

1 **Novel protein-altering variants associated with serum apolipoprotein and lipid levels**

2 Niina Sandholm*, ^{1,2,3}, Ronja Hotakainen*, ^{1,2,3}, Jani K Haukka, ^{1,2,3}, Fanny Jansson Sigfrids, ^{1,2,3},
3 Emma H Dahlström, ^{1,2,3}, Anni Antikainen, ^{1,2,3}, Erkka Valo, ^{1,2,3}, Anna Syreeni ^{1,2,3}, Elina
4 Kilpeläinen ⁴, Anastasia Kytölä ⁴, Aarno Palotie ^{4,5,6}, Valma Harjutsalo, ^{1,2,3}, Carol Forsblom, ^{1,2,3},
5 Per-Henrik Groop, ^{1,2,3,7}, on behalf of the FinnDiane Study Group

- 6 1. Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland
- 7 2. Department of Nephrology, University of Helsinki and Helsinki University Hospital,
8 Helsinki, Finland
- 9 3. Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University
10 of Helsinki, Helsinki, Finland
- 11 4. Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki,
12 Helsinki, Finland
- 13 5. Analytic and Translational Genetics Unit, Department of Medicine, Department of Neurology
14 and Department of Psychiatry Massachusetts General Hospital, Boston, MA, USA.
- 15 6. The Stanley Center for Psychiatric Research and Program in Medical and Population
16 Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA.
- 17 7. Department of Diabetes, Central Clinical School, Monash University, Melbourne, Victoria,
18 Australia

19 * These authors contributed equally to this work

20 **Corresponding authors**

21 Niina Sandholm, niina.sandholm@helsinki.fi

22 Per-Henrik Groop, per-henrik.groop@helsinki.fi

23

24 **4645 words**

25 **Tables: 3**

26 **Figures: 4**

27 **Abstract**

28 Dyslipidemia is a major risk factor for cardiovascular disease. While common genetic variants are
29 known to modestly affect the serum lipid concentrations, rare genetic mutations can cause monogenic
30 forms of hypercholesteremia and other genetic disorders of lipid metabolism. Aiming to identify low-
31 frequency protein-altering variants (PAVs) affecting lipoprotein and lipid traits, we analyzed whole-
32 exome and whole-genome sequencing data of 481 and 573 individuals with type 1 diabetes,
33 respectively. The phenotypic data consisted of 97 serum lipid, apolipoprotein, or other metabolic
34 phenotypes obtained with clinical laboratory measurements and nuclear magnetic resonance (NMR)
35 technology. Single variant analysis identified a novel association between *LIPC* p.Thr405Met
36 (rs113298164) and serum apolipoprotein-A1 levels ($p=7.8\times 10^{-8}$). In the *APOB* gene, we identified
37 novel associations at two protein-truncating variants (PTVs) resulting in lower serum apolipoprotein
38 B levels ($p=5.6\times 10^{-4}$). The burden of PAVs was significantly associated with lipid phenotypes in
39 *LIPC*, *RBM47*, *TRMT5*, and *GTF3C5* ($p<2.9\times 10^{-6}$). The *RBM47* gene is required for apolipoprotein-
40 B post-translational modifications, and in our data, the association between *RBM47* and
41 apolipoprotein C-III levels was led by a rare 21 base pair Ala496-Ala502 deletion; as replication, the
42 burden of rare deleterious variants in *RBM47* was associated with TG-to-HDLC ratio in WES of
43 20,917 individuals ($p=0.0093$). Two PAVs in *GTF3C5* were highly Finnish-enriched and associated
44 with cardiovascular phenotypes in external data, whereby the *TRMT5* p.Ser185Cys lead variant was
45 associated with stroke phenotypes. Altogether, we identified both novel variant associations in known
46 lipid genes, as well as novel genes implicated in lipoprotein metabolism.

47 **Introduction**

48 Cardiovascular disease (CVD) is the leading cause of mortality worldwide.¹ Plasma lipid
49 concentrations are key CVD risk factors, and thus, lipid-lowering medication is an essential treatment
50 option to prevent CVD. Genetic factors explain approximately 10-54% of plasma lipid levels² and
51 the largest genome-wide association study (GWAS) on plasma lipid values identified nearly 400
52 genetic loci associated with plasma low-density lipoprotein cholesterol (LDLC), triglycerides (TG),
53 total cholesterol, or high-density lipoprotein cholesterol (HDLC).³ Genome-wide analyses focusing
54 on the exonic regions of the genome have identified low-frequency or rare PAVs that contribute to
55 the previously observed common variant associations, or even explain most of the associations
56 observed for those.^{4,5} The low-frequency protein-altering variants (PAVs) can have a much stronger
57 impact on the phenotype than the disease-associated common genetic variants, which are enriched
58 for gene regulatory variants and often have moderate effect sizes.⁶ Furthermore, identification of rare
59 loss-of-function variants may reveal genes that can be targeted to prevent a disease, such as the
60 LDLC-lowering loss-of-function variants in *PCSK9* (MIM:607786), the identification of which
61 resulted in the PCSK9 inhibitors for preventing CVD.⁷

62 Lipidomic profiles consisting of more detailed lipid and lipoprotein subtypes can increase our
63 understanding of the complex lipidomic regulatory networks and, occasionally, outperform the
64 traditional lipid variables in risk prediction.⁶ In addition, apolipoprotein concentrations provide added
65 awareness of the burden of circulating lipoproteins. For example, one apolipoprotein-B (apoB)
66 molecule is embedded in each very-low-density lipoprotein (VLDL), intermediate-density lipoprotein
67 (IDL), low-density lipoprotein (LDL), and lipoprotein(a) (Lp[a]) particle, and apoB seems to estimate
68 the atherogenic risk more accurately than the traditional LDLC⁸ or even multivariable data-driven
69 sub-grouping of lipoprotein subtypes.⁹ Furthermore, apolipoprotein C-III (apoC-III) – found
70 particularly in the triglyceride-rich lipoproteins (TRLs) – has been recently implicated as a CVD risk

71 factor both in the general population and in individuals with type 1 diabetes.^{10, 11} Genetic studies of
72 these refined lipid phenotypes have revealed common variants contributing e.g. to apoB levels¹², but
73 also identified rare genetic factors with drastic impact e.g. on apoC-III levels, reflected on the CVD
74 risk.¹³

75 Individuals with type 1 diabetes carry a considerable CVD risk burden, with a 7.5-fold incidence ratio
76 for coronary artery disease (CAD) *vs.* the general population; in the presence of diabetic kidney
77 disease (DKD), this ratio is up to 27-fold.¹⁴ Dyslipidemia – particularly hypertriglyceridemia – is an
78 established risk factor for both CAD and DKD in these individuals.¹⁵ The incidence of CAD increases
79 already below the currently recommended TG cut-off of 1.7 mmol/L, suggesting that the additional
80 risk imposed by lipids is pronounced in diabetes. Therefore, genetic studies on lipids in type 1 diabetes
81 are of particular importance. Furthermore, the Finnish population provides advantages and increased
82 statistical power for studying rare variants, as some deleterious rare variants are present at higher
83 frequencies in Finnish subjects due to population isolation and recent genetic bottlenecks.¹⁶
84 Therefore, using whole-exome and whole-genome sequencing (WES and WGS, respectively), we
85 aim to identify novel PAVs and protein-truncating variants (PTVs, as putative loss-of-function
86 variants) affecting serum lipid and lipoprotein measurements, complemented with serum nuclear
87 magnetic resonance (NMR) measurements in Finnish individuals with type 1 diabetes.

88

89 **Material & Methods**

90 **Patients**

91 The Finnish Diabetic Nephropathy Study (FinnDiane) is an ongoing nationwide prospective study,
92 established in 1997 to pinpoint risk factors for long-term diabetic complications. The study currently
93 comprises more than 5,200 individuals with type 1 diabetes. The study protocol was approved by the
94 Ethical Committee of the Helsinki and Uusimaa Hospital District and local ethics committees and

95 participants gave their informed consent before recruitment. This study was performed following the
96 Declaration of Helsinki.

97 **Phenotypes**

98 Type 1 diabetes was defined as an onset of diabetes before the age of 40 and the initiation of
99 permanent insulin treatment during the first year after diagnosis. The clinical characterization of the
100 participants and the recruitment has been described earlier.¹⁷ In brief, data on diabetic complications,
101 history of cardiovascular event(s), and prescribed medications were registered using standardized
102 questionnaires, and blood and urine samples were collected during a standard visit to the attending
103 physician. DNA was extracted from blood.

104 Blood serum lipid and apolipoprotein concentrations were determined at the central research
105 laboratory (CL) of Helsinki University Central Hospital, Finland,¹⁵ with detailed methods in
106 **Supplementary Table S1.**

107 Proton NMR spectroscopy was utilized to quantify numerous lipoprotein subclasses and their
108 contents along with several metabolites from the blood serum of 3,544 participants from the
109 FinnDiane Study. Lipoproteins were classified according to their diameter into VLDL, IDL, LDL,
110 and HDL particles. These were further subdivided as described earlier.¹⁸ The spectroscopy was
111 tailored to target three molecular windows: the lipoprotein lipids, low molecular weight compounds,¹⁹
112 and serum lipid extracts.²⁰ The NMR spectroscopy was performed in four different batches, and some
113 metabolites, e.g., amino acids, were determined only in one batch.

114 Lipid-lowering medication use, defined as statin use, was accounted for by using a similar approach
115 previously adopted by others.^{4,21} We divided total cholesterol by 0.8 to account for the 20% reduction
116 in serum total cholesterol induced by statins.²² We used this adjusted value to calculate LDLC with
117 the Friedewald formula.²³ We divided the subgroups of NMR-measured LDLC by 0.7 to account for
118 a 30% reduction in LDLC. As statins also affect VLDL particles, the NMR VLDL cholesterol

119 measurements were divided by 0.8.²⁴ We left serum TG and HDL cholesterol (HDLC) measurements
120 unadjusted, as heritability estimates do not significantly improve when adjusting for statin use.²⁵ All
121 other lipid variables were left unadjusted, as the exact effect of statins remains unclear.

122 The kidney status was based on albuminuria status, and subjects were classified as having normal
123 albumin excretion rate (AER <20 µg/min), microalbuminuria (20–199 µg/min), macroalbuminuria
124 (\geq 200 µg/min), or renal failure requiring dialysis or kidney transplant. CAD was based on Statistics
125 Finland and the National Care Register for Health Care using the ICD-10 codes I21, I22, and I23 for
126 myocardial infarction; and the Nordic Classification of Surgical Procedure codes for coronary bypass
127 surgery or coronary balloon angioplasty.

128 **Whole exome- and whole-genome sequencing data**

129 WES data were available for 481 individuals with either normal AER despite long (\geq 32 years)
130 diabetes duration, or with severe DKD, i.e. macroalbuminuria and/or renal failure. The sequencing
131 process, variant calling, annotation, and quality control have been described earlier.^{26, 27} In brief,
132 sequencing was performed with Illumina HiSeq2000 platform at the University of Oxford, UK, with
133 an average requirement of 20x target capture with an above 80% coverage, resulting in mean
134 sequencing depth of 54.97 bases per position. Variant calling was performed with Genome analysis
135 toolkit (GATK) v2.1, with human genome assembly GRCh37 as a reference. Variants were updated
136 to the GRCh38 assembly using the UCSC liftOver tool with default parameters and a hg19 to hg38
137 chain file.

138 WGS data were available for 598 unrelated participants from the FinnDiane Study, non-overlapping
139 with the WES individuals. Similar to WES, the WGS data included 292 controls with normal AER
140 and long diabetes duration (\geq 35 years) and 291 cases with severe DKD. The sequencing was
141 performed using an Illumina HiSeq X platform (Macrogen Inc., Rockville, MD, USA). Variant
142 calling was done using Broad Institute's best practices guidelines with GATK v4.²⁸ The human

143 genome assembly GRCh38 was used as a reference. Variants were filtered to those with variant call
144 rate >98% and in Hardy Weinberg equilibrium (HWE; p -value $>10^{-10}$, or $>10^{-50}$ in HLA region, as all
145 had type 1 diabetes). The final data included 21.92 million variants. A total of 573 samples passed
146 the quality control filters, including the percentage of mapped de-duplicated reads and excess
147 heterozygosity. Principal component analysis indicated no population outliers. Lipid-related
148 phenotypes were available for 474 individuals.

149 All WGS and WES variants were annotated for their functional effects with SnpEff v4.3²⁹ and
150 GrCh38.86 database. Variants classified by SnpEff as PTV (exon loss, frameshift, stop or start gained
151 or lost, splice acceptor, and donor variants) and PAVs (PTV plus missense variants, and inframe
152 insertions or deletions) were included in the analyses.

153 **Single variant analysis for WES and WGS variants**

154 All PAVs were tested for association with the lipid, apolipoprotein, or other metabolic phenotypes,
155 separately for WES and WGS data sets, using the Rvtests score test.³⁰ Analyses were adjusted for
156 sex, age, and the two first genetic principal components. The NMR measured phenotypes were
157 additionally adjusted for the NMR measurement batch. Inverse normal transformation was performed
158 for all trait residuals. Finally, single-variant meta-analysis of WES and WGS cohorts was performed
159 with RAREMETAL³¹ (**Figure 1**). Exome-wide significance was defined as $<4.3 \times 10^{-7}$, adjusted for
160 116,567 tested variants (Bonferroni correction for multiple testing with $\alpha=0.05$ significance level). P-
161 values $< 1 \times 10^{-5}$ were considered suggestive. Detailed single-variant statistical analyses, including
162 survival models for CVD phenotypes, were performed in R using the survival package.³²

163 **Single variant replication**

164 Variants with a P-value $<1 \times 10^{-5}$ from the single variant meta-analysis were chosen for replication in
165 FinnDiane GWAS data with 6,449 individuals, genotyped with Illumina HumanCoreExome Bead
166 arrays, genotypes called with zCall algorithm, and initial quality control performed at the University

167 of Virginia.³³ Genotyping data were lifted over to build version 38 (GRCh38/hg38), and data from
168 the four genotyping batches were merged. In sample-wise quality control, individuals with high
169 genotype missingness (>5%), excess heterozygosity (± 4 standard deviations), and non-Finnish
170 ancestry (none) were removed. In variant-wise quality control variants with high missingness (>2%),
171 low HWE p-value ($< 10^{-6}$), or minor allele count [MAC] < 3 were removed. Chip genotyped samples
172 were pre-phased with Eagle 2.3.5, and genotype imputation was performed with Beagle 4.1 (version
173 08Jun17.d8b) based on the population-specific SISu v3 imputation reference panel with WGS data
174 for 3,775 Finnish individuals; only variants with high imputation quality of $r^2 > 0.8$ were included.
175 Depending on the phenotype, data were available for 820 (tyrosine) to 4,653 (total cholesterol)
176 individuals after excluding the FinnDiane WES and WGS individuals to ensure independent
177 replication. Rvtests software was used, and analyses were adjusted for sex, age, and the kinship
178 matrix. The variants were further tested for association with any DKD (micro- or macroalbuminuria
179 or renal failure vs. normal AER), severe DKD (macroalbuminuria or renal failure vs. normal AER),
180 and renal failure vs normal AER. For CAD, we used pre-calculated GWAS results from FinnDiane.³⁴
181 We used Sanger sequencing to confirm the 21bp deletion in the *RBM47* gene in seven heterozygotes
182 with lipid data. We designed the primers with Primer3 software³⁵ and ordered them from Sigma-
183 Aldrich Company Ltd (Haverhill, UK), and sequencing was performed at FIMM (Institute for
184 Molecular Medicine Finland, Helsinki, Finland).

185 **WES and WGS gene-based analysis**

186 Gene-based tests were performed for WES and WGS data using optimized sequence kernel
187 association test (SKAT-O).³⁶ We analyzed the burden of PAVs or PTVs with MAF $< 5\%$ using
188 Rvtests (v. 2019-02-09) --kernel skato option. Analyses were adjusted for age, sex, and genetic
189 principal components. NMR phenotypes were further adjusted for the measurement batch. Statistical
190 significance for the burden of PAVs and PTVs were defined as 2.9×10^{-6} and 1.0×10^{-5} , respectively
191 (adjusted for up to 17,022 genes with PAVs, and 4,810 genes with PTVs in the WES-WGS meta-

192 analysis; Bonferroni correction with $\alpha=0.05$). Significant WES SKAT-O results were internally
193 replicated with WGS SKAT-O results, and *vice versa* (**Figure 1**). Significant replication was defined
194 as $P<0.05$.

195 Meta-analysis of the gene-based enrichment of PAVs and PTVs in WES and WGS data was
196 performed with SKAT³⁷ implemented in RAREMETAL,³¹ and based on the single-variant score test
197 results (described above) and covariance matrices from Rvtests.³⁰ The pooled variants were re-
198 annotated with the anno tool in RAREMETAL before analysis. Again, variants were limited to those
199 with MAF <5% and analyzed for all PAVs, or PTV variants only. Significant burden of PAVs or
200 PTVs were defined as for WES and WGS SKAT-O analysis (**Figure 1**).

201 **Replication of gene aggregate findings**

202 Replication for gene aggregate findings was sought from the AMP type 2 diabetes (T2D)-GENES
203 exome sequencing study on various metabolic traits including the traditional lipid variables.³⁸

204 Variant associations with cardiovascular endpoints for the single variants within the gene aggregate
205 findings were sought from the FinnGen study GWAS results based on 109 “Diseases of the
206 circulatory system” phenotypes constructed from ICD codes for 176,899 Finnish individuals (freeze
207 4, accessed 11 March 2021), imputed with a Finnish reference panel with 3,775 individuals. Variant
208 enrichment estimates in the Finnish population vs. the gnomAD non-Finnish-non-Estonian European
209 samples were available in the same data.

210 **Functional annotation**

211 Ensembl Variant Effect Predictor was used to predict the effect of the identified variants, based on
212 SIFT and PolyPhen scoring. Gene expression in various tissues was used to annotate identified genes
213 and studied in Human Protein Atlas.

214

215 **Results**

216 The WES and WGS data included 42,682 and 101,718 PAVs, respectively, available for participants
217 with lipid data (**Supplementary Table S2**); 79-82% were low-frequency variants with MAF<5%. A
218 total of 2,240 and 9,577 variants in WES and WGS, respectively, were annotated as PTV likely to
219 disrupt the protein structure; defined here as frameshift, stop or start gained or lost, exon loss, or
220 splice site acceptor and donor variants. The vast majority, 82-90% of the PTVs, had MAF<5%.

221 **Single variant association analysis**

222 In the WES-WGS meta-analysis, a missense variant rs113298164 (Thr405Met, MAF 1.7%) in *LIPC*
223 (MIM:151670) was associated with higher serum apoA-1 levels ($p=7.8\times 10^{-8}$; **Table 1, Figure 2A**).
224 In Thr405Met carriers (n=31), the median serum apoA-1 was 163 mg/dl (inter-quartile range [IQR]
225 145 - 183) mg/dl, vs. 138 (IQR 121 - 153) mg/dl in the non-carriers (multivariate ANOVA $p=1.46\times 10^{-9}$).
226 However, in Cox proportional-hazard models, Thr405Met was not associated with CAD, nor with
227 stroke (**Figures 2B, 2C**).

228 Furthermore, 26 variants were suggestively associated with lipid, apolipoprotein, lipoprotein and,
229 metabolite phenotypes ($p<1\times 10^{-5}$; **Supplementary Table S3**). One of the variants was a 21bp inframe
230 deletion in *RBM47* (MIM:619104; Ala496-Ala502del, rs564837143, MAF=1.0%, $p=2.5\times 10^{-6}$) found
231 in the WGS data only, and associated with lower serum apoC-III concentrations, with median apoC-
232 III of 3.74 (IQR=3.38-4.69) mg/dl in the six Ala496-Ala502del carriers vs. 7.79 (IQR=5.62-10.51)
233 mg/dl in the non-carriers (**Figure 2D**). In the subsequent analysis of the full WGS data with nine
234 Ala496-Ala502del carriers (with or without apoC-III available), three experienced a CAD event
235 during the full study period, not significantly different from the non-carriers (**Figure 2E-F**).

236 **Replication of putative associations**

237 The FinnDiane GWAS dataset contained 26 of the 27 lead variants with high imputation quality.³⁹
238 Two of these were replicated with nominal significance: p.Thr1017Ala (rs45604939) in *FNDC3A*
239 (MIM:615794) was associated with higher total cholesterol (MAF=0.063, p=0.04); and p.Ala382Val
240 (rs202207045) in *GTF3C5* (MIM:604890) with lower LDLC and non-HDLc (MAF 0.008, p=0.02
241 for both; **Table 1, Supplementary Table S3**). In the independent FinnGen GWAS data of 106
242 circulatory phenotypes for 176,899 Finnish individuals, rs45604939 in *FNDC3A* was nominally
243 associated (p<0.05) with multiple circulatory phenotypes.

244 **WES and WGS gene-based analysis**

245 We performed gene aggregate tests to identify genes enriched for low frequency (MAF≤5%) PAVs
246 and PTVs. In WES, PAVs in *AKAP3* (MIM:604689) were significantly associated (p<2.9×10⁻⁶,
247 adjusted for 17,022 genes) with the TG content of the extremely large VLDL particles (p=1.4×10⁻⁷;
248 **Table 2**). Furthermore, PTVs in *PTGER3* (MIM:176806) were significantly associated (p<1.0×10⁻⁵,
249 adjusted for 4,810 genes) with free cholesterol in medium-sized HDL particles (p=9.8×10⁻⁶). Two
250 additional genes reached a suggestive p-value <1×10⁻⁵ for PAVs (**Table 2**). In WGS, SKAT-O
251 analysis revealed that PAVs in *RBM47* were associated with serum apoC-III concentrations
252 (p=2.2×10⁻⁶). Of note, the association was driven by the 21 bp inframe deletion of the *RBM47* gene
253 identified in the WGS single variant analysis. Furthermore, in WGS, PTVs in *SBDS* (MIM:607444)
254 were also associated with serum apoC-III concentrations (stop gain, and a splice donor variant;
255 p=5.0×10⁻⁶). Finally, a splice donor PTV in the *DEFTIP/DEFTIP2* genes was associated with
256 phospholipids in extra-large VLDL particles (p=1.3×10⁻⁶). Four additional genes had PAVs
257 suggestively associated with lipid phenotypes (p<1×10⁻⁵; **Table 2**).

258 We sought replication of the suggestive SKAT-O results by performing an internal replication
259 between the two data sets. The PAVs of the *TRMT5* (MIM: 611023) gene were associated in WGS
260 with free cholesterol in IDL particles (p=6.8×10⁻⁶) and with phospholipids in extra small VLDL
261 particles (p=5.9×10⁻⁶), and these associations were replicated in WES (p=0.019 and p=0.015,

262 respectively; **Table 2**). In addition, the association between PAVs in *CYP3A43* (MIM:606534), and
263 cholesterol esters in large LDL particles in WGS ($p=8.7\times 10^{-6}$), was replicated in WES ($p=0.038$).

264 Finally, to increase the statistical power, we performed gene aggregate analysis in the combined WES
265 and WGS data by applying SKAT meta-analysis implemented in RAREMETAL. The burden of
266 PAVs was significantly associated ($p<2.9\times 10^{-6}$) with lipid phenotypes in four genes (*LIPC*, *RBM47*,
267 *TRMT5*, and *GTF3C5* **Table 3**): PAVs in the *LIPC* gene – including rs113298164 from the single
268 variant meta-analysis – were associated with serum apoA-1 levels ($p=1.48\times 10^{-7}$). The PAVs in
269 *RBM47* were associated with serum apoC-III concentrations also in the WES-WGS SKAT meta-
270 analysis ($p=1.33\times 10^{-6}$), and PAVs in *TRMT5* were associated with phospholipids in extra small
271 VLDL particles ($p=7.87\times 10^{-7}$). The *TRMT5* PAVs were nominally associated also with multiple IDL
272 phenotypes (**Figure 3**). Finally, PAVs found in the *GTF3C5* gene were associated with total
273 cholesterol, LDLC, and non-HDLC. SKAT meta-analysis on DKD and CVD for the lead genes
274 suggested that PAVs in *CYP3A43* are associated with DKD ($p=0.004$, rank 43/17,578 genes i.e. top
275 0.3%: **Supplementary Table S4**).

276 Among the 33 PAVs found in these lead genes, 20 were found with good imputation quality in the
277 FinnDiane GWAS data. In addition to the above-mentioned *GTF3C5* rs202207045 variant
278 association ($p=0.02$ for LDLC and non-HDLC), a *LIPC* p.Phe368Leu (rs3829462) variant was
279 associated with higher apoA-1, along with a rare (MAF=0.0002, MAC=1.5) low imputation quality
280 (0.37) *LIPC* p.Ser301Phe variant ($p=0.04$; **Supplementary Table S5**).

281 Validation of the gene aggregate results was sought from the AMP T2D-GENES exome sequence
282 gene-level analysis of rare variants.⁴⁰ Variants in *LIPC* were associated with HDLC ($p=2.72\times 10^{-14}$)
283 and other lipid traits ($p<0.05$); variants in *RBM47* with TG-to-HDLC ratio ($p=0.0093$ for variants
284 predicted deleterious by 5 methods [“5/5”]); and variants in *PTGER3* with total cholesterol (5/5
285 $p=0.0033$) and LDLC (5/5 $p=0.0045$).

286 We further sought replication for the individual variants contributing to the gene aggregate
287 associations from the GWAS data of the FinnGen study. The strongest evidence of replication was
288 found for a rare (MAF=0.004), 80-fold Finnish-enriched deleterious start-lost variant rs189383196
289 in *GTF3C5*, associated with non-ischemic cardiomyopathy ($p=2.8\times 10^{-5}$), hypertension ($p=6.7\times 10^{-4}$),
290 and 18 other circulatory phenotypes ($p<0.05$; **Supplementary Table S6**). Also, another rare
291 (MAF=0.001), 77-fold Finnish-enriched, deleterious rs369889499 (p.Tyr347Cys) variant in *GTF3C5*
292 was associated with multiple phenotypes, including angina pectoris ($p=9.20\times 10^{-5}$) and ischemic heart
293 disease ($p=6.10\times 10^{-4}$). In the *TRMT5* gene, the variant with the strongest individual association,
294 rs115400838 (p.Ser185Cys), was associated with multiple stroke phenotypes, e.g. “Stroke, excluding
295 subarachnoid hemorrhage”, $p=1.90\times 10^{-4}$.

296 **Association for genes causing monogenic forms of dyslipidemia**

297 Previously, rare variants in multiple genes have been associated with severe monogenic forms of
298 dyslipidemia. We studied the PAV and PTV burden in eight genes causing hypercholesteremia, five
299 genes causing monogenic hypertriglyceridemia, and seven genes causing genetic disorders of HDL
300 metabolism, including the *LIPC* gene (**Supplementary Table S7**).⁴¹ In the hypercholesteremia-
301 causing *APOB* (MIM:107730), we identified two frameshift PTVs in exon 26/29, associated with
302 very low serum apoB concentration ($p=5.6\times 10^{-4}$, **Supplementary Figure S2**), as well as with LDLC
303 ($p=9.5\times 10^{-4}$), non-HDLc ($p=4.8\times 10^{-4}$) and TG content in small VLDL particles ($p=0.001$; **Figure 4**,
304 **Supplementary Table S7**); the PTVs have not been previously associated with lipid traits. In addition
305 to the *LIPC* PAV association with serum apoA-1 levels described above ($p=1.48\times 10^{-7}$), the PAVs in
306 *LIPC* were associated with total HDLC and five other lipid phenotypes ($p<0.05/19$ genes=0.0026;
307 **Figure 4**, **Supplementary Table S7**). In the *CETP* (MIM:118470) gene, known for genetic disorders
308 of the HDL metabolism, PAVs were associated with serum apoA-1 levels ($p=6.9\times 10^{-5}$), total HDLC
309 ($p=4.0\times 10^{-5}$), and seven other lipid measurements in HDL particles, driven by two low-frequency
310 missense variants, rs5880 and rs1800777 previously associated with low HDLC.⁴² PAVs in the

311 hypercholesterolemia-associated *APOE* (MIM:107741) gene were associated with apoB ($p=3.5\times 10^{-4}$), total HDLC ($p=8.0\times 10^{-4}$), and total cholesterol and cholesterol esters in LDL particles, with large
312 negative effects observed for the previously reported rare p.Glu57Lys (rs201672011) variant.⁴³
313 Finally, the three previously reported PAVs in the *PCSK9* gene, including the protective rs11591147
314 loss-of-function mutation⁴⁴ were associated with total cholesterol ($p=3.0\times 10^{-4}$), LDLC ($p=0.0014$),
315 non-HDLc ($p=4.8\times 10^{-4}$) and, surprisingly, with serum isoleucine ($p=2.9\times 10^{-4}$).

317

318 Discussion

319 Dyslipidemia is a considerable risk factor for CVD and DKD.⁴⁵ In addition to the standard clinical
320 lipid laboratory measurements, here we have used apolipoproteins as well as NMR lipid, lipoprotein,
321 and metabolite measurements, combined with WES and WGS to identify genetic variants associated
322 with a total of 97 studied phenotypes. We identified novel associations in genes already implicated
323 in lipid metabolism (e.g. rs113298164 in *LIPC*, two PTVs in *APOB*), as well as multiple novel genes
324 for lipid phenotypes, e.g. *RBM47* and *SBDS* for apoC-III levels, *GTF3C5* for LDLc, and *TRMT5* for
325 phospholipids in VLDL particles.

326 The lead variant in the single-variant analysis, rs113298164 (*LIPC* p.Thr405Met), was associated
327 with elevated apoA-1 levels ($p=7.8\times 10^{-8}$). *LIPC* encodes the hepatic lipase, which is the enzyme
328 responsible for TG hydrolysis in IDL particles and, thus, the conversion of IDL to LDL particles.
329 p.Thr405Met is predicted deleterious or probably damaging by SIFT and PolyPhen, and functional
330 studies show that p.Thr405Met reduces hepatic lipase activity.^{46,47} With 1.7% MAF, it is over 4-fold
331 enriched in the Finnish population. Previously, p.Thr405Met has been identified to cause hepatic
332 lipase deficiency in a compound heterozygous state with another rare mutation in *LIPC*, causing
333 elevated total cholesterol, TG and TG-enriched VLDL and LDL particles, followed by premature
334 atherosclerosis⁴⁸; in our GWAS data, the same rs121912502 (p.Ser301Phe) variant was nominally

335 associated ($p=0.04$) with apoA-1 despite low imputation quality (0.37) and low MAF (0.0002). In
336 addition, three other rare *LIPC* variants (two of them singletons) were found in the WES/WGS data
337 in individuals with markedly higher apoA-1 levels (p -value non-significant) and two of them were
338 predicted deleterious (**Supplementary Table S6**); rs3829462 (p.Phe368Leu) was associated with
339 apoA-1 also in the GWAS data ($p=0.02$). ApoA-1 is a key structural component of HDL particles –
340 generally associated with lower risk of CVD. While association with higher apoA-1 and HDLC may
341 seem contradictory to the association with high total cholesterol and hypertriglyceridemia, severe
342 hepatic lipase deficiency is characterized by an increase in apoA-1, HDLC, and HDL TG content⁴⁹,
343 all seen in our data as well.

344 Common variants in the *LIPC* gene are strongly associated with serum HDLC and apoA-1 levels.¹²
345 In a recent Mendelian Randomization analysis, variants associated with elevated apoA-1 levels were
346 associated with lower risk of CAD in the univariate analysis; however, this effect disappeared when
347 accounted for variants affecting apoB levels.¹² Altogether, these findings suggest that rare *LIPC*
348 variants, rs113298164 (p.Thr405Met) in particular, affect hepatic lipase activity leading to elevated
349 apoA-1 levels, but likely without direct consequence to CVD risk.

350 Importantly, we identified two PTVs in *APOB* associated with drastically low serum apoB
351 concentrations (**Supplementary Figure 2**); to our knowledge, these variants have not been previously
352 associated with lipid traits. However, with only three individuals, we do not see any association with
353 CVD endpoints.

354 In SKAT-O gene burden tests, we showed that *RBM47* was associated with lower apoC-III levels.
355 This association was driven by rs564837143, a 21 bp inframe deletion found in the WGS data, located
356 in the 6th exon. Even though the association at rs564837143 (Ala496-Ala502del) did not replicate in
357 our GWAS data, we obtained external validation, as the burden of rare deleterious variants in *RBM47*
358 was associated with TG-to-HDLC ratio ($p=0.0093$) in WES of 20,917 individuals.³⁸ Furthermore,
359 another rare missense variant in the 7th exon was recently shown to have a large impact on blood

360 pressure in a large meta-analysis.⁵⁰ *RBM47* encodes an RNA-binding protein essential for post-
361 transcriptional modification of the apoB mRNA in particular. This modification creates a premature
362 stop codon in the transcript, resulting in the production of the shorter intestinal isoform apoB-48
363 instead of the longer isoform apoB-100 produced by the liver.⁵¹ Of note, we have previously shown
364 that apoB-48 is elevated in individuals with type 1 diabetes both at fasting and postprandially.⁵² In
365 this study, however, the identified *RBM47* deletion was associated with lower serum apoC-III levels.
366 Whereas one copy of apoB is firmly embedded within the surface of each TRL (i.e., chylomicrons,
367 VLDL, and IDL) and LDL particle, apoC-III is dynamically redistributed between these and HDL
368 particles in the circulation.⁵³ ApoC-III is an important regulator of TG metabolism that, through
369 multiple pathways, impairs the clearance of the atherosclerotic, apoB-containing TRLs and their
370 remnants. One key action of apoC-III is the inhibition of lipoprotein lipase, and to some extent, also
371 hepatic lipase encoded by the *LIPC* gene.⁵⁴ There is increasing evidence – also from genetic studies
372 of a rare *APOC3* loss-of-function variant^{55, 56} – that apoC-III is an independent cardiovascular risk
373 factor, and clinical trials on apoC-III lowering therapies have yielded positive results in those with
374 high TG. ApoC-III is an important CVD risk factor also in individuals with type 1 diabetes⁵⁷. In
375 contrast to that, the *RBM47* Ala496-Ala502 carriers with lower serum apoC-III concentrations were
376 not significantly protected from CVD in our data (**Figure 2E**); however, the low number of observed
377 carriers does not allow us to draw definite conclusions.

378 PAVs in *GTF3C5* were associated with total cholesterol, LDLC, and non-HDLC. Among the eight
379 PAVs, six were predicted deleterious by SIFT and/or PolyPhen. One of them, chr9:133042147_C/T
380 (p.His72Tyr), is a novel variant, with one heterozygous carrier found in our data (verified as good
381 quality from the aligned BAM-file). Another variant, rs189383196, is either a high impact start-lost
382 variant, or a missense variant (p.Met126Thr), depending on the transcript, with over 80-fold
383 enrichment in Finns. The association for the strongest individual variant, rs202207045
384 (p.Ala382Val), was also replicated in the GWAS data (p=0.02 for LDLC and non-HDLC). The PAVs

385 in this gene were associated with multiple circulatory phenotypes, e.g. non-ischemic cardiomyopathy
386 ($p=2.8\times 10^{-5}$) in the independent FinnGen data. Interestingly, the strongest association within the
387 *GTF3C5* region in the FinnGen GWAS data was at rs671412, 28 kbp downstream, with statin
388 medication ($p=3.4\times 10^{-7}$). *GTF3C5* encodes a DNA-binding general transcription factor IIIC subunit
389 5, expressed in all tissues, and little is known about the function of this gene.

390 PAVs in *TRMT5* were associated with phospholipids in extra small VLDL particles, both in WES
391 and WGS separately, as well as in WES-WGS SKAT-O meta-analysis. Among the eight variants,
392 five were predicted deleterious. As supporting evidence, the deleterious missense variant with the
393 strongest association with lower phospholipids in VLDL particles, was associated with higher risk of
394 stroke in the FinnGen data ($p=1.90\times 10^{-4}$). *TRMT5* encodes a tRNA methyltransferase 5 involved in
395 mitochondrial tRNA methylation.

396 Other novel findings worth mentioning are PTVs in *SBDS*, as well as PAVs in *CYP3A43*, *PTGER3*,
397 and *AKAP3*. Loss-of-function variants in *SBDS* cause autosomal recessive Shwachman-Diamond
398 Syndrome 1, characterized by exocrine pancreatic dysfunction among other symptoms⁵⁸; our
399 observed association between heterozygous *SBDS* PTVs and apoC-III may be affected by a similar
400 pathway. PAVs in *CYP3A43* were associated with LDL cholesterol esters in WGS and replicated in
401 WES; *CYP3A43* was the only gene with evidence of association with clinical outcome in our WES-
402 WGS data (SKAT $p=0.004$ for DKD, rank 43/17,578 genes). While little is known about the gene, it
403 encodes one of the cytochrome P450 proteins, which are involved in the synthesis of cholesterol,
404 steroids and other lipids and, importantly, metabolize most of the drugs and can cause toxic drug-
405 drug interactions, e.g. with the statins.⁵⁹

406 This study represents the first comprehensive analysis of PAVs associated with detailed lipid,
407 apolipoprotein, and lipoprotein phenotypes in individuals with type 1 diabetes. While some of the
408 observed associations may be specific to individuals with diabetes e.g. through disturbances in insulin
409 signaling, we hypothesize that many of the associations observed in this high risk population may be

410 generalized to the wider population, as was seen in our replication attempts both for the single variant
411 and gene level findings in the general population data sets. Despite a limited number of samples, we
412 were able to identify multiple novel genetic associations. Of note, many of the identified variants
413 were markedly enriched in the Finnish population, e.g. the 80-fold Finnish-enriched *GTF3C5* PAVs,
414 providing one potential explanation why these variants have not been detected in earlier studies. It is
415 of note that many previous, larger studies were either based on chip genotyping,⁶⁰ or included only
416 the standard clinical lipid measurements such as total cholesterol, LDLC, HDLC and TG levels
417 (quantitative trait analysis in a subset of *Flannick et al.*).⁴⁰ While limited evidence of replication was
418 found for the single-variant associations in the FinnDiane GWAS data, many of the identified PAVs
419 or genes were associated with relevant metabolic traits and clinical endpoints in larger external data
420 sets.

421 Previous studies suggest that apoC-III is an important, independent risk factor for CVD. While we
422 identified a seven amino acid deletion in *RBM47* associated with lower apoC-III levels, the variant
423 was not associated with CVD in longitudinal analysis, and further studies are needed to elucidate the
424 biological mechanism that it exerts on the apolipoprotein levels. To conclude, in addition to the
425 previously known associations, we identified multiple novel genes involved in lipid metabolism.

426

427 **Supplemental Information description:**

428 Supplemental material (.docx), including Tables S1-S7 and Figures S1 and S2.

429

430 **Acknowledgements**

431 We want to acknowledge the participants and investigators of the FinnGen study for the look-ups of
432 the lead findings; and ELIXIR Finland node hosted at CSC – IT Center for Science for ICT resources

433 that enabled the WES and WGS data processing. We further acknowledge the physicians and nurses
434 at each FinnDiane center participating in the collection of the patient data.

435

436 **Funding**

437 The study was supported by grants from Folkhälsan Research Foundation, Wilhelm and Else
438 Stockmann Foundation, “Liv och Hälsa” Society, Sigrid Juselius Foundation, Helsinki University
439 Central Hospital Research Funds (EVO TYH2018207), Novo Nordisk Foundation (NNF
440 OC0013659), Academy of Finland (299200, and 316664), European Foundation for the Study of
441 Diabetes (EFSD) Young Investigator Research Award funds, and an EFSD award supported by
442 EFSD/Sanofi European Diabetes Research Programme in Macrovascular Complications.

443

444 **Declaration of Interests**

445 P-H G has received investigator-initiated research grants from Eli Lilly and Roche, is an advisory
446 board member for AbbVie, Astellas, AstraZeneca, Bayer, Boehringer Ingelheim, Cebix, Eli Lilly,
447 Janssen, Medscape, Merck Sharp & Dohme, Mundipharma, Nestlé, Novartis, Novo Nordisk and
448 Sanofi; and has received lecture fees from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Elo Water,
449 Genzyme, Merck Sharp & Dohme, Medscape, Novartis, Novo Nordisk, PeerVoice and Sanofi. Other
450 authors declare no competing interests.

451

452

453 **Web Resources**

454 AMP T2D-GENES exome sequencing, <https://t2d.hugeamp.org/>

455 Beagle, https://faculty.washington.edu/browning/beagle/b4_1.html

456 dbSNP, <https://www.ncbi.nlm.nih.gov/snp/>

457 Eagle, <https://data.broadinstitute.org/alkesgroup/Eagle/>
458 FinnGen study GWAS results, freeze 3: <https://r3.finnngen.fi/>
459 FinnGen study GWAS results, freeze 4: <https://r4.finnngen.fi/>
460 GATK, <https://gatk.broadinstitute.org/hc/en-us>
461 Human Protein Atlas, <https://www.proteinatlas.org/>
462 OMIM, <https://www.omim.org/>
463 Primer3, <https://primer3.org/>
464 R survival package: <https://cran.r-project.org/web/packages/survival/>
465 RAREMETAL, https://genome.sph.umich.edu/wiki/RAREMETAL_Documentation
466 Rvtests, <http://zhanxw.github.io/rvtests/>
467 SnpEff, <https://pcingola.github.io/SnpEff/>
468 UCSC LiftOver program, <https://genome-store.ucsc.edu/>
469 Variant effect predictor VEP, <https://www.ensembl.org/Tools/VEP>

470

471 **Data and Code Availability**

472 The sequencing data supporting the current study have not been deposited in a public repository
473 because of restrictions due to the study consent and local regulations.

474

475 **References**

- 476 1. Naghavi, M., Abajobir, A.A., Abbafati, C., Abbas, K.M., Abd-Allah, F., Abera, S.F., Aboyans,
477 V., Adetokunboh, O., Afshin, A., Agrawal, A. et al. (2017). Global, regional, and national age-
478 sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global
479 Burden of Disease Study 2016. *The Lancet* 390, 1151-1210.
- 480 2. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R.,
481 DuVall, S.L., Li, J., Peloso, G.M. et al. (2018). Genetics of blood lipids among ~300,000 multi-
482 ethnic participants of the Million Veteran Program. *Nat. Genet.* 50, 1514-1523.

- 483 3. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R.,
484 DuVall, S.L., Li, J., Peloso, G.M. et al. (2018). Genetics of blood lipids among ~300,000 multi-
485 ethnic participants of the Million Veteran Program. *Nat. Genet.* 50, 1514-1523.
- 486 4. Liu, D.J., Peloso, G.M., Yu, H., Butterworth, A.S., Wang, X., Mahajan, A., Saleheen, D., Emdin,
487 C., Alam, D., Alves, A.C. et al. (2017). Exome-wide association study of plasma lipids in
488 >300,000 individuals. *Nat. Genet.* 49, 1758-1766.
- 489 5. Surakka, I., Horikoshi, M., Mägi, R., Sarin, A., Mahajan, A., Lagou, V., Marullo, L., Ferreira, T.,
490 Miraglio, B., Timonen, S. et al. (2015). The impact of low-frequency and rare variants on lipid
491 levels. *Nat. Genet.* 47, 589-597.
- 492 6. Tabassum, R., Ripatti, S. (2021). Integrating lipidomics and genomics: emerging tools to
493 understand cardiovascular diseases. *Cellular and Molecular Life Sciences*.
- 494 7. Cohen, J.C., Boerwinkle, E., Mosley, T.H., Hobbs, H.H. (2006). Sequence Variations in PCSK9,
495 Low LDL, and Protection against Coronary Heart Disease. *N. Engl. J. Med.* 354, 1264-1272.
- 496 8. Sniderman, A.D., Thanassoulis, G., Glavinovic, T., Navar, A.M., Pencina, M., Catapano, A.,
497 Ference, B.A. (2019). Apolipoprotein B Particles and Cardiovascular Disease: A Narrative
498 Review. *JAMA cardiology* 4, 1287-1295.
- 499 9. Ohukainen, P., Kuusisto, S., Kettunen, J., Perola, M., Järvelin, M., Mäkinen, V., Ala-Korpela, M.
500 (2020). Data-driven multivariate population subgrouping via lipoprotein phenotypes versus
501 apolipoprotein B in the risk assessment of coronary heart disease. *Atherosclerosis* 294, 10-15.
- 502 10. Kanter, J.E., Shao, B., Kramer, F., Barnhart, S., Shimizu-Albergine, M., Vaisar, T., Graham,
503 M.J., Crooke, R.M., Manuel, C.R., Haeusler, R.A. et al. (2019). Increased apolipoprotein C3
504 drives cardiovascular risk in type 1 diabetes. *J. Clin. Invest.* 129, 4165-4179.
- 505 11. Taskinen, M., Packard, C.J., Borén, J. (2019). Emerging Evidence that ApoC-III Inhibitors
506 Provide Novel Options to Reduce the Residual CVD. *Curr. Atheroscler. Rep.* 21, 27.
- 507 12. Richardson, T.G., Sanderson, E., Palmer, T.M., Ala-Korpela, M., Ference, B.A., Davey Smith,
508 G., Holmes, M.V. (2020). Evaluating the relationship between circulating lipoprotein lipids and
509 apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation
510 analysis. *PLoS medicine* 17.
- 511 13. Jørgensen, A.B., Frikke-Schmidt, R., Nordestgaard, B.G., Tybjaerg-Hansen, A. (2014). Loss-of-
512 Function Mutations in APOC3 and Risk of Ischemic Vascular Disease. *N. Engl. J. Med.* 371,
513 32-41.
- 514 14. Harjutsalo, V., Thomas, M.C., Forsblom, C., Groop, P., On Behalf of the FinnDiane, Study
515 Group. (2018). Risk of coronary artery disease and stroke according to sex and presence of
516 diabetic nephropathy in type 1 diabetes. *Diabetes Obes. Metab.* 20, 2759-2767.
- 517 15. Tolonen, N., Forsblom, C., Thorn, L., Wadén, J., Rosengård-Bärlund, M., Saraheimo, M.,
518 Heikkilä, O., Pettersson-Fernholm, K., Taskinen, M.-., Groop, P.-. et al. (2008). Relationship
519 between lipid profiles and kidney function in patients with type 1 diabetes. *Diabetologia* 51, 12-
520 20.

- 521 16. Lim, E.T., Würtz, P., Havulinna, A.S., Palta, P., Tukiainen, T., Rehnström, K., Esko, T., Mägi,
522 R., Inouye, M., Lappalainen, T. et al. (2014). Distribution and Medical Impact of Loss-of-
523 Function Variants in the Finnish Founder Population. *PLOS Genetics* 10, e1004494.
- 524 17. Thorn, L.M., Forsblom, C., Fagerudd, J., Thomas, M.C., Pettersson-Fernholm, K., Saraheimo,
525 M., Wadén, J., Rönneback, M., Rosengård-Bärlund, M., Björkstén, C.a. et al. (2005). Metabolic
526 Syndrome in Type 1 Diabetes. *Diabetes Care* 28, 2019-2024.
- 527 18. Mäkinen, V.-., Soininen, P., Kangas, A.J., Forsblom, C., Tolonen, N., Thorn, L.M., Viikari, J.,
528 Raitakari, O.T., Savolainen, M., Groop, P.-. et al. (2013). Triglyceride-cholesterol imbalance
529 across lipoprotein subclasses predicts diabetic kidney disease and mortality in type 1 diabetes:
530 the FinnDiane Study. *J. Intern. Med.* 273, 383-395.
- 531 19. Mäkinen, V., Soininen, P., Forsblom, C., Parkkonen, M., Ingman, P., Kaski, K., Groop, P., Ala-
532 Korpela, M. (2008). ¹H NMR metabonomics approach to the disease continuum of diabetic
533 complications and premature death. *Mol Syst Biol* 4, 167.
- 534 20. Inouye, M., Kettunen, J., Soininen, P., Silander, K., Ripatti, S., Kumpula, L.S., Hämäläinen, E.,
535 Jousilahti, P., Kangas, A.J., Männistö, S. et al. (2010). Metabonomic, transcriptomic, and
536 genomic variation of a population cohort. *Molecular Systems Biology* 6, 441.
- 537 21. Bentley, A., Sung, Y., Brown, M., Winkler, T., Kraja, A., Ntalla, I., Schwander, K., Chasman,
538 D., Lim, E., Deng, X. et al. (2019). Multi-ancestry genome-wide gene-smoking interaction
539 study of 387,272 individuals identifies new loci associated with serum lipids. *Nat. Genet.* 51,
540 636-648.
- 541 22. Baigent, C., Keech, A., Kearney, P.M., Blackwell, L., Buck, G., Pollicino, C., Kirby, A.,
542 Sourjina, T., Peto, R., Collins, R. et al. (2005). Efficacy and safety of cholesterol-lowering
543 treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of
544 statins. *Lancet (London, England)* 366, 1267-1278.
- 545 23. Friedewald, W.T., Levy, R.I., Fredrickson, D.S. (1972). Estimation of the concentration of low-
546 density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin.*
547 *Chem.* 18, 499-502.
- 548 24. Arad, Y., Ramakrishnan, R., Ginsberg, H.N. (1992). Effects of lovastatin therapy on very-low-
549 density lipoprotein triglyceride metabolism in subjects with combined hyperlipidemia:
550 Evidence for reduced assembly and secretion of triglyceride-rich lipoproteins. *Metabolism* 41,
551 487-493.
- 552 25. Wu, J., Province, M.A., Coon, H., Hunt, S.C., Eckfeldt, J.H., Arnett, D.K., Heiss, G., Lewis,
553 C.E., Ellison, R.C., Rao, D.C. et al. (2007). An investigation of the effects of lipid-lowering
554 medications: genome-wide linkage analysis of lipids in the HyperGEN study. *BMC Genetics* 8,
555 60.
- 556 26. Sandholm, N., Van Zuydam, N., Ahlqvist, E., Juliusdottir, T., Deshmukh, H.A., Rayner, N.W.,
557 Di Camillo, B., Forsblom, C., Fadista, J., Ziemek, D. et al. (2017). The Genetic Landscape of
558 Renal Complications in Type 1 Diabetes. *J. Am. Soc. Nephrol.* 28, 557.

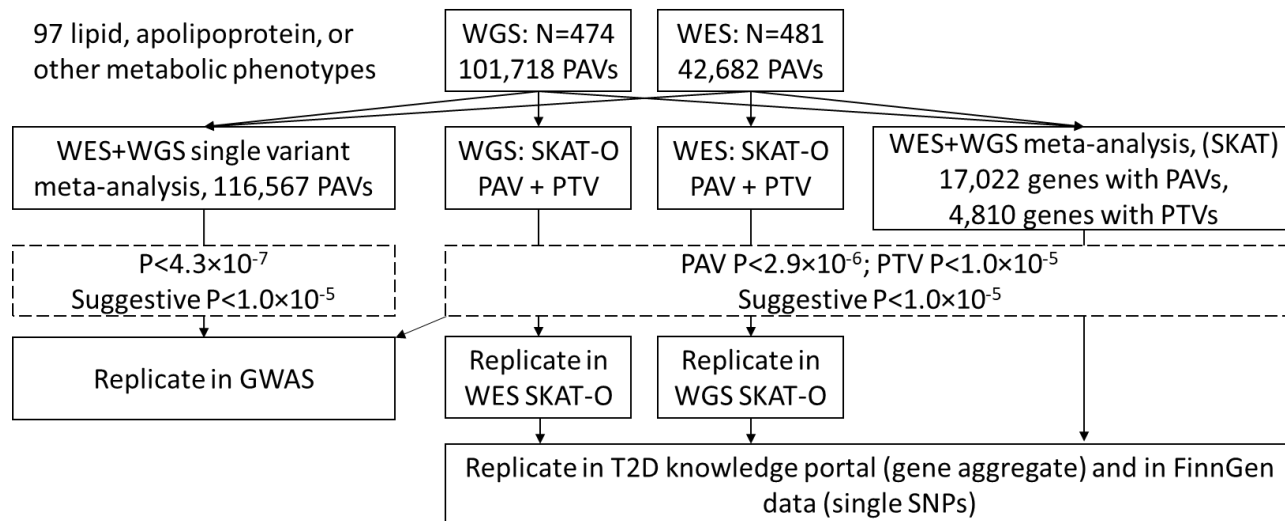
- 559 27. Sandholm, N., Haukka, J.K., Toppila, I., Valo, E., Harjutsalo, V., Forsblom, C., Groop, P.
560 (2018). Confirmation of GLRA3 as a susceptibility locus for albuminuria in Finnish patients
561 with type 1 diabetes. *Scientific Reports* 8, 12408.
- 562 28. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera,
563 Geraldine A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D. et al. (2018).
564 Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178.
- 565 29. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X.,
566 Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide
567 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2;
568 iso-3. *Fly* 6, 80-92.
- 569 30. Zhan, X., Hu, Y., Li, B., Abecasis, G.R., Liu, D.J. (2016). RVTESTS: an efficient and
570 comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics*
571 32, 1423-1426.
- 572 31. Feng, S., Liu, D., Zhan, X., Wing, M.K., Abecasis, G.R. (2014). RAREMETAL: fast and
573 powerful meta-analysis for rare variants. *Bioinformatics* 30, 2828-2829.
- 574 32. Therneau, Terry, M. A Package for Survival Analysis in R, [https://CRAN.R-](https://CRAN.R-project.org/package=survival)
575 [project.org/package=survival](https://CRAN.R-project.org/package=survival).
- 576 33. Salem, R.M., Todd, J.N., Sandholm, N., Cole, J.B., Chen, W., Andrews, D., Pezzolesi, M.G.,
577 McKeigue, P.M., Hiraki, L.T., Qiu, C. et al. (2019). Genome-Wide Association Study of
578 Diabetic Kidney Disease Highlights Biology Involved in Glomerular Basement Membrane
579 Collagen. *Journal of the American Society of Nephrology : JASN* 30, 2000-2016.
- 580 34. Antikainen, A.A.V., Sandholm, N., Tréguouët, D., Charmet, R., McKnight, A.J., Ahluwalia, T.S.,
581 Syreeni, A., Valo, E., Forsblom, C., Gordin, D. et al. (2020). Genome-wide association study
582 on coronary artery disease in type 1 diabetes suggests beta-defensin 127 as a risk locus.
583 *Cardiovasc. Res.*
- 584 35. Koressaar, T., Remm, M. (2007). Enhancements and modifications of primer design program
585 Primer3. *Bioinformatics* 23, 1289-1291.
- 586 36. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., Exome,
587 N.G., Christiani, D.C., Wurfel, M.M., Lin, X. (2012). Optimal unified approach for rare-variant
588 association testing with application to small-sample case-control whole-exome sequencing
589 studies. *Am. J. Hum. Genet.* 91, 224-237.
- 590 37. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., Lin, X. (2011). Rare-variant association testing
591 for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* 89, 82-93.
- 592 38. Flannick, J., Mercader, J.M., Fuchsberger, C., Udler, M.S., Mahajan, A., Wessel, J., Teslovich,
593 T.M., Caulkins, L., Koesterer, R., Barajas-Olmos, F. et al. (2019). Exome sequencing of 20,791
594 cases of type 2 diabetes and 24,440 controls. *Nature* 570, 71-76.

- 595 39. Syreeni, A., Sandholm, N., Cao, J., Toppila, I., Maahs, D.M., Rewers, M.J., Snell-Bergeon, J.,
596 Costacou, T., Orchard, T.J., Caramori, M.L. et al. (2019). Genetic Determinants of Glycated
597 Hemoglobin in Type 1 Diabetes. *Diabetes* 68, 858-867.
- 598 40. Flannick, J., Mercader, J.M., Fuchsberger, C., Udler, M.S., Mahajan, A., Wessel, J., Teslovich,
599 T.M., Caulkins, L., Koesterer, R., Barajas-Olmos, F. et al. (2019). Exome sequencing of 20,791
600 cases of type 2 diabetes and 24,440 controls. *Nature* 570, 71-76.
- 601 41. Musunuru, K. (2015). *Dyslipidemias: Pathophysiology, Evaluation and Management*: A.Garg,
602 Humana Press, Totowa, NJ).
- 603 42. Agerholm-Larsen, B., Tybjærg-Hansen, A., Schnohr, P., Steffensen, R., Nordestgaard, B.G.
604 (2000). Common Cholesteryl Ester Transfer Protein Mutations, Decreased HDL Cholesterol,
605 and Possible Decreased Risk of Ischemic Heart Disease. *Circulation* 102, 2197-2203.
- 606 43. Lohse, P., Brewer, H.B., Meng, M.S., Skarlatos, S.I., LaRosa, J.C., Brewer, H.B. (1992).
607 Familial apolipoprotein E deficiency and type III hyperlipoproteinemia due to a premature stop
608 codon in the apolipoprotein E gene. *J. Lipid Res.* 33, 1583-1590.
- 609 44. Cai, G., Yu, L., Huang, Z., Li, L., Fu, X. (2018). Serum PCSK9 levels, but not PCSK9
610 polymorphisms, are associated with CAD risk and lipid profiles in southern Chinese Han
611 population. *Lipids in Health and Disease* 17, 213.
- 612 45. Tolonen, N., Forsblom, C., Mäkinen, V., Harjutsalo, V., Gordin, D., Feodoroff, M., Sandholm,
613 N., Thorn, L.M., Wadén, J., Taskinen, M. et al. (2014). Different Lipid Variables Predict
614 Incident Coronary Artery Disease in Patients With Type 1 Diabetes With or Without Diabetic
615 Nephropathy: The FinnDiane Study. *Diabetes Care* 37, 2374.
- 616 46. Knudsen, P., Antikainen, M., Ehnholm, S., Uusi-Oukari, M., Tenkanen, H., Lahdenperä, S.,
617 Kahri, J., Tilly-Kiesi, M., Bensadoun, A., Taskinen, M.R. et al. (1996). A compound
618 heterozygote for hepatic lipase gene mutations Leu334→Phe and Thr383→Met: correlation
619 between hepatic lipase activity and phenotypic expression. *J. Lipid Res.* 37, 825-834.
- 620 47. Hegele, R.A., Little, J.A., Connelly, P.W. (1991). Compound heterozygosity for mutant hepatic
621 lipase in familial hepatic lipase deficiency. *Biochem. Biophys. Res. Commun.* 179, 78-84.
- 622 48. Hegele, R.A., Little, J.A., Connelly, P.W. (1991). Compound heterozygosity for mutant hepatic
623 lipase in familial hepatic lipase deficiency. *Biochem. Biophys. Res. Commun.* 179, 78-84.
- 624 49. Connelly, P.W., Hegele, R.A. (1998). Hepatic Lipase Deficiency. *Crit. Rev. Clin. Lab. Sci.* 35,
625 547-572.
- 626 50. Surendran, P., Drenos, F., Young, R., Warren, H., Cook, J.P., Manning, A.K., Grarup, N., Sim,
627 X., Barnes, D.R., Witkowska, K. et al. (2016). Trans-ancestry meta-analyses identify rare and
628 common variants associated with blood pressure and hypertension. *Nat. Genet.* 48, 1151-1161.
- 629 51. Fossat, N., Tourle, K., Radziewicz, T., Barratt, K., Liebhold, D., Studdert, J., Power, M., Jones,
630 V., Loebel, D., Tam, P. (2014). C to U RNA editing mediated by APOBEC1 requires RNA-
631 binding protein RBM47. *EMBO Rep.* 15.

- 632 52. Lassenius, M., Mäkinen, V., Fogarty, C., Peräneva, L., Jauhiainen, M., Pussinen, P., Taskinen,
633 M., Kirveskari, J., Vaarala, O., Nieminen, J. et al. (2014). Patients with type 1 diabetes show
634 signs of vascular dysfunction in response to multiple high-fat meals. *Nutr. Metab.* 11.
- 635 53. Ooi, E.M.M., Barrett, P.H., Chan, D.C., Watts, G.F. (2008). Apolipoprotein C-III:
636 understanding an emerging cardiovascular risk factor. *Clin Sci (Lond)* 114, 611-624.
- 637 54. Taskinen, M., Packard, C.J., Borén, J. (2019). Emerging Evidence that ApoC-III Inhibitors
638 Provide Novel Options to Reduce the Residual CVD. *Curr. Atheroscler. Rep.* 21, 27.
- 639 55. Anonymous (2014). Loss-of-Function Mutations in APOC3, Triglycerides, and Coronary
640 Disease. *N. Engl. J. Med.* 371, 22-31.
- 641 56. Jørgensen, A.B., Frikke-Schmidt, R., Nordestgaard, B.G., Tybjærg-Hansen, A. (2014). Loss-of-
642 Function Mutations in APOC3 and Risk of Ischemic Vascular Disease. *N. Engl. J. Med.* 371,
643 32-41.
- 644 57. Kanter, J.E., Shao, B., Kramer, F., Barnhart, S., Shimizu-Albergine, M., Vaisar, T., Graham,
645 M.J., Crooke, R.M., Manuel, C.R., Haeusler, R.A. et al. (2019). Increased apolipoprotein C3
646 drives cardiovascular risk in type 1 diabetes. *J. Clin. Invest.* 129, 4165-4179.
- 647 58. Nelson, A.S., Myers, K.C. (2018). Diagnosis, Treatment, and Molecular Pathology of
648 Shwachman-Diamond Syndrome. *Hematol. Oncol. Clin. North Am.* 32, 687-700.
- 649 59. Lynch, T., Price, A. (2007). The effect of cytochrome P450 metabolism on drug response,
650 interactions, and adverse effects. *Am. Fam. Physician* 76, 391-396.
- 651 60. Liu, D.J., Peloso, G.M., Yu, H., Butterworth, A.S., Wang, X., Mahajan, A., Saleheen, D.,
652 Emdin, C., Alam, D., Alves, A.C. et al. (2017). Exome-wide association study of plasma lipids
653 in >300,000 individuals. *Nat. Genet.* 49, 1758-1766.
- 654

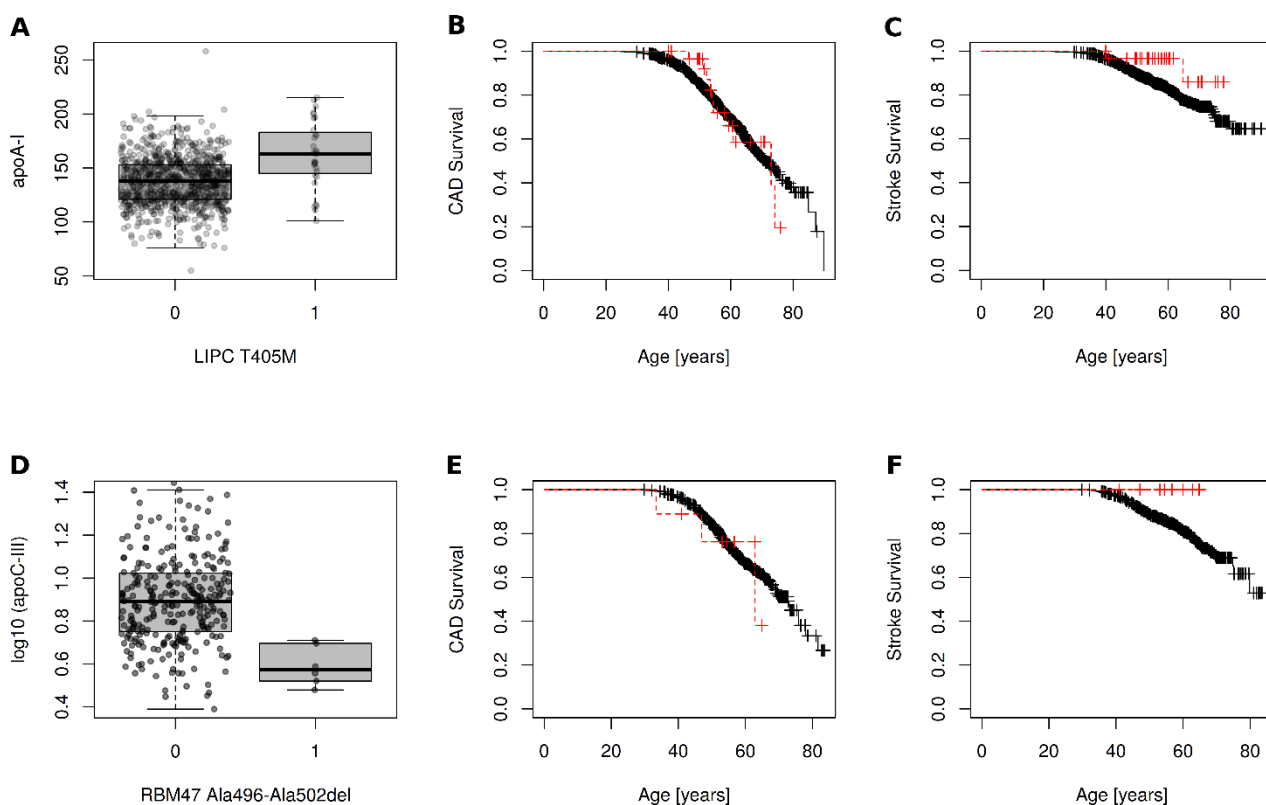
655 **Figure titles and legends:**

656 **Figure 1: Flow-chart of the study design.** PTVs: protein truncating variants i.e. exon loss,
657 frameshift, stop or start gained or lost, splice acceptor, and donor variants. PAVs: protein altering
658 variants, defined as PTV plus missense variants, inframe insertions and deletions.



