

# Supplement

## *ANGPTL8* protein-truncating variant and the risk of coronary disease, type 2 diabetes and adverse effects

### Table of contents

Supplementary methods	2
I.    Genotyping	2
II.   Sample and variant level quality control	2
III.  Statistical analysis	4
IV.   References	6
Supplementary tables	7
Online tables	20
Supplementary figures	21

## Supplementary methods

### Genotyping

27,294 Finrisk Study[1] samples were genotyped using the following arrays: the HumanCoreExome BeadChip, the Human610-Quad BeadChip, the Affymetrix6.0, and the Infinium HumanOmniExpress (Illumina Inc., San Diego, CA, USA). The genotype calling of variants on Human610-Quad BeadChip and HumanOmniExpress were performed with Illuminus and the variants on the HumanCoreExome chip with GenomeStudio and ZCall[2]

([www.github.com/jigold/zCall](http://www.github.com/jigold/zCall)) at the Institute for Molecular Medicine Finland (FIMM).

194,181 FinnGen samples were genotyped using the custom AxiomGT1 Affymetrix array or various Illumina arrays. The individuals not genotyped on the AxiomGT1 Affymetrix array came from 29 merged datasets from the Finrisk, Botnia, H2000/2011, Generisk and Psychiatric Family Collections, Auria, Borealis, DIME, FT17, HBP, Iddmgen, VPU and YA cohorts (Table S1).

### *Genotype calls and imputation*

The genotype calling of Human610-Quad BeadChip and Infinium HumanOmniExpress variants were performed with Illuminus, version 2 ([www.github.com/wtsi-npg/Illuminus/blob/master/illuminus.cc](http://www.github.com/wtsi-npg/Illuminus/blob/master/illuminus.cc)), and the variants on HumanCoreExome BeadChip with GenomeStudio Software, version 2011.1, and ZCall.

After array genotyping the Finrisk samples' haplotype phases were estimated using Eagle2[3], version 2.3.5 ([data.broadinstitute.org/alkesgroup/Eagle/](http://data.broadinstitute.org/alkesgroup/Eagle/)). Genotype imputation was carried out using IMPUTE2[4], version 2.3.2 ([mathgen.stats.ox.ac.uk/impute/impute\\_v2.html](http://mathgen.stats.ox.ac.uk/impute/impute_v2.html)), and a Finnish ancestry-specific reference panel consisting of 2,690 deep-coverage (25-30x) whole-genome sequences and 5,092 whole-exome sequences. With the `-merge_ref_panels` option in IMPUTE2 we were able to combine the whole-genome and whole-exome sequences into one genotype imputation reference panel. Imputed variants were filtered with the IMPUTE2 genotype information score > 0.7.

The estimation of haplotype phases of FinnGen samples were carried out with Eagle2. The genotype imputation in the FinnGen data was carried out using Beagle[5], version 4.1 ([faculty.washington.edu/browning/beagle/beagle.html](http://faculty.washington.edu/browning/beagle/beagle.html)), and a Finnish ancestry-specific reference panel consisting of 3,775 deep-coverage (25-30x) whole-genome sequences. The four lipid-associated PTVs were directly genotyped on the AxiomGT1 Affymetrix array ([www.finnngen.fi/en/researchers/genotyping](http://www.finnngen.fi/en/researchers/genotyping)) and the average IMPUTE2 genotype information score metric exceeded 0.92 for all four variants. The number of samples before and after genotype imputation, as well as the number of chip genotyped variants for each genotyping batch are listed in Table S1. The full genotyping and imputation protocol for FinnGen data is described at [dx.doi.org/10.17504/protocols.io.xbgfijw](https://doi.org/10.17504/protocols.io.xbgfijw).

### Sample and variant level quality of control

#### *The Finrisk Study*

In the Finrisk sample level quality control (QC) we excluded 890 samples with an ambiguous sex, a genotype missingness of more than 5% or excess heterozygosity (beyond  $\pm 4$  standard deviation units from the mean or of non-European ancestry). An additional 2,212 samples were excluded due to relatedness. During the Finrisk data variant-wise QC we excluded variants with a missingness greater than 2%, Hardy-Weinberg  $P < 1 \times 10^{-6}$ , a minor-allele count less than 3 and those located in the MHC region located in chromosome 6p21 or in low complexity regions of the

genome. An additional 757 samples were excluded due to missing phenotype information. In the remaining 23,435 samples we filtered the high-confident PTVs within a minor-allele frequency (MAF) range of 0.1% and 5% using LOFTEE[6] ([www.github.com/konradjk/loftee](https://www.github.com/konradjk/loftee)) and PLINK[7], version v1.90b3.45 ([www.cog-genomics.org/plink/1.9/](https://www.cog-genomics.org/plink/1.9/)), which left us with 1,377 variants to analyze.

### *The FinnGen Study*

The FinnGen genotypes first underwent basic variant and sample-level QC. Variants with a call rate less than 97%, and samples with genotyping missingness greater than 5% or with ambiguous sex were removed. Sample mix ups and extraneous duplicate copies of lower genotyping quality were manually removed. Then individuals with an ambiguous genotype-determined sex (F-score > 0.3), a genotyping success rate below 95%, excess heterozygosity (more than 4 standard deviation units from the mean or non-European ancestry), outside the population structure as specified by multidimensional scaling (maximum of five iterations) or by the first two principal components (beyond 4 standard deviation units from the mean with a maximum of five iterations) or contaminated samples (pi-hat linkage  $\geq 0.1$  with at least 14 samples) were removed. These steps were iterated through until all samples met all the criteria.

### Genetic sample outlier detection

We performed population outlier detection using principal component analysis (PCA) and a Bayesian algorithm using 36,944 independent and common variants with a high genotype probability and low missingness according to the following filters:

- Autosomal chromosomes,
- Variants with a genotype information score  $\geq 0.95$ ,
- Variants with a missingness  $\leq 0.01$  (according to GP [genotype probability])
- Variants with a minor-allele frequency  $\geq 0.05$ ,
- Linkage disequilibrium (LD) pruning with a window size of 500kb, step size of 50kb and a  $r^2$  filter < 0.1

Then, these same variants were selected from the 1000 Genomes Project[8] (1000G) and merged with that of the FinnGen data. As a result of the outlier detection, 4,686 samples were removed of which 2,292 were FinnGen Study samples. Fig. S2 shows the scatter plots of the genetic components of each sample projected onto the first three principal components. This routine successfully detected all the 1000G samples with non-European and Southern European ancestry, but failed to exclude all 1000G samples with Western European origin. The cluster of Western Europeans classified as Finns was too small to perform a second round of population outlier detection, using the PCA and Bayesian algorithm routine, without detecting substructures of the Finnish population. Therefore, another PCA on the remaining FinnGen samples was performed. The European and Finnish 1000G genotypes were then projected onto the new three-dimensional space and two clusters computed (Fig. S3). Then, we calculated the squared Mahalanobis distances of the FinnGen samples to the centroid of each cluster. Since the Mahalanobis squared distance is a sum of variables with unit variance we can see it as a sum of three independent variables and thus generate a  $\chi^2$  probability distribution with 3 degrees of freedom. In this way, we were able to calculate the probability of belonging to each cluster for every sample. Samples

with more than a 0.95 probability were classed as belonging to the FinnGen cluster and were included in further analyses. As a result, 588 outliers fell below this threshold.

#### Kinship determination

To maximize the number of unrelated samples used in our association tests we first determined all the pairs of FinnGen samples up to the second degree. The distribution of kinship values of the FinnGen Study samples is shown in Fig. S4. Next, the 588 samples with less than a 0.95 probability of belonging to the FinnGen cluster were removed. Thereafter, we used two algorithms from the network Python package ([networkx.github.io/](https://networkx.github.io/)) to flag samples up to second degree kinship in the remaining FinnGen Study samples:

1. Greedy, that removes the highest degree node from the network of relations until no more links in network remain
2. Native, performed on each subgraph of the network

These two algorithms separated the samples into three sets:

1. 131,863 unrelated samples with Finnish ancestry
2. 46,916 non-duplicate samples with Finnish ancestry but who are related to the samples in the first set
3. 4,915 samples who are either of non-Finnish ancestry, are twins/duplicates related to other samples, which were excluded.

A PCA was performed for the 131,863 unrelated samples after which the 46,916 samples in the second set and were projected onto the same multidimensional space yielding population covariates for 178,779 samples.

#### Phenotype information

Of the 178,779 non-duplicate population inlier samples, we excluded 1,880 samples with missing minimum phenotype information or a mismatch between imputed sex and the reported sex in the registry data. Thus, a total of 176,899 samples were used for the association tests. statistics of the phenome-wide association studies of the lipid-associated PTVs respectively.

### Statistical analysis

#### *Primary Analyses*

For the single-variant analysis on lipids, we considered a genome-wide significance (two-sided P value less than  $5 \times 10^{-8}$ ) to be significant to account for the testing of 1,377 PTVs in 1,209 genes. For this analysis we used the Finrisk Study cohorts, whose baseline data is shown in Table S2. The genetic association analyses were performed using PLINK[7], version v1.90b3.45 ([www.cog-genomics.org/plink/1.9/](https://www.cog-genomics.org/plink/1.9/)), Python, version 3.6 ([www.python.org](https://www.python.org/)), and the statsmodels Python package, version 0.8.0 ([www.statsmodels.org](https://www.statsmodels.org/)). The conditional tests shown in Tables S3-S6 of the lipid-associated PTVs were performed with previously identified genome-wide significant variants[9-12]. Moreover, the 95% credible sets for each lipid-associated locus are shown in Tables S7-S10.

### *Secondary and tertiary analyses*

In the analyses between cardiometabolic disease risk and the lipid-associated PTVs, we considered a two-sided P value of less than 0.05 to be significant. We also tested if these PTVs were associated with the traditional non-lipid cardiometabolic risk factors: hypertension and statin medication in the FinnGen Study (Table S11). In the phenome-wide scans we used a significance threshold of a two-sided P value less than  $2.2 \times 10^{-5}$  (Bonferroni-corrected threshold for 2,264 traits) to account for statistical significance. All the 2,264 endpoints that we used in our study had at least 100 disease cases among the 176,899 samples. Online Tables 1-3 include the tested disease case and control definitions and Online Tables 4-6 the detailed association statistics. For the association analyses we used a mixed model logistic regression R/C++ package called SAIGE, version 0.35.8.8 ([www.github.com/weizhouUMICH/SAIGE/releases/tag/0.35.8.8](http://www.github.com/weizhouUMICH/SAIGE/releases/tag/0.35.8.8)).

### *SAIGE null models*

For computing the null model for each disease endpoint, we used age, sex, 10 principal components and the genotyping batch as covariates. To avoid convergence issues, a genotyping batch was included as a covariate for an endpoint if the batch contained at least 10 cases and controls. One genotyping batch was not included as a covariate in the model to avoid the saturation of covariate values. We excluded the AxiomGT1\_b16 genotyping batch as it was not enriched for any particular disease endpoints.

For calculating the genetic relationship matrix (GRM), we used the genotype dataset where genotypes with GP < 0.95 were set as missing. Only variants with an IMPUTE2 information score > 0.95 in all genotyping batches were used. Variants with more than 3% missing genotypes or a MAF below 1% were excluded. The remaining variants were LD-pruned with a 1Mb window and  $r^2 < 0.1$ . These operations resulted in a set of 58,888 well-imputed and non-rare variants for computing the GRM. The SAIGE options in place for computing the GRM were:

- LOCO false
- numMarkers 30
- traceCVcutoff 0.0025
- ratioCVcutoff 0.001.

