American Journal of Epidemiology **Submitted Manuscript**

Special Collection: N/A

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

ORIGINAL UNEDITED MANUSCRIPT

Title: Understanding lipidomics associations and the lipoprotein-related caveats in population epidemiology

Authors: Siyu Zhao1,2,3, Pauli Ohukainen1,2,3, Johannes Kettunen1,2,3,4, Marjo-Riitta Järvelin2,5,6, Mika Kähönen7, Terho Lehtimäki8, Jorma Viikari9,10, Olli T. Raitakari10,11,12,13, Ville-Petteri Mäkinen1,2,3, Mika Ala-Korpela1,2,3,14,*

ORCiD IDs: Mika Ala-Korpela https://orcid.org/0000-0001-5905-1206

Correspondence Address: *Systems Epidemiology, Faculty of Medicine, University of Oulu, 90014 Oulu, Finland

Joint Authorship: N/A

© The Author(s) 2024. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Affiliations: [1]Systems Epidemiology, Faculty of Medicine, University of Oulu, Oulu, Finland; [2]Research Unit of Population Health, Faculty of Medicine, University of Oulu, Oulu, Finland; [3]Biocenter Oulu, University of Oulu, Oulu, Finland; [4]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland; [5]Department of Epidemiology and Biostatistics, MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK; [6]Department of Life Sciences, College of Health and Life Sciences, Brunel University London, UK; [7]Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; [8]Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; [9]Department of Medicine, University of Turku, Turku, Finland; [10]Division of Medicine, Turku University Hospital, Turku, Finland; [11]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; [12]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland; [13]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; [14]NMR Metabolomics Laboratory, School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland.

Clinical trial registration number and website: N/A

Key words: Bias, body mass index, cholesterol, clinical lipids, confounding, HDL, LDL, lipidomics, lipoproteins, mediation, overadjustment, pleiotropy, triglycerides

Acknowledgments[1]: N/A

Funding: The NFBC has been supported, for instance, by EU, Academy of Finland, PREcisE, Joint Programming Initiative a Healthy Diet for a Healthy Life (no. 655), UK Medical Research

[1] Study investigators, conference presentations, preprint publication information, thanks.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Council, Biotechnology and Biological Sciences Research Council (MR/S03658X/1), and European Regional Development Fund Grant no. 539/2010 A31592. The Young Finns Study has been financially supported by the Academy of Finland: grants 356405, 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117797 (Gendi), and 141071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS and grant 848146 for To Aition); European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation; Finnish Society of Clinical Chemistry; the Cancer Foundation Finland; pBETTER4U_EU (Preventing obesity through Biologically and bEhaviorally Tailored inTERventions for you; project number: 101080117); CVDLink and the Jane and Aatos Erkko Foundation. M.A.-K. was supported by a research grant from the Sigrid Juselius Foundation, the Finnish Foundation for Cardiovascular Research, and the Research Council of Finland (grant no. 357183).

Conflict of Interest: The authors declare no conflicts of interest.

Disclaimer: N/A

Data Availability Statement: The datasets used in the current study are available from the cohorts through application process for researchers who meet the criteria for access to confidential data: https://www.oulu.fi/nfbc/ (NFBC66) and http://youngfinnsstudy.utu.fi

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

(YFS). Regarding the YFS data the Ethics committee has concluded that under applicable law, the data from this study cannot be stored in public repositories or otherwise made publicly available. The data controller may permit access on case-by-case basis for scientific research, not however to individual participant level data, but aggregated statistical data, which cannot be traced back to the individual participants' data.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Understanding lipidomics associations and the lipoprotein-related caveats in population epidemiology

Siyu Zhao[1,2,3], Pauli Ohukainen[1,2,3], Johannes Kettunen[1,2,3,4], Marjo-Riitta Järvelin[2,5,6], Mika Kähönen[7], Terho Lehtimäki[8], Jorma Viikari[9,10], Olli T. Raitakari[10,11,12,13], Ville-Petteri Mäkinen[1,2,3], Mika Ala-Korpela[1,2,3,14,*]

[1]Systems Epidemiology, Faculty of Medicine, University of Oulu, Oulu, Finland; [2]Research Unit of Population Health, Faculty of Medicine, University of Oulu, Oulu, Finland; [3]Biocenter Oulu, University of Oulu, Oulu, Finland; [4]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland; [5]Department of Epidemiology and Biostatistics, MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK; [6]Department of Life Sciences, College of Health and Life Sciences, Brunel University London, UK; [7]Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; [8]Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; [9]Department of Medicine, University of Turku, Turku, Finland; [10]Division of Medicine, Turku University Hospital, Turku, Finland; [11]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; [12]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland; [13]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; [14]NMR Metabolomics Laboratory, School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland.

