²²Pfizer, New York, NY, USA; ²³Sanofi, Paris, France; ²⁴HiLIFE, University of Helsinki, Helsinki, Finland; ²⁵Auria Biobank/University of Turku/Hospital District of Southwest Finland, Turku, Finland; ²⁶THL Biobank/The National Institute of Health and Welfare Helsinki, Helsinki, Finland; ²⁷Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland; ²⁸Hospital District of Helsinki and Uusimaa, Helsinki, Finland; ²⁹Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland; ³⁰Finnish Clinical Biobank Tampere/University of Tampere/Pirkanmaa Hospital District, Tampere, Finland; ³¹Biobank of Eastern Finland/University of Eastern Finland/Northern Savo Hospital District, Kuopio, Finland; ³²Central Finland Biobank/University of Jyväskylä/Central Finland Health Care District, Jyväskylä, Finland; ³³Business Finland, Helsinki, Finland; ³⁴Northern Savo Hospital District, Kuopio, Finland; ³⁵Northern Ostrobothnia Hospital District, Oulu, Finland; ³⁶Pirkanmaa Hospital District, Tampere, Finland; ³⁷Hospital District of Southwest Finland, Turku, Finland; ³⁸The National Institute of Health and Welfare Helsinki, Helsinki, Finland; ³⁹Central Finland Health Care District, Jyväskylä, Finland; ⁴⁰University of Helsinki, Helsinki, Finland; ⁴¹University of Stanford, Stanford, CA, USA; ⁴²University of Tampere, Tampere, Finland; ⁴³Auria Biobank, Turku, Finland; ⁴⁴THL Biobank, Helsinki, Finland; ⁴⁵Blood Service Biobank, Helsinki, Finland; ⁴⁶Helsinki Biobank, Helsinki, Finland; ⁴⁷Northern Finland Biobank Borealis, Oulu, Finland; ⁴⁸Tampere Biobank, Tampere, Finland; ⁴⁹Biobank of Eastern Finland, Kuopio, Finland; ⁵⁰Central Finland Biobank, Jyväskylä, Finland

Data sharing and declaration Full summary statistics of the multi-variate GWAS on the 12 inflammatory biomarkers are available via the NHGRI-EBI GWAS Catalog, accession number [GCST90000584](https://www.ebi.ac.uk/gwas/studies/GCST90000584). The FinnGen data may be accessed through Finnish Biobanks' FinnBB portal (www.finnbb.fi) and THL Biobank data through THL Biobank (<https://thl.fi/en/web/thl-biobank>).

Compliance with ethical standards

Conflict of interest VS has received honoraria from Novo Nordisk and Sanofi for consultations and has ongoing research collaboration with Bayer AG (all unrelated to this study). All other authors have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Ahola-Olli AV, Würtz P, Havulinna AS, Aalto K, Pitkänen N, Lehtimäki T, et al. Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *Am J Human Genet.* 2017;100:40–50.
- Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell.* 2016;167:1415–29. e19.
- Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in > 300,000 individuals. *Nat Genet.* 2017;49:1758.
- Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars NJ, Aguirre M, Venkataraman GR, et al. Genetics of 38 blood and urine biomarkers in the UK Biobank. 2019. Preprint at <https://www.biorxiv.org/content/10.1101/660506v1>.
- Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun.* 2017;8:14357.
- Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018;558:73.
- Kettunen J, Demirkan A, Würtz P, Draisma HH, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun.* 2016;7:11122.
- Inouye M, Ripatti S, Kettunen J, Lyytikäinen L, Oksala N, Laurila P, et al. Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet.* 2012;8:e1002907.
- Leitsalu L, Haller T, Esko T, Tammesoo M, Alavere H, Snieder H, et al. Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *Int J Epidemiol.* 2014;44:1137–47.
- Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203.
- Kim S, Xing EP. Statistical estimation of correlated genome associations to a quantitative trait network. *PLoS Genet.* 2009;5:e1000587.
- Ferreira MA, Purcell SM. A multivariate test of association. *Bioinformatics.* 2008;25:132–3.
- O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FC, Elliott P, Jarvelin M, et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PloS One.* 2012;7:e34861.
- Cichonska A, Rousu J, Marttinen P, Kangas AJ, Soininen P, Lehtimäki T, et al. metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. *Bioinformatics.* 2016;32:1981–9.
- Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics.* 2016;32:1493–501.
- Lin J, Tabassum R, Ripatti S, Pirinen M. MetaPhat: Detecting And Decomposing Multivariate Associations From Univariate Genome-wide Association Statistics. *Front Genet.* 2020;11:431.
- McInnes IB, Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat Rev Immunol.* 2007;7:429.
- Martins TB, Rose JW, Jaskowski TD, Wilson AR, Husebye D, Seraj HS, et al. Analysis of proinflammatory and anti-inflammatory cytokine serum concentrations in patients with multiple sclerosis by using a multiplexed immunoassay. *Am J Clin Pathol.* 2011;136:696–704.
- Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. *Nature.* 2000;407:249.
- Borodulin K, Vartiainen E, Peltonen M, Jousilahti P, Juolevi A, Laatikainen T, et al. Forty-year trends in cardiovascular risk factors in Finland. *Eur J Public Health.* 2014;25:539–46.
- Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* 2014;10:e1004494.
- Hail Team. Hail 0.2.13-81ab564db2b4. <https://github.com/hail-is/hail/releases/tag/0.2.13> <https://doi.org/10.5281/zenodo.2646680>.
- Benner C, Havulinna A, Salomaa V, Ripatti S, Pirinen M. Refining fine-mapping: effect sizes and regional heritability. 2018. Preprint at <https://www.biorxiv.org/content/10.1101/318618v1>.
- Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet.* 2018;50:1335.
- Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, et al. Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Medical Informatics.* 2019;7:e14325.
- Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2018;47:D1005–D1012.
- Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science.* 2018;361:769–73.
- Sasayama D, Hattori K, Ogawa S, Yokota Y, Matsumura R, Teraishi T, et al. Genome-wide quantitative trait loci mapping of the human cerebrospinal fluid proteome. *Hum Mol Genet.* 2016;26:44–51.
- Reeve MP, Kirby A, Wierzbowski J, Daly M, Hutz J. Target Gene Notebook: Connecting genetics and drug discovery. 2019. Preprint at <https://www.biorxiv.org/content/10.1101/757690v1>.
- Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics.* 2014;10:1–15.
- Stephens M. A unified framework for association analysis with multiple related phenotypes. *PloS ONE.* 2013;8:1–19.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337:1190–5.
- Huang H, Fang M, Jostins L, Mirkov MU, Boucher G, Anderson CA, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature.* 2017;547:173.

34. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–43.
35. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812–4.
36. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols Hum Genet*. 2013;76:7.20. 1–7.20. 41.
37. Andrae N, Kirches E, Hartig R, Haase D, Keilhoff G, Kalinski T, et al. Sunitinib targets PDGF-receptor and Flt3 and reduces survival and migration of human meningioma cells. *Eur J Cancer*. 2012;48(12):1831–41.
38. Kaley TJ, Wen P, Schiff D, Ligon K, Haidar S, Karimi S, et al. Phase II trial of sunitinib for recurrent and progressive atypical and anaplastic meningioma. *Neuro Oncol*. 2014;17:116–21.
39. Todo T, Adams EF, Rafferty B, Fahlbusch R, Dingermann T, Werner H. Secretion of interleukin-6 by human meningioma cells: possible autocrine inhibitory regulation of neoplastic cell growth. *J Neurosurg*. 1994;81:394–401.
40. Yang S, Xu G. Expression of PDGF and its receptor as well as their relationship to proliferating activity and apoptosis of meningiomas in human meningiomas. *J Clin Neurosci*. 2001;8:49–53.
41. Lamszus K, Lengler U, Schmidt NO, Stavrou D, Ergün S, Westphal M. Vascular endothelial growth factor, hepatocyte growth factor/scatter factor, basic fibroblast growth factor, and placenta growth factor in human meningiomas and their relation to angiogenesis and malignancy. *Neurosurgery*. 2000;46:938–48.
42. Preusser M, Hassler M, Birner P, Rudas M, Acker T, Plate KH, et al. Microvascularization and expression of VEGF and its receptors in recurring meningiomas: pathobiological data in favor of anti-angiogenic therapy approaches. *Clin Neuropathol*. 2012;31:352–60.
43. Raizer JJ, Grimm SA, Rademaker A, Chandler JP, Muro K, Helenowski I, et al. A phase II trial of PTK787/ZK 222584 in recurrent or progressive radiation and surgery refractory meningiomas. *J Neurooncol*. 2014;117:93–101.
44. Bouton M, Boulaftali Y, Richard B, Arocas V, Michel J, Jandrot-Perrus M. Emerging role of serpinE2/protease nexin-1 in hemostasis and vascular biology. *Blood J Am Soc Hematol*. 2012;119:2452–7.
45. DeMeo DL, Mariani TJ, Lange C, Srisuma S, Litonjua AA, Celedón JC, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease. *Am J Hum Genet*. 2006;78:253–64.
46. Bergman BL, Scott RW, Bajpai A, Watts S, Baker JB. Inhibition of tumor-cell-mediated extracellular matrix destruction by a fibroblast proteinase inhibitor, protease nexin I. *Proc Natl Acad Sci*. 1986;83:996–1000.
47. Li X, Zhao D, Guo Z, Li T, Qili M, Xu B, et al. Overexpression of SerpinE2/protease nexin-1 contribute to pathological cardiac fibrosis via increasing collagen deposition. *Sci Rep*. 2016;6:37635.
48. Nuutila K, Siltanen A, Peura M, Harjula A, Nieminen T, Vuola J, et al. Gene expression profiling of negative-pressure-treated skin graft donor site wounds. *Burns*. 2013;39:687–93.
49. Ghazawi FM, Zargham R, Gilardino MS, Sasseville D, Jafarian F. Insights into the pathophysiology of hypertrophic scars and keloids: how do they differ? *Adv Skin Wound Care*. 2018;31:582–95.
50. Brissett AE, Sherris DA. Scar contractures, hypertrophic scars, and keloids. *Facial Plastic Surg*. 2001;17:263–72.
51. Lian N, Li T. Growth factor pathways in hypertrophic scars: molecular pathogenesis and therapeutic implications. *Biomed Pharmacotherapy*. 2016;84:42–50.
52. Bonner JC. Regulation of PDGF and its receptors in fibrotic diseases. *Cytokine Growth Factor Rev*. 2004;15:255–73.
