Journal of the American Heart Association

ORIGINAL RESEARCH

Genetic Profile of Endotoxemia Reveals an Association With Thromboembolism and Stroke

Jaakko Leskelä (iD), DDS*; Iiro Toppila, MSc*; Mari-Anne Härma, MSc*; Teemu Palviainen (iD), MSc; Aino Salminen, DDS, PhD; Niina Sandholm (iD), PhD; Milla Pietiäinen (iD), PhD; Elisa Kopra, DDS, PhD; Jean-Paul Pais de Barros, PhD; FinnGen;† Mariann I. Lassenius, PhD; Anmol Kumar (iD), PhD; Valma Harjutsalo (iD), PhD; Kajsa Roslund (iD), MSc; Carol Forsblom (iD), MD, PhD; Anu Loukola, PhD; Aki S. Havulinna (iD), PhD; Laurent Lagrost, PhD; Veikko Salomaa (iD), MD, PhD; Per-Henrik Groop, MD, PhD; Markus Perola (iD), MD, PhD; Jaakko Kaprio (iD), MD, PhD; Markku Lehto, PhD; Pirkko J. Pussinen (iD), PhD

BACKGROUND: Translocation of lipopolysaccharide from gram-negative bacteria into the systemic circulation results in endotoxemia. In addition to acute infections, endotoxemia is detected in cardiometabolic disorders, such as cardiovascular diseases and obesity.

METHODS AND RESULTS: We performed a genome-wide association study of serum lipopolysaccharide activity in 11 296 individuals from 6 different Finnish study cohorts. Endotoxemia was measured by limulus amebocyte lysate assay in the whole population and by 2 other techniques (Endolisa and high-performance liquid chromatography/tandem mass spectrometry) in subpopulations. The associations of the composed genetic risk score of endotoxemia and thrombosis-related clinical end points for 195 170 participants were analyzed in FinnGen. Lipopolysaccharide activity had a genome-wide significant association with 741 single-nucleotide polymorphisms in 5 independent loci, which were mainly located at genes affecting the contact activation of the coagulation cascade and lipoprotein metabolism and explained 1.5% to 9.2% of the variability in lipopolysaccharide activity levels. The closest genes included *KNG1*, *KLKB1*, *F12*, *SLC34A1*, *YPEL4*, *CLP1*, *ZDHHC5*, *SERPING1*, *CBX5*, and *LIPC*. The genetic risk score of endotoxemia was associated with deep vein thrombosis, pulmonary embolism, pulmonary heart disease, and venous thromboembolism.

CONCLUSIONS: The biological activity of lipopolysaccharide in the circulation (ie, endotoxemia) has a small but highly significant genetic component. Endotoxemia is associated with genetic variation in the contact activation pathway, vasoactivity, and lipoprotein metabolism, which play important roles in host defense, lipopolysaccharide neutralization, and thrombosis, and thereby thromboembolism and stroke.

Key Words: coagulation ■ contact activation ■ endotoxin ■ gene ■ genome-wide association study ■ lipopolysaccharide

L ipopolysaccharide, also known as endotoxin, is an important virulence factor for gram-negative bacteria. The structural differences of the lipopolysaccharide molecules between bacterial species can have a major effect on their functional properties and biological activity. Lipopolysaccharide can act as an immunostimulator or immunomodulator, thereby contributing to the virulence of various bacterial species.[1]

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?
- The biological activity of lipopolysaccharide in the circulation (ie, endotoxemia) has a small but highly significant genetic component.
- The 5 genetic loci, which associate with endotoxemia, are mainly located at genes affecting the contact activation of the coagulation cascade and lipoprotein metabolism.
- The genetic risk score of endotoxemia is associated with deep venous thrombosis, pulmonary embolism, venous thromboembolism, and ischemic stroke.

What Are the Clinical Implications?
- The analyses suggest that endotoxemia may be one of the causal factors in thromboembolism and stroke.
- The results indicate that the microbiome/host interactions play a role in thromboembolism and stroke risk.

Nonstandard Abbreviations and Acronyms

LAL	limulus amebocyte lysate
LPS-GRS	genetic risk score of endotoxemia
MR	Mendelian randomization

Translocation of lipopolysaccharide in the circulation, endotoxemia, can occur in the interface of host mucosal microbiota and the bloodstream (eg, in the gut).[1] Endotoxemia is associated with an increased risk of cardiometabolic disorders, including incident cardiovascular disease events, obesity, metabolic syndrome, and diabetes.[2–4] In addition, serum lipopolysaccharide activity is associated with many noncommunicable disease risk factors: it is inversely associated with high-density lipoprotein (HDL) cholesterol concentrations and directly with triglyceride, cholesterol, CRP (C-reactive protein), fasting glucose, insulin, glycated hemoglobin concentrations, and body mass index.[5] Overall, endotoxemia is associated with a highly adverse metabolic profile of inflammatory character.[6]

Previous candidate gene studies have demonstrated the importance of innate immune system pathways in host responsiveness to administered lipopolysaccharide.[7] In addition, a genome-wide association study (GWAS) for fever after evoked endotoxemia identified a genetic locus that modulates clinical responses in trauma and sepsis.[8] However, the genetic determinants of human endotoxemia have not been investigated previously. The aim of this work was to assess the genetic profile of serum lipopolysaccharide activity using a GWAS approach, and to determine whether this profile has an association with cardiovascular risk.

METHODS

Data Availability

The authors declare that the data supporting the findings of the study are available within the article, its Supplementary Material, and on request from the corresponding author. Genome-wide summary-level statistics are available to download from GWAS Catalog with study accession GCST90032674 (ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90032001-GCST90033000/GCST90032674/). The FINRISK data are available from the THL Biobank (https://thl.fi/en/web/thl-biobank/for-researchers) based on a written application and following the relevant Finnish legislation. For the FinnDiane study and the twin samples, individual-level data cannot be shared because of restrictions in patient and participant consent. FinnGen Data Freeze 5-summary level data are publicly available at the FinnGen website (https://www.finngen.fi/en/access_results). UK biobank summary-level data are available at http://www.nealelab.is/uk-biobank/ and Megastroke at https://www.megastroke.org/.

Study Population

Participants from 3 studies, FinnDiane, FINRISK, and Finnish Twin Cohort, were used in the GWAS analyses. The FinnDiane cohort consists of participants with type 1 diabetes (FD-T1D) and individuals with unclassified diabetes; the FINRISK studies consist of population-based surveys in Finland; and, finally, the Finnish Twin Cohort study consists of Finnish adult twins. From each study, 2 independent cohorts were analyzed separately: from the FinnDiane, FD-T1D (n=3940) and Finn Diane cohort of individuals with unclassified diabetes (n=302); from the FINRISK, FINRISK92 (n=656) and FINRISK97 (n=5667); and from the Finnish Twin Cohort, FinnTwin16 (n=451) and VpEpi (n=280). Altogether, the GWAS analyses included 11 296 unique samples. Independent from the GWAS population, FinnGen Study (https://www.finngen.fi/) Data Freeze 5, including 195 170 participants, was used[9] to analyze the associations of the designed genetic risk score of endotoxemia (LPS-GRS) with disease end points. Every participant provided written informed consent, and the study was approved by the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa. Additional lead single-nucleotide polymorphism (SNP) association lookups and Mendelian randomization (MR) analysis used UK Biobank and Megastroke populations. Data S1 contains more detailed descriptions of study populations.

Table 1.   Clinical Characteristics of the Subjects in the 6 Cohorts

Characteristic	Cohort					
	FD-T1D	FD-rest	FINRISK92	FINRISK97	FT16	VpEpi
Total No.	3940	302	656	5667	451	280
Log(lipopolysaccharide)	−0.65 (0.48)	−0.60 (0.49)	0.23 (0.55)	−0.61 (0.52)	−0.47 (0.30)	−0.61 (0.21)
Women, n (%)	1901 (48.2)	133 (44.0)	250 (38.1)	2826 (49.9)	233 (51.7)	172 (61.4)
Age, y	40.58 (12.7)	53.78 (11.9)	53.66 (8.7)	53.07 (10.8)	24.40 (0.8)	61.72 (4.2)
BMI, kg/m^2	25.37 (3.8)	26.28 (4.1)	27.59 (4.5)	27.15 (4.4)	22.72 (2.9)	27.77 (4.9)
Total cholesterol, mmol/L	4.81 (0.9)	4.87 (1.0)	5.99 (1.1)	5.68 (1.0)	4.94 (0.9)	4.68 (0.9)
HDL cholesterol, mmol/L	1.42 (0.4)	1.38 (0.4)	1.32 (0.4)	1.39 (0.4)	1.78 (0.4)	1.53 (0.4)
Log(triglycerides), mmol/L	0.08 (0.5)	0.13 (0.5)	0.45 (0.6)	0.29 (0.5)	0.14 (0.4)	0.14 (0.4)

Values are given as mean (SD) or number (percentage). Log refers to logarithmic transformation of the values. All clinical variables differed between the cohorts ($P<0.001$, 1-way ANOVA or χ^2 test, where appropriate). BMI indicates body mass index; FD-T1D, FinnDiane cohort of participants with type 1 diabetes; FD-rest, FinnDiane cohort of individuals with unclassified diabetes; FINRISK92 and FINRISK97, FINRISK-cohorts enrolled in 1992 and 1997, respectively; FT16 and VpEpi, younger and older subpopulations of the Finnish Twin Cohort; and HDL high-density lipoprotein.

Lipopolysaccharide Activity Measures

Endotoxin activities were determined with a limulus amebocyte lysate (LAL) assay on 1:5 diluted serum samples (HyCult Biotechnology b.v., Uden, the Netherlands), and the results were log transformed (natural logarithm) because of skewed distributions. Data S1 contains more details.

Genetic Analysis

Cohorts were genotyped with various genotyping platforms and went through rigorous quality control. Imputation was performed using 1000 Genomes Project phase 3 reference genotypes. Single-marker association analysis was performed with linear mixed model to correct for the effect of cryptic relatedness and close relatives. The models were adjusted for empirical kinship matrix, sex, age, body mass index, total cholesterol, HDL cholesterol, triglycerides (log transformed), and study-specific covariates (in FinnDiane,

sample freeze time and genotyping batch; in FINRISK, genotyping batch and recruitment region; and in Finnish Twin Cohort, sample freeze time, genotyping batch, and additionally pregnancy status in FinnTwin16). The results were meta-analyzed using inverse variance weighted fixed effect meta-analysis. Genome-wide significance level was set to $P<5\times10^{-8}$. More detailed specification of genetic analyses, expression quantitative trait loci analysis, gene set enrichment analysis, genetic risk score, MR, post hoc GWAS, and conditioned analyses can be found in Data S1.

Endotoxemia Measured by Using Other Techniques

A subpopulation of FinnDiane was used to determine endotoxemia by mass spectrometry–based method (n=363), as previously described,[10] and a commercially available Endolisa assay (n=326) (609033; Hyglos GmbH, Bernried, Germany). Selection of participants
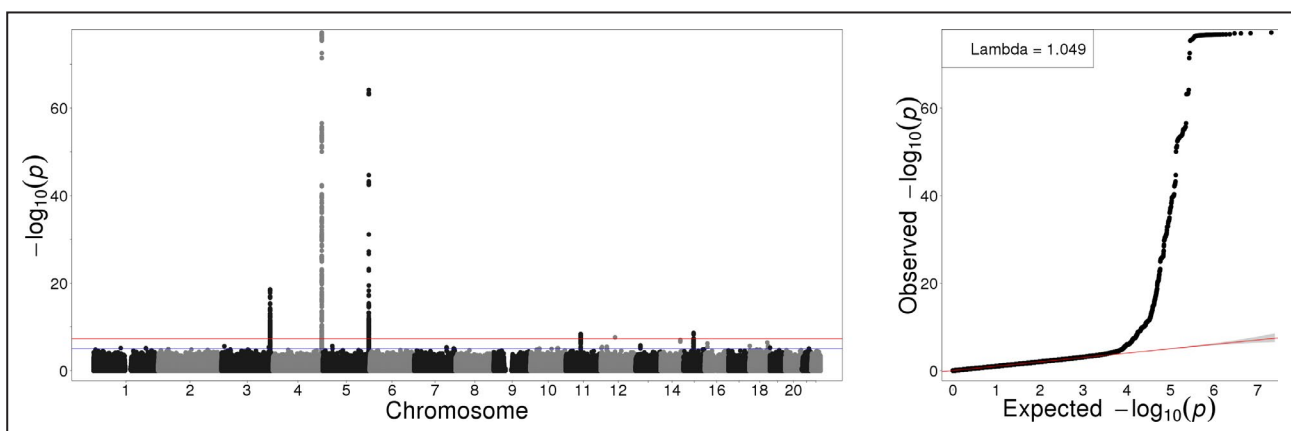
Figure 1.   Manhattan and QQ plots of genome-wide association study (GWAS) results combined in fixed-effects meta-analysis.
We performed a GWAS of endotoxemia, measured by limulus amebocyte lysate assay in 11 296 individuals with Finnish ancestry. The horizontal red line represents genome-wide significance ($P<5\times10^{-8}$). Single-nucleotide polymorphisms in 5 independent loci available in all cohorts passed the genome-wide significance threshold. Inflation of the P values is presented in the QQ plot (λ_{qc}=1.049).

for mass spectrometry was based on the *F12* SNP *rs1801020* genotype, obtaining 121 TT-homozygous and 242 CC-homozygous participants. For Endolisa assay, population genotype for *rs1801020* was distributed as follows: TT, 21; TC, 124; and CC, 181. Endolisa measures were log transformed, and the correlations with LAL assay results were computed using Pearson product-moment correlation. The SNP association with endotoxemia was analyzed with linear regression, applying an additive genetic model.

RESULTS

Genetic Association Analyses

The genetic factors associated with endotoxemia, measured using the LAL assay, were analyzed in 6 Finnish cohorts using the GWAS setting. The genetic analyses contained 8.1 to 9.9 million genotyped or imputed genetic markers passing quality control per cohort. The clinical characteristics of participants are presented in Table 1.

When the cohort-wise results were combined in fixed-effects meta-analysis composed of 11 296 participants, 741 markers clustered at 5 independent loci (in chromosomes 3, 4, 5, 11, and 15) were genome-wide significantly ($P<5×10^{-8}$) associated with endotoxemia (Figure 1 and Table 2). In addition, a single SNP, *rs77601517* on chromosome 12, reached genome-wide significance in the FINRISK-97 cohort. However, the SNP was additionally available only in the FINRISK-92 cohort, demonstrating no association, and was not studied further. All the significant SNPs (Table S1) and the Manhattan and QQ plots (Figure S1) are available in the Supplementary Material.

The cohort-wise forest plots for the lead variants are presented in Figure 2. Stepwise conditional regression with GCTA–conditional and joint analysis did not indicate independent secondary signals in any of the 5 loci after accounting for the lead SNPs. QQ plot of the GWAS meta-analysis with the significant SNPs omitted (Figure S2) demonstrated good adherence to the diagonal for the nonsignal SNPs. Figure S3 presents the Manhattan and QQ plots of GWAS conditioned on the lead SNPs. In addition, the leave-one-out analysis showed that single cohort is not excessively driving the result of the meta-analysis (Table S2 and Figure S4).

Linking SNPs to Genes

The regional plots of the 5 loci identified (Figure S5) and their position in relation to nearby genes (Table S1) are presented in the supplemental data. To identify biological pathways and processes behind endotoxemia, we performed a pathway enrichment analysis. The complete list of top pathways enriched with $P<0.01$ is presented in Table S3. The strongest gene set enrichment was observed in the biological processes "intrinsic pathway" (23 genes; $P=5×10^{-4}$), followed by "chromatin packaging and remodeling" (237 genes; $P=5×10^{-4}$). However, none of the associations in the enrichment analysis reached statistically significant false discovery rate (<0.05).

The expression quantitative trait locus analysis identified 24 genes affected by our GWAS significant markers (Table 2; full list in Tables S4 and S5 for blood[11] and genotype-tissue expression portal, respectively). Loci in chromosomes 3, 4, 5, and 11 had expression quantitative trait loci associations with several genes' expression (*KNG1*, *F11/KLKB1*, *F12*, and *SERPING1*, respectively) playing a role in the contact activation of the intrinsic pathway of coagulation. Figure 3 illustrates the relation of these genes and the intrinsic pathway of coagulation.

Post Hoc Regression Analyses

To evaluate the proportion of the endotoxemia variance explained by the genetic markers, we performed additional regression analyses in the 2 largest cohorts (FD-T1D and FINRISK97). The clinical covariates, age, sex, body mass index, and total cholesterol, HDL cholesterol, and triglyceride concentration, used in the GWAS analyses explained 50.8% of the lipopolysaccharide variability in the FinnDiane cohort and 27.8% of the lipopolysaccharide variability in the FINRISK cohort (Tables S6 and S7). Adding the 5 lead SNPs representing each loci to the models increased the proportions by 1.5 percentage points in FinnDiane and by 9.2 percentage points in FINRISK, resulting in 52.3% and 37.1% of the variance explained, respectively.

Alternative Methods to Measure Lipopolysaccharide

Endotoxemia was also measured in 2 separate subsamples of the FinnDiane cohort by determining the lipopolysaccharide mass and by using the Endolisa assay. The original LAL results, which measure the biological activity of lipopolysaccharide, had a modest but significant correlation with both the lipopolysaccharide mass (correlation=0.23; $P=7.5×10^{-6}$) and the Endolisa (correlation=0.19; $P=7.1×10^{-4}$) results. In both subsamples, the lead SNP of the FinnDiane cohort (*rs1801020*) was strongly associated with endotoxemia, when measured using the LAL assay ($\beta=0.083$ [$P=7.48×10^{-5}$] and $\beta=0.072$ [$P=0.016$], respectively). However, the lead SNP did not associate significantly with either lipopolysaccharide mass ($P=0.30$) or the Endolisa results ($P=0.7$). The measures and differences by genotype are presented in Figure S6.

Genetic Risk Score and MR

In the Megastroke population, single lead SNPs were associated with the risk of "ischemic stroke," "any

Table 2. Lead SNPs in Each Locus Associated With a Genome-Wide Significant Level With Endotoxemia

Marker	rs5030082	rs71640036	rs1801020	rs2081361	rs10152355
Closest gene	KNG1 (intron)	KLKB1 (intron)	F12 (5'UTR)	YPEL4	LIPC
Chromosomal position	3:186458949	4:187161120	5:176836532	11:57411742	15:58671178
A1/A2	A/G	T/G	A/G	T/C	A/C
A2 frequency	0.38	0.42	0.75	0.7	0.54
β (SE)	0.12 (0.01)	−0.25 (0.01)	0.26 (0.02)	0.09 (0.01)	−0.08 (0.01)
P value	2.95×10^{-19}	5.41×10^{-78}	6.62×10^{-65}	4.37×10^{-9}	2.51×10^{-9}
Direction	+−−+++	−−−−−−	++−++	+++++−+	−−−−−
I²	70.9	92.9	90	1.2	0
Heterogeneity P value	0.00415	8.85×10^{-14}	1.25×10^{-09}	0.409	0.919
eQTL associations	KNG1	CYP4V2, F11, F11-AS1, FAM149A, RPSAP70, KLKB1, FLJ38576, TLR3	F12, FGFR4, LMAN2, MXD3, RAB24, RGS14, PRELID1, SLC34A1	AP000662.4, MED19, TIMM10, ZDHHC5, SERPING1	LIPC, ADAM10

Fixed-effects meta-analysis, full list of 741 SNPs is presented in Table S1. The associations between all genome-wide association study significant SNPs and eQTL were considered and grouped under the lead SNP of the locus. Full list of eQTL associations is available in Table S7; a false discovery rate of <0.01 was required (full list in Table S2). 5'UTR indicates 5' untranslated region (regulatory region); A1, reference allele; A2 frequency, allele frequency of the effect allele in the combined sample; A2, alternative allele; β, estimated effect size for each copy of effect allele (increase in SD of normalized residual unexplained by other covariates); direction, sign of estimates (? for not available) for each cohort in order FinnDiane cohort of participants with type 1 diabetes, Finn Diane cohort of individuals with unclassified diabetes, FINRISK92, FINRISK97, FinnTwin16, and VpEpi; eQTL, expression quantitative trait loci; I², I² estimate for heterogeneity; and SNP, single-nucleotide polymorphism.

stroke," "TOAST small artery occlusion," "TOAST cardioaortic embolism," or "intracranial aneurysm" (Figure 4A). We composed a genetic risk score for endotoxemia (LPS-GRS) from GWAS-significant SNPs and analyzed its association with designed end points in an independent population, the FinnGen. LPS-GRS was significantly associated with deep vein thrombosis, pulmonary embolism, and venous thromboembolism (Figure 4B and Table S8). Next, MR using the lead SNPs was conducted on the UK Biobank data, which displayed associations with deep vein thrombosis, pulmonary embolism, and venous thromboembolism, and Megastroke populations, which revealed an association with ischemic stroke (Figure 4C and Figure S7). More detailed information is available in Tables S9 through S11.

DISCUSSION

We identified 5 genetic loci that displayed significant association with serum endotoxin activity levels in multiple cohorts. According to the expression quantitative trait loci, several of these SNPs are associated with expression of nearby genes that affect the contact activation of coagulation and lipoprotein metabolism. These both play important roles in host defense against infectious organisms, including induction of inflammatory responses and lipopolysaccharide neutralization. Furthermore, the composed genetic profile and the MR results indicated associations of endotoxemia with thromboembolism and stroke. The results further link microbiomes with cardiovascular diseases.

Endotoxemia is associated with increased risk of cardiovascular events in the largest population of the present study, the FINRISK, and other studies.[2,12–14] It may also contribute to stroke,[12–14] and metabolic endotoxemia has been suggested as a novel therapeutic target to improve stroke outcome.[14] Dysbiosis may maintain an inflammatory environment, which has a significant impact on cerebrovascular risk and stroke severity through the microbiota-gut-brain axis.[15] Dysbiosis, including increased abundance of gram-negative *Enterobacteriaceae*-family members, has been reported in patients with large-artery atherosclerotic ischemic stroke and transient ischemic attack compared with asymptomatic people.[16] Therefore, in addition to a genotype predisposing to endotoxemia, patients with stroke may have an ample source of lipopolysaccharide. A more recent study showed that interaction between lipopolysaccharide and SARS-CoV-2 S protein resulted in a hyperinflammatory effect,[17] which was hypothesized to contribute to the activation of the coagulation and complement system observed in severe COVID-19 disease. Future research will show whether endotoxemia plays a role in COVID-19.
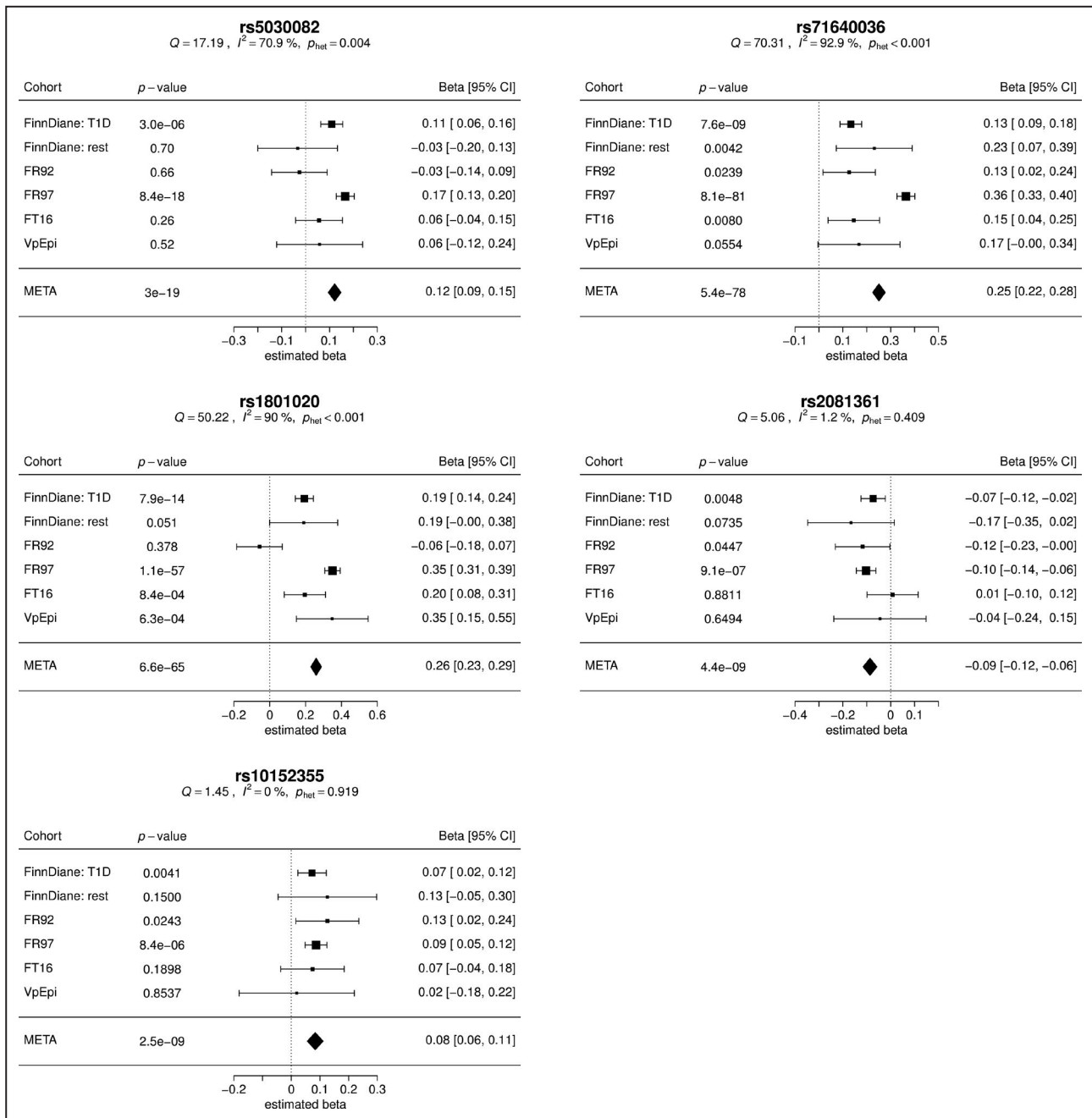
Figure 2. Forest plots of the lead single-nucleotide polymorphisms of loci reaching genome-wide significance in meta-analysis of the genome-wide association study for endotoxemia.
The meta-analysis included 11 296 samples. Presented point size is proportional to the inverse of SE of the estimate. DinnDiane: rest, FinnDiane cohort of individuals with unclassified diabetes; FinnDiane: T1D, FinnDiane cohort of participants with type 1 diabetes; FR indicates FINRISK; and FT16 and VpEpi, subpopulations of Finnish Twin Cohort.

A large portion of significant polymorphisms associated with endotoxemia in the present study were located at genes affecting the contact activation of the coagulation cascade. The dynamics of this pathway are complex but relatively well described, and the currently detected genetic associations fit well with an overall phenotype of increased coagulation. The lead SNP in the chromosome 5 locus (*rs1801020*) is a known functional variant; the minor allele results in an additional upstream open reading frame for gene *F12* at the sequence level, leading to increased expression and activity.[11,18] The *KNG1*, *KLKB1/F11*, and *SERPING1* loci also contained SNPs that affect the gene/protein expression levels or the protein activity.[11,19] The lead SNP on chromosome 11, *rs2081361*, has been previously

Figure 3. Hypothesized connection between intrinsic pathway of coagulation and genetic variants associating with endotoxemia.
Genetic associations connect endotoxemia to increased coagulation: via gene expression (expression quantitative trait loci [eQTL]), protein expression, and protein activity. Two possible mechanisms are hypothesized to explain the association. Bradykinin is cleaved from kininogen and can affect intestinal permeab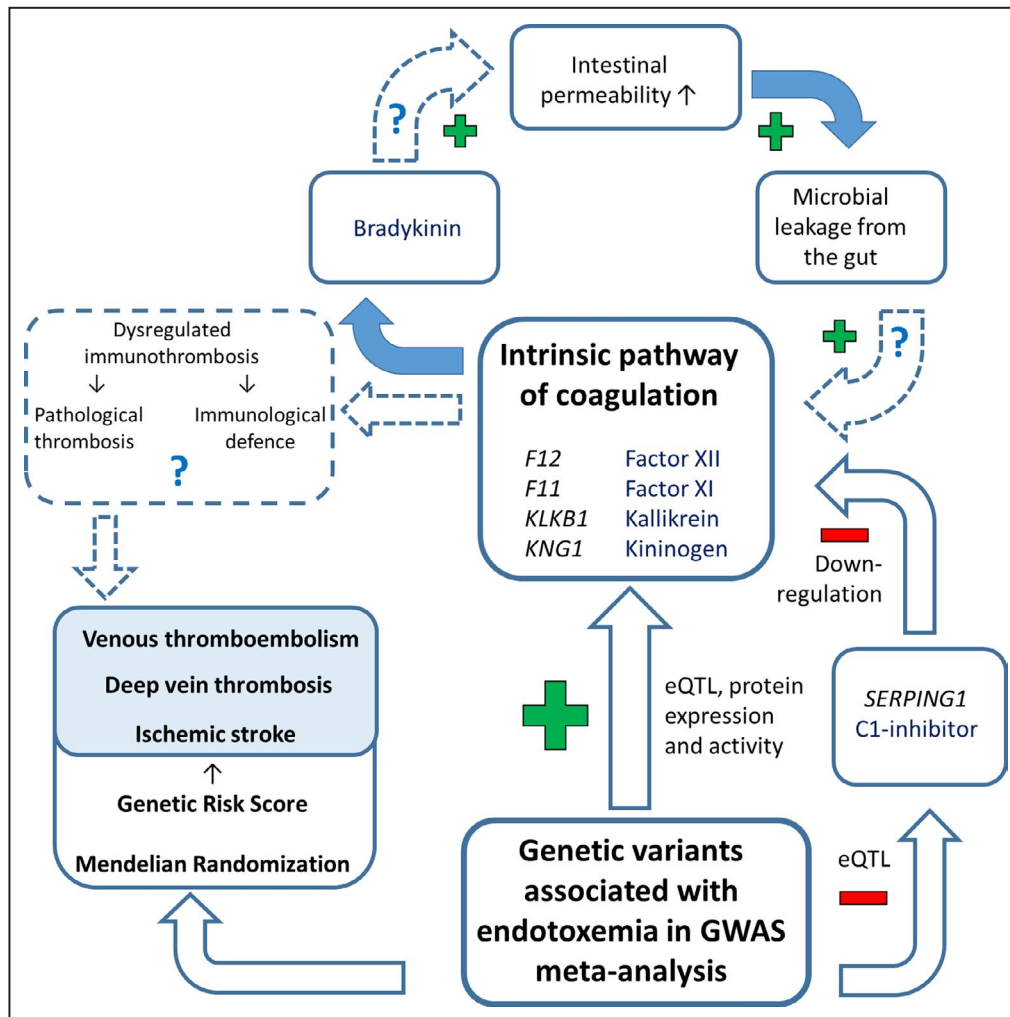ility by allowing increased microbial leakage from the gut. Immunothrombosis, a mechanism proposed to be involved in normal immunology, is hypothesized to alter both immunological defense and formation of thrombosis when dysregulated. Genetic risk score and Mendelian randomization analysis connect endotoxemia to venous thromboembolism, deep vein thrombosis, and stroke. GWAS indicates genome-wide association study.

associated with high *SERPING1* expression,[11] which may affect the levels of the central complement system regulator, C1 inhibitor, and thereby FXII activity. Furthermore, in chromosome 3 locus, *rs5030049* in the intronic region of *KNG1* is associated with higher protein levels of high-molecular-weight kininogen,[19] which is the nonenzymatic cofactor in the contact system. In the same locus, *rs710446* and *rs2304456* have also been associated with FXI plasma levels.[20] In addition, endotoxemia was associated with *rs4253238* and *rs4253417* in chromosome 4, linking it with high activities of plasma kallikrein[21] and FXI,[22] which are important activators

of coagulation (ie, FX and FIX). In addition to the complement system, all these proteins play a major role in the intrinsic pathway. "Intrinsic cascade" was also recognized as the top pathway enriched in the present study.

In post hoc analysis, lead SNPs explained a larger percentage of endotoxemia variance in the population-based cohort (FINRISK97) than in the cohort consisting of patients with diabetes (FD-T1D). Subjects with diabetes have higher serum lipopolysaccharide levels compared with subjects without diabetes because of hyperglycemia, hyperinsulinemia, hypertriglyceridemia, and low HDL cholesterol concentrations.[5] All
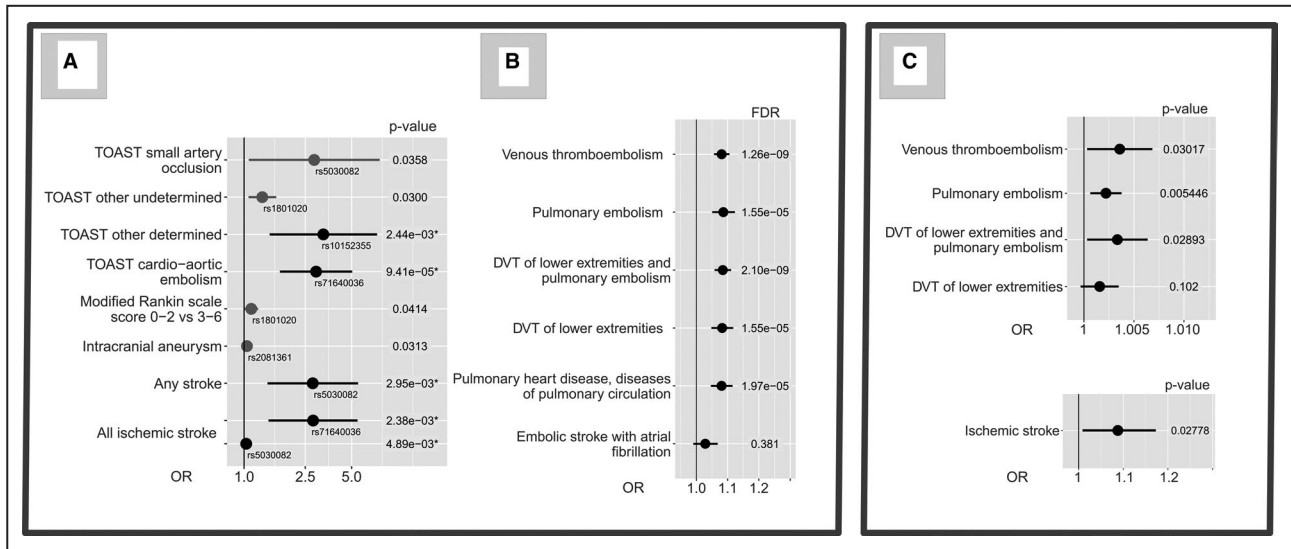
Figure 4. Associations of the cardiovascular disease end points with 5 endotoxemia-associated single-nucleotide polymorphisms (SNPs) and the composed genetic risk score (GRS).
A, Significant associations of the lead SNPs with stroke events in the Megastroke population. *P* value for logistic regression analysis is presented on the right side. The asterisk marking represents *P*<0.01. **B**, GRS of endotoxemia (LPS-GRS) is calculated from the 5 lead SNP genotypes, based on stepwise conditional regression (GCTA–conditional and joint analysis [COJO]) of lipopolysaccharide genome-wide associated study results weighted with effect sizes. False discovery rate (accounting for total of 46 end points tested) is presented on the right side. Population consists of 195 170 individuals from the FinnGen study. Full list of analyzed phenotypes is in Table S8. **C**, Association between endotoxemia and prothrombotic end points is analyzed by Mendelian randomization in Megastroke (for ischemic stroke) and UK Biobank populations using MR-base platform. The 5 lead SNPs, based on GCTA-COJO, were used in the analysis. DVT indicates deep vein thrombosis; OR, odds ratio; and TOAST, classification of 5 subtypes of ischemic stroke.

these metabolic features are associated with impaired clearance of lipopolysaccharide,[6] and the role of genetics in determining the endotoxemia levels may be smaller. Indeed, the covariates explained a larger portion of the endotoxemia variance in the cohort of diabetic subjects than in the population-based cohort. However, the variance explained may be notably overestimated because of a phenomenon called "winner's curse," which is characteristic for large-scale quantitative trait studies.[23]

Several SNPs in the endotoxemia-associated loci discovered have been earlier associated with thrombosis in general: they include SNPs in genes *F11*,[24] *KNG1*,[24] and *KLKB*.[25] The LPS-GRS of the present study also showed associations with conditions linked to blood coagulation alterations, which further validates the association between endotoxemia and the contact activation pathway. Although the genetic background of stroke has been intensively studied and heritability has been evaluated to range between 16% and 40%,[26] our results are novel: the MR results suggested a causal role of endotoxemia in thromboembolism and stroke. The MR approach uses the genetic risk score SNPs (effectively the lead SNPs of the lipopolysaccharide GWAS) as "instrumental variables," which allows us to study the causality between endotoxemia and clinical end points in comparable manner as in a randomized controlled trial. Using genetic variants as

instrumental variables, which do not have endogenous issues, such as reverse causation or missing confounders, provides us with consistent estimates from a regression. However, one of the MR assumptions is that the genetic variant does not have any other effect to the outcome other than through the exposure (endotoxemia). In the present study, this assumption was not totally fulfilled, because it can be assumed that the SNPs located at genes affecting the coagulation cascade could in addition have more direct associations with thromboembolism and stroke.

The observed signal of causality between thrombosis-related end points and endotoxemia may be hypothesized to be a sign of imbalanced "immunothrombosis" (Figure 3), a term that describes a thrombosis in microvessels triggered by inflammation.[27] In this concept, proposed by Engelmann and Massberg, immunothrombosis is a naturally occurring process that suppresses pathogen invasion, but potentially leads to pathological thrombosis if not carefully in balance. We hypothesize that LPS-GRS may partly reflect defective immunothrombosis, which may affect both the formation of pathogenic thrombosis and the elevation of lipopolysaccharide activity in circulation.

In addition to immunothrombosis, it is plausible that the contact activation pathway has a significant impact on gut barrier function (Figure 3). Negatively charged molecules (eg, heparin, dextran sulfate, and

endotoxins) activate the kallikrein-kinin system, which eventually leads to production of the vasodilator, bradykinin.[28] Earlier studies support the view that aberrant activation of kallikrein-kinin pathway could be associated with decreased intestinal barrier function, which may eventually lead to higher circulating endotoxin levels and thereby increase the risk of organ damage.[29]

The fifth locus associated with endotoxemia is next to *LIPC*, encoding hepatic lipase. Hepatic lipase also has plausible links to the kinetics of the contact activation cascade, because it is characterized by its ability to bind heparin and heparan sulfate in the endothelium.[30] Heparin or heparan sulfate interact with antithrombin (III),[31] potentiating its inhibitory effect on both FXII and FXI activity. Most important, however, genetic variation in *LIPC* affects lipoprotein particle distribution and composition, which might have a direct effect on endotoxemia.[32] Hepatic lipase is recognized in HDL metabolism and reverse cholesterol transport by hydrolyzing phospholipids and triglycerides, resulting in particles that are more susceptible to clearance. Most of the lipopolysaccharide activity in the circulation is bound to lipoproteins, especially HDL, which contributes to the neutralization of lipopolysaccharide activity.[33] However, lipoprotein distribution is different during inflammation, infection, or metabolic diseases.[5,6] Therefore, detoxification of lipopolysaccharide and the following net endotoxemia is dependent on the inflammatory and metabolic state, lipoprotein and apolipoprotein profile, and concentrations of lipopolysaccharide binding proteins, which may all be disturbed in stroke.[33]

We also measured endotoxemia using different techniques, Endolisa and a mass spectrometry–based method. The former technique is based on the binding of lipopolysaccharide to a recombinant bacteriophage protein, and the latter quantifies the most abundant hydroxylated fatty acid of the lipid A moiety of most lipopolysaccharide molecules. The genetic associations seem to be restricted to the biological activity of lipopolysaccharide determined by the LAL assay, because with the 2 other methods we did not find significant associations with the lead SNP (*rs1801020*). Indeed, correlations between the results obtained with different methods were only modest. It is known that structural variations and enzymatic modifications of lipopolysaccharide molecules lead to differing immunologic responses.[1] In relation to systemic inflammation, the level of endotoxin activity has been considered a more important determinant than the total endotoxin mass.

The observed associations between lead SNPs and coagulation-related end points should be interpreted with caution because of multiple testing of various end points and the 5 lead SNPs. Another obvious limitation of the work is that we were not able to target the mechanisms behind the observed associations. Traditionally, endotoxemia has been considered an acquired characteristic of an individual, but as the mechanisms of the endotoxemia-disease associations are still nonconclusive, a more complex connection is probable. This would include, as proposed in this study, underlying genetics, which can act as a mutual risk factor for both disease incidence and mechanisms affecting lipopolysaccharide processing or access into the circulation.

The current data convincingly show that the genetic variation makes a highly significant contribution to endotoxemia. Assuming that genetic factors can modify the translocation or the neutralization of endotoxins, the results suggest a novel part in the puzzle of host/microbiome interactions. More important, the results further characterize the concept of tight interaction between immunity and coagulation.

Disclosures

Supplementary Material

Appendix S1
Data S1
Tables S1–S11
Figures S1–S7
References 34–49

REFERENCES

1. Manco M, Putignani L, Bottazzo GF. Gut microbiota, lipopolysaccharides, and innate immunity in the pathogenesis of obesity and cardiovascular risk. *Endocr Rev*. 2010;31:817–844. doi: 10.1210/er.2009-0030

2. Kallio KAE, Hätönen KA, Lehto M, Salomaa V, Männistö S, Pussinen PJ. Endotoxemia, nutrition, and cardiometabolic disorders. *Acta Diabetol*. 2015;52:395–404. doi: 10.1007/s00592-014-0662-3

3. Pussinen PJ, Havulinna AS, Lehto M, Sundvall J, Salomaa V. Endotoxemia is associated with an increased risk of incident diabetes. *Diabetes Care*. 2011;34:392–397. doi: 10.2337/dc10-1676

4. Simonsen JR, Järvinen A, Harjutsalo V, Forsblom C, Groop P, Lehto M. The association between bacterial infections and the risk of coronary heart disease in type 1 diabetes. *J Intern Med*. 2020;288:711–724. doi: 10.1111/joim.13138

5. Gomes JMG, Costa JdA, Alfenas RdCG. Metabolic endotoxemia and diabetes mellitus: a systematic review. *Metabolism*. 2017;68:133–144. doi: 10.1016/j.metabol.2016.12.009

6. Määttä A, Salminen A, Pietiäinen M, Leskelä J, Palviainen T, Sattler W, Sinisalo J, Salomaa V, Kaprio J, Pussinen PJ. Endotoxemia is associated with an adverse metabolic profile. *Innate Immun*. 2021;27:3–14. doi: 10.1177/1753425920971702

7. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, Frees K, Watt JL, Schwartz DA. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet*. 2000;25:187–191. doi: 10.1038/76048

8. Ferguson JF, Meyer NJ, Qu L, Xue C, Liu Y, DerOhannessian SL, Rushefski M, Paschos GK, Tang S, Schadt EE, et al. Integrative genomics identifies 7p11.2 as a novel locus for fever and clinical stress response in humans. *Hum Mol Genet*. 2015;24:1801–1812. doi: 10.1093/hmg/ddu589

9. Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, Ahola-Olli A, Kurki M, Karjalainen J, Palta P, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med*. 2020;26:549–557. doi: 10.1038/s41591-020-0800-0

10. Pais de Barros J-P, Gautier T, Sali W, Adrie C, Choubley H, Charron E, Lalande C, Le Guern N, Deckert V, Monchi M, et al. Quantitative lipopolysaccharide analysis using HPLC/MS/MS and its combination with the limulus amebocyte lysate assay. *J Lipid Res*. 2015;56:1363–1369. doi: 10.1194/jlr.D059725

11. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45:1238–1243. doi: 10.1038/ng.2756

12. Pussinen PJ, Tuomisto K, Jousilahti P, Havulinna AS, Sundvall J, Salomaa V. Endotoxemia, immune response to periodontal pathogens, and systemic inflammation associate with incident cardiovascular disease events. *Arterioscler Thromb Vasc Biol*. 2007;27:1433–1439. doi: 10.1161/ATVBAHA.106.138743

13. Wiedermann CJ, Kiechl S, Dunzendorfer S, Schratzberger P, Egger G, Oberhollenzer F, Willeit J. Association of endotoxemia with carotid atherosclerosis and cardiovascular disease: prospective results from the Bruneck Study. *J Am Coll Cardiol*. 1999;34:1975–1981. doi: 10.1016/S0735-1097(99)00448-9

14. Klimiec E, Pasinska P, Kowalska K, Pera J, Slowik A, Dziedzic T. The association between plasma endotoxin, endotoxin pathway proteins and outcome after ischemic stroke. *Atherosclerosis*. 2018;269:138–143. doi: 10.1016/j.atherosclerosis.2017.12.034

15. Benakis C, Brea D, Caballero S, Faraco G, Moore J, Murphy M, Sita G, Racchumi G, Lingo L, Pamer EG, et al. Commensal microbiota affects ischemic stroke outcome by regulating intestinal gamma delta T cells. *Nat Med*. 2016;22:516–523.

16. Yin J, Liao S, He Y, Wang S, Xia G, Liu F, Zhu J, You C, Chen Q, Zhou L, et al. Dysbiosis of gut microbiota with reduced trimethylamine-N-oxide level in patients with large-artery atherosclerotic stroke or transient ischemic attack. *J Am Heart Assoc*. 2015;4:e002699. doi: 10.1161/JAHA.115.002699

17. Petruk G, Puthia M, Petrlova J, Samsudin F, Strömdahl A, Cerps S, Uller L, Kjellström S, Bond PJ, Schmidtchen A. SARS-CoV-2 Spike protein binds to bacterial lipopolysaccharide and boosts proinflammatory activity. *J Mol Cell Biol*. 2020;12:916–932. doi: 10.1093/jmcb/mjaa067

18. Calafell F, Almasy L, Sabater-Lleal M, Buil A, Mordillo C, Ramírez-Soriano A, Sikora M, Souto JC, Blangero J, Fontcuberta J, et al. Sequence variation and genetic evolution at the human F12 locus: mapping quantitative trait nucleotides that influence FXII plasma levels. *Hum Mol Genet*. 2010;19:517–525. doi: 10.1093/hmg/ddp517

19. Lourdusamy A, Newhouse S, Lunnon K, Proitsi P, Powell J, Hodges A, Nelson SK, Stewart A, Williams S, Kloszewska I, et al. Identification of cis-regulatory variation influencing protein abundance levels in human plasma. *Hum Mol Genet*. 2012;21:3719–3726.

20. Sabater-Lleal M, Martinez-Perez A, Buil A, Folkersen L, Souto JC, Bruzelius M, Borrell M, Odeberg J, Silveira A, Eriksson P, et al. A genome-wide association study identifies KNG1 as a genetic determinant of plasma factor XI level and activated partial thromboplastin time. *Arterioscler Thromb Vasc Biol*. 2012;32:2008–2016.

21. Portelli MA, Siedlinski M, Stewart CE, Postma DS, Nieuwenhuis MA, Vonk JM, Nurnberg P, Altmuller J, Moffatt MF, Wardlaw AJ, et al. Genome-wide protein QTL mapping identifies human plasma kallikrein as a post-translational regulator of serum uPAR levels. *FASEB J*. 2014;28:923–934. doi: 10.1096/fj.13-240879

22. Sennblad B, Basu S, Mazur J, Suchon P, Martinez-Perez A, van Hylckama Vlieg A, Truong V, Li Y, Gådin JR, Tang W, et al. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum Mol Genet*. 2017;26:637–649. doi: 10.1093/hmg/ddw401

23. Xiao R, Boehnke M. Quantifying and correcting for the winner's curse in quantitative-trait association studies. *Genet Epidemiol*. 2011;35:133–138. doi: 10.1002/gepi.20551

24. Lindström S, Wang L, Smith EN, Gordon W, van Hylckama Vlieg A, de Andrade M, Brody JA, Pattee JW, Haessler J, Brumpton BM, et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood*. 2019;134:1645–1657. doi: 10.1182/blood.2019000435

25. de Haan HG, van Hylckama Vlieg A, Lotta LA, Gorski MM, Bucciarelli P, Martinelli I, Baglin TP, Peyvandi F, Rosendaal FR, et al. Targeted sequencing to identify novel genetic risk factors for deep vein thrombosis: a study of 734 genes. *J Thromb Haemost*. 2018;16:2432–2441. doi: 10.1111/jth.14279

26. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NLM, Jackson C, Farrall M, Rothwell PM, Sudlow C, Dichgans M, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke*. 2012;43:3161–3167. doi: 10.1161/STROKEAHA.112.665760

27. Engelmann B, Massberg S. Thrombosis as an intravascular effector of innate immunity. *Nat Rev Immunol*. 2013;13:34–45. doi: 10.1038/nri3345

28. Nickel KF, Long AT, Fuchs TA, Butler LM, Renné T. Factor XII as a therapeutic target in thromboembolic and inflammatory diseases. *Arterioscler Thromb Vasc Biol*. 2017;37:13–20. doi: 10.1161/ATVBAHA.116.308595

29. Lehto M, Groop P. The gut-kidney axis: putative interconnections between gastrointestinal and renal disorders. *Front Endocrinol (Lausanne)*. 2018;9:553. doi: 10.3389/fendo.2018.00553

30. Sendak RA, Berryman DE, Gellman G, Melford K, Bensadoun A. Binding of hepatic lipase to heparin: identification of specific heparin-binding residues in two distinct positive charge clusters. *J Lipid Res*. 2000;41:260–268. doi: 10.1016/S0022-2275(20)32060-5

31. Chuang YJ, Swanson R, Raja SM, Bock SC, Olson ST. The antithrombin P1 residue is important for target proteinase specificity but not for heparin activation of the serpin: characterization of P1 antithrombin variants with altered proteinase specificity but normal heparin activation. *Biochemistry*. 2001;40:6670–6679.

32. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707–713. doi: 10.1038/nature09270

33. Leskelä J, Pietiäinen M, Safer A, Lehto M, Metso J, Malle E, Buggle F, Becher H, Sundvall J, Grau AJ, et al. Serum lipopolysaccharide neutralizing capacity in ischemic stroke. *PLoS One*. 2020;15:e0228806. doi: 10.1371/journal.pone.0228806

34. Borodulin K, Tolonen H, Jousilahti P, Jula A, Juolevi A, Koskinen S, Kuulasmaa K, Laatikainen T, Männistö S, Peltonen M, et al. Cohort profile: the National FINRISK Study. *Int J Epidemiol*. 2017;47:696–696i. doi: 10.1093/ije/dyx239

35. Huang Y, Ollikainen M, Sipilä P, Mustelin L, Wang X, Su S, Huan T, Levy D, Wilson J, Snieder H, et al. Genetic and environmental effects on gene expression signatures of blood pressure: a transcriptome-wide twin study. *Hypertension*. 2018;71:457–464. doi: 10.1161/HYPERTENSIONAHA.117.10527

36. Kaprio J, Pulkkinen L, Rose RJ. Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. *Twin Res*. 2002;5:366–371. doi: 10.1375/136905202320906101

37. Latvala A, Tuulio-Henriksson A, Dick DM, Vuoksimaa E, Viken RJ, Suvisaari J, Kaprio J, Rose RJ. Genetic origins of the association between verbal ability and alcohol dependence symptoms in young adulthood. *Psychol Med*. 2011;41:641–651. doi: 10.1017/S0033291710001194

38. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, van der Laan SW, Gretarsdottir S, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524–537. doi: 10.1038/s41588-018-0058-3

39. 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi: 10.1038/nature15393

40. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284–1287. doi: 10.1038/ng.3656

41. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5:e1000529. doi: 10.1371/journal.pgen.1000529

42. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics*. 2016;32:1423–1426.

43. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191. doi: 10.1093/bioinformatics/btq340

44. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36:1–48.

45. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44:369–373. doi: 10.1038/ng.2213

46. Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet*. 2010;6:e1001058. doi: 10.1371/journal.pgen.1001058

47. Elsworth B, Matthew L, Alexander T, Liu Y, Matthews P, Hallett J, Bates P, Palmer T, Haberland V, Smith GD, et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv*. 2020.

48. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. doi: 10.7554/eLife.34408

49. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet*. 2018;50:1593–1599. doi: 10.1038/s41588-018-0248-z

SUPPLEMENTAL MATERIAL

Contributors of FinnGen

Steering Committee

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

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

Pharmaceutical companies

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

University of Helsinki & Biobanks

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

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

Other Experts/ Non-Voting Members

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

Scientific Committee

Pharmaceutical companies

Jeff Waring	Abbvie, Chicago, IL, United States
Bridget Riley-Gillis	Abbvie, Chicago, IL, United States

Athena Matakidou   Astra Zeneca, Cambridge, United Kingdom

Heiko Runz   Biogen, Cambridge, MA, United States

Jimmy Liu   Biogen, Cambridge, MA, United States

Shameek Biswas   Celgene, Summit, NJ, United States

Julie Hunkapiller   Genentech, San Francisco, CA, United States

Dawn Waterworth   GlaxoSmithKline, Brentford, United Kingdom

Meg Ehm   GlaxoSmithKline, Brentford, United Kingdom

Dorothee Diogo   Merck, Kenilworth, NJ, United States

Caroline Fox   Merck, Kenilworth, NJ, United States

Anders Malarstig   Pfizer, New York, NY, United States

Catherine Marshall   Pfizer, New York, NY, United States

Xinli Hu   Pfizer, New York, NY, United States

Kathy Call   Sanofi, Paris, France

Kathy Klinger   Sanofi, Paris, France

Matthias Gossel   Sanofi, Paris, France

Robert Graham   Maze Therapeutics, San Francisco, CA, United States

Tim Behrens   Maze Therapeutics, San Francisco, CA, United States

Beryl Cummings   Maze Therapeutics, San Francisco, CA, United States

Wilco Fleuren   Janssen Biotech, Beerse, Belgium

University of Helsinki & Biobanks

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

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

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

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

Olli Carpén   Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki

Reetta Hinttala	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Johannes Kettunen	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Johanna Mäkelä	Finnish Clinical Biobank Tampere **/** University of Tampere / Pirkanmaa Hospital District, Tampere, Finland
Arto Mannermaa	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland
Jari Laukkanen	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland
Urho Kujala	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland

Other Experts/ Non-Voting Members

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

Clinical Groups

Neurology Group

Hilkka Soininen	Northern Savo Hospital District, Kuopio, Finland
Valtteri Julkunen	Northern Savo Hospital District, Kuopio, Finland
Anne Remes	Northern Ostrobothnia Hospital District, Oulu, Finland
Reetta Kälviäinen	Northern Savo Hospital District, Kuopio, Finland
Mikko Hiltunen	Northern Savo Hospital District, Kuopio, Finland
Jukka Peltola	Pirkanmaa Hospital District, Tampere, Finland
Pentti Tienari	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Juha Rinne	Hospital District of Southwest Finland, Turku, Finland

Adam Ziemann   Abbvie, Chicago, IL, United States

Jeffrey Waring   Abbvie, Chicago, IL, United States

Sahar Esmaeeli   Abbvie, Chicago, IL, United States

Nizar Smaoui   Abbvie, Chicago, IL, United States

Anne Lehtonen   Abbvie, Chicago, IL, United States

Susan Eaton   Biogen, Cambridge, MA, United States

Heiko Runz   Biogen, Cambridge, MA, United States

Sanni Lahdenperä   Biogen, Cambridge, MA, United States

Janet van Adelsberg   Celgene, Summit, NJ, United States

Shameek Biswas   Celgene, Summit, NJ, United States

John Michon   Genentech, San Francisco, CA, United States

Geoff Kerchner   Genentech, San Francisco, CA, United States

Julie Hunkapiller   Genentech, San Francisco, CA, United States

Natalie Bowers   Genentech, San Francisco, CA, United States

Edmond Teng   Genentech, San Francisco, CA, United States

John Eicher   Merck, Kenilworth, NJ, United States

Vinay Mehta   Merck, Kenilworth, NJ, United States

Padhraig Gormley   Merck, Kenilworth, NJ, United States

Kari Linden   Pfizer, New York, NY, United States

Christopher Whelan   Pfizer, New York, NY, United States

Fanli Xu   GlaxoSmithKline, Brentford, United Kingdom

David Pulford   GlaxoSmithKline, Brentford, United Kingdom

Gastroenterology Group

Martti Färkkilä   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Sampsa Pikkarainen   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Airi Jussila   Pirkanmaa Hospital District, Tampere, Finland

Timo Blomster   Northern Ostrobothnia Hospital District, Oulu, Finland

Mikko Kiviniemi   Northern Savo Hospital District, Kuopio, Finland

Markku Voutilainen   Hospital District of Southwest Finland, Turku, Finland

Bob Georgantas   Abbvie, Chicago, IL, United States

Graham Heap   Abbvie, Chicago, IL, United States

Jeffrey Waring   Abbvie, Chicago, IL, United States

Nizar Smaoui   Abbvie, Chicago, IL, United States

Fedik Rahimov   Abbvie, Chicago, IL, United States

Anne Lehtonen   Abbvie, Chicago, IL, United States

Keith Usiskin   Celgene, Summit, NJ, United States

Tim Lu   Genentech, San Francisco, CA, United States

Natalie Bowers   Genentech, San Francisco, CA, United States

Danny Oh   Genentech, San Francisco, CA, United States

John Michon   Genentech, San Francisco, CA, United States

Vinay Mehta   Merck, Kenilworth, NJ, United States

Kirsi Kalpala   Pfizer, New York, NY, United States

Melissa Miller   Pfizer, New York, NY, United States

Xinli Hu   Pfizer, New York, NY, United States

Linda McCarthy   GlaxoSmithKline, Brentford, United Kingdom

Rheumatology Group

Kari Eklund   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Antti Palomäki   Hospital District of Southwest Finland, Turku, Finland

Pia Isomäki   Pirkanmaa Hospital District, Tampere, Finland

Laura Pirilä   Hospital District of Southwest Finland, Turku, Finland

Oili Kaipiainen-Seppänen Northern Savo Hospital District, Kuopio, Finland

Johanna Huhtakangas Northern Ostrobothnia Hospital District, Oulu, Finland

Bob Georgantas   Abbvie, Chicago, IL, United States

Jeffrey Waring   Abbvie, Chicago, IL, United States

Fedik Rahimov	Abbvie, Chicago, IL, United States
Apinya Lertratanakul	Abbvie, Chicago, IL, United States
Nizar Smaoui	Abbvie, Chicago, IL, United States
Anne Lehtonen	Abbvie, Chicago, IL, United States
David Close	Astra Zeneca, Cambridge, United Kingdom
Marla Hochfeld	Celgene, Summit, NJ, United States
Natalie Bowers	Genentech, San Francisco, CA, United States
John Michon	Genentech, San Francisco, CA, United States
Dorothee Diogo	Merck, Kenilworth, NJ, United States
Vinay Mehta	Merck, Kenilworth, NJ, United States
Kirsi Kalpala	Pfizer, New York, NY, United States
Nan Bing	Pfizer, New York, NY, United States
Xinli Hu	Pfizer, New York, NY, United States
Jorge Esparza Gordillo	GlaxoSmithKline, Brentford, United Kingdom
Nina Mars	Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland

Pulmonology Group

Tarja Laitinen	Pirkanmaa Hospital District, Tampere, Finland
Margit Pelkonen	Northern Savo Hospital District, Kuopio, Finland
Paula Kauppi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Hannu Kankaanranta	Pirkanmaa Hospital District, Tampere, Finland
Terttu Harju	Northern Ostrobothnia Hospital District, Oulu, Finland
Nizar Smaoui	Abbvie, Chicago, IL, United States
David Close	Astra Zeneca, Cambridge, United Kingdom
Susan Eaton	Biogen, Cambridge, MA, United States
Steven Greenberg	Celgene, Summit, NJ, United States
Hubert Chen	Genentech, San Francisco, CA, United States
Natalie Bowers	Genentech, San Francisco, CA, United States

John Michon	Genentech, San Francisco, CA, United States
Vinay Mehta	Merck, Kenilworth, NJ, United States
Jo Betts	GlaxoSmithKline, Brentford, United Kingdom
Soumitra Ghosh	GlaxoSmithKline, Brentford, United Kingdom

Cardiometabolic Diseases Group

Veikko Salomaa	The National Institute of Health and Welfare Helsinki, Finland
Teemu Niiranen	The National Institute of Health and Welfare Helsinki, Finland
Markus Juonala	Hospital District of Southwest Finland, Turku, Finland
Kaj Metsärinne	Hospital District of Southwest Finland, Turku, Finland
Mika Kähönen	Pirkanmaa Hospital District, Tampere, Finland
Juhani Junttila	Northern Ostrobothnia Hospital District, Oulu, Finland
Markku Laakso	Northern Savo Hospital District, Kuopio, Finland
Jussi Pihlajamäki	Northern Savo Hospital District, Kuopio, Finland
Juha Sinisalo	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Marja-Riitta Taskinen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Tiinamaija Tuomi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Jari Laukkanen	Central Finland Health Care District, Jyväskylä, Finland
Ben Challis	Astra Zeneca, Cambridge, United Kingdom
Andrew Peterson	Genentech, San Francisco, CA, United States
Julie Hunkapiller	Genentech, San Francisco, CA, United States
Natalie Bowers	Genentech, San Francisco, CA, United States
John Michon	Genentech, San Francisco, CA, United States
Dorothee Diogo	Merck, Kenilworth, NJ, United States
Audrey Chu	Merck, Kenilworth, NJ, United States
Vinay Mehta	Merck, Kenilworth, NJ, United States
Jaakko Parkkinen	Pfizer, New York, NY, United States
Melissa Miller	Pfizer, New York, NY, United States

Anthony Muslin   Sanofi, Paris, France

Dawn Waterworth   GlaxoSmithKline, Brentford, United Kingdom

Oncology Group

Heikki Joensuu   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Olli Carpén   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Tuomo Meretoja   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Lauri Aaltonen   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Johanna Mattson   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Johanna Schleutker   University of Turku, Turku, Finland

Annika Auranen   Pirkanmaa Hospital District , Tampere, Finland

Peeter Karihtala   Northern Ostrobothnia Hospital District, Oulu, Finland

Saila Kauppila   Northern Ostrobothnia Hospital District, Oulu, Finland

Päivi Auvinen   Northern Savo Hospital District, Kuopio, Finland

Klaus Elenius   Hospital District of Southwest Finland, Turku, Finland

Relja Popovic   Abbvie, Chicago, IL, United States

Jeffrey Waring   Abbvie, Chicago, IL, United States

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

Anne Lehtonen   Abbvie, Chicago, IL, United States

Athena Matakidou   Astra Zeneca, Cambridge, United Kingdom

Jennifer Schutzman   Genentech, San Francisco, CA, United States

Julie Hunkapiller   Genentech, San Francisco, CA, United States

Natalie Bowers   Genentech, San Francisco, CA, United States

John Michon   Genentech, San Francisco, CA, United States

Vinay Mehta   Merck, Kenilworth, NJ, United States

Andrey Loboda   Merck, Kenilworth, NJ, United States

Aparna Chhibber   Merck, Kenilworth, NJ, United States

Heli Lehtonen   Pfizer, New York, NY, United States

Stefan McDonough   Pfizer, New York, NY, United States

Marika Crohns   Sanofi, Paris, France

Diptee Kulkarni   GlaxoSmithKline, Brentford, United Kingdom

Opthalmology Group

Kai Kaarniranta   Northern Savo Hospital District, Kuopio, Finland

Joni A Turunen   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Terhi Ollila   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Sanna Seitsonen   Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Hannu Uusitalo   Pirkanmaa Hospital District, Tampere, Finland

Vesa Aaltonen   Hospital District of Southwest Finland, Turku, Finland

Hannele Uusitalo-Järvinen   Pirkanmaa Hospital District, Tampere, Finland

Marja Luodonpää   Northern Ostrobothnia Hospital District, Oulu, Finland

Nina Hautala   Northern Ostrobothnia Hospital District, Oulu, Finland

Heiko Runz   Biogen, Cambridge, MA, United States

Stephanie Loomis   Biogen, Cambridge, MA, United States

Erich Strauss   Genentech, San Francisco, CA, United States

Natalie Bowers   Genentech, San Francisco, CA, United States

Hao Chen   Genentech, San Francisco, CA, United States

John Michon   Genentech, San Francisco, CA, United States

Anna Podgornaia   Merck, Kenilworth, NJ, United States

Vinay Mehta   Merck, Kenilworth, NJ, United States

Dorothee Diogo   Merck, Kenilworth, NJ, United States

Joshua Hoffman   GlaxoSmithKline, Brentford, United Kingdom

Dermatology Group

Kaisa Tasanen   Northern Ostrobothnia Hospital District, Oulu, Finland

Laura Huilaja   Northern Ostrobothnia Hospital District, Oulu, Finland

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

Odontology Group

Pirkko Pussinen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Aino Salminen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Tuula Salo	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
David Rice	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Pekka Nieminen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Ulla Palotie	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Maria Siponen	Northern Savo Hospital District, Kuopio, Finland
Liisa Suominen	Northern Savo Hospital District, Kuopio, Finland
Päivi Mäntylä	Northern Savo Hospital District, Kuopio, Finland
Ulvi Gursoy	Hospital District of Southwest Finland, Turku, Finland
Vuokko Anttonen	Northern Ostrobothnia Hospital District, Oulu, Finland
Kirsi Sipilä	Northern Ostrobothnia Hospital District, Oulu, Finland

FinnGen Analysis working group

| Justin Wade Davis | Abbvie, Chicago, IL, United States |

Bridget Riley-Gillis	Abbvie, Chicago, IL, United States
Danjuma Quarless	Abbvie, Chicago, IL, United States
Fedik Rahimov	Abbvie, Chicago, IL, United States
Sahar Esmaeeli	Abbvie, Chicago, IL, United States
Slavé Petrovski	Astra Zeneca, Cambridge, United Kingdom
Eleonor Wigmore	Astra Zeneca, Cambridge, United Kingdom
Jimmy Liu	Biogen, Cambridge, MA, United States
Chia-Yen Chen	Biogen, Cambridge, MA, United States
Paola Bronson	Biogen, Cambridge, MA, United States
Ellen Tsai	Biogen, Cambridge, MA, United States
Stephanie Loomis	Biogen, Cambridge, MA, United States
Yunfeng Huang	Biogen, Cambridge, MA, United States
Joseph Maranville	Celgene, Summit, NJ, United States
Shameek Biswas	Celgene, Summit, NJ, United States
Elmutaz Shaikho Elhaj Mohammed	Celgene, Summit, NJ, United States
Samir Wadhawan	Bristol-Meyers-Squibb
Erika Kvikstad	Bristol-Meyers-Squibb
Minal Caliskan	Bristol-Meyers-Squibb
Diana Chang	Genentech, San Francisco, CA, United States
Julie Hunkapiller	Genentech, San Francisco, CA, United States
Tushar Bhangale	Genentech, San Francisco, CA, United States
Natalie Bowers	Genentech, San Francisco, CA, United States
Sarah Pendergrass	Genentech, San Francisco, CA, United States
Dorothee Diogo	Merck, Kenilworth, NJ, United States
Emily Holzinger	Merck, Kenilworth, NJ, United States
Padhraig Gormley	Merck, Kenilworth, NJ, United States
Xing Chen	Pfizer, New York, NY, United States
Åsa Hedman	Pfizer, New York, NY, United States
Karen S King	GlaxoSmithKline, Brentford, United Kingdom

Clarence Wang   Sanofi, Paris, France

Ethan Xu   Sanofi, Paris, France

Franck Auge   Sanofi, Paris, France

Clement Chatelain   Sanofi, Paris, France

Deepak Rajpal   Sanofi, Paris, France

Dongyu Liu   Sanofi, Paris, France

Katherine Call   Sanofi, Paris, France

Tai-he Xia   Sanofi, Paris, France

Beryl Cummings   Maze Therapeutics, San Francisco, CA, United States

Matt Brauer   Maze Therapeutics, San Francisco, CA, United States

Mitja Kurki   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland /

                       Broad Institute, Cambridge, MA, United States

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

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

Juha Karjalainen   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland /

                       Broad Institute, Cambridge, MA, United States

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

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

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

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

Wei Zhou   Broad Institute, Cambridge, MA, United States

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

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

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

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

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

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

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

Hannele Laivuori   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Sina Rüeger	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mari E Niemi	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Taru Tukiainen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mary Pat Reeve	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Henrike Heyne	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Nina Mars	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Kimmo Palin	University of Helsinki, Helsinki, Finland
Javier Garcia-Tabuenca	University of Tampere, Tampere, Finland
Harri Siirtola	University of Tampere, Tampere, Finland
Tuomo Kiiskinen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Jiwoo Lee	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad Institute, Cambridge, MA, United States
Kristin Tsuo	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad Institute, Cambridge, MA, United States
Amanda Elliott	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad Institute, Cambridge, MA, United States
Kati Kristiansson	THL Biobank / The National Institute of Health and Welfare Helsinki, Finland
Mikko Arvas	Finnish Red Cross Blood Service, Helsinki, Finland
Kati Hyvärinen	Finnish Red Cross Blood Service, Helsinki, Finland
Jarmo Ritari	Finnish Red Cross Blood Service, Helsinki, Finland
Miika Koskinen	Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki
Olli Carpén	Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki
Johannes Kettunen	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Katri Pylkäs	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland
Marita Kalaoja	Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland

Minna Karjalainen — Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland

Tuomo Mantere — Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland

Eeva Kangasniemi — Finnish Clinical Biobank Tampere **/** University of Tampere / Pirkanmaa Hospital District, Tampere, Finland

Sami Heikkinen — Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland

Arto Mannermaa — Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland

Eija Laakkonen — Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland

Csilla Sipeky — University of Turku, Turku, Finland

Samuel Heron — University of Turku, Turku, Finland

Antti Karlsson — Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland

Dhanaprakash Jambulingam — University of Turku, Turku, Finland

Venkat Subramaniam Rathinakannan — University of Turku, Turku, Finland

Biobank directors

Lila Kallio — Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland

Sirpa Soini — THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Jukka Partanen — Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland

Eero Punkka — Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki

Raisa Serpi — Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland

Johanna Mäkelä	Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland
Veli-Matti Kosma	Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland
Teijo Kuopio	Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland

FinnGen Teams

Administration

Anu Jalanko	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Risto Kajanne	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mervi Aavikko	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Manuel González Jiménez	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Analysis

Mitja Kurki	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad Institute, Cambridge, MA, United States
Juha Karjalainen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad Institute, Cambridge, MA, United States
Pietro della Briotta Parola	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Sina Rüeger	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Arto Lehistö	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Wei Zhou	Broad Institute, Cambridge, MA, United States
Masahiro Kanai	Broad Institute, Cambridge, MA, United States

Clinical Endpoint Development

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

Tuomo Kiiskinen   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Communication

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

Data Management and IT Infrastructure

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

Elina Kilpeläinen   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Timo P. Sipilä   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Georg Brein   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Oluwaseun A. Dada   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Ghazal Awaisa   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Anastasia Shcherban   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Genotyping

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

Timo P. Sipilä   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Sample Collection Coordination

Anu Loukola   Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki

Sample Logistics

Päivi Laiho   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Tuuli Sistonen   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Essi Kaiharju   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Markku Laukkanen   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Elina Järvensivu   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Sini Lähteenmäki   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Lotta Männikkö   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Regis Wong   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Registry Data Operations

Hannele Mattsson   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

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

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

Tero Hiekkalinna   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Teemu Paajanen   THL Biobank / The National Institute of Health and Welfare Helsinki, Finland

Sequencing Informatics

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

Kalle Pärn   Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Trajectory Team

Tarja Laitinen   Pirkanmaa Hospital District, Tampere, Finland

Harri Siirtola   University of Tampere, Tampere, Finland

Javier Gracia-Tabuenca   University of Tampere, Tampere, Finland

Data S1.

Supplemental Methods

Study population

The FinnDiane Study

The FinnDiane study is an ongoing nationwide multi-center study consisting of adult participants with T1D. The participants are enrolled by their local attending physicians at 77 local hospitals or healthcare centers. Although FinnDiane is not a population study *per se*, the distribution of the participants closely follows that of the population density in Finland. Blood samples and basic clinical measures are taken during the visit of examination. Data has also been collected from the national registries. For this study, the FinnDiane cohort was divided into two sub-cohorts: participants fulfilling the conventional T1D criteria ("FD-T1D": age at diabetes onset <40 years, insulin dependent within one year of diabetes onset) and participants with plausible other types of diabetes not matching the aforementioned criteria ("FD-rest").

The FINRISK Study

FINRISK is a random population-based survey designed to study the prevalence of CVD risk factors in Finland and it is conducted every five years[34]. Participants undergo a physical examination and complete a questionnaire regarding CVD risk factors. In the present study, we included subsets of two of the cohorts enrolled in 1992[12] (FR92) and 1997[3] (FR97), and who had had the LPS activities measured. For these cohorts, data from blood measures and background questionnaires as well as

national registries was available. The subset of FINRISK -92 with LPS measures available was a CVD case-cohort sample[12], and thus is not a random population sample.

The Finnish Twin Cohort Study

The Finnish Twin Cohort was first established in 1974 to investigate genetic and environmental risk factors for chronic disorders. Twins and their families have been ascertained in three stages from the Central Population Register in 1974 (older like-sexed pairs), 1987 (multiple births 1968-1987) and 1995 (opposite-sex pairs 1938-1957). The older part consists of same-gender twin pairs born before 1958 with both twins alive in 1975, and four surveys have been carried out, the latest during 2011-2012. Based on this latest survey, twins from pairs concordant and discordant for blood pressure were invited to an in-person clinical study[35], with blood draws after overnight fasting. FinnTwin16 is a longitudinal study of initially adolescent twins born in 1975-1979[36]. The twins replied to a fourth survey as young adults in 2000-2002, after which a sample of pairs concordant and discordant for alcohol were invited to an in-person assessment including a fasting blood draw[37].

FinnGen Study

The Finngen study population used in the study was independent from the GWAS meta-analysis. Finngen is a growing project with samples from multiple biobanks and cohort studies: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-

US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta) and Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi).

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, older research cohorts, collected prior to the start of FinnGen (in August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol, Nr HUS/990/2017.

The FinnGen study is approved by Finnish Institute for Health and Welfare (THL), approval number THL/2031/6.02.00/2017, amendments THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, Digital and population data service agency VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3 the Social Insurance Institution (KELA) KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, and Statistics Finland TK-53-1041-17.

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 5 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154, Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018, Finnish Clinical Biobank Tampere MH0004, Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001.

UK Biobank, Megastroke, and Cerebrovascular Disease Knowledge Portal

UK Biobank is a large-scale biobank study conducted in United Kindom, including 452 264 individuals, 778 traits, and 30 million genetic variants (http://www.nealelab.is/uk-biobank/). The study began in 2006 and the round 2 results used in the present study were released in 2018.

Megastroke is an international consortium of 29 studies, including 67,162 cases and 454,450 controls[38]. Megastroke is included in the Cerebrovascular Disease Knowledge Portal (CDKP, https://cd.hugeamp.org/), along with 22 datasets, of which 18 datasets have GWAS information available. The portal has data on 48 disease traits, all connected to cerebrovascular diseases.

Genotyping and genotype imputation

The FinnDiane cohort was genotyped in three batches using Illumina Human CoreExome chips (v12.1.0, v12.1.1 and v24.1.0 correspondingly). The FINRISK genotype data has been created using multiple genotyping platforms. Parts of the FR97 and FR92 cohorts were genotyped in the same

batches. The chips used were Affymetrix 6.0, and various versions of Illumina HumanCoreExome, Illumina 610K, and Illumina Omniexpress. The FinnTwin16 subsample was genotyped in three batches using Illumina Human670-QuadCustom v1 A and Illumina HumanCoreExome-12 v1.1 A & v1.0 A chips. The Finnish Twin Cohort VpEpi subsample was genotyped in three batches using Illumina Human670-QuadCustom v1 A, Illumina Human610-Quad v1 B and Illumina HumanCoreExome-24 v1.0 A chips.

The genotypes of each cohort went through rigorous QC, removing poor quality samples (such as gender issues, low genotyping rate, extreme heterozygosity, sample mix-ups, and PCA/MDS-outliers) and poor quality markers (such as low genotyping rate, monomorphic in cohort or batch, HWE violations, MAF differing notably from 1000G reference).

All cohorts were imputed using 1000G phase3 reference genotypes[39]. FinnDiane and Finnish Twin Cohorts were imputed with minimac3 -software[40] and FINRISK cohorts with the IMPUTE2 – software[41].

Genetic analyses

Single marker association analyses were performed with rvtests software[42] using LMM-based analysis to correct for the effect of cryptic relatedness and close relatives included in the analyses. The p-value for association was computed using score test (--meta score). The models were adjusted for empirical kinship matrix, sex, age, BMI, total cholesterol, HDL cholesterol, triglycerides (log-transformed) and study-specific covariates (in FinnDiane, sample freeze time and genotyping batch; in FINRISK, genotyping batch and recruitment region; in Finnish Twin Cohort, sample freeze time,

genotyping batch and additionally pregnancy status in FinnTwin16). Missing covariate values were imputed to the mean of the cohort (by software default). Multivariable regression model was fitted to the data, and the residuals of the model were inverse normalized and used as response variables for genetic analyzes (as implemented in the --useResidualAsPhenotype and --inverseNormal flags of rvtests). Genotype dosages computed from the imputation posterior probabilities were used in the analyses. The kinship matrix was computed using vcf2kinship software (distributed with rvtests) using imputed genotype dosages of common (MAF>5%) markers. The analyses were conducted correspondingly in the six aforementioned cohorts. Cohort-wise results were filtered based on imputation quality (minimac3: R^2>0.7; IMPUTE2: information>0.7), Hardy-Weinberg equilibrium (p_{HVE}>1×10^{-10}), missing genotypes (less than 5%) and allele frequency (MAF>1%). Finally, the results were meta-analyzed using METAL-software[43] in an inverse variance weighted fixed effect meta-analysis. Markers reaching genome-wide significance (p<5×10^{-8}) in meta-analysis were considered statistically significant. Forest-plots of the results were plotted using R and metafor package[44].

A baseline linear regression model with the same covariates as the initial GWAS analyses were fitted, and the model performance (Akaike information criterion, AIC) and explanatory power (adjusted R^2) of covariates in the models were evaluated in FD-T1D and FR97 cohorts (using participants with complete covariate data available). Then each of the lead SNPs from fixed-effects meta-analyses were added onto the baseline model separately to evaluate the increase in performance, and finally all five lead SNPs were added to the model simultaneously and the model performance was evaluated. The increase in explanatory power was evaluated as an increment in adjusted R^2 of the baseline model vs. full model, and the improvement of models and informativeness of the SNPs were confirmed by a decrease in the AIC value. Additionally, stepwise conditional regression[45]

(GCTA-COJO) was performed to reveal putatively causal and conditionally independent signals from each top locus from meta-analysis.

As the associations of the lead SNPs were independent from each other in the multivariable regression models, the conditional GWAS analyses were adjusted for all the five lead markers (rs5030082, rs71640036, rs1801020, rs2081361, and rs10152355) simultaneously. In addition, the cohort-wise GWAS analyses were also re-analyzed, conditioning the analyses for the lead SNPs of genome-wide significant loci available in all cohorts in the initial fixed-effects meta-analysis, and the conditional analysis results were combined in fixed effects meta-analysis using METAL, corresponding to the initial GWAS analyses.

We performed leave-one-out analysis, where the GWAS meta-analysis was implemented leaving one of the cohorts out of the analysis at a time. In other respect, the meta-analysis used the same parameters as the initial GWAS meta-analysis.

Expression quantitative loci

The effect of significant variants (n=741) on gene expression was estimated using quantitative trait loci (eQTL) data from GTEx portal (gtexportal.org/home/eqtls) and eQTLs association meta-analysis study performed on 5,311 blood samples[11]. Data from GTEx Portal were obtained on 27/6/2020 and contained information from 838 individuals and 49 tissues (release v8). For the GTEx-data, p-value threshold was set to $p < 10^{-6}$ and normalized effect size (NES) to $|NES| > 0.2$ for eQTL association to be considered significant.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) of the fixed effects meta-analysis of GWAS results was performed with the Meta-Analysis Gene-set Enrichment of VariaNT Associations (MAGENTA) software[46] (vs 2.4) allowing the minimum gene set size of four and maximum gene set size of 2000. The number of randomly sampled gene sets for GSEA-GWAS p-value estimation was 10000, and the gene boundaries used for mapping SNPs onto genes were 110kb upstream to the most extreme gene transcript start position and 40kb downstream to the most extreme gene transcript end position.

Mendelian Randomization

MR-Base[47,48] was used to conduct mendelian randomization (MR) analysis. Outcome variables were chosen based on GSEA and GRS analysis: ischemic stroke, pulmonary embolism, deep vein thrombosis of lower extremities and pulmonary embolism, and venous thromboembolism (UK-biobank; http://www.nealelab.is/uk-biobank/) were studied. Lead SNPs (in accordance with GCTA-COJO analysis) of the GWAS meta-analysis were used in MR. The analysis was conducted with default parameters using an online tool available at http://app.mrbase.org/.

Genetic Risk Score

The 5 lead SNPs of GWAS meta-analysis (in accordance with GCTA-COJO analysis) were used in GRS calculation. The risk score was calculated as a sum of these SNP genotypes weighed with corresponding effect sizes of a GWAS fixed effect meta-analysis.

Genetic Risk Score endpoints

FinnGen uses nationwide registries to define disease endpoints, including registries for the International Classification of Diseases (ICD) revisions 8, 9, and 10 codes, which are linked to personal identification numbers assigned to all Finnish citizens and residents. ICD-10 diagnosis codes associated with the endpoints are listed in Table S8. Some endpoints have other inclusion criteria in addition to diagnosis codes: endpoints "Cerebrovascular diseases", "Aneurysms, operations, SAH" and "Endovascular/surgical operations to intracerebral aneurysms" include participants with information on surgical or endovascular operations for intracerebral aneurysms as retrieved from hospital discharge registry; "Atrial fibrillation and flutter" includes participants with dronedarone medication; "Peripheral artery operations in Hilmo" includes participants with peripheral artery operations as retrieved from hospital discharge registry. A comprehensive listing of FinnGen endpoints and control definitions is available online at https://www.finngen.fi/en/researchers/clinical-endpoints.