***Corresponding author**
Professor Mika Ala-Korpela
Systems Epidemiology
Research Unit of Population Health
Faculty of Medicine, University of Oulu
Oulu, Finland

E-mail: mika.ala-korpela@oulu.fi
Mobile: +358 40 1977 657
Twitter: @mak_sysepi
ORCID ID 0000-0001-5905-1206

Keywords: Bias, body mass index, cholesterol, clinical lipids, confounding, HDL, LDL, lipidomics, lipoproteins, mediation, overadjustment, pleiotropy, triglycerides

An original scientific article with 3,991 words of main text, abstract of 198 words, 5 figures and 2 tables, 51 references, and supplementary material.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Abstract

Mass spectrometry lipidomics is becoming customary to analyse serum/plasma samples in epidemiology. The measurables are molecular constituents of lipoprotein particles, but very little is known on the consequences of adjusting lipidomics data with lipoprotein measures. We studied two population cohorts with 5,657 and 2,036 participants. LC-MS/MS lipidomics was applied to analyse 24 molecular lipid classes and NMR spectroscopy to quantify seven lipoprotein lipids plus apolipoprotein A-I (apoA-I) and B (apoB). The associations of these measures were analysed via partial Spearman's correlations. The effects of nine different lipoprotein adjustments on these interrelationships were assessed. Multivariable regression modelling with these adjustments was also performed for the associations between the lipidomics data and BMI. These novel large-scale lipidomics data and their associations between the lipoprotein measures were coherent in both population cohorts, confirming the compatibility of the analytical approaches. Simulated data were generated to corroborate the mediation effects. The lipoprotein-related lipid-transport and metabolism inherently mediate the lipidomics associations as evident from the striking effects of the lipoprotein adjustments. These effects and their relevance to the interpretations of lipidomics data are presented and discussed in detail for the first time. The combined lipoprotein lipid adjustments appear prone to overadjustment and arbitrary biases.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Key Messages

- Here we present a detailed assessment of the fundamental molecular framework of circulatory lipoproteins (NMR spectroscopy) and their lipid constituents (LC-MS/MS) and how various adjustments with lipoprotein related measures affect associations and epidemiological regression models and their interpretations with LC-MS/MS lipidomics data.
- The lipoprotein-centred lipid transport system inherently intertwines the lipidomics data with the lipoprotein measures in a complex, non-linear way and thereby the combined lipoprotein lipid adjustments (triglycerides, cholesterol or LDL-C, and HDL-C) of regression models with LC-MS/MS lipidomics data are very difficult to interpret and prone to bias.
- Associations of the lipoprotein lipid constituents, analysed by LC-MS/MS lipidomics and adjusted with combined lipoprotein measures, should be interpreted with extreme care due to a high likelihood for biased results; instead, it would be more prudent to adjust for individual lipoprotein measures – triglycerides, various cholesterol measures, and apolipoprotein B – and stemming the interpretations from their known compositional and metabolic roles.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Introduction

Lipoproteins are heterogeneous noncovalent protein-lipid particles with key functions in circulatory lipid transport and metabolism. In serum/plasma samples separated from blood collections, they are basically the sole vehicles for all lipid molecules,[1–5] though albumin contains the bulk of free fatty acids[6] and transports also some lysophosphatidylcholines,[7,8] and likely other lyso-type of lipid molecules with one fatty acid chain.

