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

Ruotsalainen, Sanni E.

2021-02

Ruotsalainen , S E , Partanen , J J , Cichonska , A , Lin , J , Benner , C , Surakka , I , Reeve , M P , Palta , P , Salmi , M , Jalkanen , S , Ahola-Olli , A , Palotie , A , Salomaa , V , Daly , M J , Pirinen , M , Ripatti , S & Koskela , J 2021 , ' An expanded analysis framework for multivariate GWAS connects inflammatory biomarkers to functional variants and disease ' , þÿ E u r o p e a n J o u r n a l o f H u m a n G e n e t i c s , v o l . 2 9 , p p . 3 0 9 3 2 4 . h t t p s : / / d

http://hdl.handle.net/10138/341211
https://doi.org/10.1038/s41431-020-00730-8

acceptedVersion

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

Running title: Expanded multivariate GWAS of inflammatory markers

Sanni E. Ruotsalainen[1]*, Juulia J. Partanen[1]*, Anna Cichonska[1,2,3], Jake Lin[1], Christian Benner[1], Ida Surakka[4], *FinnGen*, Mary Pat Reeve[1], 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, Michigan, USA
5) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia
6) MediCity 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, Massachusetts, USA
9) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, 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

*authors contributed equally

**Corresponding author

Tel: +358 50 374 7924, +358 40 567 0826

E-mail: jukka.t.koskela@helsinki.fi, samuli.ripatti@helsinki.fi

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 ($p < 1\times10^{-4}$). 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.

Keywords: multivariate GWAS, inflammatory biomarkers, fine-mapping

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 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 (Figure 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 2,367 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 = 6,890) (Table 1, Supplementary Figure 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-1α), and one tumor necrosis factor (TNF-β) (Table 1, Supplementary Figure 1). Hierarchical clustering identified the cluster of 12 inflammatory biomarkers out of 66 quantitative traits of cardiometabolic or immunologic relevance (Supplementary Figure 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-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 3,775 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, ..., X_p) that is maximally correlated with a linear combination of the q response variables (Y_1, Y_2, ..., Y_q). If we denote the respective linear combinations by

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

and

$$Y^* = LCP = \boldsymbol{b}'\boldsymbol{y} = b_1y_1 + b_2y_2 + \cdots + b_qy_q,$$

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 $\boldsymbol{a}$ and $\boldsymbol{b}$ that maximize

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

where Σ_{xx}, Σ_{yy} and Σ_{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 ($\boldsymbol{b} = [b_1, b_2 \ldots, b_q]$) 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 1Mb to each variant reaching genome-wide significance (GWS; p-value $< 5\times10^{-8}$) 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 ±1Mb 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 ($p > 5\times10^{-8}$). 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.finngen.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 2,444 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 $< 1\times10^{-4}$ were considered significant. We tested the p-value threshold by sampling 1,000 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 Figure 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 2,367 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 (±0.5MB 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(5), Sun(6), Emilsson(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 Figure 1) measured in 6,890 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 analyses, 999 only by the multivariate analysis and two only by the univariate analysis using a Bonferroni-corrected p-value threshold of $5\times10^{-8}/12$ (Figure 2). A total of 1,189 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 (maf < 0.01, 0.01 < maf < 0.1, and maf > 0.1) separately, with rare variants not showing noticeably more inflation than more common variants (Supplementary Figure 4).

Within the 1,189 genome-wide significant variants in the multivariate analysis, we identified 11 independently associated loci (Figure 3 and Supplementary Figure 5), four of which (*F5, C1orf140, PDGFRB* and *ABO*) were not detected by univariate analyses corrected for multiple testing (Figure 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 (*C1QA, PCSK6,* and *VLDLR*), we noted that the multivariate and univariate lead variants were never the same. In the *C1QA* 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 *C1QA* locus, an association with only TNF-β of the 12 biomarkers was noted in the univariate results. The lead variant in the TNF-β univariate GWAS was chr1:g.22720394C>T (rs78655189, $p = 2.2\times10^{-24}$), 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, $p = 1.2\times10^{-73}$), a Finnish-enriched missense variant located in the *C1QA* gene. In the *PCKS6* locus both lead variants were intronic with similar multivariate p-values (multivariate lead variant chr15:g.101451543G>T (rs11637184, $p = 2.4\times10^{-68}$), univariate PDGF-BB lead variant chr15:g.101446695T>A (rs11634270, $p = 1.3\times10^{-67}$)).

In the *VLDLR* locus, where LD between the two lead variants was low, univariate fine-mapping of VEGF, 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 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-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*, *C1orf140*, *SERPINE2*, and *GP6*). 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 11.3M variants analyzed. P-values for enrichment were calculated using the χ^2-test for the number of nonsynonymous and splice-region or

297 missense variants within the genome-wide significant variants against the number of the
298 corresponding subset of variants within all variants tested. The multivariate genome-wide significant
299 variants were enriched for missense variants and missense, splice-region and frameshift variants (2.2-
300 fold, p = 0.015, and 1.9-fold, p = 8.8×10^{-4}, respectively).
301
302 **Fine-mapping multivariate GWAS results**
303 To identify the causal variants of the multivariate associations, we studied the likelihood of multiple
304 variants contributing to the association signal in the 11 associated loci using FINEMAP(23). Our
305 novel multivariate LCP-GWAS method based on linear combinations calculated for each locus using
306 multivariate metaCCA results enabled fine-mapping of the multivariate results. The number of
307 credible sets varied from one to four for the multivariate associated loci (Supplementary Table 3),
308 resulting in a total of 19 independent sets of variants considered putatively causal. All 183 variants
309 within the 19 credible sets are available in Supplementary Table 3 and posterior probabilities for
310 different numbers of causal signals for each locus are available in Supplementary Table 4.
311
312 Among each of the 19 sets, the variant with the highest causal probability was chosen to represent
313 the set as the representative variant (Table 2 and Supplementary Figure 6). The 19 representative
314 variants, included all except one (chr15g.101991748G>C, rs11637184 in the *PCSK6* locus) of the 11
315 lead variants from multivariate GWAS. Highlighting the importance of fine-mapping multivariate
316 GWAS results, one of the four representative variants (chr15:g.101339772G>A, rs111482836) in the
317 *PCSK6* locus was associated with disease in FinnGen, whereas the lead variant was not. Additionally,
318 the 19 representative variants were further enriched for both missense variants and missense, splice-
319 region and frameshift variants (37-fold, p = 1.3×10^{-17}, and 28-fold, p = 1.4×10^{-17}, respectively)
320 compared to multivariate genome-wide significant variants (2.2-fold, p = 0.015, and 1.9-fold, p =
321 8.8×10^{-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 Figure 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 Figure 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 2,367 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 ($p < 1\times10^{-4}$) 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, $p = 4.9\times10^{-5}$). 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 ($p < 1\times10^{-324}$). 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, $p = 7.5\times10^{-5}$) 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.

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 $p = 2.1\times10^{-5}$ 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 ($p = 8.5\times10^{-14}$). 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 ($p = 4.7\times10^{-8}$), visual field disturbances ($p < 6.5\times10^{-5}$), and diseases of the ear and mastoid process ($p = 4.8\times10^{-5}$). 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 ($p = 1.3\times10^{-212}$).(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 *C1QA* 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 *C1QA* 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 LPC-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 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.

DATA SHARING

Full summary statistics of the multivariate GWAS on the 12 inflammatory biomarkers are available via the NHGRI-EBI GWAS Catalog, accession number GCST90000584, URL: ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90000584. The FinnGen data may be accessed through Finnish Biobanks' FinnBB portal (www.finbb.fi) and THL Biobank data through THL Biobank (https://thl.fi/en/web/thl-biobank). The FinnGen data may be accessed through Finnish Biobanks' FinnBB portal (www.finbb.fi) and THL Biobank data through THL Biobank (https://thl.fi/en/web/thl-biobank).

ACKNOWLEDGMENTS

We would like to thank Lea Urpa for proofreading, 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 S.R., 312076 to M.P., 312074 to A.P., 312075 to M.D.]; Academy of Finland [Grant No 285380 to S.R, 288509 to M.P., 128650 to A.P.]; the Finnish Foundation for Cardiovascular Research [to S.R., V.S., and A.P.]; the Sigrid Jusélius Foundation [to S.R., M.P., and A.P.]; University of Helsinki HiLIFE Fellow grants 2017-2020 [to S.R. and M.P.]; Foundation and the Horizon 2020 Research and Innovation Programme [grant number 667301 (COSYN) to A.P.]; the Doctoral Programme in Population Health, University of Helsinki [to J.J.P. and S.E.R.]; and The Finnish Medical Foundation [to J.J.P.]. 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.ppshp.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/biobanking).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest: V.S. has received honoraria from Novo Nordisk and Sanofi for consultations and has ongoing research collaboration with Bayer AG (all unrelated to this study).

Contributors of FinnGen

Steering Committee

Aarno Palotie	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mark Daly	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Pharmaceutical companies

Howard Jacob	Abbvie, Chicago, IL, United States
Athena Matakidou	Astra Zeneca, Cambridge, United Kingdom
Heiko Runz	Biogen, Cambridge, MA, United States
Sally John	Biogen, Cambridge, MA, United States
Robert Plenge	Celgene, Summit, NJ, United States
Mark McCarthy	Genentech, San Francisco, CA, United States
Julie Hunkapiller	Genentech, San Francisco, CA, United States
Meg Ehm	GlaxoSmithKline, Brentford, United Kingdom
Dawn Waterworth	GlaxoSmithKline, Brentford, United Kingdom
Caroline Fox	Merck, Kenilworth, NJ, United States
Anders Malarstig	Pfizer, New York, NY, United States
Kathy Klinger	Sanofi, Paris, France
Kathy Call	Sanofi, Paris, France

University of Helsinki & Biobanks

Tomi Mäkelä	HiLIFE, University of Helsinki, Finland, Finland
Jaakko Kaprio	Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland, Finland
Petri Virolainen	Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland
Kari Pulkki	Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland
Terhi Kilpi	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Markus Perola	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Jukka Partanen	Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Anne Pitkäranta	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Riitta Kaarteenaho	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Seppo Vainio	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Kimmo Savinainen	Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland
Veli-Matti Kosma	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland
Urho Kujala	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland

Other Experts/ Non-Voting Members

Outi Tuovila	Business Finland, Helsinki, Finland
Minna Hendolin	Business Finland, Helsinki, Finland
Raimo Pakkanen	Business Finland, Helsinki, Finland

Scientific Committee

Pharmaceutical companies

Jeff Waring	Abbvie, Chicago, IL, United States
Bridget Riley-Gillis	Abbvie, Chicago, IL, United States
Athena Matakidou	Astra Zeneca, Cambridge, United Kingdom
Heiko Runz	Biogen, Cambridge, MA, United States
Jimmy Liu	Biogen, Cambridge, MA, United States
Shameek Biswas	Celgene, Summit, NJ, United States
Julie Hunkapiller	Genentech, San Francisco, CA, United States
Dawn Waterworth	GlaxoSmithKline, Brentford, United Kingdom

Meg Ehm	GlaxoSmithKline, Brentford, United Kingdom
Dorothee Diogo	Merck, Kenilworth, NJ, United States
Caroline Fox	Merck, Kenilworth, NJ, United States
Anders Malarstig	Pfizer, New York, NY, United States
Catherine Marshall	Pfizer, New York, NY, United States
Xinli Hu	Pfizer, New York, NY, United States
Kathy Call	Sanofi, Paris, France
Kathy Klinger	Sanofi, Paris, France
Matthias Gossel	Sanofi, Paris, France

University of Helsinki & Biobanks

Samuli Ripatti	Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland
Johanna Schleutker	Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland
Markus Perola	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Mikko Arvas	Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Olli Carpen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Reetta Hinttala	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Johannes Kettunen	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Reijo Laaksonen	Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland
Arto Mannermaa	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland
Juha Paloneva	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland
Urho Kujala	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland

Other Experts/ Non-Voting Members

Outi Tuovila	Business Finland, Helsinki, Finland
Minna Hendolin	Business Finland, Helsinki, Finland
Raimo Pakkanen	Business Finland, Helsinki, Finland

Clinical Groups

Neurology Group

Hilkka Soininen	Northern Savo Hospital District, Kuopio, Finland
Valtteri Julkunen	Northern Savo Hospital District, Kuopio, Finland
Anne Remes	Northern Ostrobothnia Hospital District, Oulu, Finland
Reetta Kälviäinen	Northern Savo Hospital District, Kuopio, Finland
Mikko Hiltunen	Northern Savo Hospital District, Kuopio, Finland
Jukka Peltola	Pirkanmaa Hospital District, Tampere, Finland
Pentti Tienari	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Juha Rinne	Hospital District of Southwest Finland, Turku, Finland
Adam Ziemann	Abbvie, Chicago, IL, United States
Jeffrey Waring	Abbvie, Chicago, IL, United States
Sahar Esmaeeli	Abbvie, Chicago, IL, United States
Nizar Smaoui	Abbvie, Chicago, IL, United States
Anne Lehtonen	Abbvie, Chicago, IL, United States
Susan Eaton	Biogen, Cambridge, MA, United States
Heiko Runz	Biogen, Cambridge, MA, United States
Sanni Lahdenperä	Biogen, Cambridge, MA, United States
Shameek Biswas	Celgene, Summit, NJ, United States
John Michon	Genentech, San Francisco, CA, United States
Geoff Kerchner	Genentech, San Francisco, CA, United States
Julie Hunkapiller	Genentech, San Francisco, CA, United States
Natalie Bowers	Genentech, San Francisco, CA, United States
Edmond Teng	Genentech, San Francisco, CA, United States
John Eicher	Merck, Kenilworth, NJ, United States

714	Vinay Mehta	Merck, Kenilworth, NJ, United States
715	Padhraig Gormley	Merck, Kenilworth, NJ, United States
716	Kari Linden	Pfizer, New York, NY, United States
717	Christopher Whelan	Pfizer, New York, NY, United States
718	Fanli Xu	GlaxoSmithKline, Brentford, United Kingdom
719	David Pulford	GlaxoSmithKline, Brentford, United Kingdom
720		
721	**Gastroenterology Group**	
722	Martti Färkkilä	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
723	Sampsa Pikkarainen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
724	Airi Jussila	Pirkanmaa Hospital District, Tampere, Finland
725	Timo Blomster	Northern Ostrobothnia Hospital District, Oulu, Finland
726	Mikko Kiviniemi	Northern Savo Hospital District, Kuopio, Finland
727	Markku Voutilainen	Hospital District of Southwest Finland, Turku, Finland
728	Bob Georgantas	Abbvie, Chicago, IL, United States
729	Graham Heap	Abbvie, Chicago, IL, United States
730	Jeffrey Waring	Abbvie, Chicago, IL, United States
731	Nizar Smaoui	Abbvie, Chicago, IL, United States
732	Fedik Rahimov	Abbvie, Chicago, IL, United States
733	Anne Lehtonen	Abbvie, Chicago, IL, United States
734	Keith Usiskin	Celgene, Summit, NJ, United States
735	Joseph Maranville	Celgene, Summit, NJ, United States
736	Tim Lu	Genentech, San Francisco, CA, United States
737	Natalie Bowers	Genentech, San Francisco, CA, United States
738	Danny Oh	Genentech, San Francisco, CA, United States
739	John Michon	Genentech, San Francisco, CA, United States
740	Vinay Mehta	Merck, Kenilworth, NJ, United States
741	Kirsi Kalpala	Pfizer, New York, NY, United States
742	Melissa Miller	Pfizer, New York, NY, United States
743	Xinli Hu	Pfizer, New York, NY, United States

Linda McCarthy	GlaxoSmithKline, Brentford, United Kingdom

Rheumatology Group

Kari Eklund	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Antti Palomäki	Hospital District of Southwest Finland, Turku, Finland
Pia Isomäki	Pirkanmaa Hospital District, Tampere, Finland
Laura Pirilä	Hospital District of Southwest Finland, Turku, Finland
Oili Kaipiainen-Seppänen	Northern Savo Hospital District, Kuopio, Finland
Johanna Huhtakangas	Northern Ostrobothnia Hospital District, Oulu, Finland
Bob Georgantas	Abbvie, Chicago, IL, United States
Jeffrey Waring	Abbvie, Chicago, IL, United States
Fedik Rahimov	Abbvie, Chicago, IL, United States
Apinya Lertratanakul	Abbvie, Chicago, IL, United States
Nizar Smaoui	Abbvie, Chicago, IL, United States
Anne Lehtonen	Abbvie, Chicago, IL, United States
David Close	Astra Zeneca, Cambridge, United Kingdom
Marla Hochfeld	Celgene, Summit, NJ, United States
Natalie Bowers	Genentech, San Francisco, CA, United States
John Michon	Genentech, San Francisco, CA, United States
Dorothee Diogo	Merck, Kenilworth, NJ, United States
Vinay Mehta	Merck, Kenilworth, NJ, United States
Kirsi Kalpala	Pfizer, New York, NY, United States
Nan Bing	Pfizer, New York, NY, United States
Xinli Hu	Pfizer, New York, NY, United States
Jorge Esparza Gordillo	GlaxoSmithKline, Brentford, United Kingdom
Nina Mars	Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland

Pulmonology Group

Tarja Laitinen	Pirkanmaa Hospital District, Tampere, Finland
Margit Pelkonen	Northern Savo Hospital District, Kuopio, Finland

Paula Kauppi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Hannu Kankaanranta	Pirkanmaa Hospital District, Tampere, Finland
Terttu Harju	Northern Ostrobothnia Hospital District, Oulu, Finland
Nizar Smaoui	Abbvie, Chicago, IL, United States
David Close	Astra Zeneca, Cambridge, United Kingdom
Steven Greenberg	Celgene, Summit, NJ, United States
Hubert Chen	Genentech, San Francisco, CA, United States
Natalie Bowers	Genentech, San Francisco, CA, United States
John Michon	Genentech, San Francisco, CA, United States
Vinay Mehta	Merck, Kenilworth, NJ, United States
Jo Betts	GlaxoSmithKline, Brentford, United Kingdom
Soumitra Ghosh	GlaxoSmithKline, Brentford, United Kingdom

Cardiometabolic Diseases Group

Veikko Salomaa	The National Institute of Health and Welfare Helsinki, Finland
Teemu Niiranen	The National Institute of Health and Welfare Helsinki, Finland
Markus Juonala	Hospital District of Southwest Finland, Turku, Finland
Kaj Metsärinne	Hospital District of Southwest Finland, Turku, Finland
Mika Kähönen	Pirkanmaa Hospital District, Tampere, Finland
Juhani Junttila	Northern Ostrobothnia Hospital District, Oulu, Finland
Markku Laakso	Northern Savo Hospital District, Kuopio, Finland
Jussi Pihlajamäki	Northern Savo Hospital District, Kuopio, Finland
Juha Sinisalo	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Marja-Riitta Taskinen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Tiinamaija Tuomi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Jari Laukkanen	Central Finland Health Care District, Jyväskylä, Finland
Ben Challis	Astra Zeneca, Cambridge, United Kingdom
Andrew Peterson	Genentech, San Francisco, CA, United States
Julie Hunkapiller	Genentech, San Francisco, CA, United States
Natalie Bowers	Genentech, San Francisco, CA, United States

John Michon	Genentech, San Francisco, CA, United States
Dorothee Diogo	Merck, Kenilworth, NJ, United States
Audrey Chu	Merck, Kenilworth, NJ, United States
Vinay Mehta	Merck, Kenilworth, NJ, United States
Jaakko Parkkinen	Pfizer, New York, NY, United States
Melissa Miller	Pfizer, New York, NY, United States
Anthony Muslin	Sanofi, Paris, France
Dawn Waterworth	GlaxoSmithKline, Brentford, United Kingdom

Oncology Group

Heikki Joensuu	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Tuomo Meretoja	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Olli Carpen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Lauri Aaltonen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Annika Auranen	Pirkanmaa Hospital District, Tampere, Finland
Peeter Karihtala	Northern Ostrobothnia Hospital District, Oulu, Finland
Saila Kauppila	Northern Ostrobothnia Hospital District, Oulu, Finland
Päivi Auvinen	Northern Savo Hospital District, Kuopio, Finland
Klaus Elenius	Hospital District of Southwest Finland, Turku, Finland
Relja Popovic	Abbvie, Chicago, IL, United States
Jeffrey Waring	Abbvie, Chicago, IL, United States
Bridget Riley-Gillis	Abbvie, Chicago, IL, United States
Anne Lehtonen	Abbvie, Chicago, IL, United States
Athena Matakidou	Astra Zeneca, Cambridge, United Kingdom
Jennifer Schutzman	Genentech, San Francisco, CA, United States
Julie Hunkapiller	Genentech, San Francisco, CA, United States
Natalie Bowers	Genentech, San Francisco, CA, United States
John Michon	Genentech, San Francisco, CA, United States
Vinay Mehta	Merck, Kenilworth, NJ, United States
Andrey Loboda	Merck, Kenilworth, NJ, United States

Aparna Chhibber	Merck, Kenilworth, NJ, United States
Heli Lehtonen	Pfizer, New York, NY, United States
Stefan McDonough	Pfizer, New York, NY, United States
Marika Crohns	Sanofi, Paris, France
Diptee Kulkarni	GlaxoSmithKline, Brentford, United Kingdom

Opthalmology Group

Kai Kaarniranta	Northern Savo Hospital District, Kuopio, Finland
Joni Turunen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Terhi Ollila	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Sanna Seitsonen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Hannu Uusitalo	Pirkanmaa Hospital District, Tampere, Finland
Vesa Aaltonen	Hospital District of Southwest Finland, Turku, Finland
Hannele Uusitalo-Järvinen	Pirkanmaa Hospital District, Tampere, Finland
Marja Luodonpää	Northern Ostrobothnia Hospital District, Oulu, Finland
Nina Hautala	Northern Ostrobothnia Hospital District, Oulu, Finland
Heiko Runz	Biogen, Cambridge, MA, United States
Erich Strauss	Genentech, San Francisco, CA, United States
Natalie Bowers	Genentech, San Francisco, CA, United States
Hao Chen	Genentech, San Francisco, CA, United States
John Michon	Genentech, San Francisco, CA, United States
Anna Podgornaia	Merck, Kenilworth, NJ, United States
Vinay Mehta	Merck, Kenilworth, NJ, United States
Dorothee Diogo	Merck, Kenilworth, NJ, United States
Joshua Hoffman	GlaxoSmithKline, Brentford, United Kingdom

Dermatology Group

Kaisa Tasanen	Northern Ostrobothnia Hospital District, Oulu, Finland
Laura Huilaja	Northern Ostrobothnia Hospital District, Oulu, Finland
Katariina Hannula-Jouppi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Teea Salmi	Pirkanmaa Hospital District, Tampere, Finland
Sirkku Peltonen	Hospital District of Southwest Finland, Turku, Finland
Leena Koulu	Hospital District of Southwest Finland, Turku, Finland
Ilkka Harvima	Northern Savo Hospital District, Kuopio, Finland
Kirsi Kalpala	Pfizer, New York, NY, United States
Ying Wu	Pfizer, New York, NY, United States
David Choy	Genentech, San Francisco, CA, United States
John Michon	Genentech, San Francisco, CA, United States
Nizar Smaoui	Abbvie, Chicago, IL, United States
Fedik Rahimov	Abbvie, Chicago, IL, United States
Anne Lehtonen	Abbvie, Chicago, IL, United States
Dawn Waterworth	GlaxoSmithKline, Brentford, United Kingdom

FinnGen Teams

Administration Team

Anu Jalanko	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Risto Kajanne	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Ulrike Lyhs	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Communication

Mari Kaunisto	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Analysis Team

Justin Wade Davis	Abbvie, Chicago, IL, United States
Bridget Riley-Gillis	Abbvie, Chicago, IL, United States
Danjuma Quarless	Abbvie, Chicago, IL, United States
Slavé Petrovski	Astra Zeneca, Cambridge, United Kingdom
Jimmy Liu	Biogen, Cambridge, MA, United States
Chia-Yen Chen	Biogen, Cambridge, MA, United States
Paola Bronson	Biogen, Cambridge, MA, United States

894	Robert Yang	Celgene, Summit, NJ, United States
895	Joseph Maranville	Celgene, Summit, NJ, United States
896	Shameek Biswas	Celgene, Summit, NJ, United States
897	Diana Chang	Genentech, San Francisco, CA, United States
898	Julie Hunkapiller	Genentech, San Francisco, CA, United States
899	Tushar Bhangale	Genentech, San Francisco, CA, United States
900	Natalie Bowers	Genentech, San Francisco, CA, United States
901	Dorothee Diogo	Merck, Kenilworth, NJ, United States
902	Emily Holzinger	Merck, Kenilworth, NJ, United States
903	Padhraig Gormley	Merck, Kenilworth, NJ, United States
904	Xulong Wang	Merck, Kenilworth, NJ, United States
905	Xing Chen	Pfizer, New York, NY, United States
906	Åsa Hedman	Pfizer, New York, NY, United States
907	Kirsi Auro	GlaxoSmithKline, Brentford, United Kingdom
908	Clarence Wang	Sanofi, Paris, France
909	Ethan Xu	Sanofi, Paris, France
910	Franck Auge	Sanofi, Paris, France
911	Clement Chatelain	Sanofi, Paris, France
912	Mitja Kurki	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad
913	Institute, Cambridge, MA, United States	
914	Samuli Ripatti	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
915	Mark Daly	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
916	Juha Karjalainen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad
917	Institute, Cambridge, MA, United States	
918	Aki Havulinna	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
919	Anu Jalanko	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
920	Kimmo Palin	University of Helsinki, Helsinki, Finland
921	Priit Palta	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
922	Pietro Della Briotta Parolo	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
923	Wei Zhou	Broad Institute, Cambridge, MA, United States

Susanna Lemmelä	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Manuel Rivas	University of Stanford, Stanford, CA, United States
Jarmo Harju	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Aarno Palotie	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Arto Lehisto	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Andrea Ganna	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Vincent Llorens	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Antti Karlsson	Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland
Kati Kristiansson	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Mikko Arvas	Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Kati Hyvärinen	Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Jarmo Ritari	Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Tiina Wahlfors	Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Miika Koskinen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland BB/HUS/Univ Hosp Districts
Olli Carpen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland BB/HUS/Univ Hosp Districts
Johannes Kettunen	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Katri Pylkäs	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Marita Kalaoja	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Minna Karjalainen	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Tuomo Mantere	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland

Eeva Kangasniemi	Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland
Sami Heikkinen	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland
Arto Mannermaa	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland
Eija Laakkonen	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland
Juha Kononen	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland

Sample Collection Coordination

Anu Loukola	Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Sample Logistics

Päivi Laiho	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Tuuli Sistonen	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Essi Kaiharju	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Markku Laukkanen	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Elina Järvensivu	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Sini Lähteenmäki	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Lotta Männikkö	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Regis Wong	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Registry Data Operations

Kati Kristiansson	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Hannele Mattsson	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Susanna Lemmelä	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Tero Hiekkalinna	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Manuel González Jiménez	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Genotyping

Kati Donner	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Sequencing Informatics

Priit Palta	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Kalle Pärn	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Javier Nunez-Fontarnau	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Data Management and IT Infrastructure

Jarmo Harju	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Elina Kilpeläinen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Timo P. Sipilä	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Georg Brein	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Alexander Dada	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Ghazal Awaisa	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Anastasia Shcherban	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Tuomas Sipilä	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Clinical Endpoint Development

Hannele Laivuori	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Aki Havulinna	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Susanna Lemmelä	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Tuomo Kiiskinen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Trajectory Team

Tarja Laitinen	Pirkanmaa Hospital District, Tampere, Finland
Harri Siirtola	University of Tampere, Tampere, Finland
Javier Gracia Tabuenca	University of Tampere, Tampere, Finland

Biobank Directors

Lila Kallio	Auria Biobank, Turku, Finland
Sirpa Soini	THL Biobank, Helsinki, Finland
Jukka Partanen	Blood Service Biobank, Helsinki, Finland
Kimmo Pitkänen	Helsinki Biobank, Helsinki, Finland
Seppo Vainio	Northern Finland Biobank Borealis, Oulu, Finland
Kimmo Savinainen	Tampere Biobank, Tampere, Finland
Veli-Matti Kosma	Biobank of Eastern Finland, Kuopio, Finland
Teijo Kuopio	Central Finland Biobank, Jyväskylä, Finland

REFERENCES

(1) 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. The American Journal of Human Genetics 2017;100(1):40-50.

(2) 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(5):1415-1429. e19.

(3) 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(12):1758.

(4) 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. BioRxiv 2019:660506.

(5) 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. Nature communications 2017;8:14357.

(6) Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature 2018;558(7708):73.

(7) 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. Nature communications 2016;7:11122.

(8) 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 genetics 2012;8(8):e1002907.

(9) 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(4):1137-1147.

(10) 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(7726):203.

(11) Kim S, Xing EP. Statistical estimation of correlated genome associations to a quantitative trait network. PLoS genetics 2009;5(8):e1000587.

(12) Ferreira MA, Purcell SM. A multivariate test of association. Bioinformatics 2008;25(1):132-133.

(13) 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(5):e34861.

(14) 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(13):1981-1989.

(15) 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(10):1493-1501.

(16) Lin J, Tabassum R, Ripatti S, Pirinen M. MetaPhat: Detecting and Decomposing Multivariate Associations From Univariate Genome-Wide Association Statistics. Frontiers in Genetics 2020;11:431.

(17) McInnes IB, Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. Nature Reviews Immunology 2007;7(6):429.

(18) 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(5):696-704.

