Genetic architecture of human plasma lipidome and its link to cardiovascular disease

Rubina Tabassum et al.

Understanding genetic architecture of plasma lipidome could provide better insights into lipid metabolism and its link to cardiovascular diseases (CVDs). Here, we perform genome-wide association analyses of 141 lipid species (n = 2,181 individuals), followed by phenome-wide scans with 25 CVD related phenotypes (n = 511,700 individuals). We identify 35 lipid-species-associated loci (P < 5 × 10^{-8}), 10 of which associate with CVD risk including five new loci—COL5A1, GLTPD2, SPTLC3, MBOAT7 and GALNT16 (false discovery rate<0.05). We identify loci for lipid species that are shown to predict CVD e.g., SPTLC3 for CER(d18:1/24:1). We show that lipoprotein lipase (LPL) may more efficiently hydrolyze medium length triacylglycerides (TAGs) than others. Polyunsaturated lipids have highest heritability and genetic correlations, suggesting considerable genetic regulation at fatty acids levels. We find low genetic correlations between traditional lipids and lipid species. Our results show that lipidomic profiles capture information beyond traditional lipids and identify genetic variants modifying lipid levels and risk of CVD.

Correspondence and requests for materials should be addressed to S.R. (email: samuli.ripatti@helsinki.fi). A full list of authors and their affiliations appears at the end of the paper.
Cardiovascular diseases (CVDs) encompass many pathologic conditions of impaired heart function, vascular structure and circulatory system. CVDs are the leading cause of mortality and morbidity worldwide, necessitating the need for better preventive and predictive strategies. Plasma lipids, the well-established heritable risk factors for CVDs, are routinely monitored to assess CVD risk. Standard lipid profiling measures traditional lipids (referred to LDL-C, HDL-C, total triglycerides and total cholesterol), but does not capture the functionally and chemically diverse molecular components—the lipid species. These molecular lipid species may independently and specifically affect different manifestations of CVD, such as ischemic heart disease and stroke. Lipid species including cholesterol esters (CEs), lyso-phosphatidylincholines (LPCs), phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), ceramides (CERs), sphingomyelins (SMs) and triacylglycerols (TAGs) potentially improve CVD risk assessment over traditional lipids.

Understanding of the genetic architecture and genetic regulation of these lipid species could help guide tool development for CVD risk prediction and treatment. Genetic studies of traditional lipids have identified over 250 genomic loci and improved our understanding of CVD pathophysiology. For the majority of the lipid loci, however, their effects on detailed lipidome beyond traditional lipids are unknown. Only a few studies have reported genetic associations for lipid species either through studies on subsets of the lipidome or GWASs on metabolomes. In light of the limited information about the genetics of lipid species and their relationship with CVDs, we carried out a GWAS of lipidomic profiles of 2181 individuals using ~9.3 million genetic markers followed by PhEWas including 25 CVD-related phenotypes in up to 511,700 individuals (Fig. 1). We aimed to (1) determine heritability of lipid species and their genetic correlations; (2) identify genetic variants influencing the plasma levels of lipid species; (3) test the relationship between identified lipid–species–associated variants and CVD manifestations; and (4) gain mechanistic insights into established lipid variants. We find that lipid species are heritable, suggesting a considerable role of endogenous regulation in lipid metabolism. We report association of new genomic loci with lipid species and CVD risk in humans. In addition to enhancing the current understanding of genetic regulation of circulating lipids, our study emphasises the need of lipidomic profiling in identifying additional variants influencing lipid metabolism.

**Results**

**Heritability of lipid species.** First, we determined SNP-based heritability for each of the lipid species and traditional lipids using genetic relationship matrix for all the study participants. The demographic characteristics of the study participants are provided in Supplementary Table 1. SNP-based heritability estimates ranged from 0.10 to 0.54 (Fig. 2a; Supplementary Table 2), showing considerable variation across lipid classes (Fig. 2b), with similar trends as reported previously. CERs showed the greatest estimated heritability (median = 0.39, range = 0.35–0.40), whereas phosphatidylinositosols (PIs) showed the least heritability (median = 0.19, range = 0.11–0.31). Sphingolipids had higher heritability than glycerolipids ranging from 0.24 to 0.41 (Fig. 2b), which is similar to a previous study that reported higher heritability for sphingolipids ranging from 0.28 to 0.53 estimated based on pedigrees. Lipids containing polyunsaturated fatty acids, particularly C20:4, C20:5 and C22:6, had significantly higher heritability compared with other lipid species (Fig. 2c). For instance, PC (17:0;0–18:2;0) had the lowest heritability (median = 0.50), whereas PC (16:0;0–16:1;0) and LPC (22:6;0) had the highest heritability (> 0.50), whereas PC (16:0;0–16:1;0) and PI (16:0;0–18:2;0) had the lowest heritability estimates (< 0.12) (Supplementary Table 2).

**Genetic correlations between lipid species.** Longer, polyunsaturated lipids (those with four or more double bonds) had stronger genetic correlations with each other than with other lipid species (Supplementary Fig. 1, Supplementary Data 1). This can be seen in the hierarchical clustering based on genetic correlations that segregate TAG subspecies into two clusters based on carbon content and degree of unsaturation (Fig. 2d). These patterns were not seen in phenotypic correlations that were estimated based on plasma levels of lipid species (Supplementary Fig. 2).

We observed low phenotypic and genetic correlation between traditional lipids and molecular lipid species, except strong positive genetic correlations of triglycerides with TAGs and DAGs (average $r = 0.88$) (Fig. 3). However, triglycerides had low genetic correlation with other lipid species (average (abs) $r = 0.26$). HDL-C and LDL-C levels had low genetic and phenotypic correlations with other lipid species (average (abs) $r = 0.35$) and imputation. The Finnish population