## References

1. Vartiainen, E., et al., *Thirty-five-year trends in cardiovascular risk factors in Finland*. Int J Epidemiol, 2010. **39**(2): p. 504-18.
2. Goldstein, J.I., et al., *zCall: a rare variant caller for array-based genotyping: genetics and population analysis*. Bioinformatics, 2012. **28**(19): p. 2543-5.
3. Loh, P.R., P.F. Palamara, and A.L. Price, *Fast and accurate long-range phasing in a UK Biobank cohort*. Nat Genet, 2016. **48**(7): p. 811-6.
4. Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies*. PLoS Genet, 2009. **5**(6): p. e1000529.
5. Browning, B.L. and S.R. Browning, *Genotype Imputation with Millions of Reference Samples*. Am J Hum Genet, 2016. **98**(1): p. 116-26.
6. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans*. Nature, 2016. **536**(7616): p. 285-91.
7. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets*. Gigascience, 2015. **4**: p. 7.
8. Genomes Project, C., et al., *A map of human genome variation from population-scale sequencing*. Nature, 2010. **467**(7319): p. 1061-73.
9. Klarin, D., et al., *Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program*. Nat Genet, 2018. **50**(11): p. 1514-1523.
10. Liu, D.J., et al., *Exome-wide association study of plasma lipids in >300,000 individuals*. Nat Genet, 2017. **49**(12): p. 1758-1766.
11. Lu, X., et al., *Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease*. Nat Genet, 2017. **49**(12): p. 1722-1730.
12. Willer, C.J., et al., *Discovery and refinement of loci associated with lipid levels*. Nat Genet, 2013. **45**(11): p. 1274-1283.

## Supplementary tables

**Table S1, The FinnGen Study samples data release 4, version 1.0.<sup>a</sup>**

Dataset	Chip type	No. of samples before imputation QC	No. imputed samples
AxiomGT1_b01	Axiom_FinnGen1.r1	4,661	4,355
AxiomGT1_b02	Axiom_FinnGen1.r1	4,709	4,497
AxiomGT1_b03	Axiom_FinnGen1.r1	4,701	4,471
AxiomGT1_b04	Axiom_FinnGen1.r1	4,683	4,408
AxiomGT1_b05	Axiom_FinnGen1.r1	4,697	4,192
AxiomGT1_b06	Axiom_FinnGen1.r1	4,692	4,400
AxiomGT1_b07	Axiom_FinnGen1.r1	4,714	4,023
AxiomGT1_b08	Axiom_FinnGen1.r1	4,714	4,488
AxiomGT1_b09	Axiom_FinnGen1.r1	4,640	4,425
AxiomGT1_b10	Axiom_FinnGen1.r2	4,687	4,393
AxiomGT1_b11	Axiom_FinnGen1.r2	4,642	4,465
AxiomGT1_b12	Axiom_FinnGen1.r2	1,925	1,789
AxiomGT1_b13	Axiom_FinnGen1.r2	4,010	3,790
AxiomGT1_b14	Axiom_FinnGen1.r2	4,063	3,867
AxiomGT1_b15	Axiom_FinnGen1.r2	4,061	3,910
AxiomGT1_b16	Axiom_FinnGen1.r2	4,072	3,911
AxiomGT1_b17	Axiom_FinnGen1.r2	3,980	3,779
AxiomGT1_b18	Axiom_FinnGen1.r2	3,870	3,706
AxiomGT1_b19	Axiom_FinnGen1.r2	3,998	3,911
AxiomGT1_b20	Axiom_FinnGen1.r2	3,960	3,892
AxiomGT1_b21	Axiom_FinnGen1.r2	4,038	3,866
AxiomGT1_b22	Axiom_FinnGen1.r2	4,044	3,919
AxiomGT1_b23	Axiom_FinnGen1.r2	3,995	3,802
AxiomGT1_b24	Axiom_FinnGen1.r2	4,045	3,894
AxiomGT1_b25	Axiom_FinnGen1.r2	4,087	3,937
AxiomGT1_b26	Axiom_FinnGen1.r2	3,994	3,820
AxiomGT1_b27	Axiom_FinnGen1.r2	4,053	3,930
AxiomGT1_b28	Axiom_FinnGen1.r2	4,061	3,814
AxiomGT1_b29	Axiom_FinnGen1.r2	4,037	3,937
R4_Corogene_sanger	Illumina Human610-Quadv1_B	2,220	2,193

Dataset	Chip Type	No. of samples before imputation QC	No. imputed samples
R4_BOTNIA_DGI	Affymetrix_250KNsp_250KSty	2,167	1,385
R4_FINRISK_Corogene	Human610-Quadv1_B	1,878	1,748
R4_FINRISK_Engage	HumanCoreExome-12v1-0_A	5,548	5,197
R4_FINRISK_FR02_BROAD	GSAMD-24v1-0_20011747_A1	1,536	1,504
R4_FINRISK_FR12	Illumina_HumanCoreExome-24-v1.1	3,495	3,380
R4_FINRISK_Finpcga	Illumina_HumanCoreExome-12v1.1a	7,506	7,339
R4_FINRISK_MRPREP	Illumina_HumanCoreExome-24_v1.0	2,883	2,406
R4_FINRISK_PalotieCoreExome	HumanCoreExome-12v1-1_A	3,496	3,371
R4_FINRISK_SUMMIT	HumanOmniExpress-12v1_H	659	650
R4_FINRISK_bf	Illumina_HumanCoreExome-24_v1.0	847	550
R4_GENERISK	Illumina_HumanCoreExome-24v1-0_A	7,271	7,066
R4_H2000_Broad	Broad_GWAS_supplemental_15061359_A1	4,668	4,458
R4_H2000_FIMM	Illumina_HumanCoreExome-24v1-1_A	1,004	855
R4_H2000_Genmets	Human610-Quadv1_B	2,173	2,130
R4_Migraine_1	PsychChip_15048346_B-_24	6,244	5,939
R4_Migraine_2	HumanCoreExome-12v1-0_A	2,222	1,994
R4_BOTNIA_T2DGo	Illumina_HumanOmni2.5-4v1_B_SNV_array	447	434
R4_FINRISK_predictCVD_Corogene_Tarto	HumanOmniExpress-12v1_A	2,144	2,092
R4_super_1	GSAMD-24v1-0_20011747_A1	4,588	4,478
R4_super_2	GSAMD-24v1-0_20011747_A1	2,343	2,299
R4_twins1	Illumina Human670/Human610	2,611	2,494
R4_twins2	IlluminaCoreExome	4,398	4,141
<b>Total</b>		<b>194,181</b>	<b>183,694</b>
<sup>a</sup> AxiomGT1 batches 9 and 11 had possibly contaminated samples identified by excessive relatedness (pihat linkage cutoff $\geq 0.1$ for more than 30 samples). Therefore, in the QC step before imputation, an additional 83 and 50 samples were removed from batches 9 and 11 respectively.			



**Table S2. Baseline characteristics of the 23,435 Finrisk Study participants.<sup>a</sup>**

Mean age ± standard deviation units, yrs	48.8 ± 13.46
Female Sex - no. (%)	12,623 (52.5)
Mean waist-hip ratio ± standard deviation units	0.89 ± 0.97
Current smoker - no. (%)	6,057 (25.19)
Hypertension - no. (%)	11,402 (45.40)
Lipid-lowering therapy - no. (%)	1,923 (8.00)

<sup>a</sup> Shown in the table are the baseline characteristics of the Finrisk Study participants used for finding the PTVs associated with blood lipid levels.

**Table S3. Joint testing of previously associated variants in *CETP* locus, rs751916721-T and high-density lipoprotein cholesterol levels.<sup>a</sup>**

Previously known variant		Known variant association in the Finrisk Study		rs751916721-T association in conditional test	
Variant	MAF in Finrisk %	Effect mg/dL	P value	Effect mg/dL	P value
rs1529927	3.18	-1.8	$9.1 \times 10^{-7}$	16.8	$1.9 \times 10^{-21}$
rs36049418	0.94	0.7	0.28	16.9	$1.6 \times 10^{-21}$
rs11643718	12.62	1.1	$6.3 \times 10^{-9}$	15.7	$5.9 \times 10^{-19}$
rs2217332	13.75	-0.5	0.04	17.1	$6.9 \times 10^{-22}$
rs9989419	36.94	-1.6	$1.2 \times 10^{-32}$	17.4	$1.7 \times 10^{-22}$
rs173539	28.29	3.2	$1.8 \times 10^{-121}$	17.5	$4.4 \times 10^{-24}$
rs247616	27.86	3.3	$2.3 \times 10^{-125}$	17.5	$4.8 \times 10^{-24}$
rs3764261	27.84	3.3	$1.6 \times 10^{-126}$	17.6	$4.5 \times 10^{-24}$
rs821840 <sup>b</sup>	-	-	-	-	-
rs1800775	49.64	-2.3	$1.2 \times 10^{-74}$	17.9	$3.6 \times 10^{-24}$
rs34119551	< 0.1	-	-	-	-
rs34065661	< 0.1	-	-	-	-
rs711752	44.93	-2.4	$5.8 \times 10^{-83}$	17.8	$4.9 \times 10^{-24}$
rs201790757	< 0.1	-	-	-	-
rs9939224	21.53	-2.8	$9.0 \times 10^{-76}$	16.3	$1.3 \times 10^{-20}$
rs7205804	42.01	2.5	$2.6 \times 10^{-85}$	17.8	$1.1 \times 10^{-24}$
rs1532624	41.93	2.5	$7.3 \times 10^{-86}$	17.9	$1.1 \times 10^{-24}$
rs7499892	16.70	-3.2	$1.2 \times 10^{-81}$	16.2	$5.4 \times 10^{-21}$
rs12708980	30.49	-0.5	$1.1 \times 10^{-4}$	16.6	$3.1 \times 10^{-21}$
rs5880	2.30	-4.0	$1.4 \times 10^{-22}$	16.7	$1.6 \times 10^{-21}$
rs5882	37.39	1.1	$4.0 \times 10^{-18}$	17.1	$1.4 \times 10^{-22}$
rs2303790	< 0.1	-	-	-	-
rs16965039	6.76	0.2	0.30	16.7	$1.5 \times 10^{-21}$
rs1672867	0.43	-2.5	$8.2 \times 10^{-3}$	16.6	$1.8 \times 10^{-21}$
rs289723	22.16	0.0	0.95	16.7	$1.7 \times 10^{-21}$

Previously known variant		Known variant association in the Finrisk Study		rs751916721-T association in conditional test	
Variant	MAF in Finrisk %	Effect mg/dL	P value	Effect mg/dL	P value
rs7206703	9.71	-0.1	0.78	16.6	$1.7 \times 10^{-21}$
rs117587884	3.87	1.4	$9.1 \times 10^{-6}$	16.7	$1.2 \times 10^{-21}$

<sup>a</sup> This table shows the association statistics in Finrisk of previously reported genome-wide associations in the *CETP* locus and the association statistics when the rs751916721-T variant is tested in a joint model with the earlier associations. Missing results are due to the rarity of a previously known variant in the Finrisk genotype data.

<sup>b</sup> rs821840 is located in a low complexity region and was thus excluded from the Finrisk genotype data; hence the missing data.

**Table S4. Joint testing of previously associated variants in *LIPG* locus, rs200435657-A and high-density lipoprotein cholesterol Levels.<sup>a</sup>**

Previously known variant		Known variant association in the Finrisk Study		rs200435657-A association in conditional test	
Variant	MAF in Finrisk %	Effect mg/dL	P value	Effect mg/dL	P value
rs2000813	27.1	0.5	$1.6 \times 10^{-3}$	9.9	$3.6 \times 10^{-12}$
rs201922257	0.18	12.4	$5.8 \times 10^{-18}$	10.1	$6.9 \times 10^{-13}$
rs142545730	< 0.1	-	-	-	-
rs77960347	0.67	2.4	$1.7 \times 10^{-3}$	10.2	$4.9 \times 10^{-13}$
rs117623631	< 0.1	-	-	-	-
rs7240405	16.2	-1.0	$1.3 \times 10^{-8}$	9.9	$1.5 \times 10^{-12}$
rs7241918	16.4	-1.0	$1.5 \times 10^{-8}$	9.9	$1.4 \times 10^{-12}$
rs4939883	16.7	-1.0	$2.3 \times 10^{-9}$	10.0	$1.6 \times 10^{-12}$

<sup>a</sup> This table shows the association statistics in Finrisk of previously reported genome-wide associations in the *LIPG* locus and the association statistics when the rs200435657-A variant is tested in a joint model with the earlier associations. Missing results are due to the rarity of a previously known variant in the Finrisk genotype data.