53. Kichaev G, Bhatia G, Loh P, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet*. 2019;104:65–75.
54. Tsuda H, Yamasaki H. Type I and type II T-cell profiles in aplastic anemia and refractory anemia. *Am J Hematol*. 2000;64:271–4.
55. OuYang Z, Hirota Y, Osuga Y, Hamasaki K, Hasegawa A, Tajima T, et al. Interleukin-4 stimulates proliferation of endometriotic stromal cells. *Am J Pathol*. 2008;173:463–9.
56. Hsu C, Yang B, Wu M, Huang K. Enhanced interleukin-4 expression in patients with endometriosis. *Fertil Steril*. 1997;67:1059–64.
57. Watanabe K, Arumugam S, Sreedhar R, Thandavarayan RA, Nakamura T, Nakamura M, et al. Small interfering RNA therapy against carbohydrate sulfotransferase 15 inhibits cardiac remodeling in rats with dilated cardiomyopathy. *Cell Signal*. 2015;27:1517–24.
58. Laudanski P, Charkiewicz R, Kuzmicki M, Szamatowicz J, Świątecka J, Mroczko B, et al. Profiling of selected angiogenesis-related genes in proliferative eutopic endometrium of women with endometriosis. *Eur J Obstet Gynecol Reprod Biol*. 2014;172:85–92.
59. Hannoush ZC, Puerta H, Bauer MS, Goldberg RB. New JAG1 mutation causing Alagille syndrome presenting with severe hypercholesterolemia: case report with emphasis on genetics and lipid abnormalities. *J Clin Endocrinol Metabol*. 2016;102:350–3.
60. van Schaarenburg RA, Daha NA, Schonkeren JJ, Levarht EN, van Gijlswijk-Janssen DJ, Kurreeman FA, et al. Identification of a novel non-coding mutation in C1qB in a Dutch child with C1q deficiency associated with recurrent infections. *Immunobiology*. 2015;220:422–7.
61. Nath AP, Ritchie SC, Grinberg NF, Tang HH, Huang QQ, Teo SM, et al. Multivariate Genome-wide Association Analysis of a Cytokine Network Reveals Variants with Widespread Immune, Haematological, and Cardiometabolic Pleiotropy. *Am J Hum Genet*. 2019;105:1076–90.
62. Surakka I, Kristiansson K, Anttila V, Inouye M, Barnes C, Moutsianas L, et al. Founder population-specific HapMap panel increases power in GWA studies through improved imputation accuracy and CNV tagging. *Genome Res*. 2010;20:1344–51.
63. Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet*. 2017;25:869.

Affiliations

Sanni E. Ruotsalainen¹ · Juulia J. Partanen ¹ · Anna Cichonska ^{1,2,3} · Jake Lin¹ · Christian Benner¹ · Ida Surakka⁴ · FinnGen · Mary Pat Reeve¹ · Priit Palta ^{1,5} · Marko Salmi^{6,7} · Sirpa Jalkanen^{6,7} · Ari Ahola-Olli ^{1,8,9} · Aarno Palotie^{1,8,9} · Veikko Salomaa ¹⁰ · Mark J. Daly ^{1,8,9} · Matti Pirinen ^{1,11,12} · Samuli Ripatti ^{1,8,11} · Jukka Koskela ^{1,8}

¹ Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland

² Department of Computer Science, Helsinki Institute for Information Technology HIIT, Aalto University, Espoo, Finland

³ Department of Future Technologies, University of Turku, Turku, Finland

⁴ Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

⁵ Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

⁶ MedCity Research Laboratory, University of Turku, Turku, Finland

⁷ Institute of Biomedicine, University of Turku, Turku, Finland

⁸ The Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁹ Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

¹⁰ Finnish Institute for Health and Welfare, Helsinki, Finland

¹¹ Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland

¹² Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland