

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

3 Sanni E. Ruotsalainen^{1*}, Juulia J. Partanen^{1*}, Anna Cichonska^{1,2,3}, Jake Lin¹, Christian Benner¹, Ida Surakka⁴,
4 *FinnGen*, Mary Pat Reeve¹, Priit Palta^{1,5}, Marko Salmi^{6,7}, Sirpa Jalkanen^{6,7}, Ari Ahola-Olli^{1,8,9}, Aarno
5 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**}

6

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

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

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

11 4) Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA

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

13 6) MediCity Research Laboratory, University of Turku, Turku, Finland

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

15 8) The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

16 9) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School,
17 Boston, Massachusetts, USA

18 10) Finnish Institute for Health and Welfare, Helsinki, Finland

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

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

22

23 *authors contributed equally

24 **Corresponding author

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

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

27

28 **ABSTRACT**

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

47

48

49 **INTRODUCTION**

50 Genome-wide association studies (GWAS) of biomarkers have been highly successful in identifying
51 novel biological pathways and their impact on health and disease. Biomarkers increase statistical
52 power in GWAS, compared to disease diagnoses, due to their quantitative nature and lack of errors

53 due to subjectivity, such as misclassification. Thus, biomarker GWAS have identified thousands of
54 biomarker-associated loci and elucidated the mechanisms underlying numerous disease associations¹⁻
55 ³. A recent study on 38 biomarkers in the UK Biobank (UKBB) identified over 1,800 independent
56 genetic associations with causal roles in several diseases⁴. Proteomics and metabolomics integrated
57 with genomics has also revealed causal molecular pathways connecting the genome to multiple
58 diseases, e.g. autoimmune disorders and cardiovascular disease⁵⁻⁸. Although biomarkers are more
59 closely related to pathophysiology, a single biomarker is usually an inaccurate estimator of complex
60 disease due to phenotypic heterogeneity and individual variation. Therefore, combinations of
61 biomarkers provide a more robust predictive molecular signature. Studies examining combinations
62 of biomarkers are increasingly feasible given the availability of biobank resources around the globe
63 with deep phenotyping, i.e. precise and comprehensive data on phenotypic variation including
64 quantitative measures such as biomarkers^{9,10}.

65

66 Multivariate GWAS of correlated traits increases statistical power compared to univariate analysis,
67 especially in the case of complex biological processes and correlated traits^{8,11,12}. This leads to
68 identifying multivariate associations that are otherwise missed by univariate analysis^{8,13}. Efficient
69 software programs are available for performing multivariate GWAS such as metaCCA¹⁴, yet
70 multivariate analyses currently have shortcomings in interpreting the arising signals. Follow-up tools
71 for fine-mapping causal variants within the associated loci are lacking and the subset of tested traits
72 that drive the association signals have not been identified. These shortcomings are largely due to the
73 lack of a multivariate counterpart to the univariate regression coefficients (beta estimates). Lack of
74 these necessary follow-up tools has hindered the utilization of multivariate methods.

75

76 In this study, we developed a novel computational workflow for multivariate GWAS discovery and
77 follow-up analyses including fine-mapping and identification of driver traits (Figure 1). Our

78 workflow includes 1) a customized version of the metaCCA software that overcomes the problem of
79 missing beta estimates by turning each multivariate association into its optimal univariate Linear
80 Combination Phenotype (LCP), enabling an LCP-GWAS, 2) fine-mapping, i.e. identifying putative
81 causal variants underlying each association using summary statistics from the LCP-GWAS and a
82 multivariate extension to FINEMAP¹⁵, and 3) determining the traits driving each multivariate
83 association using a newly developed tool, MetaPhat¹⁶ that efficiently decomposes the multivariate
84 associations into a smaller set of underlying driver traits. Taken together, we present to our knowledge
85 the first comprehensive framework to map multivariate associations into individual causal variants
86 and a subset of driver traits. We demonstrate the potential of our workflow in a Finnish population-
87 based cohort with 12 inflammatory biomarkers implicated in the pathogenesis of autoimmune
88 disorders and cancer¹⁷⁻¹⁹. This set of highly-correlated biomarkers is particularly advantageous for
89 multivariate analysis as high correlation between traits increases the boost in statistical power
90 achieved by multivariate methods. Using multivariate analysis, we identify additional hits compared
91 to univariate analysis, totaling 11 independent associations. We follow them up in a phenome-wide
92 association study (PheWAS) in the FinnGen study (n = 176,899) across 2,367 disease endpoints and
93 in the UKBB (n = 408,910)¹⁰. We discover multiple disease associations, as well as identify
94 orthogonal evidence for the biological impact of the causal variants through several expression
95 quantitative trait loci (eQTLs) and protein quantitative trait loci (pQTLs) within the multivariate loci.

96

97

98 **MATERIALS AND METHODS**

99 **Study cohort and data**

100 We studied 12 highly correlated inflammatory biomarkers in the population-based national FINRISK
101 Study collected in 1997 (n = 6,890)²⁰ (Table 1, Supplementary Figure 1). The FINRISK Study is a
102 large Finnish population survey of risk factors for chronic, non-communicable diseases, and it has

103 been collected by independent random population sampling every five years beginning in 1972 with
104 multiple recruiting waves. The 12 inflammatory biomarkers included five interleukins (IL-4, IL-6,
105 IL-10, IL-12p70, IL-17), three growth factors (FGF2, PDGF-BB, VEGF-A), one colony-stimulating
106 factor (G-CSF), one interferon (IFN- γ), one chemokine (SDF-1 α), and one tumor necrosis factor
107 (TNF- β) (Table 1, Supplementary Figure 1. Hierarchical clustering identified the cluster of 12
108 inflammatory biomarkers out of 66 quantitative traits of cardiometabolic or immunologic relevance
109 (Supplementary Figure 2 Supplementary Table 1 and Supplementary Methods). The 66 quantitative
110 traits were measured as previously described^{11,20,21}.

111

112 **Genotyping, imputation and quality control**

113 Samples were genotyped using multiple different genotyping chips, for which the quality control
114 (QC), phasing and imputation were done in multiple chip-wise batches. Imputation of the genotypes
115 was done utilizing a Finnish population-specific reference panel of 3,775 high-coverage whole-
116 genome sequences. Genotype imputation was followed by post-imputation sample QC
117 (Supplementary Methods) and variant QC (imputation INFO > 0.8, minor allele frequency > 0.002
118 and Hardy-Weinberg equilibrium p-value > 1×10^{-6}). A total of 26,717 samples and 11,329,225
119 variants passed this rigorous quality control.

120

121 **Univariate and multivariate GWAS**

122 Univariate genome-wide association analyses for the biomarkers were performed using a linear mixed
123 model implemented in Hail²², adjusting for age, sex, genotyping chip, first ten principal components
124 of genetic structure and the genetic relationship matrix (GRM) (Supplementary Methods). The GRM
125 was estimated using 73K independent high-quality genotyped variants (Supplementary Methods).
126 We performed multivariate GWAS on the biomarkers using metaCCA¹⁴, software that performs

127 multivariate analysis by implementing Canonical Correlation Analysis (CCA) for a set of univariate
128 GWAS summary statistics.

129

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

133

$$134 \quad X^* = \mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \dots + a_px_p$$

135 and

$$136 \quad Y^* = LCP = \mathbf{b}'\mathbf{y} = b_1y_1 + b_2y_2 + \dots + b_qy_q,$$

137

138 then finding the linear combination of the predictor variables that are maximally correlated with the
139 linear combination of the response variables corresponds to finding vectors \mathbf{a} and \mathbf{b} that maximize

140

$$141 \quad r = \frac{(X\mathbf{a})'(Y\mathbf{b})}{\|X\mathbf{a}\|\|Y\mathbf{b}\|} = \frac{\mathbf{a}'\Sigma_{xy}\mathbf{b}}{\sqrt{\mathbf{a}'\Sigma_{xx}\mathbf{a}}\sqrt{\mathbf{b}'\Sigma_{yy}\mathbf{b}}}$$

142

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

147

148 **Novel multivariate LCP-GWAS method**

149 To enable follow-up analyses of multivariate GWAS results, such as fine-mapping, we developed a
150 novel method to produce linear combination phenotypes (LCP) at the single variant level by

151 extending the functionality of metaCCA. The updated metaCCA is available online at:
152 <https://github.com/acichonska/metaCCA>.

153

154 LCPs were constructed as the weighted sum of the trait residuals, where the weights ($\mathbf{b} = [b_1, b_2 \dots,$
155 $b_q]$) were chosen to maximize the correlation between the resulting linear combination of traits and
156 the genotypes at the variant. We determined association regions by adding 1Mb to each variant
157 reaching genome-wide significance (GWS; $p\text{-value} < 5 \times 10^{-8}$) in the multivariate analysis and joining
158 overlapping regions. We constructed LCPs for the lead variant, i.e. the variant with the smallest p -
159 value, in each of these regions, as a univariate representation of the multivariate association in that
160 region. Next, we performed chromosome-wide LCP-GWAS for the constructed LCPs in a similar
161 manner as for each of the biomarkers.

162

163 **Fine-mapping multivariate associations**

164 We used FINEMAP^{15,23} on the LCP-GWAS summary statistics to identify causal variants underlying
165 the multivariate associations. FINEMAP analyses were restricted to a $\pm 1\text{Mb}$ region around the GWS
166 variants from the LCP-GWAS.

167

168 We assessed variants in the top 95% credible sets, i.e. the sets of variants encompassing at least 95%
169 of the probability of being causal (causal probability) within each causal signal conditional on other
170 causal signals in the genomic region. Within these sets we excluded those sets that did not clearly
171 represent one signal, determined by low minimum linkage disequilibrium (LD, $r^2 < 0.1$).

172

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

176

177 **Identifying driver traits**

178 We determined the traits driving the multivariate associations for the putative causal variants
179 suggested by fine-mapping using the MetaPhat software developed in-house¹⁶. MetaPhat determines
180 the set of driver traits for each multivariate association by performing multivariate testing using
181 metaCCA iteratively on subsets of the traits, excluding one trait at a time until a single trait remains.
182 At each iteration, the trait to be excluded is the one whose exclusion leads to the highest p-value for
183 the remaining subset of traits. The driver traits are determined as a set of traits that have been removed
184 when the multivariate p-value becomes non-significant ($p > 5 \times 10^{-8}$). The interpretation is that the
185 driver traits make the multivariate association significant.

186

187 **Phenome-wide association testing in FinnGen and UKBB**

188 We performed a PheWAS in the FinnGen study for variants suggested to be causal by multivariate
189 fine-mapping and for multivariate GWS functional variants (Table 2 and Supplementary Table 2.
190 FinnGen (<https://www.finnngen.fi/en>) is a large biobank study that aims to genotype 500,000 Finns
191 and combine this data with longitudinal registry data, including national hospital discharge, death,
192 and medication reimbursement registries, using unique national personal identification numbers.
193 FinnGen includes prospective epidemiological and disease-based cohorts as well as hospital biobank
194 samples. A total of 176,899 samples from FinnGen Data Freeze 4 with 2,444 disease endpoints were
195 analyzed using Scalable and Accurate Implementation of Generalized mixed model (SAIGE), which
196 uses saddlepoint approximation (SPA) to calibrate unbalanced case-control ratios²⁴. Additional
197 details and information on the genotyping and imputation are provided in the Supplementary Material
198 and contributors of FinnGen are listed in the Acknowledgements.

199

200 FinnGen disease associations with p-values $< 1 \times 10^{-4}$ were considered significant. We tested the p-
201 value threshold by sampling 1,000 allele frequency-matched sets of n variants, where n represents
202 the number of variants of interest, from 8.2 million non-coding variants and determining a null
203 distribution of the number of FinnGen associations passing the p-value threshold. We confirmed the
204 validity of the p-value threshold by comparing the observed number of FinnGen associations passing
205 the p-value threshold to the null distribution (Supplementary Figure 3). We excluded disease
206 endpoints within the ICD-10 (International Statistical Classification of Diseases and Related Health
207 Problems 10th Revision) chapters XXI and XXII from PheWAS analyses, resulting in 2,367 disease
208 endpoints analyzed. To assess the relevance of the putative causal variants and the functional variants
209 for their disease associations in FinnGen, the disease associations were conditioned on the variant
210 with the strongest FinnGen disease association within the locus (± 0.5 MB of the putative causal
211 variant or functional variant). Finally, we assessed replication of the disease associations in the
212 UKBB, where associations with p-values < 0.05 were considered replicated given that the direction
213 of effects were coherent. Phecodes from the UKBB were mapped to ICD-10 diagnosis codes using
214 the PheCode map 1.2²⁵. The NHGRI-EBI GWAS Catalog²⁶ was used for assessing the novelty of the
215 observed genetic associations.

216

217 We also explored whether the fine-mapped putative causal variants or variants in LD with them ($r^2 >$
218 0.6) had previously been reported as eQTLs or pQTLs. For eQTLs, we looked at overlap with LD-
219 pruned associations derived from the Genotype-Tissue Expression (GTEx) Portal and pQTLs were
220 included from studies by Suhre⁵, Sun⁶, Emilsson²⁷ and Sasayama²⁸; regional overlap and architecture
221 were visualized in Target Gene Notebook²⁹.

222

223

224 **RESULTS**

225 **Comparison of multivariate and univariate GWAS of 12 inflammatory biomarkers**

226 We first tested for genome-wide associations of 12 highly correlated inflammatory biomarkers (Table
227 1, Supplementary Figure 1) measured in 6,890 FINRISK study participants using both multivariate
228 and univariate methods. Pearson correlations between the biomarkers ranged from 0.64 to 0.93, with
229 a mean of 0.80. Out of the 11,329,225 variants tested, 190 were significantly associated using both
230 univariate and multivariate analyses, 999 (in 11 loci) only by the multivariate analysis and two only
231 by the univariate analysis using a Bonferroni-corrected p-value threshold of $5 \times 10^{-8}/12$ (Figure 2).
232 The two variants that were significant only in the univariate analysis were both located in a locus
233 (*JMJD1C*) that was found to be significant also by the multivariate analysis. A total of 1,189 variants
234 reached the significance threshold in the multivariate analysis compared to only 192 in the univariate
235 analysis, reflecting a considerable increase in statistical power achieved by the multivariate analysis
236 while preserving the Type I error rate (Supplementary Figure 4). The corresponding genomic inflation
237 factor λ was 1.036, with no evidence of concerning genomic inflation due to multivariate analysis
238 using Canonical Correlation Analysis.

239
240 Within the 1,189 genome-wide significant variants in the multivariate analysis, we identified 11
241 independently associated loci (Figure 3 and Supplementary Figure 5), four of which (*F5*, *C1orf140*,
242 *PDGFRB* and *ABO*) were not detected by univariate analyses corrected for multiple testing (Figure
243 3). Eight of the 11 loci had previously been associated with at least one of the 12 biomarkers in the
244 NHGRI-EBI GWAS catalog while three loci (*F5*, *C1orf140* and *PDGFRB*) were novel.

245
246 At several loci, multivariate analysis revealed more plausible candidates for causal variants than the
247 univariate analyses. For example, in the *CIQA* locus, an association with only one of the 12
248 biomarkers, TNF- β , was noted in the univariate results. The lead variant in the TNF- β univariate
249 GWAS was rs78655189 ($p = 2.2 \times 10^{-24}$), an intronic variant in the *EPHB2* gene. In contrast, the lead

250 variant for the same locus in the multivariate analysis was rs17887074 ($p = 1.2 \times 10^{-73}$), a Finnish-
251 enriched missense variant located in the *CIQA* gene. The *CIQA* gene has been previously associated
252 with immunologic diseases, such as immunodeficiency and systemic lupus erythematosus³⁰. Our
253 multivariate analysis may point towards a plausible mechanism underlying these associations via
254 TNF- β levels. Further, in the *VLDLR* locus, univariate fine-mapping of VEGF, the only associated
255 biomarker, suggested that the lead variant rs2375981 from the multivariate analysis was more likely
256 causal than the lead variant rs10967570 from the VEGF univariate analysis (posterior probabilities
257 1.0 and 0.025, respectively).

258

259 **Fine-mapping multivariate GWAS results**

260 To identify the causal variants of the multivariate associations, we studied the likelihood of multiple
261 variants contributing to the association signal in the 11 associated loci using FINEMAP²³. Our novel
262 multivariate LCP-GWAS method based on linear combinations calculated for each locus using
263 multivariate metaCCA results enabled fine-mapping of the multivariate results. The number of
264 credible sets varied from one to four for the multivariate associated loci (Supplementary Table 3),
265 resulting in a total of 19 independent sets of variants considered putatively causal. All 183 variants
266 within the 19 credible sets are available in Supplementary Table 3 and posterior probabilities for
267 different numbers of causal signals for each locus are available in Supplementary Table 4.

268

269 Among each of the 19 sets, the variant with the highest the causal probability (initial representative
270 variant) was chosen to represent the set, unless the set contained a functional variant (missense, splice-
271 region or frameshift) in high LD ($r^2 > 0.95$) with the initial representative variant, in which case the
272 functional variant was chosen (Table 2 and Supplementary Figure 6). This was the case for one
273 credible set in the *F5* locus where the missense variant rs9332701 (causal probability 46.1%) replaced
274 the initial representative non-coding variant rs61808983 (causal probability 53.3%) as they were in

275 high LD ($r^2 = 0.996$). We also assessed whether the causal probabilities changed in the *F5* credible
276 set if the LCP was generated for the missense variant rs9332701 rather than the lead variant
277 rs61808983. This had no notable effects on the causal probabilities (46.1% vs. 48.5%, 53.3% vs.
278 51.5% for rs9332701 and rs61808983, respectively).

279

280 The 19 representative variants, hereon referred to as the putative causal variants, included all except
281 one (rs11637184 in *PCSK6* locus) of the 11 lead variants from multivariate GWAS. In the *PCSK6*
282 locus one of the four putative causal variants (rs111482836) was associated with disease in FinnGen,
283 whereas the lead variant was not, highlighting the importance of fine-mapping multivariate GWAS
284 results.

285

286 Fine-mapping suggested at least as many causal signals as there were conditional rounds in stepwise
287 conditional analysis ($n = 16$), thus verifying the results from FINEMAP. Further, 13 of the 19 (68,4%)
288 putative causal variants were also conditioned on in the conditional analysis (Supplementary Table
289 5). The main benefit of fine-mapping is the probabilistic quantification of causality for each variant
290 in the region, which is crucial information when there are several plausible candidates for causal
291 variants. Such metrics are not available from stepwise conditional analysis.

292

293 **Functional coding variants**

294 GWAS hits are generally non-coding, although concentrated in regulatory regions³¹, and enrichment
295 of functional coding variants has been seen mainly only after fine-mapping e.g. in inflammatory
296 bowel disease³². We, however, observed enrichment of functional coding variants in the multivariate
297 GWAS hits already prior to fine-mapping. Two of the 19 putative causal variants were missense
298 variants (rs17887074 and rs199588110, in the *CIQA* and *GP6* loci respectively). These two variants
299 (2/19, 10.5%) were enriched (>1.5-fold) in Finns compared to non-Finnish, Swedish, Estonian

300 Europeans (NFSEE) in the gnomAD genome reference database³³. Considering all genome-wide
301 significant variants in the multivariate GWAS, we found 13 functional variants (missense, splice-
302 region and frameshift variants) with at least one functional variant in five of the 11 multivariate loci
303 (*CIQA*, *F5*, *C1orf140*, *SERPINE2*, and *GP6*; Supplementary Table 2). Out of the 13 functional
304 variants, 11 were missense variants, one was a splice region variant and one a frameshift variant. A
305 total of six (46.2%) of the 13 variants were enriched in the Finnish population, highlighting the
306 potential of utilizing isolated populations in GWAS.

307

308 We studied whether the multivariate genome-wide significant variants and variants identified by fine-
309 mapping were enriched for functional variants including missense, splice-region and frameshift
310 variants compared to the 11.3M variants analyzed. P-values for enrichment were calculated using the
311 χ^2 -test for the number of functional or missense variants within the variants assessed against the
312 number of the corresponding subset of variants within all variants tested. The multivariate genome-
313 wide significant variants were enriched for missense variants and functional variants including
314 missense, splice-region and frameshift variants (2.2-fold, $p = 0.015$, and 1.9-fold, $p = 8.8 \times 10^{-4}$,
315 respectively). The 19 putative causal variants were further enriched for both missense variants and
316 the broader set of functional variants (37-fold, $p = 1.3 \times 10^{-17}$, and 28-fold, $p = 1.4 \times 10^{-17}$, respectively)
317 as were the 183 variants in the credible sets (3.9-fold, $p = 0.050$, and 2.9-fold, $p = 0.050$, respectively).

318

319 **Identifying driver traits**

320 Next, we studied which traits were driving the multivariate associations in each of the 11 loci using
321 metaPhat¹⁶. The number of driver traits for each of the 11 loci varied between one and all 12. The
322 driver traits were very much in line with the univariate results; the most significantly associated
323 biomarkers in the univariate GWAS were typically included among the driver traits (Table 2). In loci
324 with multiple putative causal variants, driver traits for the variants were generally subsets of the lead

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

329

330 **Disease implications of the multivariate loci**

331 Finally, we tested how the 19 putative causal variants as well as the 13 genome-wide significant
332 functional variants in the 11 loci associated with disease risk among 2,367 disease endpoints defined
333 in FinnGen. Altogether, 53 disease associations were observed with seven putative causal variants.
334 Two of these variants did not lead the multivariate associations at the 11 loci and thus would have
335 gone unnoticed without fine-mapping. Five genome-wide significant functional variants not
336 overlapping with the putative causal variants had an additional 35 disease associations.

337

338 To assess the relevance of the putative causal variants and the functional variants for their disease
339 associations in FinnGen, the disease associations were conditioned on the variant with the strongest
340 FinnGen disease association within the locus. In 13 of the 53 FinnGen disease associations with the
341 putative causal variants, the putative causal variant or a variant in near perfect LD ($r^2 > 0.95$) led the
342 association signal or remained significant after conditioning. Correspondingly, for the functional
343 variants not overlapping with the putative causal variants 18 of the 35 disease associations were either
344 led by the functional variant or a variant in near perfect LD or remained significant after conditioning.
345 We also tested the disease associations in the UKBB, where associations with p-values < 0.05 were
346 considered replicated given that the direction of effects were coherent (Supplementary Table 6).

347

348 In addition to disease associations, we explored whether the putative causal variants or variants in
349 LD with them ($r^2 > 0.6$) had previously been reported as eQTLs or pQTLs. Several reported eQTLs

350 and pQTLs⁵ in the 11 loci provided orthogonal evidence for the biologically relevant functions of the
351 putative causal variants (Supplementary Table 7).

352

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

358

359 ***GP6* gene locus**

360 *Multivariate association and FinnGen disease associations*

361 The Finnish enriched rare missense variant rs199588110 (AF = 0.33%, 3.7-fold enrichment),
362 predicted deleterious by SIFT³⁴ and probably damaging by Polyphen³⁵, was suggested causal in the
363 *GP6* locus. In FinnGen it led the association with benign neoplasms of meninges (OR = 6.4, $p =$
364 4.9×10^{-5}). The association was not replicated in the UKBB, although this may be due to impaired
365 power as the AF of the Finnish enriched variant in the UKBB (0.036%) was roughly a tenth of its AF
366 in FinnGen, and an inadequate match of the discovery and replication phenotypes as UKBB
367 phenotype definitions included all benign neoplasms of the brain and spinal cord and were not
368 restricted to neoplasms of the meninges.

369

370 *Driver traits*

371 All 12 biomarkers were considered driver traits of the multivariate association. Cytokines, including
372 many of the 12 biomarkers studied (e.g. IL-6, IL-4, PDGF-BB and VEGF-A), have been implicated
373 in the autocrine regulation of meningioma cell proliferation and motility³⁶⁻³⁹. Further, higher
374 expression levels of both PDGF-BB and VEGF occur in atypical and malignant meningiomas than in

375 benign meningiomas^{39,40} and microvascular density regulated by VEGF has been linked with time to
376 recurrence⁴¹. Several phase II clinical trials have tested therapies targeting VEGF and PDGF-BB
377 signaling pathways as treatments for recurrent or progressive meningiomas³⁷ with promising results
378 for two multifunctional tyrosine kinase inhibitors, sunitinib and PTK787/ZK 222584 that inhibit both
379 VEGF and PDGF receptors^{37,42}.

380

381 ***SERPINE2* gene locus**

382 *Multivariate association and FinnGen disease associations*

383 The *SERPINE2* locus was the locus with the most significant association in the multivariate analysis
384 ($p < 1 \times 10^{-324}$). Three variants (rs13412535, rs58116674 and rs7578029) were suggested causal
385 (putative causal variants). One of them, the intronic lead variant rs13412535 from the multivariate
386 analysis, increased the risk of hypertrophic scars (OR = 1.3, $p = 7.5 \times 10^{-5}$) and was in very high LD
387 with the variant that led the disease association in FinnGen (rs68066031, $r^2 = 0.99$). The association
388 was not replicated in the UKBB and had not been previously reported at gene-level. Nonetheless, the
389 variant in question had an association with another hypertrophic skin disorder, acquired keratoderma
390 (OR = 1.5, $P = 0.02$) in the UKBB.

391

392 *Previous knowledge of gene function and driver traits*

393 The *SERPINE2* gene encodes protease nexin-1, a protein in the serpin family of proteins that inhibits
394 serine proteases, especially thrombin, and has therefore been implicated in coagulation and tissue
395 remodeling⁴³. The gene has been associated with chronic obstructive pulmonary disease and
396 emphysema⁴⁴. As previously reported, *SERPINE2* has been shown to inhibit extracellular matrix
397 degradation⁴⁵ and overexpression of *SERPINE2* has been shown to contribute to pathological cardiac
398 fibrosis in mice⁴⁶. Additionally, serine protease inhibitor genes including *SERPINE2* have been noted
399 to be heavily induced during wound healing⁴⁷. According to GTEx the *SERPINE2* gene is most highly

400 expressed in fibroblasts. Further, inflammation plays an important role in hypertrophic scar formation
401 and cytokines including PDGF and VEGF are dysregulated in hypertrophic scars⁴⁸. The lead variant
402 had genome-wide significant associations with 11 of the 12 biomarkers and all 12 were regarded as
403 driver traits of the association.

404

405 *eQTLs and pQTLs*

406 The lead variant (rs13412535) is a pQTL impacting one of the driver traits, PDGF-BB levels, and an
407 intronic variant rs68066031 in high LD ($r^2 = 0.99$) with the lead variant is a pQTL for SERPINE2^{6,27}
408 . PDGF is considered essential in wound repair⁴⁹ and growth factors including PDGF are considered
409 key players in the pathogenesis of hypertrophic scars⁵⁰. PDGF enhances pathologic fibrosis in several
410 tissues such as skin, lung, liver and kidney by means of mitogenic and chemoattractant actions on the
411 principal collagen-producing cell type, myofibroblasts, as well as stimulation of collagen
412 production⁵¹.

413

414 ***ABO* gene locus**

415 *Multivariate association and FinnGen disease associations*

416 An association with the *ABO* locus was only detected by multivariate analysis (minimum univariate
417 $p = 2.1 \times 10^{-5}$ for the lead variant from multivariate analysis). One variant, the intronic lead variant
418 rs550057 (aka rs879055593) from multivariate analysis ($p = 8.5 \times 10^{-14}$) was suggested causal and was
419 associated with 45 endpoints in FinnGen, such as endometriosis, heart failure and statin usage. Most
420 of these associations resulted from LD to other stronger regional associations, however, nine
421 remained significant after conditioning on other lead variants within the *ABO* locus, including risk an
422 increasing effect on anemias, for which rs550057 lead the genome-wide significant association signal
423 ($p = 4.7 \times 10^{-8}$), visual field disturbances ($p < 6.5 \times 10^{-5}$) and diseases of the ear and mastoid process (p
424 $= 4.8 \times 10^{-5}$). Replication of only two of the nine associations (other anemias and visual field defects)

425 could be attempted in the UKBB due to poor phenotype matching and did not replicate; however,
426 bearing relevance to the genome-wide significant finding in anemia, rs550057 led the association
427 with red blood cell count in the UKBB ($p = 1.3 \times 10^{-212}$).⁵²

428

429 *Driver traits*

430 IL-4 was the only driver trait of the multivariate association and has been implicated in the
431 pathogenesis of many of the diseases associated with the locus. Aplastic anemia is considered to result
432 primarily from immune-mediated bone marrow failure and an imbalance in Type I versus Type II T-
433 cells that secrete IL-4 among other cytokines has been reported⁵³. In endometriosis, IL-4 levels have
434 been shown to be upregulated and induce the proliferation of endometriotic stromal cells^{54,55}.

435

436 *eQTLs and pQTLs*

437 The lead variant rs550057 is a pQTL impacting the levels of four proteins: ALPI, CHST15,
438 FAM177A1 and JAG1⁶. Two of these proteins, carbohydrate sulfotransferase 15 (CHST15) and
439 Jagged1 (JAG1), have been implicated in the pathogenesis of diseases associated with the locus. A
440 small-interfering RNA targeting CHST15 improved myocardial function as well as reduced cardiac
441 fibrosis, hypertrophy and secretion of proinflammatory cytokines in rats with chronic heart failure⁵⁶.
442 Upregulation of JAG1 has been reported in the endometrium of patients with endometriosis compared
443 to controls⁵⁷. Alagille Syndrome mainly caused by mutations in the JAG1 gene, is accompanied by
444 congenital heart defects and varying degrees of hypercholesterolemia⁵⁸

445

446 ***F5 gene locus***

447 *Multivariate association*

448 An association with the *F5* locus was only detected by multivariate analysis (minimum univariate p
449 = 1.1×10^{-3} for the lead variant from multivariate analysis) and the locus had not been previously

450 associated with any of the biomarkers. The locus included two putative causal variants, rs3820060
451 and rs9332701, out of which the former was the lead variant from multivariate analysis ($p = 6.15 \times 10^{-20}$).
452

453

454 *FinnGen disease associations*

455 Three genome-wide significant missense variants in the *F5* locus (rs4524, rs4525, rs6032), all in high
456 LD with one another ($r^2 > 0.98$), were associated with nine diseases in FinnGen with four of these
457 associations remaining significant after conditioning. Three of the four associations were protective
458 for venous thromboembolism (VTE)-related endpoints ($p < 6.9 \times 10^{-5}$) and one increased the risk of
459 fluid and electrolyte balance disruption, more specifically hypo-osmolality and hyponatraemia ($p =$
460 9.5×10^{-5} , Supplementary Table 2). We replicated a previously reported protective effect of the
461 missense variant rs4524 on VTE⁵⁹ that remained significant after conditioning on factor V Leiden
462 (rs6025; $p = 1.5 \times 10^{-11}$), a missense variant with a well-known risk-increasing effect on VTE⁶⁰, while
463 the hypo-osmolality and hyponatraemia association was novel. The VTE-related associations were
464 replicated in the UKBB. A fourth missense variant in the locus (rs6027) increased the risk of four
465 VTE-related diseases ($p < 2.4 \times 10^{-5}$), all of which remained significant after conditioning on the
466 variant with the strongest association in the locus. These associations were not replicated in the
467 UKBB.

468

469 *Driver traits*

470 The multivariate association in this locus had two driver traits: IL-4 and IL-12 both of which are
471 relevant for coagulation as IL-12 has been shown to activate coagulation⁶¹ and cross-talk between the
472 inflammatory and coagulation systems is extensive⁶².

473

474 *eQTLs and pQTLs*

475 The rs3820060 variant was an eQTL for the *F5* and *NME7* genes and was in the same LD-block (r^2
476 > 0.6) as many pQTLs affecting *SEC13*, *NPTX2*, *SIG11*, *CAMK1*, and *TFPI* levels. This block also
477 included the three highly-correlated genome-wide significant missense variants mentioned above.
478 Tissue factor pathway inhibitor (TFPI) is a major antithrombotic protein that inhibits thrombin and
479 the external coagulation pathway. Low levels of TFPI increase the risk of venous thrombosis⁶³ and
480 TFPI has been shown to interact with the two driver traits IL-4 and IL-12^{64,65}. The other causal
481 variant rs9332701 was a pQTL for *F5*⁵ and was in high LD ($r^2 = 0.97$) with an eQTL for *NME7* and
482 a pQTL for *EHBP1*⁶.

483

484

485 **DISCUSSION**

486 We developed a novel method for multivariate GWAS follow-up analyses and demonstrated the
487 considerable boost in power provided by multivariate GWAS using 12 highly correlated
488 inflammatory markers. In total, four out of 11 genome-wide significant loci were detected only by
489 multivariate analysis when adjusting univariate GWAS for multiple testing. At several loci,
490 multivariate analysis also seemed to highlight more plausible candidates for causal variants than the
491 univariate analyses. For example, in the *CIQA* locus, the lead variant in the univariate GWAS of the
492 driver trait TNF- β was an intronic variant in the *EPHB2* gene, whereas the lead variant for the locus
493 in the multivariate analysis was a Finnish-enriched missense variant located in the *CIQA* gene which
494 has been previously associated with immunologic diseases. Our multivariate analysis may point
495 towards a plausible mechanism underlying these associations via TNF- β levels.

496

497 Although both univariate and multivariate scans have previously been applied to these biomarkers^{1,66},
498 these studies have suffered from the lack of essential follow-up analyses due to the absence of beta
499 estimates in multivariate summary statistics. Our novel method enables two key follow-up analyses

500 for multivariate GWAS: fine-mapping and trait prioritization. Our method solves the problem of
501 missing effect sizes and standard errors required for fine-mapping by an extension of metaCCA
502 followed by LCP-GWAS. This process allows for the transformation of CCA-based multivariate
503 GWAS results into univariate summary statistics and thus extends the use of FINEMAP and other
504 summary statistics-based tools to multivariate GWAS. Fine-mapping complex multivariate
505 associations allows for assessing causality of the variants within the associated loci. This has not been
506 previously feasible. We also further describe the multivariate associations by determining the traits
507 driving the associations using MetaPhat. This workflow allows the identification of both the variants
508 and traits underlying the multivariate associations.

509

510 Our study also elucidates the advantage of multivariate analysis combined with large biobank-based
511 phenome-wide screening by discovering multiple novel disease associations. For example, in the *GP6*
512 locus we observe a novel risk-increasing association between the Finnish enriched rare missense
513 variant rs199588110 and benign neoplasms of meninges. Altogether, a majority of the observed
514 disease associations were for variants in the *F5* and *ABO* loci that were only detected by multivariate
515 GWAS. All these associations, including a genome-wide significant association with anemia that
516 replicated in the UKBB as an effect on red blood cell count, would have gone undetected had we used
517 univariate GWAS. In addition to disease association discovery, our workflow promotes increasing
518 insight into the pathophysiology underlying the associations by identifying the biomarkers driving
519 the associations. Detailed exploration of biological evidence including eQTLs and pQTLs in the *GP6*,
520 *SERPINE2*, *ABO*, and *F5* loci orthogonally supports our evidence of causal variants and driver traits.
521 For example, in the *SERPINE2* locus one of the three putative causal variants rs13412535 increased
522 the risk of hypertrophic skin disorders in FinnGen and was a pQTL for PDGF-BB⁶ that is considered
523 a key player in the pathogenesis of hypertrophic scars⁵⁰, increasing evidence of the biologically
524 relevant functions of this variant.

525

526 These methodological development and novel findings notwithstanding, our study has some
527 limitations. First, our newly developed workflow for multivariate fine-mapping requires individual
528 level genotype and phenotype data, problematic for some analysis settings. Additionally, the LCPs
529 are optimized for the lead variants, potentially resulting in overestimation of the causal probability of
530 these variants. We did not, however, see evidence of this in the *F5* locus where we constructed LCPs
531 for two missense variants in addition to the lead variant with no significant changes in the causal
532 probabilities of the variants. We also acknowledge that the credible sets we chose for follow-up may
533 not encompass all causal signals within the multivariate associations. The credible sets excluded due
534 to low LD may arise from multiple signals included in the same set, resulting in small LD within the
535 set. Further, some disease associations require replication and follow-up analyses.

536

537 On the other hand, our study has many strengths. First, a prospective cohort study was used to assess
538 deep phenotype data rarely available at large scale. Second, we are the first to present phenome-wide
539 results from FinnGen, a very large and well-phenotyped Finnish biobank study, and also make use of
540 the UKBB, in disease association follow-up, ensuring enough power for disease association detection.
541 Finland has a public healthcare system and national health registries, which enable the vast and
542 accurate phenotyping in FinnGen. Besides FinnGen, an additional advantage to performing the study
543 in Finns is that deleterious variants are enriched in the Finnish population due to population history²¹.
544 Furthermore, our reference panel for genotype imputation is from the same population as our
545 discovery and follow-up data sets, which, as demonstrated also by others^{67,68}, allows us to study
546 variants that are enriched (and often unique) in the study-specific population.

547

548 In conclusion, we developed a novel workflow for multivariate GWAS discovery and follow-up
549 analyses, including fine-mapping and identification of driver traits, and thus promote the
550 advancement of powerful multivariate methods in genomic analyses. We demonstrate the benefit of

551 applying this workflow by identifying novel associations and further describing previously reported
552 associations with both biomarkers and diseases using a set of inflammatory markers. We show that
553 compared to univariate analyses, multivariate analysis of biomarker data combined with large
554 biobank-based PheWAS reveals a considerably increased number of novel genetic associations with
555 several diseases.

556

557

558 **ACKNOWLEDGMENTS**

559 We would like to thank Lea Urpa for proofreading, and Sari Kivikko, Huei-Yi Shen and Ulla
560 Tuomainen for management assistance. We would like to thank all participants of the FINRISK,
561 FinnGen and UKBB studies for their generous participation. The FINRISK data used for the research
562 were obtained from THL Biobank. This research has been conducted using the UK Biobank Resource
563 with application number 22627.

564

565 This work was supported by the Academy of Finland Center of Excellence in Complex Disease
566 Genetics [Grant No 312062 to S.R., 312074 to A.P., 312075 to M.D.]; Academy of Finland [Grant
567 No 285380 to S.R., 128650 to A.P.]; the Finnish Foundation for Cardiovascular Research [to S.R.,
568 V.S., and A.P.]; the Sigrid Jusélius Foundation [to S.R. and A.P.]; University of Helsinki HiLIFE
569 Fellow grants 2017-2020 [to S.R.]; Foundation and the Horizon 2020 Research and Innovation
570 Programme [grant number 667301 (COSYN) to A.P.]; the Doctoral Programme in Population Health,
571 University of Helsinki [to J.J.P. and S.E.R.]; and The Finnish Medical Foundation [to J.J.P.]. The
572 FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH
573 4386/31/2016) and nine industry partners (AbbVie, AstraZeneca, Biogen, Celgene, Genentech, GSK,
574 MSD, Pfizer and Sanofi). Following biobanks are acknowledged for collecting the FinnGen project
575 samples: Auria Biobank (<https://www.auria.fi/biopankki/en>), THL Biobank ([23](https://thl.fi/fi/web/thl-</p></div><div data-bbox=)

576 [biopankki](#)), Helsinki Biobank (<https://www.terveyskyla.fi/helsinginbiopankki/en>), Northern Finland
577 Biobank Borealis (<https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki>), Finnish Clinical Biobank
578 Tampere ([https://www.tays.fi/en-](https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)
579 [US/Research_and_development/Finnish_Clinical_Biobank_Tampere](https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Biobank of Eastern Finland
580 (<https://ita-suomenbiopankki.fi/>), Central Finland Biobank ([https://www.ksshp.fi/fi-](https://www.ksshp.fi/fi-FI/Potilaalle/Biopankki)
581 [FI/Potilaalle/Biopankki](https://www.ksshp.fi/fi-FI/Potilaalle/Biopankki)), Finnish Red Cross Blood Service Biobank
582 (<https://www.bloodservice.fi/Research%20Projects/biobanking>).

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

585

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

588

589 **Contributors of FinnGen**

590 **Steering Committee**

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

593

594 **Pharmaceutical companies**

595 Howard Jacob Abbvie, Chicago, IL, United States
596 Athena Matakidou Astra Zeneca, Cambridge, United Kingdom
597 Heiko Runz Biogen, Cambridge, MA, United States
598 Sally John Biogen, Cambridge, MA, United States
599 Robert Plenge Celgene, Summit, NJ, United States
600 Mark McCarthy Genentech, San Francisco, CA, United States
601 Julie Hunkapiller Genentech, San Francisco, CA, United States
602 Meg Ehm GlaxoSmithKline, Brentford, United Kingdom

603	Dawn Waterworth	GlaxoSmithKline, Brentford, United Kingdom
604	Caroline Fox	Merck, Kenilworth, NJ, United States
605	Anders Malarstig	Pfizer, New York, NY, United States
606	Kathy Klinger	Sanofi, Paris, France
607	Kathy Call	Sanofi, Paris, France
608		
609	University of Helsinki & Biobanks	
610	Tomi Mäkelä	HiLIFE, University of Helsinki, Finland, Finland
611	Jaakko Kaprio	Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland, Finland
612	Petri Virolainen	Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland
613	Kari Pulkki	Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland
614	Terhi Kilpi	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
615	Markus Perola	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
616	Jukka Partanen	Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank,
617	Helsinki, Finland	
618	Anne Pitkäranta	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
619	Riitta Kaarteenaho	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
620	District, Oulu, Finland	
621	Seppo Vainio	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
622	District, Oulu, Finland	
623	Kimmo Savinainen	Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District,
624	Tampere, Finland	
625	Veli-Matti Kosma	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District,
626	Kuopio, Finland	
627	Urho Kujala	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District,
628	Jyväskylä, Finland	
629		
630	Other Experts/ Non-Voting Members	
631	Outi Tuovila	Business Finland, Helsinki, Finland
632	Minna Hendolin	Business Finland, Helsinki, Finland

633 Raimo Pakkanen Business Finland, Helsinki, Finland

634

635 **Scientific Committee**

636 **Pharmaceutical companies**

637 Jeff Waring Abbvie, Chicago, IL, United States

638 Bridget Riley-Gillis Abbvie, Chicago, IL, United States

639 Athena Matakidou Astra Zeneca, Cambridge, United Kingdom

640 Heiko Runz Biogen, Cambridge, MA, United States

641 Jimmy Liu Biogen, Cambridge, MA, United States

642 Shameek Biswas Celgene, Summit, NJ, United States

643 Julie Hunkapiller Genentech, San Francisco, CA, United States

644 Dawn Waterworth GlaxoSmithKline, Brentford, United Kingdom

645 Meg Ehm GlaxoSmithKline, Brentford, United Kingdom

646 Dorothee Diogo Merck, Kenilworth, NJ, United States

647 Caroline Fox Merck, Kenilworth, NJ, United States

648 Anders Malarstig Pfizer, New York, NY, United States

649 Catherine Marshall Pfizer, New York, NY, United States

650 Xinli Hu Pfizer, New York, NY, United States

651 Kathy Call Sanofi, Paris, France

652 Kathy Klinger Sanofi, Paris, France

653 Matthias Gossel Sanofi, Paris, France

654

655 **University of Helsinki & Biobanks**

656 Samuli Ripatti Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland

657 Johanna Schleutker Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland

658 Markus Perola THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

659 Mikko Arvas Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank,

660 Helsinki, Finland

661 Olli Carpen Hospital District of Helsinki and Uusimaa, Helsinki, Finland

662	Reetta Hinttala	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
663	District, Oulu, Finland	
664	Johannes Kettunen	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
665	District, Oulu, Finland	
666	Reijo Laaksonen	Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District,
667	Tampere, Finland	
668	Arto Mannermaa	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital
669	District, Kuopio, Finland	
670	Juha Paloneva	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District,
671	Jyväskylä, Finland	
672	Urho Kujala	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District,
673	Jyväskylä, Finland	
674		
675	Other Experts/ Non-Voting Members	
676	Outi Tuovila	Business Finland, Helsinki, Finland
677	Minna Hendolin	Business Finland, Helsinki, Finland
678	Raimo Pakkanen	Business Finland, Helsinki, Finland
679		
680	Clinical Groups	
681	Neurology Group	
682	Hilkka Soininen	Northern Savo Hospital District, Kuopio, Finland
683	Valtteri Julkunen	Northern Savo Hospital District, Kuopio, Finland
684	Anne Remes	Northern Ostrobothnia Hospital District, Oulu, Finland
685	Reetta Kälviäinen	Northern Savo Hospital District, Kuopio, Finland
686	Mikko Hiltunen	Northern Savo Hospital District, Kuopio, Finland
687	Jukka Peltola	Pirkanmaa Hospital District, Tampere, Finland
688	Pentti Tienari	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
689	Juha Rinne	Hospital District of Southwest Finland, Turku, Finland
690	Adam Ziemann	Abbvie, Chicago, IL, United States
691	Jeffrey Waring	Abbvie, Chicago, IL, United States

692	Sahar Esmaeeli	Abbvie, Chicago, IL, United States
693	Nizar Smaoui	Abbvie, Chicago, IL, United States
694	Anne Lehtonen	Abbvie, Chicago, IL, United States
695	Susan Eaton	Biogen, Cambridge, MA, United States
696	Heiko Runz	Biogen, Cambridge, MA, United States
697	Sanni Lahdenperä	Biogen, Cambridge, MA, United States
698	Shameek Biswas	Celgene, Summit, NJ, United States
699	John Michon	Genentech, San Francisco, CA, United States
700	Geoff Kerchner	Genentech, San Francisco, CA, United States
701	Julie Hunkapiller	Genentech, San Francisco, CA, United States
702	Natalie Bowers	Genentech, San Francisco, CA, United States
703	Edmond Teng	Genentech, San Francisco, CA, United States
704	John Eicher	Merck, Kenilworth, NJ, United States
705	Vinay Mehta	Merck, Kenilworth, NJ, United States
706	Padhraig Gormley	Merck, Kenilworth, NJ, United States
707	Kari Linden	Pfizer, New York, NY, United States
708	Christopher Whelan	Pfizer, New York, NY, United States
709	Fanli Xu	GlaxoSmithKline, Brentford, United Kingdom
710	David Pulford	GlaxoSmithKline, Brentford, United Kingdom
711		
712	Gastroenterology Group	
713	Martti Färkkilä	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
714	Sampsa Pikkarainen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
715	Airi Jussila	Pirkanmaa Hospital District, Tampere, Finland
716	Timo Blomster	Northern Ostrobothnia Hospital District, Oulu, Finland
717	Mikko Kiviniemi	Northern Savo Hospital District, Kuopio, Finland
718	Markku Voutilainen	Hospital District of Southwest Finland, Turku, Finland
719	Bob Georgantas	Abbvie, Chicago, IL, United States
720	Graham Heap	Abbvie, Chicago, IL, United States
721	Jeffrey Waring	Abbvie, Chicago, IL, United States

722	Nizar Smaoui	Abbvie, Chicago, IL, United States
723	Fedik Rahimov	Abbvie, Chicago, IL, United States
724	Anne Lehtonen	Abbvie, Chicago, IL, United States
725	Keith Usiskin	Celgene, Summit, NJ, United States
726	Joseph Maranville	Celgene, Summit, NJ, United States
727	Tim Lu	Genentech, San Francisco, CA, United States
728	Natalie Bowers	Genentech, San Francisco, CA, United States
729	Danny Oh	Genentech, San Francisco, CA, United States
730	John Michon	Genentech, San Francisco, CA, United States
731	Vinay Mehta	Merck, Kenilworth, NJ, United States
732	Kirsi Kalpala	Pfizer, New York, NY, United States
733	Melissa Miller	Pfizer, New York, NY, United States
734	Xinli Hu	Pfizer, New York, NY, United States
735	Linda McCarthy	GlaxoSmithKline, Brentford, United Kingdom
736		
737	Rheumatology Group	
738	Kari Eklund	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
739	Antti Palomäki	Hospital District of Southwest Finland, Turku, Finland
740	Pia Isomäki	Pirkanmaa Hospital District, Tampere, Finland
741	Laura Pirilä	Hospital District of Southwest Finland, Turku, Finland
742	Oili Kaipainen-Seppänen	Northern Savo Hospital District, Kuopio, Finland
743	Johanna Huhtakangas	Northern Ostrobothnia Hospital District, Oulu, Finland
744	Bob Georgantas	Abbvie, Chicago, IL, United States
745	Jeffrey Waring	Abbvie, Chicago, IL, United States
746	Fedik Rahimov	Abbvie, Chicago, IL, United States
747	Apinya Lertratanakul	Abbvie, Chicago, IL, United States
748	Nizar Smaoui	Abbvie, Chicago, IL, United States
749	Anne Lehtonen	Abbvie, Chicago, IL, United States
750	David Close	Astra Zeneca, Cambridge, United Kingdom
751	Marla Hochfeld	Celgene, Summit, NJ, United States

752	Natalie Bowers	Genentech, San Francisco, CA, United States
753	John Michon	Genentech, San Francisco, CA, United States
754	Dorothee Diogo	Merck, Kenilworth, NJ, United States
755	Vinay Mehta	Merck, Kenilworth, NJ, United States
756	Kirsi Kalpala	Pfizer, New York, NY, United States
757	Nan Bing	Pfizer, New York, NY, United States
758	Xinli Hu	Pfizer, New York, NY, United States
759	Jorge Esparza Gordillo	GlaxoSmithKline, Brentford, United Kingdom
760	Nina Mars	Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland
761		
762	Pulmonology Group	
763	Tarja Laitinen	Pirkanmaa Hospital District, Tampere, Finland
764	Margit Pelkonen	Northern Savo Hospital District, Kuopio, Finland
765	Paula Kauppi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
766	Hannu Kankaanranta	Pirkanmaa Hospital District, Tampere, Finland
767	Terttu Harju	Northern Ostrobothnia Hospital District, Oulu, Finland
768	Nizar Smaoui	Abbvie, Chicago, IL, United States
769	David Close	Astra Zeneca, Cambridge, United Kingdom
770	Steven Greenberg	Celgene, Summit, NJ, United States
771	Hubert Chen	Genentech, San Francisco, CA, United States
772	Natalie Bowers	Genentech, San Francisco, CA, United States
773	John Michon	Genentech, San Francisco, CA, United States
774	Vinay Mehta	Merck, Kenilworth, NJ, United States
775	Jo Betts	GlaxoSmithKline, Brentford, United Kingdom
776	Soumitra Ghosh	GlaxoSmithKline, Brentford, United Kingdom
777		
778	Cardiometabolic Diseases Group	
779	Veikko Salomaa	The National Institute of Health and Welfare Helsinki, Finland
780	Teemu Niiranen	The National Institute of Health and Welfare Helsinki, Finland
781	Markus Juonala	Hospital District of Southwest Finland, Turku, Finland

782	Kaj Metsärinne	Hospital District of Southwest Finland, Turku, Finland
783	Mika Kähönen	Pirkanmaa Hospital District, Tampere, Finland
784	Juhani Junttila	Northern Ostrobothnia Hospital District, Oulu, Finland
785	Markku Laakso	Northern Savo Hospital District, Kuopio, Finland
786	Jussi Pihlajamäki	Northern Savo Hospital District, Kuopio, Finland
787	Juha Sinisalo	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
788	Marja-Riitta Taskinen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
789	Tiinamaija Tuomi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
790	Jari Laukkanen	Central Finland Health Care District, Jyväskylä, Finland
791	Ben Challis	Astra Zeneca, Cambridge, United Kingdom
792	Andrew Peterson	Genentech, San Francisco, CA, United States
793	Julie Hunkapiller	Genentech, San Francisco, CA, United States
794	Natalie Bowers	Genentech, San Francisco, CA, United States
795	John Michon	Genentech, San Francisco, CA, United States
796	Dorothee Diogo	Merck, Kenilworth, NJ, United States
797	Audrey Chu	Merck, Kenilworth, NJ, United States
798	Vinay Mehta	Merck, Kenilworth, NJ, United States
799	Jaakko Parkkinen	Pfizer, New York, NY, United States
800	Melissa Miller	Pfizer, New York, NY, United States
801	Anthony Muslin	Sanofi, Paris, France
802	Dawn Waterworth	GlaxoSmithKline, Brentford, United Kingdom
803		
804	Oncology Group	
805	Heikki Joensuu	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
806	Tuomo Meretoja	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
807	Olli Carpen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
808	Lauri Aaltonen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
809	Annika Auranen	Pirkanmaa Hospital District, Tampere, Finland
810	Peeter Karihtala	Northern Ostrobothnia Hospital District, Oulu, Finland
811	Saila Kauppila	Northern Ostrobothnia Hospital District, Oulu, Finland

812	Päivi Auvinen	Northern Savo Hospital District, Kuopio, Finland
813	Klaus Elenius	Hospital District of Southwest Finland, Turku, Finland
814	Relja Popovic	Abbvie, Chicago, IL, United States
815	Jeffrey Waring	Abbvie, Chicago, IL, United States
816	Bridget Riley-Gillis	Abbvie, Chicago, IL, United States
817	Anne Lehtonen	Abbvie, Chicago, IL, United States
818	Athena Matakidou	Astra Zeneca, Cambridge, United Kingdom
819	Jennifer Schutzman	Genentech, San Francisco, CA, United States
820	Julie Hunkapiller	Genentech, San Francisco, CA, United States
821	Natalie Bowers	Genentech, San Francisco, CA, United States
822	John Michon	Genentech, San Francisco, CA, United States
823	Vinay Mehta	Merck, Kenilworth, NJ, United States
824	Andrey Loboda	Merck, Kenilworth, NJ, United States
825	Aparna Chhibber	Merck, Kenilworth, NJ, United States
826	Heli Lehtonen	Pfizer, New York, NY, United States
827	Stefan McDonough	Pfizer, New York, NY, United States
828	Marika Crohns	Sanofi, Paris, France
829	Diptee Kulkarni	GlaxoSmithKline, Brentford, United Kingdom
830		
831	Ophthalmology Group	
832	Kai Kaarniranta	Northern Savo Hospital District, Kuopio, Finland
833	Joni Turunen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
834	Terhi Ollila	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
835	Sanna Seitsonen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
836	Hannu Uusitalo	Pirkanmaa Hospital District, Tampere, Finland
837	Vesa Aaltonen	Hospital District of Southwest Finland, Turku, Finland
838	Hannele Uusitalo-Järvinen	Pirkanmaa Hospital District, Tampere, Finland
839	Marja Luodonpää	Northern Ostrobothnia Hospital District, Oulu, Finland
840	Nina Hautala	Northern Ostrobothnia Hospital District, Oulu, Finland
841	Heiko Runz	Biogen, Cambridge, MA, United States

842	Erich Strauss	Genentech, San Francisco, CA, United States
843	Natalie Bowers	Genentech, San Francisco, CA, United States
844	Hao Chen	Genentech, San Francisco, CA, United States
845	John Michon	Genentech, San Francisco, CA, United States
846	Anna Podgornaia	Merck, Kenilworth, NJ, United States
847	Vinay Mehta	Merck, Kenilworth, NJ, United States
848	Dorothee Diogo	Merck, Kenilworth, NJ, United States
849	Joshua Hoffman	GlaxoSmithKline, Brentford, United Kingdom
850		
851	Dermatology Group	
852	Kaisa Tasanen	Northern Ostrobothnia Hospital District, Oulu, Finland
853	Laura Huilaja	Northern Ostrobothnia Hospital District, Oulu, Finland
854	Katariina Hannula-Jouppi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
855	Teea Salmi	Pirkanmaa Hospital District, Tampere, Finland
856	Sirkku Peltonen	Hospital District of Southwest Finland, Turku, Finland
857	Leena Koulu	Hospital District of Southwest Finland, Turku, Finland
858	Ilkka Harvima	Northern Savo Hospital District, Kuopio, Finland
859	Kirsi Kalpala	Pfizer, New York, NY, United States
860	Ying Wu	Pfizer, New York, NY, United States
861	David Choy	Genentech, San Francisco, CA, United States
862	John Michon	Genentech, San Francisco, CA, United States
863	Nizar Smaoui	Abbvie, Chicago, IL, United States
864	Fedik Rahimov	Abbvie, Chicago, IL, United States
865	Anne Lehtonen	Abbvie, Chicago, IL, United States
866	Dawn Waterworth	GlaxoSmithKline, Brentford, United Kingdom
867		
868	FinnGen Teams	
869	Administration Team	
870	Anu Jalanko	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
871	Risto Kajanne	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

872 Ulrike Lyhs Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
873

874 **Communication**

875 Mari Kaunisto Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
876

877 **Analysis Team**

878 Justin Wade Davis Abbvie, Chicago, IL, United States
879 Bridget Riley-Gillis Abbvie, Chicago, IL, United States
880 Danjuma Quarless Abbvie, Chicago, IL, United States
881 Slavé Petrovski Astra Zeneca, Cambridge, United Kingdom
882 Jimmy Liu Biogen, Cambridge, MA, United States
883 Chia-Yen Chen Biogen, Cambridge, MA, United States
884 Paola Bronson Biogen, Cambridge, MA, United States
885 Robert Yang Celgene, Summit, NJ, United States
886 Joseph Maranville Celgene, Summit, NJ, United States
887 Shameek Biswas Celgene, Summit, NJ, United States
888 Diana Chang Genentech, San Francisco, CA, United States
889 Julie Hunkapiller Genentech, San Francisco, CA, United States
890 Tushar Bhangale Genentech, San Francisco, CA, United States
891 Natalie Bowers Genentech, San Francisco, CA, United States
892 Dorothee Diogo Merck, Kenilworth, NJ, United States
893 Emily Holzinger Merck, Kenilworth, NJ, United States
894 Padhraig Gormley Merck, Kenilworth, NJ, United States
895 Xulong Wang Merck, Kenilworth, NJ, United States
896 Xing Chen Pfizer, New York, NY, United States
897 Åsa Hedman Pfizer, New York, NY, United States
898 Kirsi Auro GlaxoSmithKline, Brentford, United Kingdom
899 Clarence Wang Sanofi, Paris, France
900 Ethan Xu Sanofi, Paris, France
901 Franck Auge Sanofi, Paris, France

902 Clement Chatelain Sanofi, Paris, France

903 Mitja Kurki Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad
904 Institute, Cambridge, MA, United States

905 Samuli Ripatti Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

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

907 Juha Karjalainen Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad
908 Institute, Cambridge, MA, United States

909 Aki Havulinna Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

910 Anu Jalanko Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

911 Kimmo Palin University of Helsinki, Helsinki, Finland

912 Priit Palta Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

913 Pietro Della Briotta Parolo Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

914 Wei Zhou Broad Institute, Cambridge, MA, United States

915 Susanna Lemmelä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

916 Manuel Rivas University of Stanford, Stanford, CA, United States

917 Jarmo Harju Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

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

919 Arto Lehisto Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

920 Andrea Ganna Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

921 Vincent Llorens Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

922 Antti Karlsson Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland

923 Kati Kristiansson THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

924 Mikko Arvas Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank,
925 Helsinki, Finland

926

927 Kati Hyvärinen Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank,
928 Helsinki, Finland

929 Jarmo Ritari Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank,
930 Helsinki, Finland

931	Tiina Wahlfors	Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank,
932	Helsinki, Finland	
933	Miika Koskinen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland BB/HUS/Univ Hosp Districts
934	Olli Carpen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland BB/HUS/Univ Hosp Districts
935	Johannes Kettunen	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
936	District, Oulu, Finland	
937	Katri Pylkäs	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
938	District, Oulu, Finland	
939	Marita Kalaoja	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
940	District, Oulu, Finland	
941	Minna Karjalainen	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
942	District, Oulu, Finland	
943	Tuomo Mantere	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital
944	District, Oulu, Finland	
945	Eeva Kangasniemi	Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District,
946	Tampere, Finland	
947	Sami Heikkinen	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital
948	District, Kuopio, Finland	
949	Arto Mannermaa	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District,
950	Kuopio, Finland	
951	Eija Laakkonen	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District,
952	Jyväskylä, Finland	
953	Juha Kononen	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District,
954	Jyväskylä, Finland	
955		
956	Sample Collection Coordination	
957	Anu Loukola	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
958		
959	Sample Logistics	
960	Päivi Laiho	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

961	Tuuli Sistonen	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
962	Essi Kaiharju	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
963	Markku Laukkanen	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
964	Elina Järvensivu	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
965	Sini Lähteenmäki	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
966	Lotta Männikkö	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
967	Regis Wong	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
968		
969	Registry Data Operations	
970	Kati Kristiansson	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
971	Hannele Mattsson	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
972	Susanna Lemmelä	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
973	Tero Hiekkalinna	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
974	Manuel González Jiménez	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
975		
976		
977	Genotyping	
978	Kati Donner	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
979		
980	Sequencing Informatics	
981	Priit Palta	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
982	Kalle Pärn	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
983	Javier Nunez-Fontarnau	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
984		
985	Data Management and IT Infrastructure	
986	Jarmo Harju	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
987	Elina Kilpeläinen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
988	Timo P. Sipilä	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
989	Georg Brein	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
990	Alexander Dada	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

991 Ghazal Awaisa Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
992 Anastasia Shcherban Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
993 Tuomas Sipilä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

994

995 **Clinical Endpoint Development**

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

1000

1001 **Trajectory Team**

1002 Tarja Laitinen Pirkanmaa Hospital District, Tampere, Finland
1003 Harri Siirtola University of Tampere, Tampere, Finland
1004 Javier Gracia Tabuena University of Tampere, Tampere, Finland

1005

1006 **Biobank Directors**

1007 Lila Kallio Auria Biobank, Turku, Finland
1008 Sirpa Soini THL Biobank, Helsinki, Finland
1009 Jukka Partanen Blood Service Biobank, Helsinki, Finland
1010 Kimmo Pitkänen Helsinki Biobank, Helsinki, Finland
1011 Seppo Vainio Northern Finland Biobank Borealis, Oulu, Finland
1012 Kimmo Savinainen Tampere Biobank, Tampere, Finland
1013 Veli-Matti Kosma Biobank of Eastern Finland, Kuopio, Finland
1014 Teijo Kuopio Central Finland Biobank, Jyväskylä, Finland

1015 **REFERENCES**

- 1016 1. Ahola-Olli AV, Würtz P, Havulinna AS, et al. Genome-wide association study identifies
1017 27 loci influencing concentrations of circulating cytokines and growth factors. *The American*
1018 *Journal of Human Genetics*. 2017;100(1):40-50.
- 1019 2. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation
1020 and links to common complex disease. *Cell*. 2016;167(5):1415-1429. e19.
- 1021 3. Liu DJ, Peloso GM, Yu H, et al. Exome-wide association study of plasma lipids in
1022 300,000 individuals. *Nat Genet*. 2017;49(12):1758.
- 1023 4. Sinnott-Armstrong N, Tanigawa Y, Amar D, et al. Genetics of 38 blood and urine
1024 biomarkers in the UK biobank. *BioRxiv*. 2019:660506.
- 1025 5. Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points
1026 through the human blood plasma proteome. *Nature communications*. 2017;8:14357.
- 1027 6. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome.
1028 *Nature*. 2018;558(7708):73.
- 1029 7. Kettunen J, Demirkan A, Würtz P, et al. Genome-wide study for circulating metabolites
1030 identifies 62 loci and reveals novel systemic effects of LPA. *Nature communications*.
1031 2016;7:11122.
- 1032 8. Inouye M, Ripatti S, Kettunen J, et al. Novel loci for metabolic networks and multi-tissue
1033 expression studies reveal genes for atherosclerosis. *PLoS genetics*. 2012;8(8):e1002907.
- 1034 9. Leitsalu L, Haller T, Esko T, et al. Cohort profile: Estonian biobank of the estonian
1035 genome center, university of tartu. *Int J Epidemiol*. 2014;44(4):1137-1147.

- 1036 10. Bycroft C, Freeman C, Petkova D, et al. The UK biobank resource with deep phenotyping
1037 and genomic data. *Nature*. 2018;562(7726):203.
- 1038 11. Kim S, Xing EP. Statistical estimation of correlated genome associations to a quantitative
1039 trait network. *PLoS genetics*. 2009;5(8):e1000587.
- 1040 12. Ferreira MA, Purcell SM. A multivariate test of association. *Bioinformatics*.
1041 2008;25(1):132-133.
- 1042 13. O'Reilly PF, Hoggart CJ, Pomyen Y, et al. MultiPhen: Joint model of multiple
1043 phenotypes can increase discovery in GWAS. *PloS one*. 2012;7(5):e34861.
- 1044 14. Cichonska A, Rousu J, Marttinen P, et al. metaCCA: Summary statistics-based
1045 multivariate meta-analysis of genome-wide association studies using canonical correlation
1046 analysis. *Bioinformatics*. 2016;32(13):1981-1989.
- 1047 15. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP:
1048 Efficient variable selection using summary data from genome-wide association studies.
1049 *Bioinformatics*. 2016;32(10):1493-1501.
- 1050 16. Lin J, Tabassum R, Ripatti S, Pirinen M. MetaPhat: Detecting and decomposing
1051 multivariate associations from univariate genome-wide association statistics. *bioRxiv*.
1052 2019:661421.
- 1053 17. McInnes IB, Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nature*
1054 *Reviews Immunology*. 2007;7(6):429.

- 1055 18. Martins TB, Rose JW, Jaskowski TD, et al. Analysis of proinflammatory and anti-
1056 inflammatory cytokine serum concentrations in patients with multiple sclerosis by using a
1057 multiplexed immunoassay. *Am J Clin Pathol*. 2011;136(5):696-704.
- 1058 19. Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. *Nature*.
1059 2000;407(6801):249.
- 1060 20. Borodulin K, Vartiainen E, Peltonen M, et al. Forty-year trends in cardiovascular risk
1061 factors in finland. *The European Journal of Public Health*. 2014;25(3):539-546.
- 1062 21. Lim ET, Würtz P, Havulinna AS, et al. Distribution and medical impact of loss-of-
1063 function variants in the finnish founder population. *PLoS genetics*. 2014;10(7):e1004494.
- 1064 22. Hail Team. Hail 0.2.13-81ab564db2b4. <https://Github.com/hail-is/hail/releases/tag/0.2.13>
1065 <http://Doi.org/10.5281/zenodo.2646680> .
- 1066 23. Benner C, Havulinna A, Salomaa V, Ripatti S, Pirinen M. Refining fine-mapping: Effect
1067 sizes and regional heritability. *bioRxiv*. 2018:318618.
- 1068 24. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance
1069 and sample relatedness in large-scale genetic association studies. *Nat Genet*.
1070 2018;50(9):1335.
- 1071 25. Wu P, Gifford A, Meng X, et al. Developing and evaluating mappings of ICD-10 and
1072 ICD-10-CM codes to phecodes. *BioRxiv*. 2019:462077.
- 1073 26. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS catalog of
1074 published genome-wide association studies, targeted arrays and summary statistics 2019.
1075 *Nucleic Acids Res*. 2018;47(D1):D1005-D1012.

- 1076 27. Emilsson V, Ilkov M, Lamb JR, et al. Co-regulatory networks of human serum proteins
1077 link genetics to disease. *Science*. 2018;361(6404):769-773.
- 1078 28. Sasayama D, Hattori K, Ogawa S, et al. Genome-wide quantitative trait loci mapping of
1079 the human cerebrospinal fluid proteome. *Hum Mol Genet*. 2016;26(1):44-51.
- 1080 29. Reeve MP, Kirby A, Wierzbowski J, Daly M, Hutz J. Target gene notebook: Connecting
1081 genetics and drug discovery. *bioRxiv*. 2019:757690.
- 1082 30. van Schaarenburg RA, Daha NA, Schonkeren JJ, et al. Identification of a novel non-
1083 coding mutation in C1qB in a dutch child with C1q deficiency associated with recurrent
1084 infections. *Immunobiology*. 2015;220(3):422-427.
- 1085 31. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-
1086 associated variation in regulatory DNA. *Science*. 2012;337(6099):1190-1195.
- 1087 32. Huang H, Fang M, Jostins L, et al. Fine-mapping inflammatory bowel disease loci to
1088 single-variant resolution. *Nature*. 2017;547(7662):173.
- 1089 33. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and
1090 genomes reveals the spectrum of loss-of-function intolerance across human protein-coding
1091 genes. *BioRxiv*. 2019:531210.
- 1092 34. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function.
1093 *Nucleic Acids Res*. 2003;31(13):3812-3814.
- 1094 35. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense
1095 mutations using PolyPhen-2. *Current protocols in human genetics*. 2013;76(1):7.20. 1-7.20.
1096 41.

- 1097 36. Andrae N, Kirches E, Hartig R, et al. Sunitinib targets PDGF-receptor and Flt3 and
1098 reduces survival and migration of human meningioma cells. *Eur J Cancer*.
1099 2012;48(12):1831-1841.
- 1100 37. Kaley TJ, Wen P, Schiff D, et al. Phase II trial of sunitinib for recurrent and progressive
1101 atypical and anaplastic meningioma. *Neuro-oncology*. 2014;17(1):116-121.
- 1102 38. Todo T, Adams EF, Rafferty B, Fahlbusch R, Dingermann T, Werner H. Secretion of
1103 interleukin-6 by human meningioma cells: Possible autocrine inhibitory regulation of
1104 neoplastic cell growth. *J Neurosurg*. 1994;81(3):394-401.
- 1105 39. Yang S, Xu G. Expression of PDGF and its receptor as well as their relationship to
1106 proliferating activity and apoptosis of meningiomas in human meningiomas. *Journal of*
1107 *Clinical Neuroscience*. 2001;8(4):49-53.
- 1108 40. Lamszus K, Lengler U, Schmidt NO, Stavrou D, Ergün S, Westphal M. Vascular
1109 endothelial growth factor, hepatocyte growth factor/scatter factor, basic fibroblast growth
1110 factor, and placenta growth factor in human meningiomas and their relation to angiogenesis
1111 and malignancy. *Neurosurgery*. 2000;46(4):938-948.
- 1112 41. Preusser M, Hassler M, Birner P, et al. Microvascularization and expression of VEGF and
1113 its receptors in recurring meningiomas: Pathobiological data in favor of anti-angiogenic
1114 therapy approaches. *Clin Neuropathol*. 2012;31(5):352-360.
- 1115 42. Raizer JJ, Grimm SA, Rademaker A, et al. A phase II trial of PTK787/ZK 222584 in
1116 recurrent or progressive radiation and surgery refractory meningiomas. *J Neurooncol*.
1117 2014;117(1):93-101.

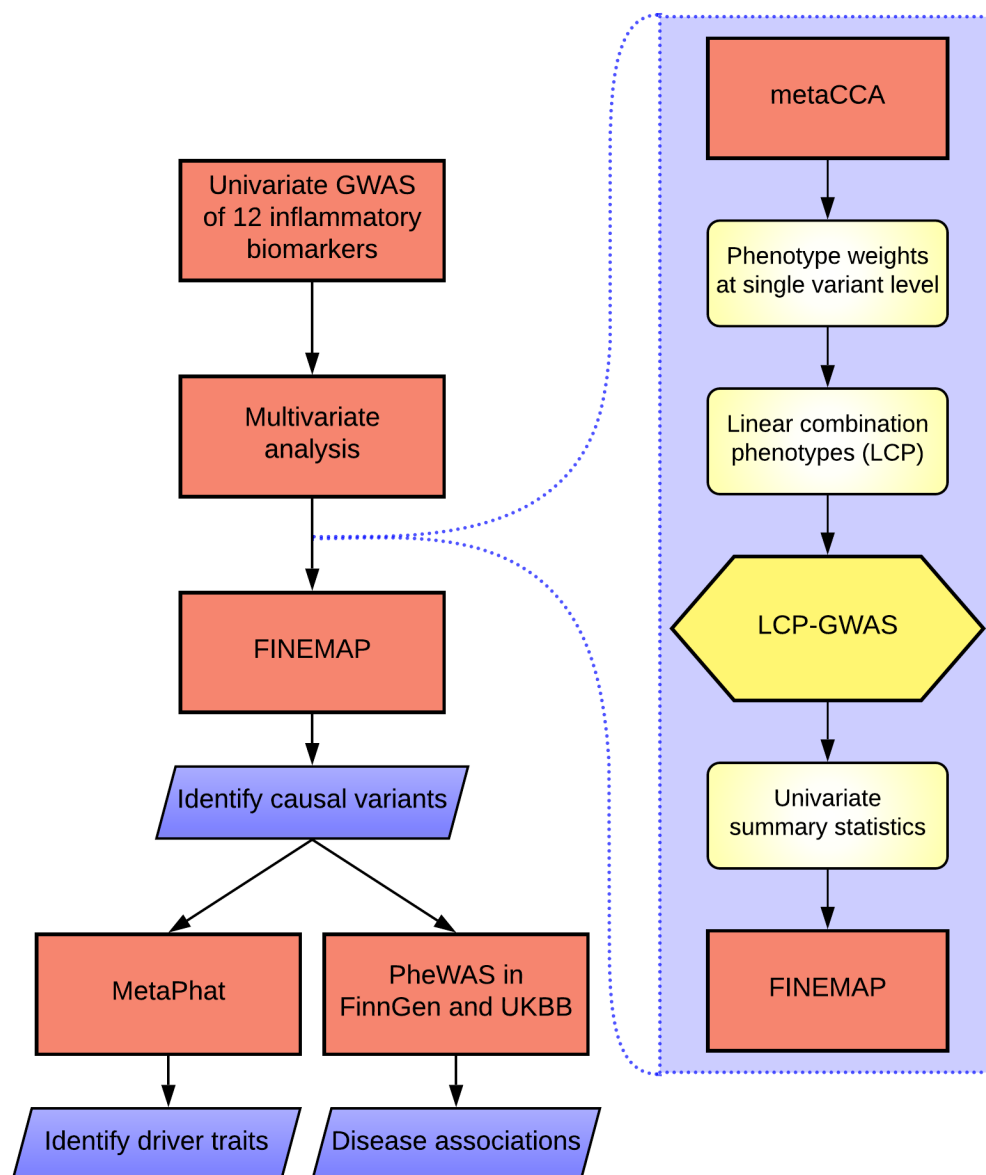
- 1118 43. Bouton M, Boulaftali Y, Richard B, Arocas V, Michel J, Jandrot-Perrus M. Emerging
1119 role of serpinE2/protease nexin-1 in hemostasis and vascular biology. *Blood, The Journal of*
1120 *the American Society of Hematology*. 2012;119(11):2452-2457.
- 1121 44. DeMeo DL, Mariani TJ, Lange C, et al. The SERPINE2 gene is associated with chronic
1122 obstructive pulmonary disease. *The American Journal of Human Genetics*. 2006;78(2):253-
1123 264.
- 1124 45. Bergman BL, Scott RW, Bajpai A, Watts S, Baker JB. Inhibition of tumor-cell-mediated
1125 extracellular matrix destruction by a fibroblast proteinase inhibitor, protease nexin I.
1126 *Proceedings of the National Academy of Sciences*. 1986;83(4):996-1000.
- 1127 46. Li X, Zhao D, Guo Z, et al. Overexpression of SerpinE2/protease nexin-1 contribute to
1128 pathological cardiac fibrosis via increasing collagen deposition. *Scientific reports*.
1129 2016;6:37635.
- 1130 47. Nuutila K, Siltanen A, Peura M, et al. Gene expression profiling of negative-pressure-
1131 treated skin graft donor site wounds. *Burns*. 2013;39(4):687-693.
- 1132 48. Ghazawi FM, Zargham R, Gilardino MS, Sasseville D, Jafarian F. Insights into the
1133 pathophysiology of hypertrophic scars and keloids: How do they differ? *Adv Skin Wound*
1134 *Care*. 2018;31(1):582-595.
- 1135 49. Brissett AE, Sherris DA. Scar contractures, hypertrophic scars, and keloids. *Facial plastic*
1136 *surgery*. 2001;17(04):263-272.
- 1137 50. Lian N, Li T. Growth factor pathways in hypertrophic scars: Molecular pathogenesis and
1138 therapeutic implications. *Biomedicine & Pharmacotherapy*. 2016;84:42-50.

- 1139 51. Bonner JC. Regulation of PDGF and its receptors in fibrotic diseases. *Cytokine Growth*
1140 *Factor Rev.* 2004;15(4):255-273.
- 1141 52. Kichaev G, Bhatia G, Loh P, et al. Leveraging polygenic functional enrichment to
1142 improve GWAS power. *The American Journal of Human Genetics.* 2019;104(1):65-75.
- 1143 53. Tsuda H, Yamasaki H. Type I and type II t-cell profiles in aplastic anemia and refractory
1144 anemia. *Am J Hematol.* 2000;64(4):271-274.
- 1145 54. OuYang Z, Hirota Y, Osuga Y, et al. Interleukin-4 stimulates proliferation of
1146 endometriotic stromal cells. *The American journal of pathology.* 2008;173(2):463-469.
- 1147 55. Hsu C, Yang B, Wu M, Huang K. Enhanced interleukin-4 expression in patients with
1148 endometriosis. *Fertil Steril.* 1997;67(6):1059-1064.
- 1149 56. Watanabe K, Arumugam S, Sreedhar R, et al. Small interfering RNA therapy against
1150 carbohydrate sulfotransferase 15 inhibits cardiac remodeling in rats with dilated
1151 cardiomyopathy. *Cell Signal.* 2015;27(7):1517-1524.
- 1152 57. Laudanski P, Charkiewicz R, Kuzmicki M, et al. Profiling of selected angiogenesis-
1153 related genes in proliferative eutopic endometrium of women with endometriosis. *European*
1154 *Journal of Obstetrics & Gynecology and Reproductive Biology.* 2014;172:85-92.
- 1155 58. Hannoush ZC, Puerta H, Bauer MS, Goldberg RB. New JAG1 mutation causing alagille
1156 syndrome presenting with severe hypercholesterolemia: Case report with emphasis on
1157 genetics and lipid abnormalities. *The Journal of Clinical Endocrinology & Metabolism.*
1158 2016;102(2):350-353.

- 1159 59. Germain M, Chasman DI, De Haan H, et al. Meta-analysis of 65,734 individuals
1160 identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism.
1161 *The American Journal of Human Genetics*. 2015;96(4):532-542.
- 1162 60. Bertina RM, Koeleman BP, Koster T, et al. Mutation in blood coagulation factor V
1163 associated with resistance to activated protein C. *Nature*. 1994;369(6475):64.
- 1164 61. Lauw FN, Dekkers PE, te Velde AA, et al. Interleukin-12 induces sustained activation of
1165 multiple host inflammatory mediator systems in chimpanzees. *J Infect Dis*. 1999;179(3):646-
1166 652.
- 1167 62. Levi M, van der Poll T. Two-way interactions between inflammation and coagulation.
1168 *Trends Cardiovasc Med*. 2005;15(7):254-259.
- 1169 63. Dahm A, van Hylckama Vlieg A, Bendz B, Rosendaal F, Bertina RM, Sandset PM. Low
1170 levels of tissue factor pathway inhibitor (TFPI) increase the risk of venous thrombosis. *Blood*.
1171 2003;101(11):4387-4392.
- 1172 64. Paysant J, Soria C, Cornillet-Lefèbvre P, et al. Long-term incubation with IL-4 and IL-10
1173 oppositely modifies procoagulant activity of monocytes and modulates the surface expression
1174 of tissue factor and tissue factor pathway inhibitor. *Br J Haematol*. 2005;131(3):356-365.
- 1175 65. Naldini A, Aarden L, Pucci A, Bernini C, Carraro F. Inhibition of interleukin-12
1176 expression by α -thrombin in human peripheral blood mononuclear cells: A potential
1177 mechanism for modulating Th1/Th2 responses. *Br J Pharmacol*. 2003;140(5):980-986.
- 1178 66. Nath AP, Ritchie SC, Grinberg NF, et al. Multivariate genome-wide association analysis
1179 of a cytokine network reveals variants with widespread immune, haematological and
1180 cardiometabolic pleiotropy. *bioRxiv*. 2019:544445.

- 1181 67. Surakka I, Kristiansson K, Anttila V, et al. Founder population-specific HapMap panel
1182 increases power in GWA studies through improved imputation accuracy and CNV tagging.
1183 *Genome Res.* 2010;20(10):1344-1351.
- 1184 68. Mitt M, Kals M, Pärn K, et al. Improved imputation accuracy of rare and low-frequency
1185 variants using population-specific high-coverage WGS-based imputation reference panel.
1186 *European Journal of Human Genetics.* 2017;25(7):869.
- 1187

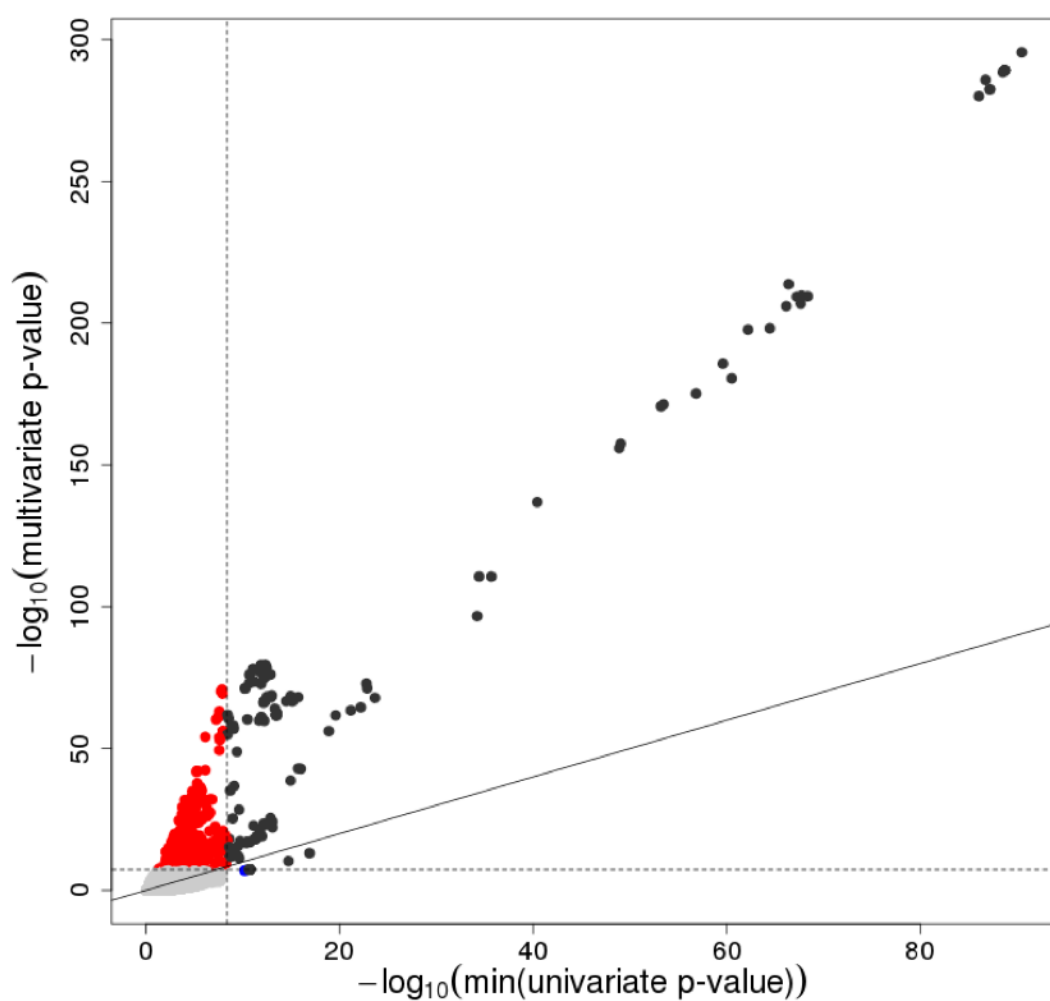
1188 FIGURES



1189

1190 **Figure 1. Study workflow.** The novel LCP-GWAS method that enables follow-up analyses

1191 such as fine-mapping for multivariate GWAS is illustrated in the violet panel on the right.



Total # variants: 11 329 225		Univariate	
		Non-significant	Significant
Multivariate	Significant	999 (0.0088 %)	190 (0.0016%)
	Non-significant	11 328 034 (99.98 %)	2 (0.000017 %)

1192

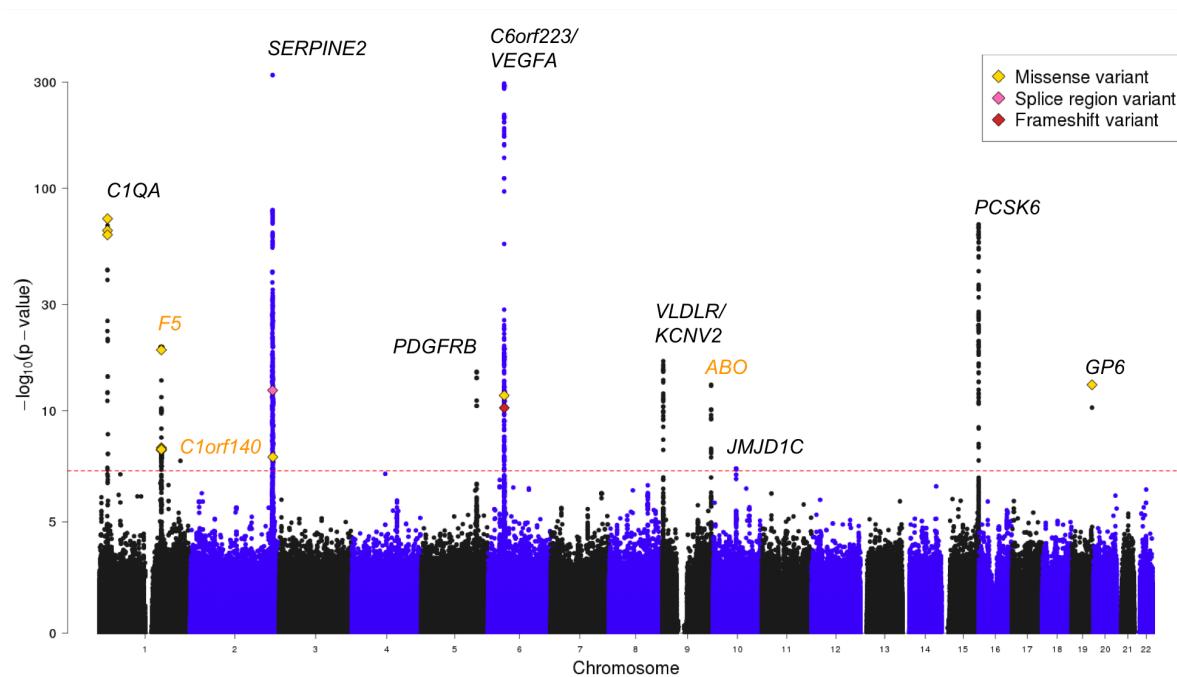
1193 **Figure 2. Power comparison between multivariate and univariate methods.** Red and blue

1194 dots represent genetic variants reaching genome-wide significance only by the multivariate or

1195 univariate method, respectively. Black dots reach the genome-wide significance threshold by

1196 both methods and grey dots do not by either method. Respective numbers are reported in the

1197 accompanying table.



1198

1199 **Figure 3. Manhattan plot of the multivariate GWAS results on 12 inflammatory**

1200 **biomarkers.** Gene names colored in orange represent associations only detected by the

1201 multivariate method while black are detected by both multivariate and univariate methods. 13

1202 genome-wide significant functional variants are denoted with diamonds.

1203 **TABLES**

1204

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-1 α)	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

1205

1206 **Table 1. Characterization of the 12 inflammatory biomarker measurements.** n = sample

1207 size, SD = standard deviation. The cytokine concentrations are pg/ml.

1208

Variant ^a	Locus	AF ^b (FIN enrichment)	Multivariate p-value	Minimum univariate p-value (biomarker)	Driver traits ^c	Previous biomarker associations ^d	QTL	Gene	p-value	FinnGen disease associations ^e	FinnGen association statistics		Novel disease association ^f
											OR	p-value	
rs17887074*	<i>C1QA</i>	1.48% (4.64)	1.21E-73	1.70E-23 (TNF-β)	TNF-β	IFN-γ, IL-17, TNF-β	—	—	—	—	—	—	—
rs3820060*	<i>F5</i>	29.60 %	6.15E-20	1.07E-3 (VEGF-A)	IL-4, IL-12p70	NOVEL	eQTL eQTL	F5 NME7	9.2E-118 1.8E-10	—	—	—	—
rs9332701**	<i>F5</i>	4.03 %	3.71E-06	3.02E-2 (VEGF-A)	—	NOVEL	pQTL	F5	1.0E-23	—	—	—	—
rs151049317	<i>C1orf140</i>	0.98 %	1.79E-08	2.79E-2 (PDGF-BB)	PDGF-BB	NOVEL	—	—	—	—	—	—	—
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	pQTL	PDGF-BB	4.6E-13	Hypertrophic scar	1.34	7.5E-5	YES
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	—	—	—	—	—	—	—
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	—	—	—	—	—	—	—
rs2304058	<i>PDGFRB</i>	37.7 %	1.08E-15	2.46E-8 (SDF-1α)	SDF-1α	NOVEL	pQTL	PDGFRB	2.3E-458	—	—	—	—
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	pQTL	VEGFA	7.8E-71	—	—	—	—
rs4714726	<i>C6orf223 / VEGFA</i>	45.5 %	1.12E-11	6.95E-5 (VEGF-A)	VEGF-A	IL-10, IL-12p70, VEGF-A	—	—	—	—	—	—	—
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	—	—	—	—	—	—	—
rs10122155	<i>VLDLR / KCNV2</i>	43.3 %	1.08E-04	5.43E-3 (VEGF-A)	—	IFN-γ, IL-10, IL-12p70, VEGF-A	—	—	—	—	—	—	—
rs550057	<i>ABO</i>	31.0 %	8.49E-14	2.08E-5 (IL-4)	IL-4	FGF2	pQTL pQTL pQTL pQTL	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 1.24 1.04 1.04	4.7E-8 4.9E-5 2.6E-5 2.9E-5 4.4E-5 9.4E-6 4.8E-5	NO YES YES
rs7080386	<i>JMJD1C</i>	38.1 %	4.04E-08	1.86E-11 (VEGF-A)	VEGF-A	IFN-γ, IL-10, IL-12p70, VEGF-A	pQTL	HB-EGF	1.60E-13	—	—	—	—
rs111482836	<i>PCSK6</i>	29.0 %	3.42E-05	0.010 (PDGF-BB)	—	PDGF-BB	—	—	—	—	—	—	—
rs12905972	<i>PCSK6</i>	21.8 %	0.035	0.027 (VEGF-A)	—	PDGF-BB	—	—	—	—	—	—	—
rs6598475	<i>PCSK6</i>	65.7 %	1.27E-54	2.63E-8 (PDGF-BB)	PDGF-BB, SDF-1α, IL-4, IL-17	PDGF-BB	—	—	—	—	—	—	—
rs11639051	<i>PCSK6</i>	24.3 %	2.71E-69	1.11E-13 (PDGF-BB)	PDGF-BB, SDF-1α, IL-4, IL-10	PDGF-BB	—	—	—	—	—	—	—
rs199588110*	<i>GP6</i>	0.33% (3.69)	8.54E-14	1.25E-17 (IL-17)	all 12 biomarkers	G-CSF	—	—	—	Benign neoplasm of meninges	6.4	4.9E-5	YES

1209

1210 **Table 2. Results of the 19 putative causal variants, i.e. the representative variants of the**
 1211 **19 credible sets.**

1212 * missense variant

1213 ▲ variant was in high linkage disequilibrium ($r^2 = 0.997$) with a missense variant (rs6030)
 1214 (Supplementary Table 2)

1215 **missense variant that replaced the initial representative variant (rs61808983) in its credible
 1216 set

1217 ^a Bolded variants are lead variants.

1218 ^b AF = allele frequency, FIN enrichment = AF in Finns compared to AF in non-Finnish,
 1219 Swedish, Estonian Europeans (NFSEE) in the gnomAD genomes database; reported if it was
 1220 at least 1.5-fold.

1221 ^c Driver traits can only be determined for those variants with a genome-wide significant
1222 association in the multivariate analysis.

1223 ^d Previous associations with the 12 biomarkers were searched for in the NHGRI-EBI GWAS
1224 Catalog within a region encompassing ± 500 kB around the variant. An association was
1225 regarded novel if no associations with any of the 12 biomarkers had been reported in this
1226 region.

1227 ^e Only associations that remain significant after conditioning are reported here. Closely related
1228 disease diagnoses are represented in a shared cell and their replication is assessed jointly.

1229 ^f Novelty of disease associations was assessed at gene-level.