Lead SNP associations in UK Biobank and CDKP populations

Online lookup tools were used to analyze lead SNP associations in UK-biobank and Cerebrovascular Disease Knowledge Portal (CDKP) populations. The Gene ATLAS[49] tool was used to achieve a phenome-wide association study approach in the UK-biobank population (Table S10). CDKP online browser (https://cd.hugeamp.org/) was used to study lead SNP associations in cerebrovascular disease phenotypes (Table S10). CDKP combines multiple populations, including Megastroke, and summarizes the result in bottom line analysis. Beta-values and corresponding confidence intervals were transformed to odds ratios using exp()-function in R.

UK Biobank population showed associations in a wide range of phenotypes, including body mass- and coagulation-related traits. The most significant associations in CDKP included "Cardio-aortic embolism (TOAST-classification)", "all ischemic stroke", "any stroke", and "TOAST other determined".

Endotoxemia measured by using other techniques

A subset of FinnDiane participants (n=369) was selected based on the *F12* SNP rs1801020 genotype, and endotoxemia was measured using a HPLC-MS/MS approach as previously described[16]. The method quantifies 3-hydroxytetradecanoic acid (3HM), and the plasma samples (50 µl) were spiked with the internal standard (3-hydroxytridecanoic acid, 4 pmol). The samples were hydrolyzed (for total 3HM) or not (for unesterified 3HM), and free fatty acids were extracted, separated using HPLC, and detected by MS/MS. The LPS-derived 3HM was calculated as the difference between total and unesterified 3HM. The homozygous participants were selected in approximately 1:2 ratio, resulting in 124 and 245 participants with the TT- and the CC-homozygous genotype, respectively. In QC, four of the samples were excluded, and the LPS-mass could not be defined in two samples. Thus 363 participants were used for the regression analyses (TT: 121, CC: 242).

For the Endolisa-assay analyses, the population consisted of 326 subjects genotyped for rs1801020 (TT: 21, TC: 124 and CC: 181). In addition to the endotoxemia determinations by using the LAL-assay similarly as in the GWAS analysis, it was also measured using Endolisa assay (609033, Hyglos GmbH,

Bernried, Germany) according to the manufacturers' protocol. All samples passed additional QC and were included in the regression analyses.

Due to skewed distribution, Endolisa measures were log-transformed (natural logarithm), but the LPS mass demonstrated visually normal distribution. The correlation was computed using Pearson's product-moment correlation. The SNP association with endotoxemia was analyzed with linear regression applying an additive genetic model. Due to the *post-hoc* nature of the additional analyses, $p<0.05$ was considered statistically significant.

Results from post-hoc analyses

Despite belonging to the same pathway, the lead SNP associations were independent of each other. This was observed in further analyses, where the estimated effect sizes and significances in models with either single or multiple SNPs were of similar magnitude (Table S6 and S7). GCTA-analysis highlighted the lead-SNP of each locus as being causal or being conditionally independent, except that rs2048 was the lead-SNP used in the GCTA-analysis for chromosome 4 since rs71640036 was not found from the LD reference data we used.

Heterogeneity of the GWAS meta-analysis was substantial in lead SNPs of the loci in chromosomes 3, 4, and 5, while lead SNPs in chromosomes 11 and 15 did not show heterogeneity. However, the leave-one-out GWAS meta-analyses showed that the result is not driven by a single cohort (Figure S3).

In post-hoc GWAS analyses conditioned on the lead SNPs, no genome-wide significant associations were identified, either at the single cohort or meta-analysis levels, (Figure S2, Table S1), suggesting there were no additional underlying independent associations in these or other loci.

Table S1. SNPs associated with a genome-wide significant level with endotoxemia. See Excel file.

Table S2. Leave-one-cohort-out GWAS meta-analysis, SNPs associated with genome-wide significant level with endotoxemia. See Excel file.

Table S3. Pathway enrichment analyses associating with the GWAS of endotoxemia. See Excel file.

Table S4. Expression quantitative trait locus (eQTL) analysis in 5,311 blood samples. See Excel file.

Table S5. Expression quantitative trait locus (eQTL) analysis in GTEx (release v8) data. See Excel file.

Table S6. The estimated effect sizes and significances of the lead SNPs associated with endotoxemia in single and multi SNP models in the FinnDiane T1D cohort.

SNP	Baseline model[1]		Single SNP model[2]					Multiple SNP model[3]				
	AIC	Adjusted R^2	beta	SE	p-value	AIC	Adjusted R^2	beta	SE	p-value	AIC	Adjusted R^2
rs5030082			0.034	0.0080	1.70E-05	2527	0.510	0.033	0.0079	2.55E-05		
rs71640036			0.046	0.0078	3.62E-09	2510	0.512	0.045	0.0077	4.93E-09		
rs1801020	2543	0.508	0.065	0.0088	1.19E-13	2490	0.515	0.063	0.0087	3.84E-13	2429	0.523
rs2081361			-0.026	0.0088	0.0029	2536	0.509	-0.028	0.0087	0.0012		
rs10152355			0.025	0.0087	0.0046	2537	0.509	0.023	0.0085	0.0059		

[1]log(LPS) ~ age + sex + BMI + Cholesterol + HDL-cholesterol + log(Triglycerides) + sample freeze time + Genotyping batch
[2]log(LPS) ~ baseline model + one of the lead SNPs
[3]log(LPS) ~ baseline model + rs5030082 + rs71640036 + rs1801020 + rs2081361 + rs10152355
Model fitted using N=3,812 participants with complete covariate data from the FD-T1D cohort

Table S7. The estimated effect sizes and significances of the lead SNPs associated with endotoxemia in single and multi SNP models in the FINRISK-97 cohort.

SNP	Baseline model[1]		Single SNP model[2]					Multiple SNP model[3]				
	AIC	Adjusted R^2	beta	SE	p-value	AIC	Adjusted R^2	beta	SE	p-value	AIC	Adjusted R^2
rs5030082			0.071	0.0085	5.12E-17	6665	0.288	0.076	0.0080	3.37E-21		
rs71640036			0.159	0.0081	1.05E-82	6363	0.325	0.156	0.0079	6.27E-85		
rs1801020	6734	0.279	0.157	0.0095	1.73E-60	6466	0.313	0.153	0.0091	1.66E-62	5974	0.371
rs2081361			-0.045	0.0091	8.94E-07	6712	0.282	-0.035	0.0085	4.35E-05		
rs10152355			0.038	0.0086	7.74E-06	6716	0.282	0.031	0.0080	1.07E-04		

[1]log(LPS) ~ age + sex + BMI + Cholesterol + HDL-cholesterol + log(Triglycerides) + recruiting area + Genotyping batch
[2]log(LPS) ~ baseline model + one of the lead SNPs
[3]log(LPS) ~ baseline model + rs5030082 + rs71640036 + rs1801020 + rs2081361 + rs10152355
Model fitted using N=5,588 participants with complete data from the FR97 cohort

Table S8. Genetic risk score (LPS-GRS) is calculated from GWAS-significant SNP genotypes, weighted with GWAS effect sizes. Association analysis was done using logistic regression model. The model was adjusted for sex, age, and 10 genetic principal components.

End-point (related ICCD-10 diagnosis	BETA	SE	P-value	OR	BETA_L	BETA_U	N_controls	N_cases	FDR	Control exclusion criteria
Cerebrovascular diseases* (I60, I61, I63, I64, I66, I67.1-I67.7, I68.1, I68.2, excl. I63.6)	0.0166	0.0100	0.097	1.017	-0.00301	0.0362	183921	11249	0.303	-
Diseases of arteries, arterioles and capillaries (I70-I72, I74)	0.0150	0.0117	0.199	1.015	-0.00791	0.0379	187010	8160	0.398	-
Atrial fibrillation and flutter* (I48)	0.0124	0.0100	0.215	1.013	-0.0072	0.0320	107200	18777	0.398	Cardiovascular diseases (excluding rheumatic etc)
Endovascular/surgical operations to intracerebral aneurysms*	0.0131	0.0422	0.757	1.013	-0.0696	0.0957	182218	568	0.841	Cerebrovascular diseases
Cerebral aneurysm, nonruptured (I67.1)	-0.0188	0.0336	0.575	0.981	-0.0846	0.0470	182218	892	0.725	Cerebrovascular diseases
Angina pectoris (I20)	-0.00296	0.00918	0.747	0.997	-0.0210	0.0150	168997	15621	0.841	Cerebrovascular diseases
Atherosclerosis, excluding cerebral, coronary and PAD (I70)	0.0174	0.0140	0.215	1.018	-0.0101	0.0449	184632	5537	0.398	Diseases of arteries, arterioles and capillaries
Cerebral atherosclerosis (67.2)	0.119	0.111	0.284	1.127	-0.0991	0.3380	182218	83	0.490	Cerebrovascular diseases
Cerebral atherosclerosis (67.2)	0.116	0.111	0.296	1.123	-0.102	0.3345	195087	83	0.493	-
Cerebrovascular diseases (I60-I69)	0.0183	0.00942	0.0522	1.018	-0.00018	0.0367	182218	12952	0.218	-
Major coronary heart disease event (I20-I22)	-0.00302	0.00859	0.725	0.997	-0.0199	0.0138	177483	17687	0.841	-
Major coronary heart disease event excluding revascularizations (I20-I22)	0.00558	0.00838	0.506	1.006	-0.0108	0.0220	168997	19800	0.722	Major coronary heart disease event
Dissection of cerebral arteries, nonruptured (I67.0)	-0.003004	0.108	0.978	0.997	-0.214	0.208	182218	86	0.978	Cerebrovascular diseases
Diseases of arteries, arterioles and capillaries (I7)	0.0159	0.0103	0.122	1.016	-0.00425	0.0361	184632	10538	0.321	-
DVT of lower extremities and pulmonary embolism (I80.2- I80.3, I26)	0.0810	0.0125	9.14E-11	1.084	0.0565	0.106	188301	6869	2.28E-09	-
Intracerebral haemmorrhage (I61)	-0.00269	0.0279	0.923	0.997	-0.0574	0.0520	180528	1302	0.962	Ischaemic Stroke (excluding all haemorrhages), TIA
Ischaemic heart disease, wide definition (I20-I25	0.00274	0.00741	0.711	1.003	-0.0118	0.0173	168469	26701	0.841	-

Nontraumatic intracranial haemmorrhage I60-I61)	-0.00143	0.0215	0.947	0.999	-0.0435	0.0407	182218	2212	0.966	Cerebrovascular diseases
Ischemic heart diseases (I20-I25)	0.00438	0.00744	0.556	1.004	-0.0102	0.0190	168997	26173	0.725	-
Death due to cardiac causes	0.0231	0.0144	0.108	1.023	-0.00508	0.0513	189855	5315	0.306	-
Myocardial infarction (I21-I22)	0.00970	0.0107	0.363	1.010	-0.0112	0.0306	168997	10813	0.567	Ischemic heart diseases
Complications following myocardial infarction (I23)	-0.187	0.106	0.0762	0.829	-0.394	0.0197	168997	86	0.272	-
Myocardial infarction, strict (I21-I22, excl. I21.90, I21.97, I22.90, I2.97)	0.00690	0.0111	0.535	1.007	-0.0149	0.0287	168997	9789	0.723	Ischemic heart diseases
Myocardial infarction, unclassifiable (I21.9, I22)	0.00989	0.0179	0.580	1.010	-0.0252	0.0449	191871	3299	0.725	-
Myocardial infarction, without ST-evelation (I21.4)	0.0254	0.0143	0.0755	1.026	-0.00262	0.0535	189853	5317	0.272	-
Other intracranial haemorrhages (I62)	0.138	0.0609	0.0239	1.147	0.0182	0.2569	182218	280	0.118	Cerebrovascular diseases
Other peripheral vascular diseases (I73)	0.0796	0.0335	0.0177	1.083	0.0138	0.1453	184632	914	0.0981	Diseases of arteries, arterioles and capillaries
Peripheral artery disease (E10.5, E11.5, E12.5, E13.5, E14.5, I70.2, I73.9)	0.0181	0.0134	0.177	1.018	-0.00822	0.0445	184632	6001	0.386	Diseases of arteries, arterioles and capillaries
Peripheral artery operations in Hilmo *	0.0328	0.0272	0.227	1.033	-0.0205	0.0861	184632	1387	0.406	Diseases of arteries, arterioles and capillaries
DVT of lower extremities (I80.2-I80.3)	0.0792	0.0164	1.35E-06	1.082	0.0471	0.1113	170316	3930	1.35E-05	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified
Status post-ami (I25.3)	-0.0281	0.0304	0.354	0.972	-0.0876	0.0314	168997	1145	0.567	Ischemic heart diseases
Pulmonary embolism (I26)	0.0828	0.0171	1.21E-06	1.086	0.0493	0.116	191560	3610	1.35E-05	-
Pulmonary heart disease, diseases of pulmonary circulation (I26-I28)	0.0776	0.0164	2.14E-06	1.081	0.0455	0.110	191249	3921	1.78E-05	-
Other pulmonary heart/vessel disease (I27, I28)	0.0310	0.0495	0.532	1.031	-0.0661	0.128	191249	413	0.723	Pulmonary heart disease, diseases of pulmonary circulation
Subarachnoid haemmorrhage (I60)	-0.00527	0.0302	0.862	0.995	-0.0645	0.0540	180547	1105	0.917	Ischaemic Stroke (excluding all haemmorrhages), TIA
Aneurysms, operations, SAH (I67.1, I60)*	-0.00613	0.0236	0.795	0.994	-0.0524	0.0401	182218	1818	0.864	Cerebrovascular diseases
Myocardial infarction, with ST-evelation (I21.1-I21.3)	0.0104	0.0203	0.607	1.010	-0.0293	0.0502	192647	2523	0.740	-

	BETA	SE		OR	BETA_L95	BETA_U95	N	N_NA	FDR	
Occlusion and stenosis of arteries, not leading to stroke (I66)	0.139	0.0826	0.0930	1.149	-0.0231	0.301	182218	151	0.303	Cerebrovascular diseases
Stroke, excluding SAH (I61, I63, I64, excl. I63.6)	0.0137	0.0108	0.205	1.014	-0.00752	0.0350	180531	9576	0.398	Ischaemic Stroke (excluding all haemorrhages), TIA
Ischaemic Stroke, excluding all haemorrhages (I63, I64, excl. I63.6)	0.0154	0.0114	0.175	1.016	-0.00685	0.0377	181368	8653	0.386	Ischaemic Stroke (excluding all haemorrhages), TIA
Stroke, including SAH (I60, I61, I63, I64)	0.0151	0.0105	0.149	1.015	-0.00541	0.0356	179847	10313	0.355	Ischaemic Stroke (excluding all haemorrhages), TIA
Other embolism and thrombosis (I82)	0.0606	0.0247	0.0143	1.062	0.0121	0.109	170316	1680	0.089	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified
Transient ischemic attack (G45, excl. G45.4)	0.0193	0.0121	0.110	1.020	-0.00439	0.0431	181368	7480	0.306	Ischaemic Stroke (excluding all haemorrhages), TIA
Unstable angina pectoris (I20.0)	-0.0106	0.0135	0.430	0.989	-0.0371	0.0158	177348	6125	0.633	Major coronary heart disease event
Venous thromboembolism (I26, I80, O87.1, O88.2, excl. I80.0)	0.0778	0.0117	2.73E-11	1.081	0.0549	0.101	187253	7917	1.36E-09	-
Embolic stroke with flimmer (I63.1, with I48)	0.0280	0.0194	0.147	1.028	-0.00989	0.0660	179847	2891	0.355	Cerebrovascular diseases

BETA, beta-value for logistic regression model; SE, standard error; OR, odds ratio; BETA_L95, lower bound of 95%-confidence interval of beta; BETA_U95, upper bound of 95%-confidence interval of beta, N_NA, number of samples with phenotype information not available; FDR, false discovery rate; PAD, pheripheral artery disease; DVT, deep vein thrombosis; SAH, subarachnoid hemorrhage; PAD, diagnosis based on biopsy; TIA, Transient ischemic attack.

*Inclusion criteria contains other clinical data in addition to ICD-10 diagnosis codes, such as national codes for surgical operations or information about medications.