(19) Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. Nature 2000;407(6801):249.

(20) Borodulin K, Vartiainen E, Peltonen M, Jousilahti P, Juolevi A, Laatikainen T, et al. Forty-year trends in cardiovascular risk factors in Finland. The European Journal of Public Health 2014;25(3):539-546.

(21) 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 genetics 2014;10(7):e1004494.

(22) Hail Team. Hail 0.2.13-81ab564db2b4. https://github.com/hail-is/hail/releases/tag/0.2.13 http://doi.org/10.5281/zenodo.2646680.

(23) Benner C, Havulinna A, Salomaa V, Ripatti S, Pirinen M. Refining fine-mapping: effect sizes and regional heritability. bioRxiv 2018:318618.

(24) 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(9):1335.

(25) Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, et al. Developing and Evaluating Mappings of ICD-10 and ICD-10-CM codes to Phecodes. BioRxiv 2019:462077.

(26) 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(D1):D1005-D1012.

(27) 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(6404):769-773.

(28) 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(1):44-51.

(29) Reeve MP, Kirby A, Wierzbowski J, Daly M, Hutz J. Target Gene Notebook: Connecting genetics and drug discovery. bioRxiv 2019:757690.

(30) 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(5).

(31) Stephens M. A unified framework for association analysis with multiple related phenotypes. PloS one 2013;8(7).

(32) 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(6099):1190-1195.

(33) 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(7662):173.

(34) Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. BioRxiv 2019:531210.

(35) Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31(13):3812-3814.

(36) Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics 2013;76(1):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):1831-1841.

(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-oncology 2014;17(1):116-121.

(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(3):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. Journal of Clinical Neuroscience 2001;8(4):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(4):938-948.

(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(5):352-360.

(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(1):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, The Journal of the American Society of Hematology 2012;119(11):2452-2457.

(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. The American Journal of Human Genetics 2006;78(2):253-264.

(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. Proceedings of the National Academy of Sciences 1986;83(4):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. Scientific reports 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(4):687-693.

(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(1):582-595.

(50) Brissett AE, Sherris DA. Scar contractures, hypertrophic scars, and keloids. Facial plastic surgery 2001;17(04):263-272.

(51) Lian N, Li T. Growth factor pathways in hypertrophic scars: molecular pathogenesis and therapeutic implications. Biomedicine & Pharmacotherapy 2016;84:42-50.

(52) Bonner JC. Regulation of PDGF and its receptors in fibrotic diseases. Cytokine Growth Factor Rev 2004;15(4):255-273.

(53) Kichaev G, Bhatia G, Loh P, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. The American Journal of Human Genetics 2019;104(1):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(4):271-274.

(55) OuYang Z, Hirota Y, Osuga Y, Hamasaki K, Hasegawa A, Tajima T, et al. Interleukin-4 stimulates proliferation of endometriotic stromal cells. The American journal of pathology 2008;173(2):463-469.

(56) Hsu C, Yang B, Wu M, Huang K. Enhanced interleukin-4 expression in patients with endometriosis. Fertil Steril 1997;67(6):1059-1064.

(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(7):1517-1524.

(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. European Journal of Obstetrics & Gynecology and Reproductive Biology 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. The Journal of Clinical Endocrinology & Metabolism 2016;102(2):350-353.

(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(3):422-427.

(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. bioRxiv 2019:544445.

(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(10):1344-1351.

(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. European Journal of Human Genetics 2017;25(7):869.

FIGURES

Figure 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.

Figure 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 grey dots do not by either method. Respective numbers are reported in the accompanying table.

Figure 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.

TABLES

Table 1. Characterization of the 12 inflammatory biomarker measurements. n = sample size, SD = standard deviation. The cytokine concentrations are pg/ml.

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

* missense variant

▲ variant was in high linkage disequilibrium (r^2 = 0.997) with a missense variant (chr1:g.169529737T>C, rs6030)

** variant 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.

[a] Bolded variants are lead variants. Genomic positions are mapped to Genome Assembly GRCh38.

[b] AF = 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.

[c] Driver traits can only be determined for those variants with a genome-wide significant association in the multivariate analysis.

[d] Previous 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.

[e] Posterior probability of being causal, determined by FINEMAP.

[f] Only associations that remain significant after conditioning are reported here. Closely related disease diagnoses are represented in a shared cell and their replication is assessed jointly.

[g] Novelty of disease associations was assessed at gene-level.