Lipoprotein-related lipid measures have played a key role in the risk assessment of cardiometabolic conditions since the 1950's.[9] Circulating triglycerides (TG), total cholesterol (TC), and high-density lipoprotein cholesterol (HDL-C) have been the routine panel of measurements for decades with TC more recently often replaced by low-density lipoprotein cholesterol (LDL-C). In addition, improved analytical techniques have shaped the contemporary convention towards a more comprehensive measurement panel in assessing cardiometabolic risk. The additional measures typically include separate assessment of very-low-density and intermediate-density lipoprotein cholesterol (VLDL-C and IDL-C, respectively) as well as apolipoprotein A-I (apoA-I) and B (apoB).[10,11] Remnant cholesterol, defined as (non-LDL, non-HDL)-C, is also often nowadays considered valuable for risk assessments.[12–14] Findings from recent genetic epidemiology analyses with Mendelian randomization[15] have been transformative, indicating that all the above-mentioned lipoprotein-related measures would be independently, and causally, related to coronary heart disease (though the causal role of HLD-C and apoA-I is still ambiguous[11,16,17]), and many of them also for multiple other diseases, for example, stroke and Alzheimer.[11]

The above background is of the essence regarding lipidomics studies of serum/plasma samples. Unlike enzymatic lipid assays or NMR spectroscopy, mass spectrometry (MS) lipidomics calls for lipid extraction of the serum/plasma samples as an integral part of the analyses as illustrated in **Figure 1**.[18–20] Thus, the preparations analysed by MS lipidomics represent pooled mixtures of all the various lipid molecules from all the circulating lipoprotein particles. All the information from the specific lipoprotein origin of a certain lipid in the circulation is lost in the extraction phase. This calls attention to the integration of lipoprotein data when interpreting serum/plasma lipidomics associations. However, this fact has not been addressed in detail in the literature. Conversely, it is often either disregarded or dealt with via adjusting analyses in a straightforward manner for

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

TG+TC+HDL-C, implicitly giving the rationale to consider the results "independent of standard clinical lipids".[21–24]

Confounding, mediation and overadjustment biases are common sources of erroneous conclusions in epidemiology of pleiotropic traits.[25] Combined lipidomics and lipoprotein data are likely to be particularly adversely affected due to the tight biophysical and metabolic relationships that lead to large sets of highly collinear measures.[5,26] In this work we present novel data from two independent population cohorts with 7,693 participants with serum samples, analysed by NMR spectroscopy for an extensive 9-measure clinical lipoprotein panel and by LC-MS/MS lipidomics for 24 molecular lipid classes. We investigate, for the first time in detail, how various individual and combined adjustments with lipoprotein measures affect the results and their interpretation with lipidomics data.

Materials and methods

Population cohorts

Two independent population cohorts were studied cross-sectionally: the Northern Finland Birth Cohort 1966 (NFBC66)[27] with 5,657 participants (median age 46 years, 56% women) and the Cardiovascular Risk in Young Finns Study (YFS)[28] with 2,036 participants (median age 43 years, 55% women). More details on these cohorts are given in the **Supplementary material**. NFBC66 is one of the biggest epidemiological cohorts with LC-MS/MS data available and together these cohorts comprise a unique combination of large-scale epidemiological data for comprehensive lipoprotein panels and lipidomics measures.

NMR spectroscopy analysis for the lipoprotein panel

We applied an NMR platform that has been widely used in epidemiology and genetic studies over the last 15 years and for which the general methodological issues have been published and discussed previously.[5,20,29–39] We included a 9-measure lipoprotein panel consisting of key lipid measures for the four major lipoprotein fractions VLDL, IDL, LDL, and HDL as well as apoA-I and apoB, that are abundant in the circulation and are the two most analysed

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

apolipoproteins in epidemiological studies.[11] **Figure 1E** and **1F** demonstrate the metabolic overlap and interrelationships of these lipoprotein measures and the absolute concentration data in both cohorts are given in **Table 1**. TG, free cholesterol (FC), cholesteryl ester (CE), and phospholipid (PL) concentration data from the NMR analysis were used to outline the lipid compositions of the four key lipoprotein fractions (**Figure 1G**) and their relative contributions for the lipid transport in the bloodstream (**Figure 1H**). A more detailed dissection of the structural and metabolic relationships of comprehensive lipoprotein data can be found from our previous work.[5]

LC-MS/MS mass spectrometry analysis for the molecular lipid classes

The LC-MS/MS experiments were performed at the Baker Heart and Diabetes Institute, Melbourne, Australia. Serum lipids were extracted as previously described[19] and the molecular lipid classes analysed using an adapted method as described earlier.[21] Both NFBC66 and YFS were run on the same methodology of which more details can be found from the **Supplementary material**. We included the 24 most abundant molecular classes into the analyses to limit the results to those most pertinent to convey the key message in this work and to prevent the common overflow of (highly correlated) results and complicated visualisations with lipidomics data (**Figure 1L-1N**). Data wise this is a minor exclusion since these lipid classes constitute 99.91% and 99.90% of total serum lipid concentrations in NFBC66 and in YFS, respectively. The absolute concentration data for these 24 lipid classes in both cohorts are given in **Table 2**. The data for both cohorts were scaled according to a set of National Institute of Standards and Technology (NIST) plasma samples.