**Table S5. Joint testing of previously associated variants in *ANGPTL4* locus, rs746226153-G and triglyceride levels.<sup>a</sup>**

Previously known variant		Known variant association in the Finrisk Study		rs746226153-G association in conditional test	
Variant	MAF in Finrisk %	Effect mg/dL	P value	Effect mg/dL	P value
rs116843064	2.586	-8.8	$9.6 \times 10^{-15}$	-15.9	$2.0 \times 10^{-9}$
rs7255436	47.51	0.3	.43	-15.9	$6.8 \times 10^{-9}$
rs2967605	14.59	0.9	.08	-15.9	$5.6 \times 10^{-9}$
rs968502	14.15	-0.9	.04	-15.0	$3.7 \times 10^{-8}$
rs199822445	< 0.1	-	-	-	-

<sup>a</sup> This table shows the association statistics in Finrisk of previously reported genome-wide associations in the *ANGPTL4* locus and the association statistics when the rs746226153-G variant is tested in a joint model with the earlier associations. Missing results are due to the rarity of a previously known variant in the Finrisk Study genotype data.

**Table S6. Joint testing of previously associated variants in *ANGPTL8* locus, rs760351239-T and triglyceride levels.<sup>a</sup>**

Previously known variant		Known variant association in the Finrisk Study		rs760351239-T association in conditional test	
Variant	MAF in Finrisk %	Effect mg/dL	P value	Effect mg/dL	P value
rs145464906	< 0.1	-	-	-	-
rs59168178	< 0.1	-	-	-	-

<sup>a</sup> This table shows the association statistics in Finrisk of previously reported genome-wide associations in the *ANGPTL8* locus and the association statistics when the rs760351239-T variant is tested in a joint model with the earlier associations. Missing results are due to the rarity of a previously known variant in the Finrisk Study genotype data.

**Table S7. Top configuration of the variants forming the 95% credible sets and their probabilities in the *CETP* locus.<sup>a</sup>**

index	cred1	prob1	cred2	prob2	cred3	prob3	cred4	prob4	cred5	prob5
1	rs3764261	0.476	rs12720922	0.473	rs5883	1.000	rs751916721	0.505	rs9924087	0.575
2	rs36229491	0.150	rs11076175	0.313	-	-	rs566571297 <sup>b</sup>	0.493	rs9936680	0.329
<b>Sum of posterior probabilities in each credible set</b>		<b>0.626</b>		<b>0.786</b>		<b>1.000</b>		<b>0.998</b>		<b>0.904</b>

<sup>a</sup> Shown in the table are the variants with the highest posterior probability in each 95% credible set in the *CETP* locus. The columns cred1-cred5 and prob1-prob5 indicate these variants and their respective posterior probabilities in each credible set.

<sup>b</sup> The noncoding rs566571297 and protein-truncating rs751916721 variants are in very high linkage disequilibrium ( $r^2=0.99$ ).

**Table S8. Top configuration of the variants forming credible sets in the *LIPG* locus.<sup>a</sup>**

index	cred1	prob1	cred2	prob2	cred3	prob3	cred4	prob4
1	rs201922257	0.496	rs4939883	0.223	rs200435657	1.000	rs80204526	0.103
Sum of posterior probabilities in each credible set		<b>0.496</b>		<b>0.223</b>		<b>1.000</b>		<b>0.103</b>

<sup>a</sup> Shown in the table are the variants with the highest posterior probability in each 95% credible set in the *LIPG* locus. The columns cred1-cred4 and prob1-prob4 indicate these variants and their respective posterior probabilities in each credible set.



**Table S9. Top configuration of the variants forming credible sets in the *ANGPTL4* locus.<sup>a</sup>**

index	cred1	prob1	cred2	prob2
1	rs116843064	1.000	rs919624228 <sup>b</sup>	0.860
2	-	-	rs746226153	0.140
<b>Sum of posterior probabilities in each credible set</b>		<b>1.000</b>		<b>1.000</b>

<sup>a</sup> Shown in the table are the variants with the highest posterior probability in each 95% credible set in the *ANGPTL4* locus. The columns cred1-cred2 and prob1-prob2 indicate these variants and their respective posterior probabilities in each credible set.

<sup>b</sup> The noncoding rs919624228 and protein-truncating rs746226153 variants are in very high linkage disequilibrium ( $r^2=0.97$ ).

**Table S10. Top configuration of the variants forming credible sets in the *ANGPTL8* locus.<sup>a</sup>**

index	cred1	prob1
1	rs760351239	0.999
<b>Sum of posterior probabilities in each credible set</b>		<b>0.999</b>

<sup>a</sup> Shown in the table are the variants with the highest posterior probability in each 95% credible set in the *ANGPTL8* locus. The columns cred1 and prob1 indicate the variant and its posterior probability in each credible set.

**Table S11. Associations between the lipid-associated protein-truncating variants, and hypertension and statin medication in the FinnGen Study.<sup>a</sup>**

Locus and variant	Hypertension; 23,338 cases				Statin medication; 53,518 cases			
	Allele frequency <sup>b</sup>		OR <sup>c</sup>	P	Allele frequency <sup>b</sup>		OR <sup>c</sup>	P
	Cases	Controls	(95% CI)		Cases	Controls	(95% CI)	
	%				%			
<i>ANGPTL4</i> <i>rs746226153-G</i>			0.80	0.002			0.83	0.006
	0.48	0.53	(0.70-2.56)		0.49	0.53	(0.73-2.62)	
<i>ANGPTL8</i> <i>rs760351239-T</i>			0.89	0.44			0.53	$1.3 \times 10^{-5}$
	0.12	0.12	(0.67-3.25)		0.10	0.13	(0.40-2.26)	
<i>CETP</i> <i>rs751916721-T</i>			0.92	0.57			0.70	0.01
	0.11	0.11	(0.69-3.36)		0.11	0.10	(0.53-2.68)	
<i>LIPG</i> <i>rs200435657-A</i>			0.67	0.003			1.10	0.46
	0.11	0.14	(0.52-2.55)		0.13	0.15	(0.85-3.91)	

<sup>a</sup> Shown in the table are the associations between the lipid-associated PTVs, and hypertension and statin medication in the FinnGen Study.

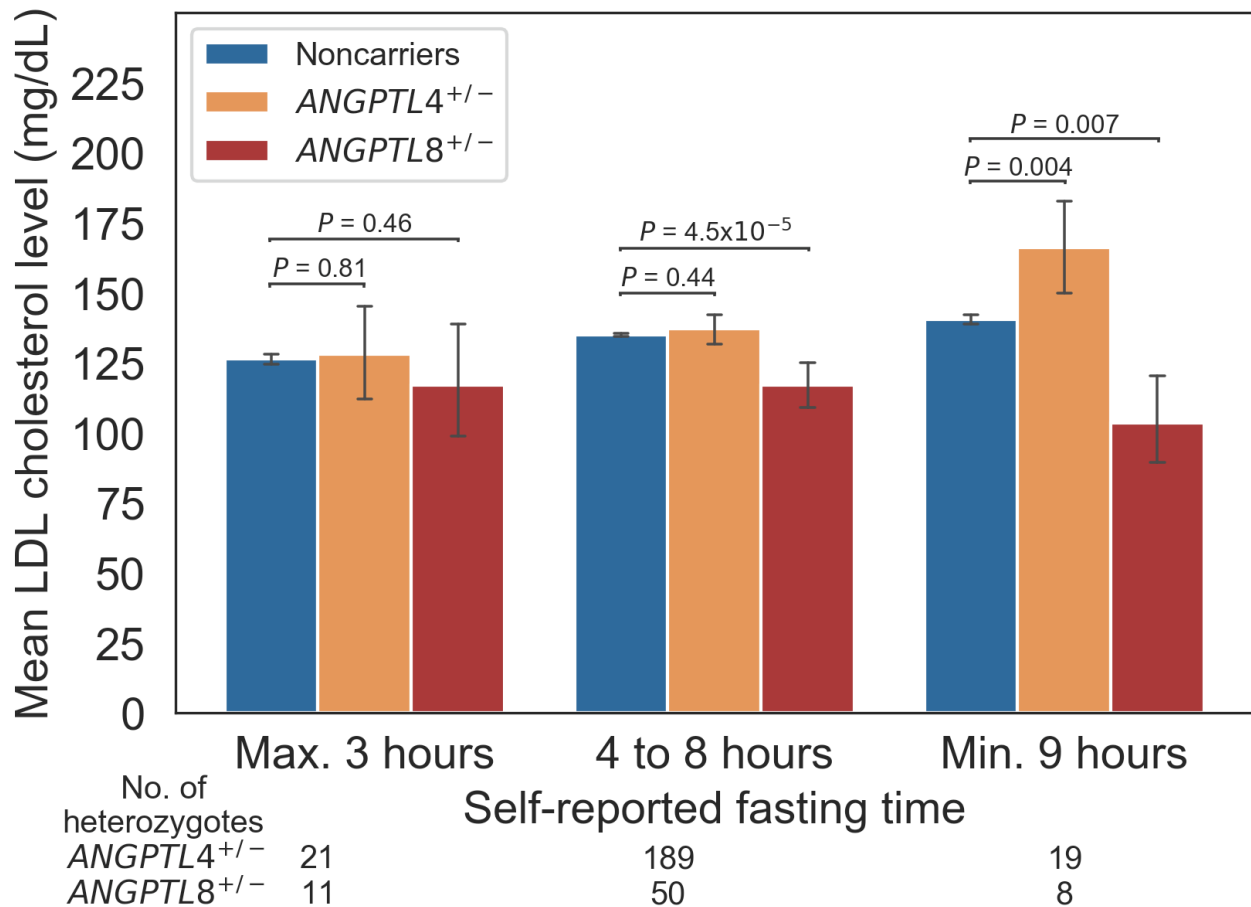
<sup>b</sup> The allele frequencies are reported in percent.

<sup>c</sup> The ORs were calculated using SAIGE saddle-point approximation-based score test after adjustment for age, sex, genotyping batch and ten principal components of ancestry in the FinnGen Study.