**Fig. 1** Study design and work flow. The figure illustrates the study design and key findings of the study.
Fig. 2 Heritability of lipidomic profiles and genetic correlations among the lipid species. **a** Histogram and kernel density curve showing the distribution of heritability estimates across all the lipid species. **b** Boxplot showing the heritability estimates in each lipid class. **c** Boxplot showing comparison of the median heritability estimates of lipid species containing C20:4, C20:5 and C22:6 acyl chains and all others. The P-values were calculated using the Wilcoxon rank-sum test. **d** Hierarchical clustering of lipid species based on genetic correlations among lipid species. Lipids containing polyunsaturated fatty acids C20:5, C20:4 and C22:6 are highlighted with black bars. The data presented in the boxplots represent the interquartile range (IQR) defined by the bounds of the box with the median (middle line of the box) and whiskers extending to the largest/smallest values no further than 1.5 times the IQR. CER ceramide, DAG diacylglyceride, LPC lysophosphatidylcholine, LPE lysophosphatidyethanolamine, PC phosphatidylcholine, PCE phosphatidylcholine-ether, PE phosphatidylethanolamine, PEO phosphatidylethanolamine-ether, PI phosphatidylinositol, CE cholesteryl ester, SM sphingomyelin, ST sterol, TAG triacylglycerol, Trad traditional lipids

Fig. 3 Lipidomic profiles capture information beyond traditional lipids. The genetic and phenotypic correlations between traditional lipids and molecular lipid species are shown in lower panel. The bar plot in the upper panel shows the heritability estimates of each lipid species (red bars) and the variance explained by all the known loci together (green bars). The lipid species are ordered based on the hierarchical clustering showing the correlations between the lipid species and traditional lipids. TC total cholesterol, TG triglycerides
correlations with most of the lipid species (Fig. 3; Supplementary Data 1). Consistently, all of the known lipid variants explained 2–21% of variances in plasma levels of various lipid species, with the least variance accounting for LPGs (Fig. 3). To rule out the possibility that lipid-lowering medications resulted in the observed low genetic correlations between traditional lipids and lipid species, we also calculated the genetic correlations after re-analysis provided the similar results as the primary study (Supplementary Fig. 3). It is to be noted that this sample size might not provide sufficient power for heritability estimations in unrelated samples. Our study also included the family samples which provides higher statistical power in heritability estimation than unrelated samples.

**Lipid species associated variants.** Next, we performed the genome-wide association analyses for 141 lipid species with ~9.3 million genetic markers. We identified 2817 associations between 518 variants located within 11 genomic loci (1MB blocks) and 42 lipid species from 10 lipid classes at study-wide significance ($P < 5 \times 10^{-8}$ accounting for 34 principal components that explain 90% of the variance in lipidome) (Table 1; Supplementary Data 2, 3). These included three new loci (ROCK1, MAF and SYT1) that are not previously reported for any lipid measure or related metabolite (Fig. 4). Among the new loci, the strongest association was an intronic variant rs151223356 near ROCK1 with short acyl-chain LPC(14:0) ($P = 1.9 \times 10^{-10}$). ROCK1 encodes for a serine/threonine kinase that plays key role in glucose metabolism23. In line with our observation of higher heritability for lipids with C20:4, C20:5 or C22:6 acyl chains, we detected associations for 15 out of 21 lipids with these acyl chains.

We also replicated the previous associations of FADS2, SYNE2, LIPC, CERS4 and MBOAT7 with the same lipid species13–20. The previously reported associations at the known loci identified in previous metabolomics GWASs are provided in Supplementary Data 4. This information was obtained from the database-SNIPA (http://www.snpedia.org) using block annotation and PhenoScanner v2 (http://www.phenoscanner.medschl.cam.ac.uk/), and were manually curated to include associations from literature search. In addition, we also identified new locus–lipid species associations at previously reported lipid loci including new associations of variants at ABCG5/8 with CE (20:20) ($P = 3.9 \times 10^{-10}$), MBOAT7 with PI (18:0;0–20:3) ($P = 3.0 \times 10^{-12}$) and GLTPD2 with SM (34:0:2) ($P = 3.4 \times 10^{-22}$) (Supplementary Data 2, 3).

Further, we systematically evaluated the associations of variants previously identified in metabolomics GWAS (126 variants from 46 loci available in our data set out of 132 reported) with 141 lipid species. Of these known variants, 76 variants from 12 loci showed previously reported lipid loci including new associations of variants at ABCG5/8 with CE (20:20) ($P = 3.9 \times 10^{-10}$), MBOAT7 with PI (18:0;0–20:3) ($P = 3.0 \times 10^{-12}$) and GLTPD2 with SM (34:0:2) ($P = 3.4 \times 10^{-22}$) (Supplementary Data 2, 3).
were located in genomic regions not previously reported for any lipid measure or related metabolite, and 8 loci were located near known loci for lipids but were independent of any previously reported variant (Table 1; Supplementary Data 3). The regional association plots for all 35 loci with $P < 5.0 \times 10^{-8}$ are presented in Supplementary Data 7, and the genotype–phenotype relationships for the lead variants in these 35 loci are provided in Supplementary Fig. 4.

**Relationship between identified variants and risk of CVD.** As many of the lipid species have previously been shown to predict CVD risk, we determined if the variants associated with lipid species affect individuals’ susceptibility to CVD-related phenotypes in FinnGen and UK Biobank cohorts. We identified 25 CVD-related phenotypes from the clinical outcomes derived from health registry data in the FinnGen and UK Biobanks (Supplementary Table 3). The follow-up PheWAS analyses included lead variants from all of the 35 independent loci that showed associations with $P < 5.0 \times 10^{-8}$ (Table 1). Overall, 10 of the 35 lipid–species variants (APOA5, ABCG5/8, BLK, LPL, FADS2, COL5A1, GALNT16, GLTPD2, MBOAT7 and SPTLC3) were associated with at least one of the CVD outcomes (FDR < 5%) (Fig. 5; Supplementary Data 8). These included novel associations of variants at COL5A1 with cerebrovascular disease ($P = 4.6 \times 10^{-5}$), GALNT16 with angina ($P = 9.3 \times 10^{-6}$), MBOAT7 with venous thromboembolism ($P = 1.3 \times 10^{-3}$), GLTPD2 with atherosclerosis ($P = 5.3 \times 10^{-6}$) and SPTLC3 with intracerebral haemorrhage ($P = 1.0 \times 10^{-3}$) (Fig. 5). FADS1-2-3 is a well-known lipid modifying locus; however, like many other known lipid loci, its effects on CVD risk has been unclear. We found an association of FADS2 rs28456-G with peripheral artery disease ($P = 2.2 \times 10^{-4}$) and aterial embolism and thrombosis ($P = 2.5 \times 10^{-4}$). BLK (rs1478898-A) was also found to be associated with decreased risk of obesity (OR = 0.97, $P = 5.6 \times 10^{-6}$) and type 2 diabetes (OR = 0.96, $P = 4.5 \times 10^{-5}$).