Statistical analyses

Partial Spearman's correlations were calculated between all the individual measures in the lipoprotein panel (**Figure 1F**) and between all the lipoprotein and lipidomics measures (**Figure 2**). Both cohorts were analysed separately and then combined via inverse variance weighted meta-analysis. All correlations were adjusted for sex in both cohorts and age in

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

YFS. The color-coded heat map in **Figure 2** was hierarchically clustered in both dimensions and the resulting order of the measures preserved in the following analyses with additional adjustments for the key lipoprotein measures and their combinations: TG, TC, LDL-C, HDL-C, TG+TC+HDL-C, TG+LDL-C+HDL-C, apoB, apoA-I, and apoB+apoA-I (**Figure 3**). Results for the individual cohorts are shown in **Supplementary Figures S1** – **S3**.

The effects of these lipoprotein adjustments were exemplified for the associations of the 24 lipidomics measures with BMI via linear regression analyses (**Figure 4**). As for the correlations, both cohorts were analysed separately and then combined via inverse variance weighted meta-analysis (results for the individual cohorts are given in **Supplementary Figures S4** and **S5**). Association magnitudes are reported in standard deviation (SD) units throughout to ease the comparison across multiple measures with markedly varying absolute concentrations.

Twenty-five principal components explained over 95% of the variation in the 24 lipid classes and the nine lipoprotein measures in both cohorts. Therefore, we set the 5% Bonferroni-adjusted type 1 error threshold at $p < 0.05/25 = 0.002$. Extreme values for both lipoprotein and lipidomics measures were truncated in all analyses to third quartile + 8 x interquartile range. Extreme values were rare (**Supplementary Table S1**) and had negligible effects on the results. All the analyses were done with the R software (version 4.2.1).

Simulated data

To illustrate that lipoprotein mediation leads to the results observed with real data, two sets of simulated data were prepared and analysed as illustrated in **Figure 5**. Firstly, we reproduced a biological system in which the circulating lipid (referring to LC-MS/MS data) is purely mediated by lipoprotein particles (referring to NMR data) (**Figure 5B Simulation 1**). This simulation model was set up to reproduce the observed pattern of correlation coefficients between BMI and lipoprotein measures and it was used to simulate a lipid concentration for 5,657 participants (as in NFBC66). The simulated lipoprotein mediation was expressed as a causality graph where BMI affects each lipoprotein measure with a specific strength (w_1 to w_9 in **Figure 5B**), and where the interactions between lipoprotein

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

ORIGINAL UNEDITED MANUSCRIPT

particles are modelled via mutually correlated random variation (ε_1 to ε_9). Both features were necessary to reproduce the observed correlation structure connecting BMI and lipoprotein measures. The simulation was implemented by drawing random values from a normal distribution and multiplying these values with w_1 to w_9 to create nine lipoprotein variables with ε_1 to ε_9 added. The final lipid concentration was an unweighted sum of the nine lipoprotein measures with measurement noise drawn from a normal distribution. In this simulation we did not introduce any direct association between the lipid and BMI.

Secondly, we reproduced a biological system in which the circulating lipid is not associated with lipoprotein metabolism (**Figure 5B Simulation 2**). This was done by drawing a random value for BMI that was directly copied as the lipid concentration but with measurement noise added. In this simulation we introduced a positive association between the lipid and BMI.