## Online tables

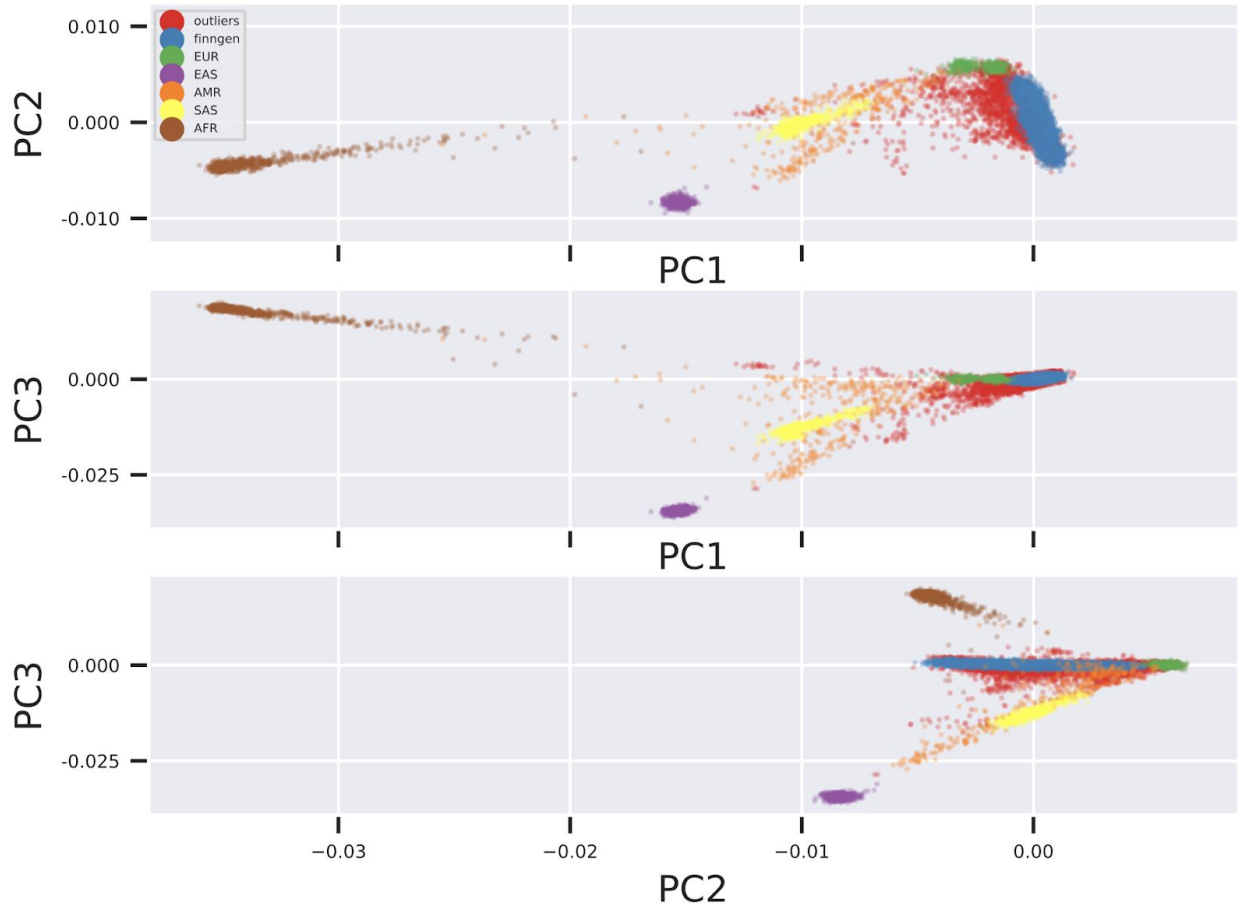
For a complete list of the considered clinical endpoints and corresponding ICD and ATC codes, see Online Tables 1-3. See Online Tables 4-6 for the associations between disease risks and the PTVs in *LIPG*, *ANGPTL4* and *ANGPTL8* in the FinnGen data.

**Figure S1. Mean LDL cholesterol levels at the self-reported fasting time of *ANGPTL4* and *ANGPTL8* PTV heterozygotes, as well as noncarriers in the Finrisk Study.<sup>a</sup>**



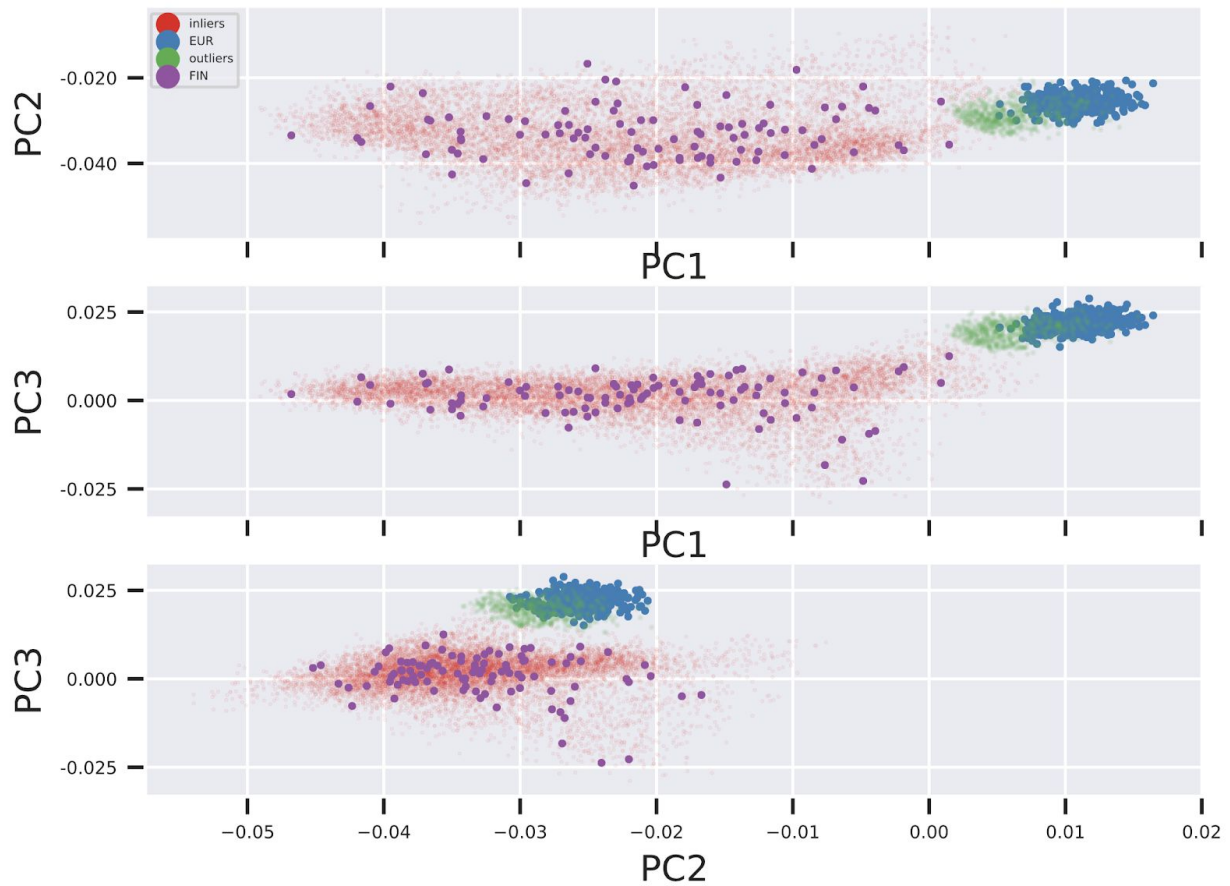
<sup>a</sup> The figure shows the mean plasma LDL cholesterol level by *ANGPTL4* and *ANGPTL8* PTV carrier status with respect to self-reported fasting time. The number of *ANGPTL4* and *ANGPTL8* heterozygotes for each fasting time interval are reported below the fasting time legend. The points indicate means and the error bars 95% confidence intervals. The P values are for the two-sided Welch's t-test between triglyceride levels of noncarriers and heterozygotes. LDL cholesterol levels of individuals with lipid-lowering therapy were divided by 0.7.