Several studies have suggested a role for sphingolipids, including CERS and SMs, in the pathogenesis of CVDs. CER (d18:1/24:0) and CER (d18:1/24:1) have been reported to be associated with the increased risk of CVD events9. We found that the CER (d18:1/24:1) decreasing variant SPTLC3 rs364585-G was associated with decreased risk of intracerebral haemorrhage, while CER (d18:1/24:0) increasing variant ZNF385D rs13070110-C was nominally associated with increased risk of intracerebral haemorrhage. Furthermore, consistent with the observation
that elevated plasma SMs levels are atherogenic, we identified association of GLTPD2 rs79202680-T (associated with reduced levels of SMs) with reduced risk of atherosclerosis.

Mechanistic insights into lipid variants. Next, we determined if the detailed lipidomic profiles could provide new mechanistic insights into the role of known lipid variants in lipid biology. We present two examples of well-established lipid variants here. First is the fatty acid desaturase (FADS) gene cluster that has been consistently reported to be associated with omega-3 and omega-6 fatty acids levels with inverse effects on different PUFAs. Its mechanism, however, has not been fully deciphered. Here, we present two examples of well-established lipid variants here. First is the fatty acid desaturase (FADS) gene cluster that has been consistently reported to be associated with omega-3 and omega-6 fatty acids levels with inverse effects on different PUFAs. Its mechanism, however, has not been fully deciphered. Here, we found that FADS2 rs28456-G was associated with increased levels of lipids with a C20:3 acyl chain and decreased levels of lipids with C20:4, C20:5 and C22:6 acyl chains (Supplementary Fig. 5). These data together explain the inverse relationship of these data together explain the inverse relationship of FADS2 variants with lipids containing different polyunsaturated fatty acids (PUFAs) (Fig. 6).

Another example is lipoprotein lipase (LPL). LPL codes for lipoprotein lipase that is the master lipolytic factor of TAGs in TAG-enriched chylomicrons and VLDL particles. We found that LPL rs11570891-T was associated with reduced levels of medium length TAGs (C50–C56), with strongest associations with TAG (52:3:0). This suggested that LPL enzyme might have different efficiency in hydrolysis of TAGs of different length. We explored this possibility by evaluating (1) the effect of LPL rs11570891-T on LPL enzymatic activity and (2) the relationship between LPL activity and plasma levels of TAGs of different length, using post-heparin LPL measured in the EUFAM cohort. We found that LPL rs11570891-T (an eQTL increasing LPL expression) was associated with increased LPL activity, which in turn was associated with TAG species with stronger effect on medium length TAGs than other TAGs (Fig. 6). Consistent with a previous report by Rhee et al., variant rs964184-C at APOA5, which codes for the activator that stimulates LPL-mediated lipolysis of TAG-rich lipoproteins and their remnants, also showed association with medium length TAGs (Fig. 6). These results provide first clues to the probable variable role of LPL and APOA5 in the hydrolysis of different TAG species.

Similarly, the association patterns of some of the newly mapped loci suggested their underlying functions. For example, SYNGR1 rs186680008-C showed strongest associations with decreased levels of lipid species with C20:3 acyl chain from different lipid classes, including CEs, PCs and PCOs (Supplementary Fig. 5), suggesting its role in PUFA metabolism (Fig. 6). PTPRN2 rs10281741-G and MIR100HG rs10790495-G showed associations with reduced levels of long polyunsaturated TAG species, suggesting their role in negative regulation of either elongation and desaturation of fatty acids or incorporation of long-chain unsaturated fatty acids during TAG biosynthesis.

Lipidomics provide higher statistical power. As intermediate phenotypes are known to provide more statistical power, we assessed whether the lipid species could help to detect genetic associations with greater power than traditional lipids using...
variants previously identified for traditional lipids (number of variants = 557; Supplementary Data 9). We found that molecular lipid species have much stronger associations than traditional lipids with the same sample size, except for well-known APOE and CETP (Fig. 7; Supplementary Data 10). The associations were several orders of magnitudes stronger for the variants in or near genes involved in lipid metabolism, such as FADS1-2-3, LPC, ABCG5/8, SGPP1 and SPTLC3. This shows that the lipidomics provides higher chances to identify lipid-modulating variants, particularly the ones with direct role in lipid metabolism, with much smaller sample size than traditional lipids.

**Discussion**

We present findings from a large-scale study that integrate lipodrome, genome and phenome revealing detailed description of genetic regulation of lipidome and its associations with CVD outcomes. In addition to enhancing the current understanding of genetic determinants of circulating lipids, our study highlights the potential of lipidomics in gene mapping for lipids and CVDs over traditional lipids. The study generates a publicly available knowledgebase of genetic associations of molecular lipid species and their relationships with thousands of clinical outcomes.
Despite the expected influence of dietary intake on the circulatory lipids, plasma levels of lipid species are found to be heritable, suggesting considerable role of endogenous regulation in lipid metabolism. Importantly, genetic mechanisms do not seem to regulate all lipid species in a lipid class in the same way, as also observed in recent mice lipidomics studies\textsuperscript{25,26}. Longer and more unsaturated lipid species from different lipid classes clearly display stronger genetic correlations. These observations are consistent with a previous study based on family pedigrees\textsuperscript{21}. Our finding is important in the light of the proposed role of lipids containing PUFAs in CVDs, diabetes and other disorders\textsuperscript{27–29}. Identification of genetic factors regulating these particular lipids is important for understanding the subtleties of lipid metabolism and devising preventive strategies including dietary interventions. Our study provides multiple leads in this direction by identifying 11 genomic loci (\textit{KLHL17, APOA5, CD33, SHTN1, FADS2, LIPC, MOBAT7, MIR100HG, PTPRN2, PDHA2} and \textit{TMEM86B}) associated with long, polyunsaturated lipids at genome-wide significance. Of these, \textit{FADS2, APOA5, LPL} and \textit{MOBAT7} variants were also associated with risk of CVDs (Fig. 5).