Table S9. Mendelian Randomization. The analysis was done using mr-base database, which utilized UK-biobank and Megastroke populations.

outcome	samplesize	SNP	beta	se	p-value
	440328	rs10152355	0.08982	0.130539	0.491407
	440328	rs1801020	0.000386	0.046368	0.993351
	440328	rs2048	0.127549	0.039984	0.001423
Ischemic stroke	440328	rs2081361	-0.04167	0.128472	0.745693
	440328	rs5030082	0.216283	0.083059	0.009215
	440328	All - Inverse variance weighted	0.084186	0.038261	0.027783
	440328	All - MR Egger	0.064933	0.100758	0.565173
	361194	rs10152355	0.004752	0.00236	0.044013
	361194	rs1801020	0.000929	0.000843	0.270067
Diagnoses - main ICD10:	361194	rs2048	0.00347	0.000755	4.36E-06
I26 Pulmonary	361194	rs2081361	-0.00235	0.002446	0.33724
embolism	361194	rs5030082	0.001847	0.001584	0.243784
	361194	All - Inverse variance weighted	0.002198	0.000791	0.005446
	361194	All - MR Egger	0.002723	0.002065	0.2788
	361194	rs10152355	0.004553	0.003192	0.153794
	361194	rs1801020	0.000121	0.00114	0.915801
DVT of lower	361194	rs2048	0.006111	0.001022	2.23E-09
extremities and	361194	rs2081361	-0.0033	0.003308	0.318297
pulmonary embolism	361194	rs5030082	0.00475	0.002143	0.026673
	361194	All - Inverse variance weighted	0.003338	0.001528	0.028928
	361194	All - MR Egger	0.004127	0.00401	0.379034
	361194	rs10152355	0.000264	0.002242	0.906254
	361194	rs1801020	-0.00049	0.000801	0.543722
DVT of lower	361194	rs2048	0.003311	0.000718	3.99E-06
extremities	361194	rs2081361	-0.00174	0.002324	0.453454
	361194	rs5030082	0.003179	0.001506	0.034759
	361194	All - Inverse variance weighted	0.001572	0.000961	0.10196
	361194	All - MR Egger	0.00221	0.002509	0.443217
	361194	rs10152355	0.004868	0.0033	0.140155
	361194	rs1801020	5.75E-05	0.001179	0.961077
	361194	rs2048	0.006462	0.001056	9.52E-10
Venous	361194	rs2081361	-0.0035	0.00342	0.306437
thromboembolism	361194	rs5030082	0.005677	0.002216	0.0104
	361194	All - Inverse variance weighted	0.003573	0.001648	0.030169
	361194	All - MR Egger	0.004194	0.004342	0.405296

se, standard error; DVT deep vein thrombosis

Table S10. Phenome-wide association study of lead SNPs in UK-Biobank population. See Excel file.

Table S11. Lead-SNP associations in Cerebrovascular Disease Knowledge Portal -material. The result is summarized using bottom line analysis.

Variant	Phenotype	P-value	Odds Ratio	CI_hi	CI_low	Sample size
rs5030082	Any stroke	2.95e-03	2.795	5.50689	1.419068	442,255
rs5030082	1 All ischemic stroke	4.89e-03	1.031	1.083287	0.982161	525,705
rs5030082	TOAST small artery occlusion	3.58e-02	2.858	7.614086	1.072508	252,646
rs71640036	TOAST cardio-aortic embolism	9.41e-05	2.939	5.04804	1.710867	322,150
rs71640036	2 All ischemic stroke	2.38e-03	2.809	5.468476	1.441955	424,295
rs1801020	TOAST other undetermined	3.00e-02	1.313	1.616074	1.067159	10,318
rs1801020	Modified Rankin scale score 0-2 vs 3-6	4.14e-02	1.113	1.233678	1.004008	5,802
rs2081361	Intracranial aneurysm	3.13e-02	1.043	1.086542	1.001001	30,745

CI_hi, upper limit of 95% confidence interval; CI_low, lower limit of 95% confidence interval; TOAST, classification of five subtypes of ischemic stroke

Figure S1. Cohort-wise Manhattan and QQ-plots of GWAS results for endotoxemia. The GWAS was performed in five cohorts, FinnDiane T1D (n=3,940), FinnDiane rest (n=302), FINRISK-92 (n=656), FINRISK-97 (n=5,667), and Twin studies TF16 (n=451) and VpEpi (n=280).The horizontal red and blue line represents genome-wide ($p<5\times10^{-8}$) and suggestive significance level ($p<1\times10^{-5}$), respectively. Inflation of the p-values is presented in the QQ-plots (λ_{qc} values).
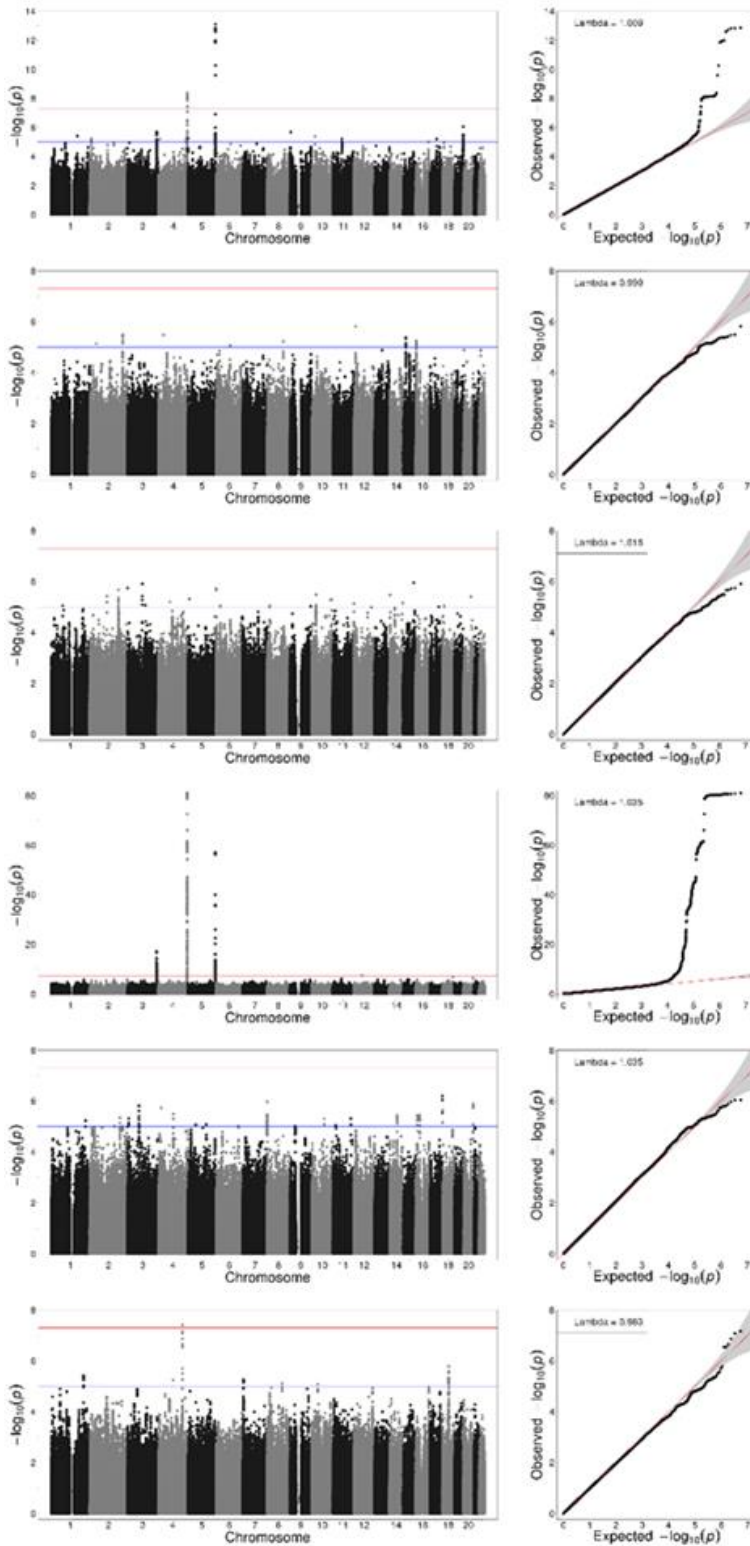

Figure S1. Cohort-wise Manhattan and QQ-plots of GWAS results for endotoxemia.

Figure S2: QQ-plot of GWAS meta-analysis results after exclusion of genome-wide significant SNPs (p-value < 5e-8). The QQ-plot demonstrated good adherence to the diagonal for the non-signal SNPs.
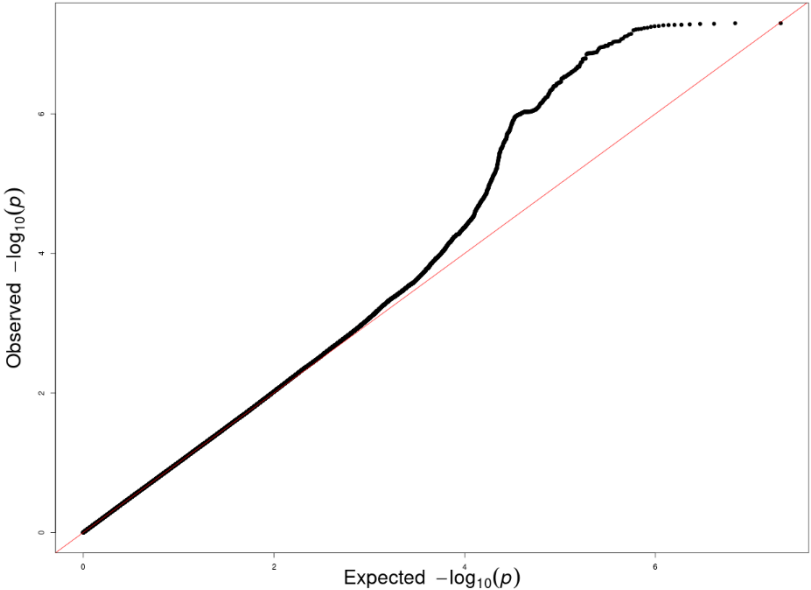
Figure S3. *Post-hoc* **GWAS analysis for endotoxemia conditioned on the lead SNPs.** Cohort-wise Manhattan and QQ-plots are presented. The GWAS was performed in five cohorts, FinnDiane T1D (n=3,940), FinnDiane rest (n=302), FINRISK-92 (n=656), FINRISK-97 (n=5,667), and Twin studies TF16 (n=451) and VpEpi (n=280).The horizontal red and blue line represents genome-wide (p<5×10^{-8}) and suggestive significance level (p<1×10^{-5}), respectively. Inflation of the p-values is presented in the QQ-plots (λ_{qc} values).
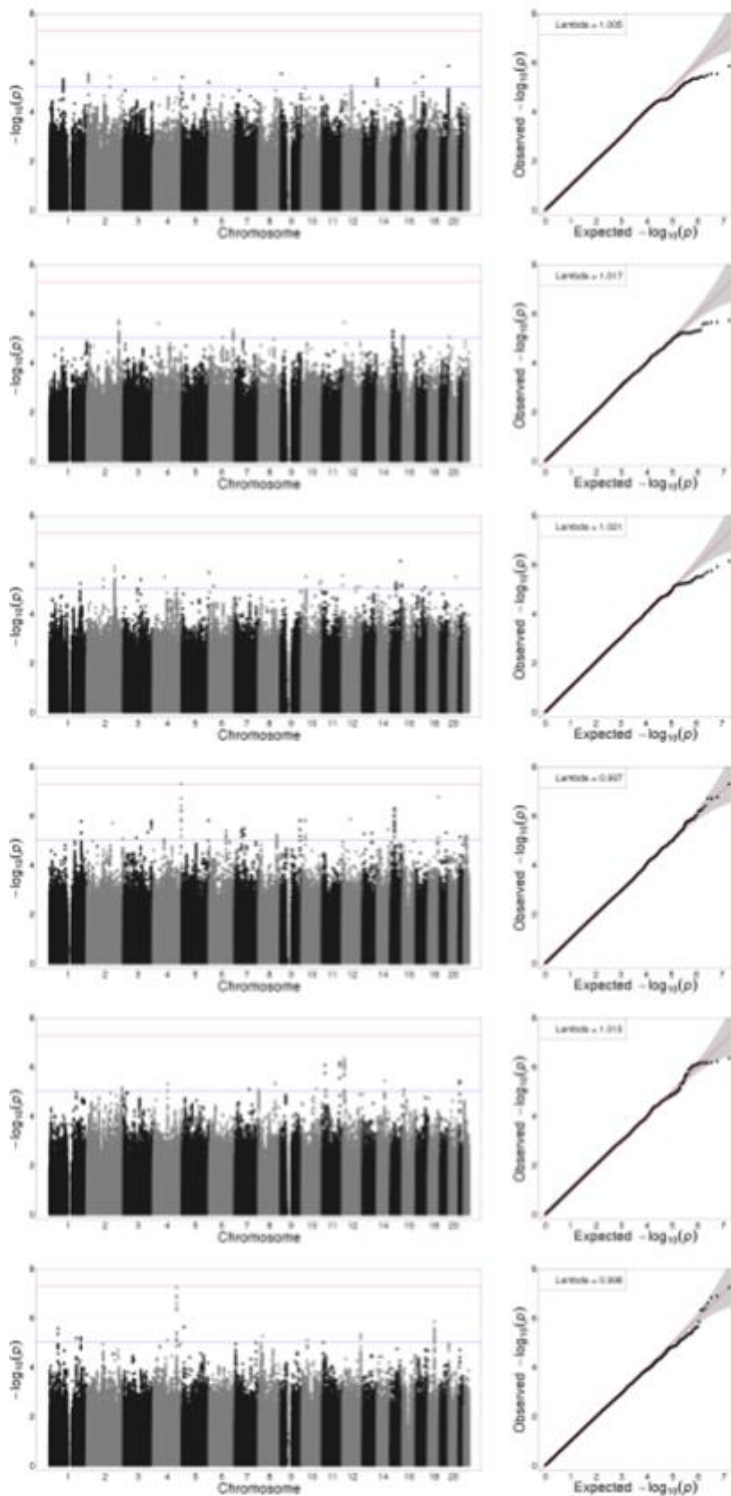
Figure S4. Leave-one-cohort-out GWAS meta-analysis for endotoxemia. GWAS meta-analysis was done leaving one cohort out at a time in following order: FinnDiane rest (n=302), FinnDiane T1D (n=3,940), FINRISK-92 (n=656), FINRISK-97 (n=5,667), and Twin studies TF16 (n=451) and VpEpi (n=280). The horizontal red and blue line represents genome-wide (p<5×10^{-8}) and suggestive significance level (p<1×10^{-5}), respectively. Inflation of the p-values is presented in the QQ-plots (λ$_{qc}$ values).
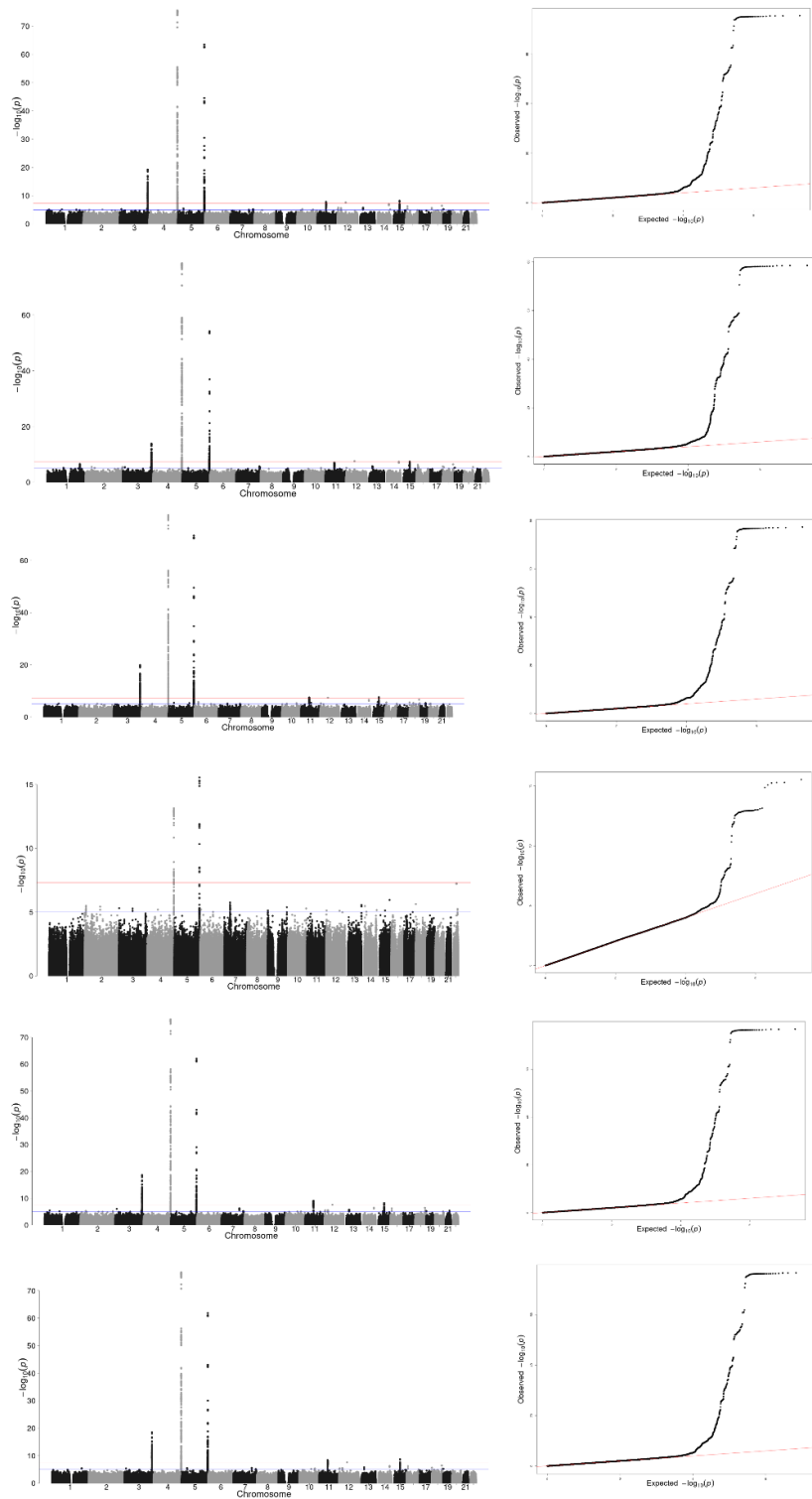
Figure S5. Regional plots of the five loci identified in GWAS of combined in fixed-effects meta-analysis. In the meta-analysis of 11,296 individuals with Finnish ancestry, a total of 740 markers at five independent loci in chromosomes 3 (A), 4 (B), 5 (C), 11 (D), and 15 (E) associated with endotoxemia. The regional plots of these loci are presented.
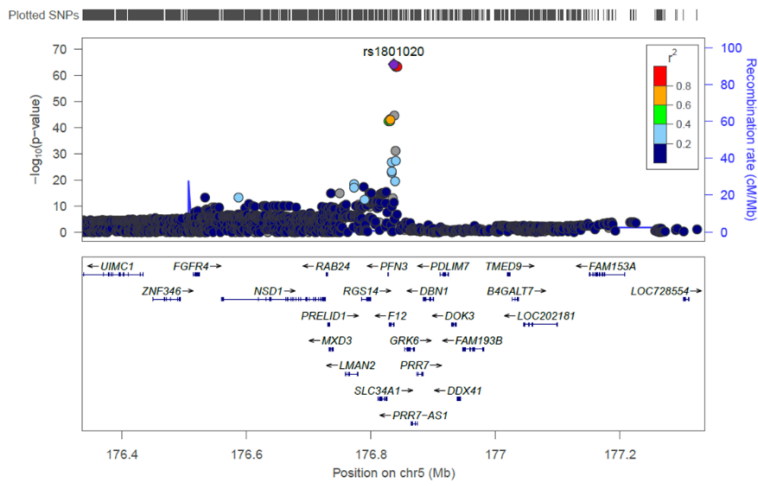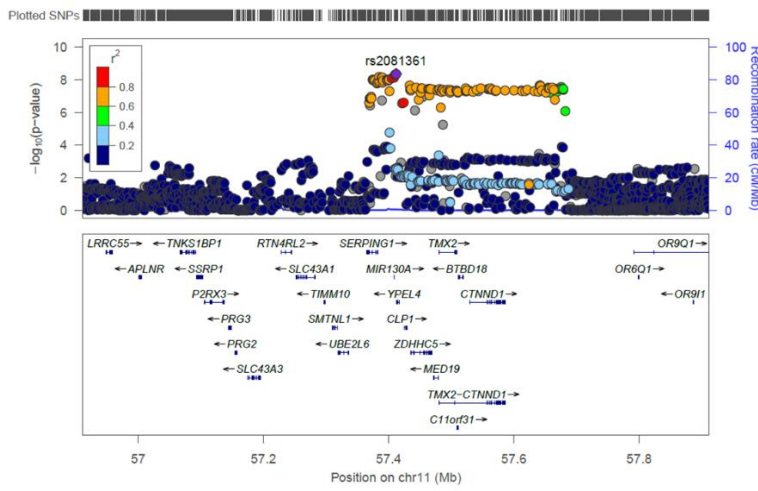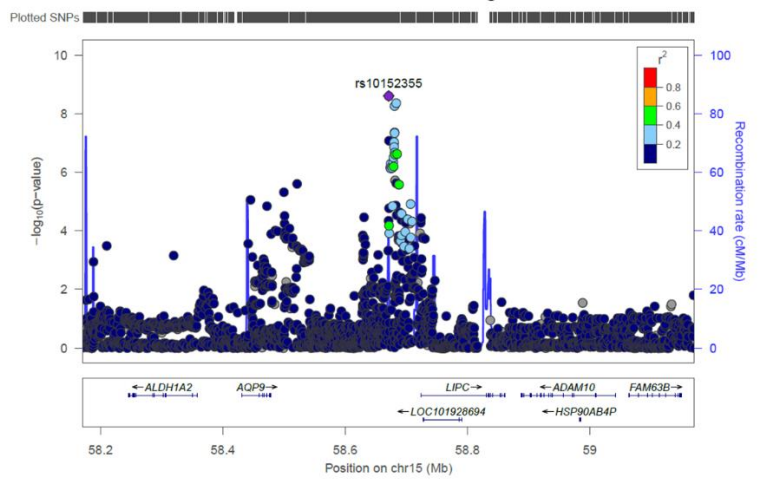
A