Results and discussion

Fundamental molecular context

Figure 1 illustrates the key factors inherently linking the blood serum samples to lipoprotein and lipidomics data. Lipoprotein analyses with NMR proceed directly with the serum sample without any particular sample pre-treatments (**Figure 1A** and **1B**).[29,30] The increased use of NMR in lipoprotein analytics comes from the fact that the technology is directly sensitive for the lipoprotein particle size and can therefore lead to comprehensive quantitative lipid data on various lipoprotein particles (**Figure 1C**).[5,40] The situation is radically different for MS lipidomics techniques. To achieve the precise molecular resolution, the individual molecular constituents must be directly available, i.e., the serum lipoprotein structures (**Figure 1B – 1D**) and heterogeneity in the original sample must be simplified. This is achieved by a lipid extraction procedure[19] (marked **I** in multiple places in **Figure 1**), resulting in a sample that consists of all the individual lipid molecules in the original serum sample as a homogeneous lipid-soluble mixture (**Figure 1J**). This sample is used in the LC-MS/MS analysis with the inherent limitation that all the information on the specific lipoprotein origin of the individual lipid molecules is lost (**Figure 1K**).

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

ORIGINAL UNEDITED MANUSCRIPT

Lipoprotein metabolism is naturally a continuum linking all apoB-containing lipoprotein particles metabolically together.[5,12,41,42] This is illustrated in **Figure 1F** that shows strong positive associations between TC and apoB with all the apoB-containing particle cholesterol measures (VLDL, IDL, LDL, and remnant). In addition, HDL metabolism is tightly interconnected with circulating TG,[1,5,16,43–45] resulting in a strong negative correlation between TG and HDL-C (and apoA-I, the main apolipoprotein component in HDL particles). However, the lipoprotein particles per se are heterogeneous, e.g., in their structure and lipid composition (**Figure 1C** and **1G**). Their contribution to the circulating lipids also varies markedly between the lipid classes, e.g., HDL particles transport only 13% of triglycerides in the bloodstream, while apoB-containing particles transport 87%, of which over 60% resides in VLDL (**Figure 1E** and **1H**). These convoluted metabolic and structural issues[5] are the fundamental source for the perplexity in understanding adjustment effects with lipoprotein measures.

The presented absolute concentration data for the 24 molecular lipid classes in two independent large population studies are unique (**Figure 1L-1N** and **Table 2**). They are in good agreement with the current view on the relative concentrations of plasma lipids.[46] The coherence of the results for the two independent epidemiological cohorts is striking. While similarity for the lipidomics data are expected due to the close correspondence of the lipoprotein profiles in both cohorts (**Table 1** and **Supplementary Figure S1**), this is a pioneering demonstration of the quantitative performance of LC-MS/MS analyses[21–23,47] in large-scale epidemiology.

NMR spectroscopy data on lipoprotein lipids are limited to four major lipid categories, namely TG, FC, CE, and PL. However, this information is available for various lipoprotein categories informing at various stages of metabolism.[5,30] LC-MS/MS lipidomics of plasma samples opens a more prolific view on circulating lipids (**Figure 1K-1N** and **Table 2**). Even so, it is an elementary biological outcome that only a few lipid classes account for most of the circulating lipids. The core of circulating lipoprotein particles is principally consisted of TG and CE, comprising around 60% of circulating lipids, while phosphatidylcholines and FC – key molecular components of lipoprotein surfaces – comprise around 25% of circulating lipids. Free fatty acids and sphingomyelins amount to some 3% each, and

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

lysophosphatidylcholines as well as diglycerides to around a percent of circulating lipids. The other molecular lipid classes contribute less than a percent per class.

In the case of epidemiological studies of serum/plasma, based on the fundamental molecular framework illustrated in **Figure 1**, it is apparent that the lipoprotein measures inevitable affect almost all the lipidomics data associations. In the following we will present novel data on the molecular associations of lipoprotein and lipidomics measures and aim to reveal the most important lipoprotein attributes that influence the lipidomics associations as well as to provide guidelines for adjustments in epidemiological studies.

Molecular clusters and associations within the lipoprotein and lipidomics data

Figure 2 illustrates how the nine key lipoprotein measures associate with the 24 lipidomics lipid classes. The general association characteristics are summarised via four and eight metabolic clusters for the lipoprotein and lipidomics data, respectively. The lipoprotein data clusters combine circulating TG and VLDL-C (cluster no. 1), apoB and remnant-C (cluster no. 2), HDL-C and apoA-I (cluster no. 3), and total, LDL, and IDL cholesterol (cluster no. 4). These clusters are as expected based on the interrelationships of these lipoprotein measures as depicted in **Figure 1F**. Fundamentally, for example, LDL-C is roughly 50% of all circulating cholesterol and remnant particles make up approximately 30% of apoB-containing lipoprotein particles and carry around 80% of TG in the bloodstream.[5] The latter fact is behind the key differences in the association behaviour of remnant-C and LDL-C, particularly with HDL-C and apoA-I.