Further, we mapped genetic variants for lipid species from several lipid classes, including CERs, CE, TAGs, SMs and PCs, that are shown to predict CVD risk\textsuperscript{3–9}. Our PheWAS analyses also suggested relationship between many of the mapped genetic variants and CVD outcomes. This knowledge can directly fuel studies on CVD prediction or drug target discovery. For instance, CERs and CE are also been reported to associate with increased risk of CVD events\textsuperscript{3–9}. Our study revealed three loci associated with CEs, including \textit{FADS2} and two novel loci-\textit{ABCG5/8} and \textit{SYNGR1}, and two loci for CERs (\textit{SPTLC3} and \textit{ZNF385D}). CER species, particularly CER (d18:1/24:0) and CER (d18:1/24:1) are recently reported to be associated with the increased risk of CVD\textsuperscript{9}. We identified two variants near \textit{SPTLC3} and \textit{ZNF385D} that modulate the plasma levels of CER (d18:1/24:1) and CER (d18:1/24:0), respectively, and risk for intracerebral haemorrhage. This information could also guide future studies to establish the causal relationship between lipid species and CVD.

The detailed lipidomic profile also provided clues towards understanding the mechanisms of effects of well-established lipid loci like \textit{FADS2} and \textit{LPL} on lipid metabolism and CVD risks. We show how the inverse effects of \textit{FADS2} rs28456-G on the expression of two desaturases (\textit{FADS2} and \textit{FADS1}) could explain its opposite effects on lipids with different PUFAs. The delta-6 desaturation by \textit{FADS2} generates gamma-linolenic acid and stearidonic acid that by elongation yield dihomo-gamma-linolenic acid and eicosatetraenoic acid (Fig. 6)\textsuperscript{30}. Further,
detecting lipid species to ensure high data quality that narrowed the spectrum of lipidomic profiles. Further advances in lipidomics platforms might help to capture more comprehensive and complete lipidomic profiles, including the position of fatty acyl chains in the glycerol backbone of TAGs and glycerophospholipids and detection of sphingosine-1-P species and several other species, that would allow to overcome these limitations.

In conclusion, our study demonstrates that lipidomics enables deeper insights into the genetic regulation of lipid metabolism than clinically used lipid measures, which in turn might help guide future biomarker and drug target discovery and disease prevention.

**Methods**

**Subjects and clinical measurements.** The study included participants from the following cohorts: EUFAM, FINRISK, FinnGen and UK Biobank. The EUFAM (The European Multicenter Study on Familial Dyslipidemias in Patients with Premature Coronary Heart Disease) study cohort is comprised of the Finnish familial combined hyperlipidemia families37. The families in EUFAM study were identified via probands admitted to Finnish university hospitals with a diagnosis of premature coronary heart disease. The probands had premature coronary heart disease and high levels of the total cholesterol, triglycerides, or both (≥90th Finnish age-specific and sex-specific population percentile), or low HDL-C levels (≤10th percentile). Invitation was extended to all the family members and spouses of the probands if at least one first-degree relative of the proband had high levels of the total cholesterol, triglycerides, or both. Venous blood samples were obtained from all the participants after overnight fasting. Triglycerides and total cholesterol were measured by enzymatic methods using an automated Cobas Mira analyser (Hoffman-La Roche, Basel, Switzerland)37,38. HDL-C was quantified by phosphotungstic acid/magnesium chloride precipitation procedures, and LDL-C was calculated using the Friedewald formula39.

The Finnish National FINRISK study is a population-based survey conducted every 5 years since 1972, and thus far samples have been collected in 1992, 1997, 2002, 2007 and 201220. Collections from the 1992, 1997, 2002, 2007 and 2012 surveys are stored in the National Institute for Health and Welfare (THL) Biobank. Lipidomic profiling was performed for 1142 participants that were randomly selected from the FINRISK 2012 survey (Supplementary Table 1). The participants were advised to fast for at least 4 h before the examination and to avoid heavy meals earlier during the day. Venous blood samples were obtained from all the participants and sera were separated. HDL-C, triglycerides and total cholesterol were measured with enzymatic methods (Abbott laboratories, Abbott Park, IL, USA) with Abbott Architect c8000 clinical chemistry analyser40.

The FinnGen data release 2 is composed of 102,739 Finnish participants. The phenotypes were derived from ICD codes in Finnish national hospital registries and cause-of-death registry as a part of FinnGen project. The quality of the CVD phenotypes in these registers have been validated in previous studies. The UK Biobank data is comprised of ≥500,000 participants based in UK and aged 40–69 years, annotated for over 2000 phenotypes46. The PhEWAS analyses in this study included 408,961 samples from white British participants.

**Ethics statement.** The study was conducted in accordance with the principles of the Helsinki declaration. Written informed consent was obtained from all the study participants. The study protocols were approved by the ethics committees of the participating centres (The Hospital District of Helsinki and Uusimaa Coordinating Ethics committees, approval No. 184/13/03/00/12). For the Finnish Institute of Health and Welfare (THL) driven FinnGen prepreatory project (here called FinnGen), all patients and control subjects had provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, older cohorts were based on study specific consents and later transferred to the THL Biobank after approval by Valvira, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Valvira. The Ethical Review Board of the Hospital District of Helsinki and Uusimaa approved the FinnGen study protocol NR HUS/990/2017. The FinnGen prepreatory project is approved by THL, approval numbers THL/2031/6.02.00.00/2017, amendments THL/341/6.02.00.00/2018, THL/222/6.02.00.00/2018 and THL/283/6.02.00.09. All DNA samples and data in this study were pseudonymized.

**Lipidomic profiling.** Mass spectrometry-based lipid analysis of 2181 participants was performed in three batches-333 and 866 EUFAM participants in two batches and the FINRISK participants in one batch at Lipotype (Dresden, Germany). Samples were analysed by direct infusion in a QExactive mass spectrometer (Thermo Scientific) equipped with a TriVersa NanoMate ion source (Advion Biosciences)46. The data were analysed using in-house developed lipid identification software based on LipidXplorer46. Post processing and normalisation of data were performed using an in-house developed data management system. Only lipids with signal-to-noise ratio ≥5 and amounts at least fivefold higher than in the corresponding blank samples were considered for further
analyses. Reproducibility of the assay was assessed by the inclusion of reference plasma samples (eight reference samples for EUFAM and three reference samples for FINRISK) to create a control dataset. Median coefficient of variation was <10% across all batches. The data were corrected for batch and drift effects. Lipid species detected in <80% of the samples in any of the batches and samples (N = 64) with low lipid contents were excluded. Among the lipid species which passed quality control, a total of 141 lipid species from 13 lipid classes (Supplementary Table 2) were detected consistently in all three batches and were included in all analyses. The total amounts of lipid classes were calculated by summing up the absolute concentrations of all lipid species belonging to each lipid class. The measured concentrations of the lipid species and calculated class total were transformed to normal distribution by rank-based inverse normal transformation.