B

C

Plotted SNPs

rs1801020

D

Plotted SNPs

rs2081361

E

Plotted SNPs

rs10152355
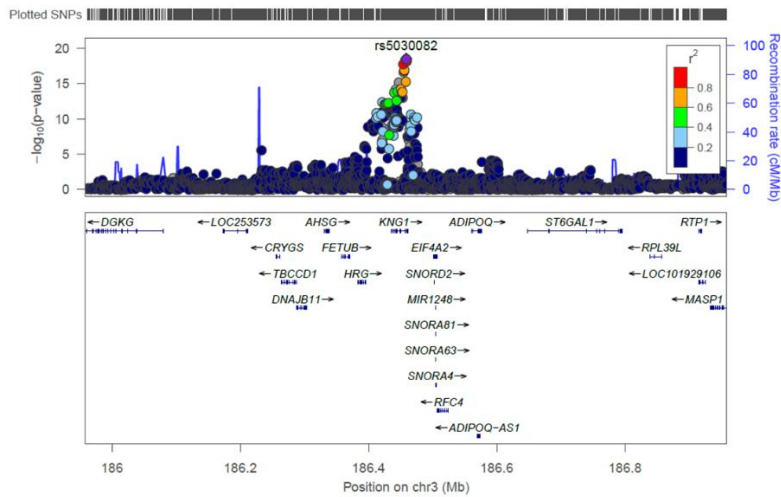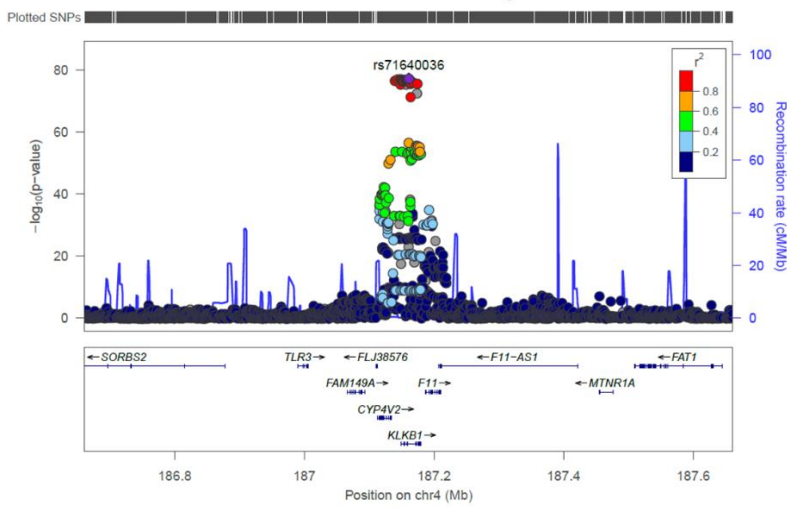
Figure S6. Correlation of LPS measured by different techniques. LPS was measured in serum samples in two subpopulations of FinnDiane population. LPS was determined by using the LAL assay (all samples), the mass spectrometry -based technology (n=363), and Endolisa-assay (n=326). Scatterplots presenting results obtained with different methods are presented. The correlation coefficients (cor) are shown above. The box plots show the mean LPS results obtained with different methods for different genotypes of the lead SNP, rs1801020. The p-values for the comparisons between the genotypes are shown above. The center line represents the mean, the box limits the standard deviation, and the whiskers show the 95% confidence intervals. The results obtained by the LAL and Endolisa assay were logarithmically transformed (log).
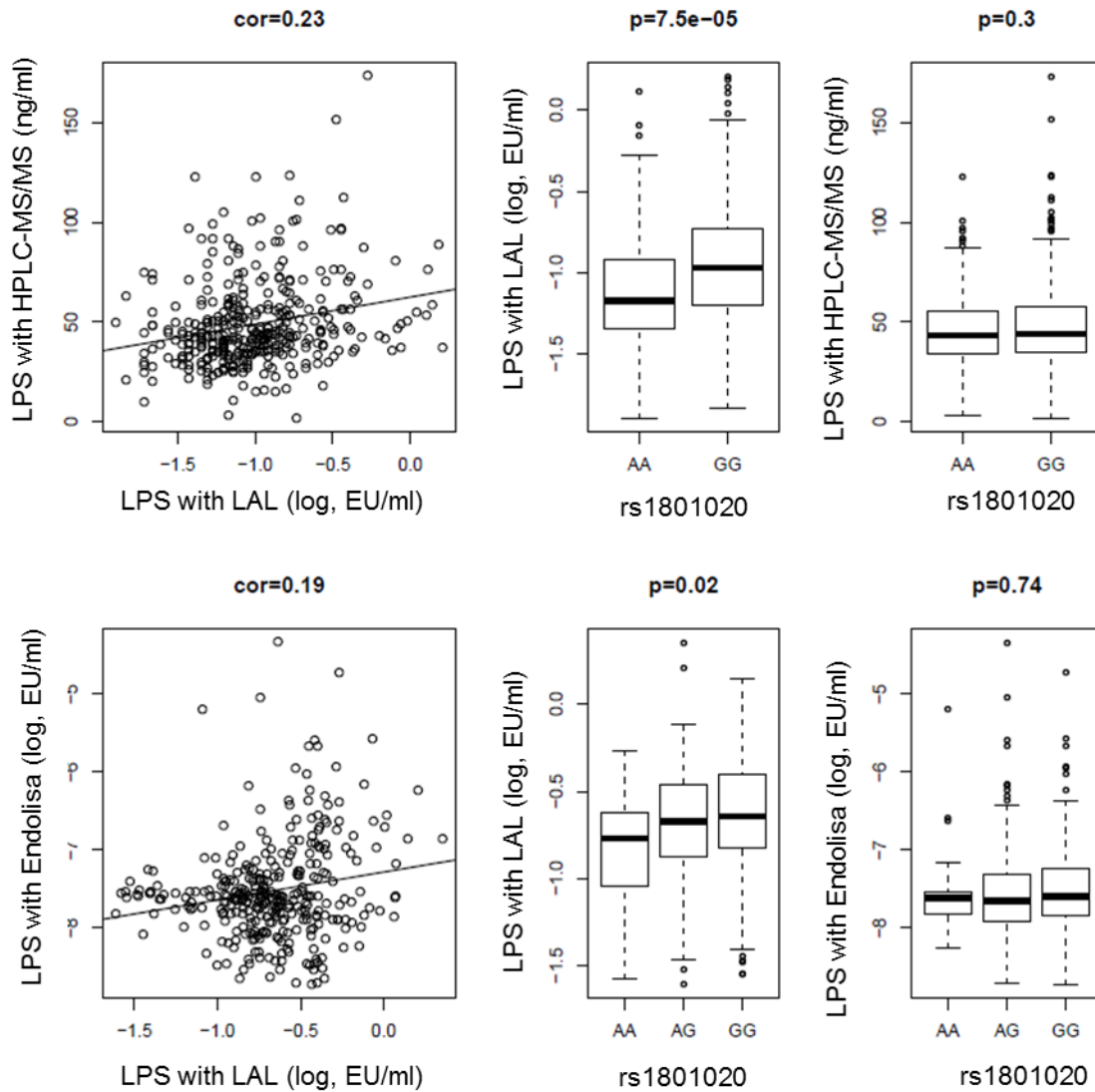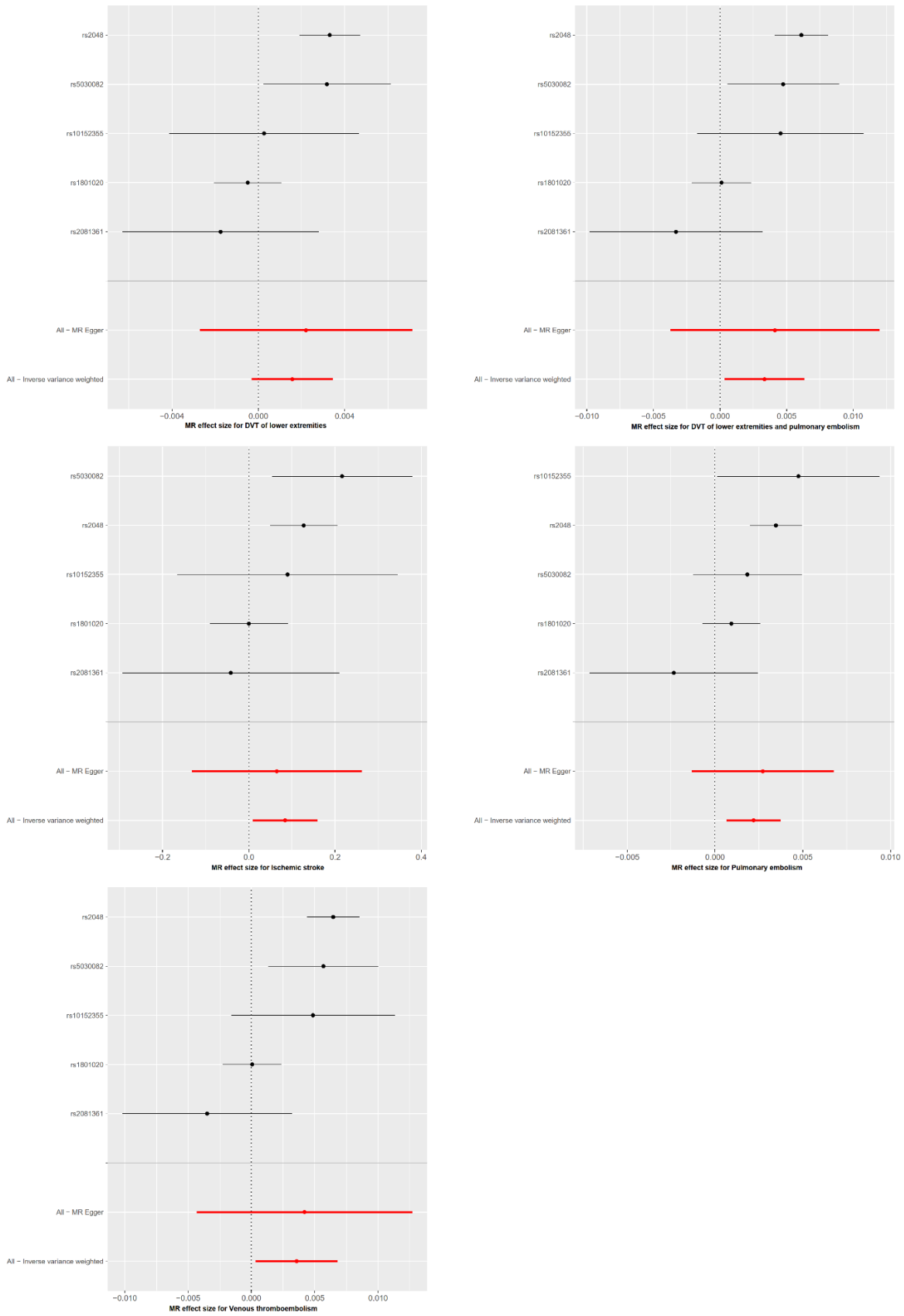
Figure S7. Forest plots for mendelian randomization of endotoxemia on various disease endpoints.

MR, mendelian randomization.