The association and clustering analyses shown in **Figure 2** for the lipoprotein and lipidomics data are novel. Three clusters for the lipid classes relate to over 90% of the circulating lipid concentrations, namely clusters no. 1 (30.5%), no. 3 (15.2%), and no. 8 (45.0%). Cluster no. 1 consists of five lipid classes: two for oxidised lipids, phosphatidylcholines, free cholesterol, and ceramides. Cluster no. 3 is di- and triglycerides, and cluster no. 8 is sphingomyelins and cholesteryl esters. All these clusters associate strongly and positively with the lipoprotein cholesterol cluster no. 4. Notably, FC and CE measures from lipidomics fundamentally make up the circulating TC and thereby represent

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

key parts of LDL-C as well as other apoB-containing lipoprotein particles. In a similar fashion, the lipidomics TG plus DG represent an independently measured circulating lipoprotein TG; thus, a very high positive correlation between them (lipidomics cluster no. 3) with lipoprotein cluster no. 1 as well as a negative correlation with HDL-C.

The association behaviour of ceramides, as part of lipidomics cluster no. 1, is important to note in relation to the recent interpretations suggesting an independent role for ceramides (and phosphatidylcholines) as biomarkers for cardiovascular disease risk.[48–50] The (lipoprotein) independent role of ceramides (and phosphatidylcholines) comes into question since they strongly and positively associate with all the seven non-HDL-related lipoprotein measures (**Figure 2**), all causal for coronary heart disease.[11]

Free fatty acids form the fourth most abundant lipidomics cluster (4.3%). Apart from a very weak association with HDL-C, they do not associate with any lipoprotein measure. The four other lipidomics clusters represent markedly rarer lipid classes than the aforementioned four (see **Supplementary material**, **Extended results and discussion**).

The associations between lipoprotein and lipidomics data have not been analysed at this level of detail before. Nevertheless, while these associations (as displayed in **Figure 2**) are presented explicitly for the first time, they are mostly what could be foreseen based on the fundamental molecular context (**Figure 1**). These data and findings demonstrate excellent methodological and epidemiological consistency and confirm the compatibility of the LC-MS/MS lipidomics and NMR spectroscopy platforms at large-scale population studies. In the following two sections we will demonstrate that the pivotal epidemiological message of these associations is in the correlation structure and in the non-uniform overlap of the various lipoprotein measures.

Lipoprotein adjustments substantially modify associations between the lipoprotein and lipidomics data

Figure 3 demonstrates how the adjustments for the lipoprotein measures and their combinations impact on the associations between the nine key lipoprotein measures and the 24 lipidomics lipid classes (numerical values are given in **Supplementary Tables S2–S4**).

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Figure 3A reiterates the base results for the associations from **Figure 2** (for the convenience of direct comparisons). Individual adjustments for each of the lipoprotein lipids result in marked and characteristic changes in the association patterns (**Figure 3B–3E**). Overall, the effects of these adjustments are logical. Adjusting for TC (**Figure 3C**) diminishes (or inverts) associations for those lipid classes that are strongly associated with various lipoprotein cholesterol measures (**Figure 3A**) and adjusting for TG (**Figure 3B**) does have the strongest effects on the di- and triglyceride associations in the lipidomics data. The effects for adjusting for HDL-C are rather limited (**Figure 3E**) and those for LDL-C (**Figure 3D**) similar to those for TC (**Figure 3C**).

Adjusting for the combination of the standard clinical lipid measures (TG+TC+HDL-C) results in pronounced weakening of all the associations (**Figure 3F**). Adjustment for TG+LDL-C-HDL-C also results in markedly weakened associations (**Figure 3G**). Adjusting for apoB (**Figure 3H**), apoA-I (**Figure 3I**) and both (**Figure 3J**) results in another type of prominent changes in the associations. The effects of apoB adjustment show characteristics from adjusting for TG as well as for TC, which is logical since the apoB-particles are the main transporters of both triglycerides and cholesterol in the circulation. Please see **Supplementary material**, **Extended results and discussion** for further details.