It is to be noted that Lipotype platform used in the study detected many additional lipid species (N = 83) that were not captured previously by other platforms. The list of the lipid species detected by different platforms and overlaps across the platforms are provided in the Supplementary Data 12 and Supplementary Fig. 7.

Genotyping and imputation. Genotyping for both EUFAM and FINRISK cohorts was performed using the HumanCoreA BeadChip (Illumina Inc., San Diego, CA, USA). The genotype calls were generated together with other available data sets using xCell at the Institute for Molecular Medicine Finland (FIMM). Genotype data underwent stringent quality control (QC) before imputation that included exclusion of samples with low call rate (<95%), sex discrepancies, excess heterozygosity (greater than 5% of all variants), low call rate (<95%), low deviation from Hardy–Weinberg Equilibrium (HWE P < 1 × 10^{-6}) were excluded. Imputation was performed using IMPUTE2, which used two population-specific reference panels of 2690 high-coverage whole-genome and 5093 high-coverage whole-exome sequence data. Variants with imputation info score <0.70 were filtered out of the genotyped lipid organisms and variants in the lipid species analyses included 2045 individuals and ~9.3 million variants with MAF >0.005 that were available in both cohorts. FinnGen samples were genotyped with Illumina and Affymetrix arrays (Thermo Fisher Scientific, Santa Clara, CA, USA). Genotype calls were made with GenCall and Cavatica methods for Illuminum and AxioM GT2 algorithm, for Affymetrix chip genotyping data. Genotyping data produced with previous chip platforms were lifted over to build version 38 (GRCh38/hg38) following the protocol described here: dx.doi.org/10.17504/protocols.io.nmndc5e. Samples with sex discrepancies, high genotype missingness (>5%), excess heterozygosity (+4SE) and non-Finnish samples with high missingness (>2%), deviation from HWE (P < 1e-6) and low minor allele count (MAC < 2) were removed. Pre-phasing of genotyping data was performed with Eagle 2.3.5 (https://data.broadinstitute.org/alkesgroup/Eagle/) with the default parameters, except the number of conditioning haplotypes was set to 20,000. Imputation was carried out by using the population-specific SISu v3 imputation reference panel with Eagle 2.3.5 (version 08un17.dbb, https://faculty.washington.edu/browning/beagle/b4_1.1.html) as described in the following protocol: [dx.doi.org/10.17504/ protocols.io.mndc5e]. SISu v3 imputation reference panel was developed using the high-coverage (25–30x) whole-genome sequencing data generated at the Broad Institute and at the McDonnell Genome Institute of Washington University; and jointly processed at the Broad Institute. Variant callset was produced with GATK HaplotypeCaller algorithm by following GATK best practices for variant calling. Genotype-, sample- and variant-wise QC was applied in an iterative manner by using the Hail framework v0.1 [https://github.com/hail-is/hail] The resulting high-quality WGS data for 3775 individuals were phased with Eagle 2.3.5 as described above. Post-imputation quality control involved excluding variants with INFO score < 0.7.

Genotyping for the majority of the UK Biobank participants was done using the Affymetrix UK Biobank Axiom Array, while a subset of participants was genotyped using the Affymetrix UK BiLEVE Axiom Array. Details about the quality control and imputation of UK Biobank cohort are described by Bycroft et al.51.

Heritability estimates and genetic correlations. For heritability and genetic correlation estimation, rank-based inverse-transformed measures of lipid species, computed separately for the EUFAM and FINRISK cohorts, were combined to increase statistical power. The residuals of inverse-transformed measures after regressing for age, sex, first ten principal components (PCs) of genetic population structure, and thyroid hormone replacement were used as covariates. Type 1 diabetes was used as a phenotype. SNP-based heritability estimates were calculated using the variance component analysis using a genetic relationship matrix (GRM) as implemented in bIMM.52 Only the good quality variants with missingness <10% and MAF >0.005 were used to generate the GRM. The GRM was generated using GCTA by setting the off-diagonal elements that are <0.05 to 0 as proposed by Zaitlen et al.53. This allows to estimate SNP-based heritability in family data without removing closely related individuals. The heritability estimates of lipid species in different groups were compared using Wilcoxon rank-sum test.

The genetic correlation between each pair of lipid species and between each lipid species and traditional lipid traits were determined using the generated GRM with bivariate linear mixed model as implemented in bIMM. The correlations based on the plasma levels (termed as phenotypic correlations) between all the pairs of the lipid species and traditional lipid classes were calculated using Pearson’s correlation coefficient. The heatmaps and hierarchical clustering based on genetic and phenotype correlations were generated using heatmap.2 in R. As lipid-lowering medications could affect the plasma levels of lipid species, all analyses were adjusted for the usage of lipid-lowering medications, and separate analyses were also performed after excluding individuals using lipid-lowering medications (N = 172).

Lipidomics GWAS. We performed univariate association tests for 141 individual lipid species, 12 total lipid classes and 4 traditional lipid measures (HDL-C, LDL-C, total cholesterol and triglycerides), in all batches to control for possible batch effects and combined the summary statistics by meta-analysis. The association analyses for the EUFAM cohort were performed using linear mixed models, including the above-mentioned covariates as fixed effects and kinship matrix as random effects as implemented in M3M.54 The kinship matrices for the GWAS analyses were computed separately for each chromosome to include the variants from the other chromosomes using directly genotyped variants with MAF >0.1 and missingness <2%. The FINRISK cohort was analysed with linear regression model adjusting for age, sex, first ten PCs, lipid medication and diabetes using METAL.55 Meta-analysis of the lipid species was performed using inverse variance weighted method for fixed effects adjusted for genomic inflation factor in METAL.55 In addition, analyses adjusting for the traditional lipids (in addition to above-mentioned covariates) were also performed for the identified variants to determine the independent effect on lipid species.