659

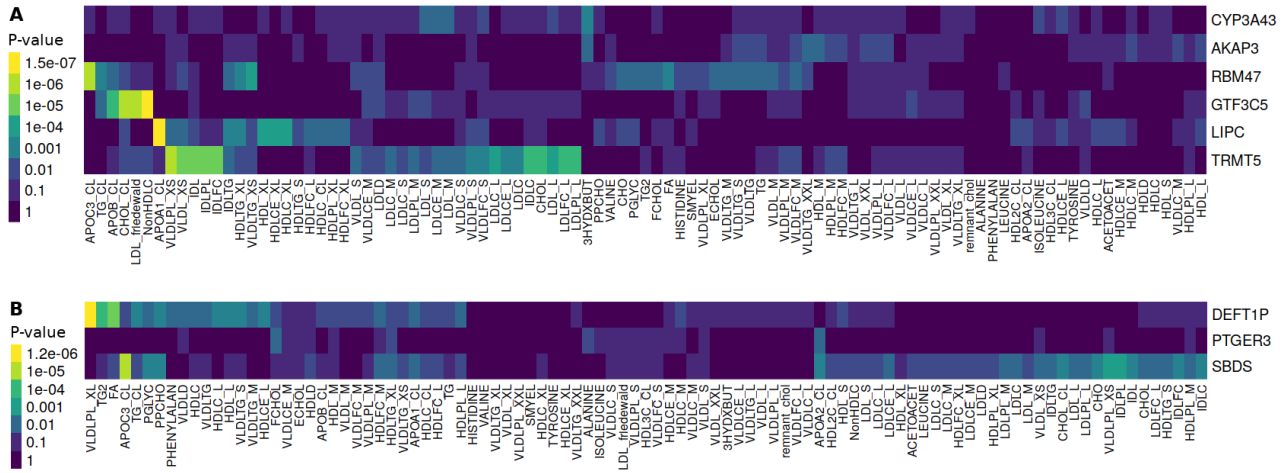
660 **Figure 2: Rare variants in *LIPC* and *RBM47* are associated with serum apolipoprotein A1 and**
661 **C3 levels, respectively, but not with cardiovascular events.** For *LIPC* Thr405Met (rs113298164)
662 data is combined for WES and WGS; *RBM47* Ala496-Ala502del is only found in the WGS data. **A:**
663 *LIPC* Thr405Met (rs113298164) is associated with higher apoA-1 ($p=7.8\times 10^{-8}$; multivariate ANOVA
664 $p=1.46\times 10^{-9}$; $N=887$ carriers, 31 non-carriers). Group number in **A** and **D** indicates the number of
665 rare variants, i.e. 0 refers to non-carriers, 1 refers to heterozygous variant carriers. **B and C:** Kaplan-
666 Meier survival plot of CAD events (**B**) and strokes (**C**), stratified by *LIPC* Thr405Met variant count:
667 black lines indicate non-carriers, red dashed line indicates variant carriers. Cox proportional hazard
668 (PH) model p -value was non-significant for both CAD and strokes. **D:** Serum apoC-III levels are
669 reduced in the *RBM47* Ala496-Ala502del carriers ($p=2.49\times 10^{-6}$, multivariate ANOVA $p=2.92\times 10^{-4}$;
670 $N=288$ non-carriers, 6 carriers). **E and F:** Kaplan-Meier survival plot of CAD events (**E**) and strokes
671 (**F**), stratified the *RBM47* Ala496-Ala502del variant number (Cox PH p -value non-significant).



672

673

674 **Figure 3: WES-WGS SKAT meta-analysis results across all the studied phenotypes for the lead**
 675 **genes.** The color indicates the strength of the gene – phenotype association (SKAT *p*-value). **A) WES-**
 676 **WGS PAV SKAT meta-analysis across phenotypes. B) WES-WGS PTV SKAT meta-analysis across**
 677 **phenotypes.** The phenotypes are ordered according to their similarity in clustering of the association
 678 data.



679

684 **Table 1: Single variant association results for variants reaching exome-wide significance**
685 **($p < 4.3 \times 10^{-7}$) or with evidence of replication ($p < 0.05$).**

Gene	Variant	MAF	SIFT/PolyPhen	Phenotype	N	P-value	Dir	P repl
<i>LIPC</i>	15:58563549 C>T p.Thr405Met rs113298164	0.017	Deleterious, probably damaging	apoA-1	918	7.8×10^{-8}	++	0.89
<i>GTF3C5</i>	9:133054787 C>T p.Ala382Val rs202207045	0.007	Tolerated, benign	LDLC	894	1.3×10^{-6}	--	0.02
<i>GTF3C5</i>	9:133054787 C>T p.Ala382Val rs202207045	0.007	Tolerated, benign	Non- HDLC	919	7.3×10^{-7}	--	0.02
<i>FNDC3A</i>	13:49201861 A>G p.Thr1017Ala rs45604939	0.064	Deleterious, probably damaging	CHOL*	748	8.8×10^{-6}	++	0.04

686 Variant: variant chromosome:base pair position with REF>ALT alleles; amino acid change, and rs
687 number. MAF: minor allele frequency. SIFT/Polyphen: Predicted effect. Dir: Effect direction in
688 WES and WGS, respectively. P repl: P-value for replication in GWAS data.

689 *rs45604939 association with total cholesterol was obtained for the NMR measured total
690 cholesterol; replication with standard laboratory total cholesterol.

691

692

693 **Table 2: WES and WGS SKAT-O gene burden test results for genes reaching a suggestive p-**
 694 **value $<1 \times 10^{-5}$, and the internal replication in the other data set.** In the exome-wide discovery, p-
 695 values $<2.9 \times 10^{-6}$ for PAVs, and p-values $<10^{-5}$ for PTV burden was defined as statistically significant,
 696 highlighted in italics; in replication, p-values <0.05 are highlighted in italics

Gene	Gene position	Type	Pheno	Discovery			Replication		
				N	Nvar	P	N	Nvar	P
Discovery study: WGS									
<i>DEFTIP/</i>									
<i>DEFTIP2</i>	8:7006280-7008824	PTV	VLDLPL XL	187	1	<i>1.3×10^{-6}</i>			
<i>SBDS</i>	7:66987676-66995696	PTV	apoC-III	294	2	<i>5.0×10^{-6}</i>			
<i>RBM47</i>	4:40423254-40629866	PAV	apoC-III	294	3	<i>2.0×10^{-6}</i>	323	1	0.398
<i>TRMT5</i>	14:60971448-60981690	PAV	CHOL	329	8	3.5×10^{-6}	419	4	0.322
<i>TRMT5</i>	14:60971448-60981690	PAV	VLDLPL XS	329	8	5.9×10^{-6}	419	4	<i>0.015</i>
<i>TRMT5</i>	14:60971448-60981690	PAV	IDLFC	329	8	6.8×10^{-6}	419	4	<i>0.019</i>
<i>TRMT5</i>	14:60971448-60981690	PAV	LDLC L	329	8	9.3×10^{-6}	419	4	0.436
<i>CCAR1</i>	10:68721143-68792377	PAV	VLDL XL	329	2	7.4×10^{-6}	419	3	0.555
<i>CYP3A43</i>	7:99828012-99866106	PAV	LDLCE M	329	6	8.7×10^{-6}	419	3	0.067
<i>CYP3A43</i>	7:99828012-99866106	PAV	LDLCE L	329	6	8.7×10^{-6}	419	3	<i>0.038</i>
<i>CYP3A43</i>	7:99828012-99866106	PAV	LDL M	329	6	9.2×10^{-6}	419	3	0.074
<i>DIPK1A</i>	1:92832728-92961522	PAV	TG	451	4	9.9×10^{-6}	469	3	0.848
Discovery study: WES									
<i>PTGER3</i>	1:70852352-71047808	PTV	HDLFC M	419	1	<i>9.8×10^{-6}</i>	329	1	0.119
<i>AKAP3</i>	12:4615507-4649047	PAV	VLDLTG XXL	304	5	<i>6.0×10^{-7}</i>	245	5	0.766
<i>TTYH1</i>	19:54415430-54436719	PAV	HDLCE L	419	3	5.7×10^{-6}			
<i>ATP4A</i>	19:35550192-35563658	PAV	HDLTG XL	419	3	8.1×10^{-6}	329	3	0.810

697 Discovery: Exome-wide analysis of either WGS, or WES data. Replication: for WGS, replication in
 698 WES data; for WES, replication in WGS data. Gene position: Chromosome number:start-end; If
 699 multiple gene isoforms exist, only one set of co-ordinates are given. Type: burden of PAVs or
 700 PTVs. Nvar: number of variants of the given variant type in the gene in discovery/ replication data.
 701 apoC-III: Serum Apolipoprotein C-III, CHOL: Total cholesterol, HDLCE L: Cholesterol ester in
 702 large HDL, HDLFC M: Free cholesterol in medium HDL, HDLTG XL: TG in extra large HDL,
 703 IDLFC: Free cholesterol in IDL particles, LDL M: Total lipids in medium LDL, LDLCE L:
 704 Cholesterol in large LDL, LDLCE M/L: Cholesterol esters in medium/large LDL, VLDL XL: Total
 705 lipids in extra large VLDL, VLDLPL XS/XL: Phospholipid in extra small/ extra large VLDL,
 706 VLDLTG XXL: TG in extremely large VLDL.

707

708

709

710

711

712 **Table 3: WES and WGS SKAT meta-analysis results for genes with significant burden of PAVs (p-**
 713 **value < 2.9×10^{-6}) or PTVs (p-value < 1×10^{-5}).** Also indicated the P-values from the separate WGS and WES
 714 SKAT-O gene burden analyses.

GENE	WES-WGS SKAT meta-analysis					WGS		WES		
	Type	Nvar	Pheno	P _{Liu}	N	Nvar	P _{SKAT-O}	N	Nvar	P _{SKAT-O}
<i>DEFT1P/DEFT1P2</i>	PTV	1	VLDLPL XL	1.23×10^{-6}	187	1	1.32×10^{-6}			
<i>SBDS</i>	PTV	2	apoC-III	5.37×10^{-6}	294	2	5.02×10^{-6}			
<i>LIPC</i>	PAV	6	apoA-1	1.48×10^{-7}	449	4	2.60×10^{-4}	469	4	1.72×10^{-4}
<i>GTF3C5</i>	PAV	8	Non-HDLC	6.86×10^{-7}	450	5	0.014	469	7	2.48×10^{-4}
<i>GTF3C5</i>	PAV	8	LDL Friedewald	9.27×10^{-7}	432	5	0.013	462	7	4.62×10^{-4}
<i>GTF3C5</i>	PAV	8	Total cholesterol	1.30×10^{-6}	451	5	0.034	469	7	3.86×10^{-4}
<i>TRMT5</i>	PAV	8	VLDLPL XS	7.87×10^{-7}	329	8	5.85×10^{-6}	419	4	0.015
<i>RBM47</i>	PAV	3	apoC-III	1.33×10^{-6}	294	3	2.01×10^{-6}	323	1	0.398

715 Type: burden of PAVs or PTVs. Nvar: number of variants of the given variant type in WES and/or WGS
 716 data. P_{Liu}: P-value for SKAT gene burden, calculated with Liu method. P_{SKAT-O}: SKAT-O P-value for
 717 association in WGS or WES data. VLDLPL XS/XL: Phospholipid in extra small/ extra large VLDL.