It would be straightforward to take the coherent depletion of the associations between the lipoprotein and lipidomics measures, as a result of the combined TG+TC+HDL-C (or TG+LDL-C-HDL-C) adjustment, as an indication of an efficient and recommendable way to adjust in lipidomics studies. However, as we will exemplify in the following, this is not the case.

Combined lipoprotein adjustments can unsystematically bias the lipidomics associations

Figure 4 shows that almost all the circulating lipid class concentrations associate with BMI if the model is adjusted only for sex (+ age in YFS). These results are in overall correspondence between those recently reported.[23] However, the detailed effects of lipoprotein metabolism

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

ORIGINAL UNEDITED MANUSCRIPT

on these associations have not been previously analysed in detail. The focus here is solely on these effects as detailed above.

The following discussion depicts the key extreme examples of the effects of lipoprotein adjustments on the outcome associations of the lipidomics measures:

- *Adjusting for an individual lipoprotein measure nullifies the association but combined lipoprotein lipid adjustments show robust associations*

The association of sphingomyelins (SM; lipidomics cluster no. 8 in **Figure 4**) with BMI is a good example for the generation of robust positive results due to the combined lipoprotein adjustments (TG+TC+HDL-C and TG+LDL-C+HDL-C). Instead, individual adjustments for TG and LDL-C abolish the association. Also, the apolipoprotein adjustments have strong effects on the association, implying different inherent association structure between lipoprotein lipid and apolipoprotein measures (**Figure 1F** and **Supplementary Figure 6** that shows the associations of the lipoprotein measures with BMI). A more detailed metabolic interpretation of the SM associations is given in the **Supplementary material**, **Extended results and discussion**.

The association of alkenylphosphatidylethanolamines (PE(P); lipidomics cluster no. 4 in **Figure 4**) is also a representative example of an adjustment leading to positive results. There is neither a baseline association for PE(P) with BMI nor for any of the individual TG, TC, or LDL-C adjustments. Nevertheless, the combined lipoprotein adjustments lead to robust positive associations for PE(P) with BMI. While with the observational data the correct answer is obscured, if a lipoprotein measure exists that alone abolishes the lipid association, it would suggest that if a combination that includes the same lipoprotein measure creates an association, that association is likely incorrect.

- *Combined lipoprotein lipid adjustments nullify the association but adjustments with individual lipoprotein measures show robust associations*

The association of phosphatidylinositols (PI; lipidomics cluster no. 2 in **Figure 4**) with BMI is a good example for the generation of negative results due to the combined lipoprotein adjustments. The baseline association and those adjusted individually for the lipoprotein

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

measures show a robust association for PI with BMI. However, the combined lipoprotein adjustments abolish the association.

The above examples pinpoint the most obvious challenges to combined lipoprotein lipid adjustments in interpreting epidemiological associations of lipidomics measures – the combined adjustments can, inconsistently, result in robust associations when already one lipoprotein measure explains the association and nullify an association that appears robust for adjustments with all the individual lipoprotein measures. **Figure 4** also illustrates various more subtle situations in which the meaning and comparison of the association effects are obscured. This is the first time that these effects are identified and compared in detail, though they might not come as a surprise in an epidemiological context.[11,25,26,45,51]

In addition to the meta-analyses for both cohorts and the adjustments with the 9-measure clinical lipoprotein panel (**Figure 4**), we also assessed the effects of adjusting for lipoprotein(a) (Lp(a)) for the associations of the 24 most abundant lipidomics lipid classes with BMI in YFS as illustrated in **Supplementary Figure S7** (Lp(a) measurements are not available in NFBC66). Lp(a) is typically not involved in the clinical lipoprotein measurements since its circulating concentration is very low in most individuals. Lp(a) concentration is also known to be independent of the main lipoprotein concentrations.[32] In concordance with these, the Lp(a) adjustments had no effect on the associations of the lipidomics lipid classes with BMI. Thus, only if a notable portion of the entire pool of a circulating lipid is mediated via certain lipoprotein particles, the corresponding adjustments will have strong effects on the associations. This is further exemplified by the results with simulated data in the next paragraph.