The top 20 analyses were remapped for the 35 lead variants with P values from 0.01 to 0.20 after correcting for 34 principal components (PCs) that explain over 90% of the variance in lipidomic profiles. Only the associations consistent in effect direction in all three batches were considered significant. Variants were designated as new if not located within 1 Mb of any previously reported variants for lipids (any of the traditional lipids and molecular lipid species) and as independent signal if located within 1 Mb but r2 < 0.20 with the previous lead variant and confirmed by conditional analysis. Variants with the strongest association in the identified lipid species loci was identified as the lead variant, and were annotated to the nearest gene for the new loci.

PhеАWAS. We identified 25 CVD-related outcomes from the derived phenotypes in the FinnGen and UK Biobanks (Supplementary Table 3). Associations between the 35 lead variants from the identified loci and 25 selected CVD phenotypes in FinnGen cohort were obtained from the ongoing analyses as a part of the FinnGen project. The associations were tested using s jade point approximation method for adjusting age, sex and first 10 PCs as implemented in SPAtest R package.56 Associations between selected binary phenotypes and 35 lead variants in UK Biobank were obtained from Zhou et al. that were tested using logistic mixed model in SAIGE with a s jade point approximation and adjusting for first four principal components, age and sex (https://www.leelabsg.org/resources). Data for four phenotypes were not available from Zhou et al. and hence were obtained from http://www.ancefallos.org/uk-biobank/. All the phenotypes were tested using linear regression models with the same covariates as mentioned above, both for Finnish and UK Biobank cohorts. Meta-analyses of both cohorts were performed using the inverse variance weighted method for fixed effects model in METAL. The P-values obtained from the meta-analyses of the two cohorts are reported for PheWAS associations. All the PheWAS associations with false discovery rate (FDR) <5% evaluated using the Benjamini–Hochberg method and consistent direction of effects were considered significant.

Variance explained. To determine the variance explained by the known loci for traditional lipids, we included all the lead variants with MAF >0.005 in 250 genotype loci that have previously been associated with one or more of the four traditional lipids. Of the 636 reported variants, 557 variants with MAF >0.005 (including lipid proxies) were analyzed using quantitative traits were tested using linear regression models with the same covariates as mentioned above, both for Finnish and UK Biobank cohorts. Meta-analyses of both cohorts were performed using the inverse variance weighted method for fixed effects model in METAL. The P-values obtained from the meta-analyses of the two cohorts are reported for PheWAS associations. All the PheWAS associations with false discovery rate (FDR) <5% evaluated using the Benjamini–Hochberg method and consistent direction of effects were considered significant.

LPL activity. The post-heparin lipoprotein lipase (LPL) after 15 min of heparin load was measured for 630 individuals in the EUFAM cohort using the ELISA method developed by Antikainen et al.57. The measured values were transformed using rank-based inverse normal transformation. Associations between the LPL activity and plasma levels of TAGs were determined using linear regression model adjusted for age, sex, lipid medication, hormone replacement therapy, thyroid hormone therapy, and the LDL variant rs1570891. As LPL activity was tested using linear mixed model adjusted for age, sex, first ten PCs of genetic population structure, lipid medication, hormone replacement therapy, thyroid hormone therapy, and the LDL variant rs1570891.
thyroid condition and type 2 diabetes as fixed effect and kinship matrix as random effect as implemented in MMM.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The full lipidomics GWAS summary level data are available on the web-based database [https://mgltl.fiml.mfi.](https://mgltl.fiml.mfi). Similarly, the PhEWS summary data can be obtained through [https://www.leelabh.org/resources](https://www.leelabh.org/resources) and [http://www.neealab.is/uk-biobank/]. The data presented in the figures and other summary level data are contained within the Supplementary Files and Supplementary Data. Other data are available through the Institute for Molecular Medicine Finland Data Access Committee on reasonable request after appropriate ethical approval.

Received: 18 April 2019 Accepted: 13 August 2019

**Published online:** 24 September 2019

**References**

1. Global Burden of Disease 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 390, 1151–1201 (2017).
2. Ference, B. A. et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A common mechanism from the European atherosclerosis society consensus panel. Eur. Heart J. 38, 2459–2472 (2017).
3. Quehenberger, O. & Dennis, E. A. The human plasma lipidome. N. Engl. J. Med. 365, 1812–1823 (2016).
4. Alshehry, Z. H. et al. Plasma lipidomic profiles of human plasma demonstrate independent and shared effects of EPA, DPA and DHA. The EPIC-InterAct Case-Control Study. PLoS Med. 13, e1002094 (2016).
5. Fry, A. et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J. Clin. Invest. 140, 1402–1411 (2011).
6. Alshehry, Z. H. et al. Plasma lipidomic profiles are a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J. Clin. Invest. 140, 1402–1411 (2011).
7. Jha, P. et al. Systems analyses reveal physiological roles and genetic regulators of liver lipid species. Cell Metab. 6, 722–733 (2016).
8. Jha, P. et al. Genetic Regulation of Plasma Lipid Species and their Association with Metabolic Phenotypes. Cell Metab. 6, 709–721 (2016).
9. Ander, B. P., Dupasquier, C. M., Prociuk, M. A. & Pierce, G. N. Polyunsaturated fatty acids and their effects on cardiovascular disease. Exp. Clin. Cardiol. 8, 164–172 (2003).
10. Forouhi, N. G. et al. Association of plasma phospholipid n-3 and n-6 polyunsaturated fatty acids with type 2 diabetes. The EPIC-InterAct Case-Control Study. PLoS Med. 13, e1002094 (2016).
11. Bellis, C. et al. Human plasma lipidome is pleiotropically associated with metabolic and cardiovascular disease. Arterioscler. Thromb. Vasc. Biol. 26, 3614–3623 (2016).
Acknowledgements