**Figure S2. Placement of FinnGen Study individuals along the three principal components together with 1000G samples.<sup>a</sup>**



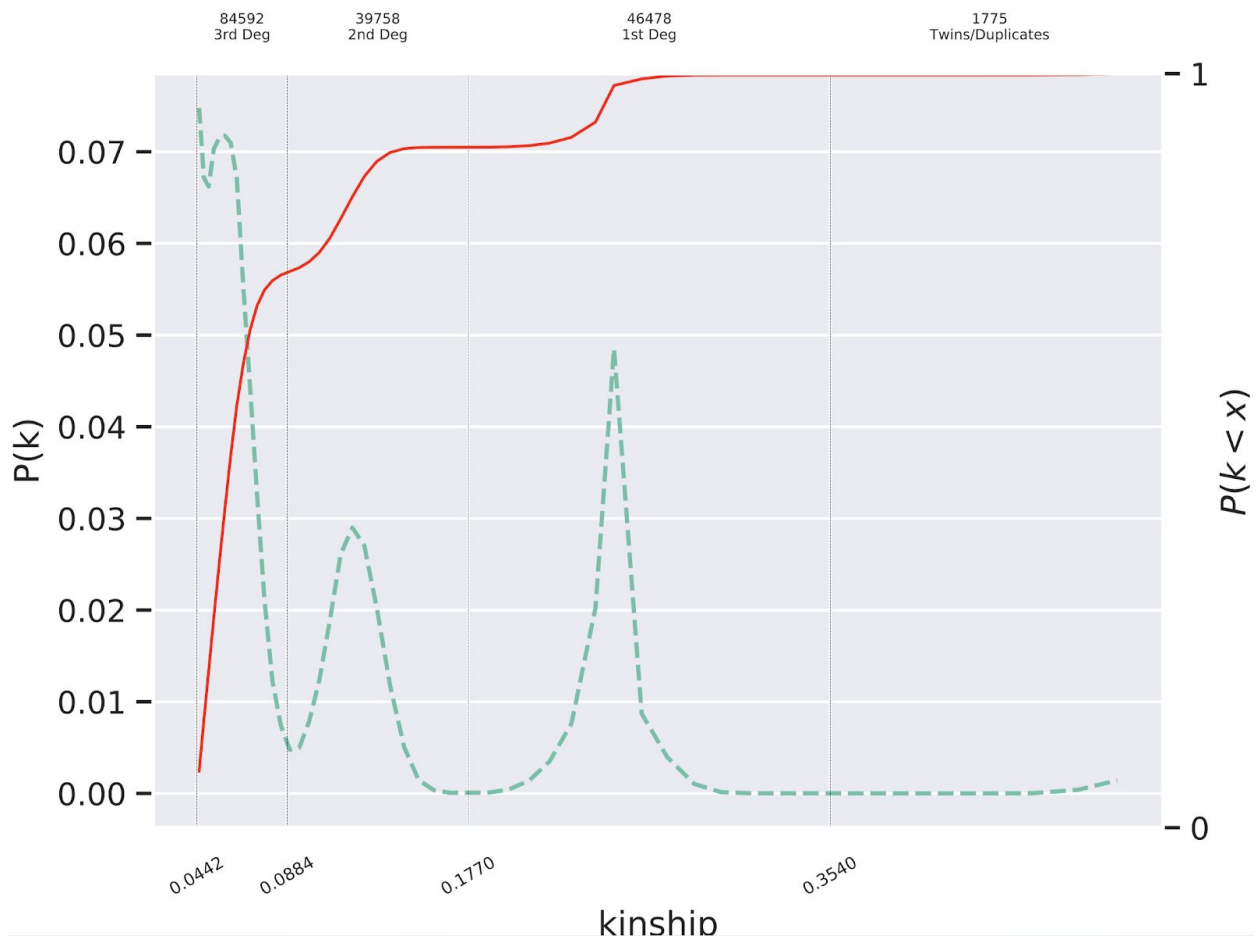
<sup>a</sup> The Finnish population outliers (separated by PCA and Bayesian algorithm) marked with red points, are separate from the blue cluster of FinnGen Study samples.

**Figure S3. Placement of FinnGen Study individuals along the three principal components together with remaining non-Finnish European 1000G samples.<sup>a</sup>**



<sup>a</sup> Finnish 1000G samples are indicated by purple points, while non-Finnish European 1000G samples by blue points. Green points represent FinnGen Study samples who are flagged as being non-Finnish (by the Bayesian algorithm), while red points indicate FinnGen Study samples who are considered Finnish.

**Figure S4. distribution of kinship values up to the third degree in the FinnGen Study.<sup>a</sup>**



<sup>a</sup> The probability and cumulative distributions of close relatedness of the individuals in the FinnGen Study cohort. The kinship is represented by the dashed green line and the left vertical axis. The cumulative distribution of the kinship is given by the solid red line and the right vertical axis.