We would like to thank Sari Kivikko, Hsei-Yi Shen and Ulla Tuomainen for management assistance. We thank all study participants of the study for their participation. The FINRISK and Finngen data used for the research were obtained from THL Biobank. We thank the THL DNA laboratory for its skilful work to produce the DNA samples used in the genotyping work, which was used in this study. Part of the genotyping was performed by the Institute for Molecular Medicine Finland FIMM Technology Centre, University of Helsinki. This research has been conducted using the UK Biobank Resource with application number 22627. This work was supported by National Institutes of Health [grant HL133135 to S.R., M.R.T., N.F. and A.P.]; Finnish Foundation for Cardiovascular Research [to S.R., V.S., M.R.T., M.J. and A.P.]; Academy of Finland Center of Excellence in Complex Disease Genetics [grant 312062 to S.R.]; Academy of Finland [285380 to S.R., 288509 to M.P.]; Jane and Aatos Erkko Foundation [to M.J.]; Sigrid Jusélius Foundation [to S.R. and M.R.T.]; Horizon 2020 Research and Innovation Programme [grant 692145 to S.R.]; EU-project RESOLVE (EU 7th Framework Programme) [grant 305707 to M.R. T.]; HiLife Fellowship [to S.R.]; Helsinki University Central Hospital Research Funds [to M.R.T.]; Magnus Ehrensvard Foundation [to M.J.]; Leducq Foundation [to M.R.T.]; Ida Montin Foundation [to P.R.]; MD-PhD Programme of the Faculty of Medicine, University of Helsinki [to I.T.R. and P.R.]; Finnish Medical Foundation [to J.T.R.]; Emil Aaltonen Foundation [to P.R.]; Biomedicum Helsinki Foundation [to I.T.R.]; Paulo Foundation [to I.T.R.]; Idman Foundation [to I.T.R.]; Veritas Foundation [to I.T.R.]; FIMM-EMBL PhD Fellowship grant [to S.H.]. The Finngen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and nine industry partners (AlbVie, AstraZeneca, Boehringer, Celgene, Genentech, GSK, Merck, Pfizer and Sanofi). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

Author contributions

R.T., S.R., M.R.T., K.S., A.P. and N.B.F. conceived and designed the study and wrote the paper; R.T. and J.T.R. performed the statistical analyses; J.T.K., M.K., I.K., J.T.T., J.K., A.S.H., H.L. and V.S. were involved in Finngen phenotype definitions and PhEWAS analyses; S.H., J.N. and P.P. performed the genotype imputation; M.J.G., C.K. and M.A.S. performed lipidomic profiling and processed the raw data; S.K.S. and M.P. provided critical inputs in statistical analyses; M.J. provided critical inputs in the interpretation of the data; P.R., S.S., N.M., N.O.S., M.J.D., V.S., N.B.F., A.P., M.R.T., K.S. and S.R. acquired the data. All authors read, revised and approved the paper.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-11954-8.

Competing interests: V.S. has participated in a conference trip sponsored by Novo Nordisk and received an honorarium from the same source for participating in an advisory board meeting. M.J.G. is an employee of Lipotype GmbH, C.K. is a shareholder and employee of Lipotype GmbH, K.S. is a shareholder and CEO of Lipotype GmbH. M.A.S. is a shareholder of Lipotype GmbH and an employee of Lukasiewicz Research Network—PORT Polish Center for Technology Development. The remaining authors have no relevant competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rubina Tabassum1, Joel T. Rämö1, Pietari Ripatti1, Jukka T. Koskela1, Mitja Kurki1,2,3, Juha Karjalainen1,4,5, Priti Palta1,6, Shabbeer Hassan1, Javier Nunez-Fontanau1, Tuomo T.J. Kiiskinen1,7, Sanni Söderlund8, Niina Matikainen8,9, Mathias J. Gerl1,10, Michael A. Surma1,10,11, Christian Klose10, Nathan O. Stitziel1,12,13,14, Hannele Laivuori15,16, Aki S. Havulinna1,17, Susan K. Service17, Veikko Salomaa1,7, Matti Pirinen1,18,19, FinnGen Project, Matti Jauhiainen7,20, Mark J. Daly1,14, Nelson B. Freimer17, Aarno Palotie1,4,21, Marja-Riitta Taskinen8, Kai Simons1,22,23 & Samuli Ripatti1,4,23

1Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland. 2Program in Medical and Population Genetics and Genetic Analysis Platform, Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. 3Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. 4Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. 5Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. 6Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. 7National Institute for Health and Welfare, Helsinki, Finland. 8Research Programs Unit, Diabetes & Obesity, University of Helsinki and Department of Internal Medicine, Helsinki University Hospital, Helsinki, Finland. 9Endocrinology, Abdominal Center, Helsinki University Hospital, Helsinki, Finland. 10Lipotype GmbH, Dresden, Germany. 11Łukasiewicz Research Network-PORT Polish Center for Technology Development, Stabłowicka 147 Str., 54-066 Wroclaw, Poland. 12Cardiovascular Division, Department of Medicine, Washington University School of Medicine, Saint Louis, MO, USA. 13Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA. 14McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, USA.
FinnGen Project

Anu Jalanko1, Jaakko Kaprio1, Kati Donner1, Mari Kaunisto1, Nina Mars1, Alexander Dada1, Anastasia Shcherban1, Andrea Ganna1, Arto Lehisto1, Elina Kilpeläinen1, Georg Brein1, Ghazal Awaisa1, Jarmo Harju1, Kalle Pärn1, Pietro Della Briotta Parolo1, Risto Kajanne1, Susanna Lemmelä1, Timo P. Sipilä1, Tuomas Sipilä1, Ulrike Lyhs1, Vincent Llorens1, Teemu Niiranen7, Kati Kristiansson7, Lotta Männikkö7, Manuel González Jiménez7, Markus Perola7, Regis Wong7, Terhi Kilpi7, Tero Hiekkalinna7, Eina Järvensivu7, Essi Kahlajr7, Hannele Mattsson7, Markku Laukkanen7, Päivi Laiho7, Sini Lähteenmäki7, Tuuli Sistonen7, Sirpa Soini7, Adam Ziemann24, Anne Lehtonen24, Apinya Lertratanakul24, Bob Georgantas24, Bridget Riley-Gillis24, Danjuma Quarless24, Fedik Rahimov24, Graham Heap24, Howard Jacob24, Jeffrey Waring24, Justin Wade Davis24, Nizar Smouei24, Relja Popovic24, Sahar ESAmeei24, Jeff Waring24, Athena Matakidou25, Ben Challis25, David Close25, Slavé Petrovski25, Antti Karlsson25, Johanna Schleutker25, Kari Pulkk25, Petri Virolainen26, Lila Kallio27, Arto Mannermaa28, Sami Heikkinen28, Veli-Matti Kosma29, Chia-Yen Chen29, Heiko Runz30, Jimmy Liu30, Paola Bronzon30, Sally John30, Sanni Lahdenper30, Susan Eaton30, Wei Zhou4, Minna Hendolin31, Outi Tuovila31, Raimo Pakkanen31, Joseph Maranville32, Keith Usiskin32, Marla Hochfeld32, Robert Plenge32, Robert Yang32, Shameek Biswas32, Steven Greenberg32, Eija laakkonen33, Juha Kononen33, Juha Paloneva33, Urho Kujala33, Teijo Kuopio34, Jari Laukkanen35, Eeva Kangasniemi36, Kimmo Savinainen36, Reijo Laaksonen36, Mikko Arvas36, Jarmo Ritari37, Jukka Partanen37, Hannele Mattsson37, Johanna Schleutker37, Tiina Wahlfors37, Andrew Peterson38, Danny Oh38, Diana Chang38, Edmond Teng38, Erich Strauss38, Geoff Kerchner38, Hao Chen38, Hubert Chen38, Jennifer Schutzman38, John Michon38, Julie Hunkapiller38, Mark McCarthy38, Natalie Bowers38, Tim Lu38, Tushar Bangale38, David Pulford39, Dawn Waterworth39, Dipette Kulkarni39, Fanli Xu39, Jo Betts39, Jorge Esparza Gordillo39, Joshua Hoffman39, Kirsii Auro39, Linda McCarthy39, Soumitra Ghosh39, Meg Ehm39, Kimmo Pitkänen40, Tomi Mäkelä41, Anu Loukola42, Heikki Joensuu42, Juha Sinisalo42, Kari Eklund42, Lauri Aalto42, Martti Färkkila42, Olli Carpen42, Paula Kauppi42, Pentti Tienari42, Terhi Ollila42, Tiina Maija Tuomi42, Tuomo Meretoja42, Anne Pitkäranta42, Anu Loukola42, Heikki Joensuu42, Juha Sinisalo42, Kari Eklund42, Lauri Aalto42, Martti Färkkila42, Olli Carpen42, Paula Kauppi42, Pentti Tienari42, Terhi Ollila42, Tiina Maija Tuomi42, Tuomo Meretoja42, Anne Pitkäranta42, Joni Turunen42, Katiarianna Hannula-Jouppi42, Sampsa Piikkarainen42, Sanna Seitsonen42, Miika Koskinen42, Antti Palomäki43, Juha Rinne43, Kai Metsärinne43, Kari Metsärinne43, Claus Elenius43, Laura Pirilä43, Leena Koulu43, Markku Voutilainen43, Markus Juonala43, Sirkku Peltonen43, Vesa Aalto43, Andrey Loboda44, Anna Podgornaia44, Aparnaa Chhibber44, Audrey Chu44, Caroline Fox44, Dorothee Diogo44, Emily Holzinger44, John Eicher44, Padhraig Gormley44, Vinay Mehta44, Xulong Wang44, Johannes Kettunen45, Katri Pylkäs45, Marita Kalajärvi45, Minna Karjalainen45, Reetta Hilttala45, Riitta Kaarteenaho45, Seppo Vainio45, Tuomo Mantere45, Seppo Vainio46, Anne Remes47, Johanna Huhtakangas47, Juhani Junttila47, Kaisa Tasanen47, Laura Huijala47, Marja Luodonpää47, Nina Hautala47, Peeter Karihtala47, Salla Kauppi47, Terttu Harju47, Tomo Blomster47, Hilkka Soininen48, Ilkka Harvima48, Jussi Pihlajamäki48, Kai Kaarniranta48, Margit Pelkonen48, Markku Laakso48, Mikko Hiltunen48, Mikko Kivinemi48, Olli Kaipiainen-Seppänen48, Päivi Auvinen48, Reetta Kälviäinen48, Valtteri Julkunen48, Anders Malarstig49, Åsa Hedman49, Catherine Marshall49, Christopher Whelan49, Heli Lehtonen49, Jaakko Parkkinen49, Kari Linden49, Kirsi Kalpalu49, Melissa Miller49, Nan Bing49.
Stefan McDonough, Xing Chen, Xinli Hu, Ying Wu, Annika Auranen, Airi Jussila, Hannele Uusitalo-Järvinen, Hannu Kankaanranta, Hannu Uusitalo, Jukka Peltola, Mika Kähönen, Pia Isomäki, Tarja Laitinen, Teet Salmi, Anthony Muslin, Clarence Wang, Clement Chatelain, Ethan Xu, Franck Auge, Kathy Call, Kathy Klinger, Marika Crohns, Matthias Gossel, Kimmo Palin, Manuel Rivas, Harri Siirtola & Javier Gracia Tabuenca

24 Abbvie, Chicago, IL, USA. 25 Astra Zeneca, Cambridge, UK. 26 Auria Biobank, University of Turku and Hospital District of Southwest Finland, Turku, Finland. 27 Auria Biobank, Turku, Finland. 28 Biobank of Eastern Finland, University of Eastern Finland and Northern Savo Hospital District, Kuopio, Finland. 29 Biobank of Eastern Finland, Kuopio, Finland. 30 Biogen, Cambridge, MA, USA. 31 Business Finland, Helsinki, Finland. 32 Celgene, Summit, NJ, USA. 33 Central Finland Biobank, University of Jyväskylä and Central Finland Health Care District, Jyväskylä, Finland. 34 Central Finland Biobank, Jyväskylä, Finland. 35 Central Finland Health Care District, Jyväskylä, Finland. 36 Finnish Clinical Biobank Tampere, University of Tampere and Pirkanmaa Hospital District, Tampere, Finland. 37 Finnish Red Cross Blood Service and Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland. 38 Genentech, San Francisco, CA, USA. 39 GlaxoSmithKline, Brentford, UK. 40 Helsinki Biobank, Helsinki, Finland. 41 HiLIFE, University of Helsinki, Helsinki, Finland. 42 Hospital District of Helsinki and Uusimaa, Helsinki, Finland. 43 Hospital District of Southwest Finland, Turku, Finland. 44 Merck, Kenilworth, NJ, USA. 45 Northern Finland Biobank Borealis, University of Oulu and Northern Ostrobothnia Hospital District, Oulu, Finland. 46 Northern Finland Biobank Borealis, Oulu, Finland. 47 Northern Ostrobothnia Hospital District, Oulu, Finland. 48 Northern Savo Hospital District, Kuopio, Finland. 49 Pfizer, New York, NY, USA. 50 Pirkanmaa Hospital District, Tampere, Finland. 51 Sanofi, Paris, France. 52 University of Helsinki, Helsinki, Finland. 53 University of Stanford, Stanford, CA, USA. 54 University of Tampere, Tampere, Finland.