Simulated data for lipoprotein mediation and non-mediation of circulating lipids

The simulated mediation scenarios depicted in **Figure 5**, i.e., pure mediation and no mediation by lipoprotein particles, clearly illustrate the complexity introduced by the highly correlated lipoprotein metabolism to the associations of their lipid constituents. For the simulated lipid 1 we did not introduce any direct association between the lipid and BMI.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

However, the results show that the lipoprotein mediation per se can create a positive association. In addition, the lipoprotein adjustment can either enhance (apoA-I) or reverse (apoB) the (non-existent) association. It is (correctly) nullified by the TG adjustment. In a real observational situation there is no way of knowing which one would be correct.

Overall, we wish to emphasize that the simulation is a pure linear mediation model while the lipoprotein system is more complex with non-linear bi-directional causality (**Figure 5A**). As we do not know the true causal structure in detail, it is unfeasible to provide exact recipes for how to handle the complexity when choosing adjustments. Observational data are also affected by the likely presence of unknown epidemiological confounding. Nevertheless, the simulated data with the marked differences between the lipoprotein adjustments for the lipoprotein mediated and non-mediated circulating lipid (**Figure 5B**) complement the observational findings and concur with the key role of lipoprotein mediation in the interpretation of lipidomics results. This calls for multiple types of adjustments and case-by-case interpretation based on metabolic characteristics of the adjusting measures.

Conclusions

These analyses demonstrate that the fundamental lipoprotein-centred transport of circulating lipids is such a complex and intertwined molecular system that it puts doubt on any "lipoprotein-independent" interpretation of lipidomics results and even for the search for such interpretations. These results call for awareness of the inherent structural and mediating role of lipoprotein particles in systemic lipid metabolism. Incorporation of lipoprotein data in an appropriate manner is crucial for correct analyses and interpretations of epidemiological (and genetic) lipidomics data. The combined lipoprotein adjustments are those that accentuate the inherent complexities. It would be more sensible to perform adjustments separately with a few key individual lipoprotein measures, e.g., TG, LDL-C, HDL-C, and apoB. Albeit often still strenuous, the effects of these rather specific adjustments can, at least in principle, be interpreted based on the known compositional characteristics of lipoprotein particles and their overall roles in lipoprotein metabolism.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Sources of funding: The NFBC has been supported, for instance, by EU, Academy of Finland, PREcisE, Joint Programming Initiative a Healthy Diet for a Healthy Life (no. 655), UK Medical Research Council, Biotechnology and Biological Sciences Research Council (MR/S03658X/1), and European Regional Development Fund Grant no. 539/2010 A31592. The Young Finns Study has been financially supported by the Academy of Finland: grants 356405, 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117797 (Gendi), and 141071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS and grant 848146 for To Aition); European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation; Finnish Society of Clinical Chemistry; the Cancer Foundation Finland; pBETTER4U_EU (Preventing obesity through Biologically and bEhaviorally Tailored inTERventions for you; project number: 101080117); CVDLink and the Jane and Aatos Erkko Foundation. M.A.-K. was supported by a research grant from the Sigrid Juselius Foundation, the Finnish Foundation for Cardiovascular Research, and the Research Council of Finland (grant no. 357183).

Role of Sponsor: The study funders were not involved in the design of the study; the collection, analysis, and interpretation of data; writing the report; and did not impose any restrictions regarding the publication of the report.

Author Contributions. S.Z. performed the statistical analyses, prepared the figures, interpreted the results, and helped in writing the manuscript. P.O. helped in statistical analyses. J.K., M.-R.J., M.K., T.L., J.V., O.T.R., and M.A.-K. provided clinical and metabolomics data. V.-P.M. interpreted the results, supervised the statistical analyses, and helped in writing the manuscript. M.A.-K. conceived, designed, and led the study, interpreted the results, supervised the statistical analyses, and wrote the manuscript. All authors discussed the results, commented, and contributed to the manuscript, and approved the final version.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

M.A.-K. is the guarantor of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analyses.

Competing Interests. The authors declare no conflicts of interest.

Data availability. The datasets used in the current study are available from the cohorts through application process for researchers who meet the criteria for access to confidential data: https://www.oulu.fi/nfbc/ (NFBC66) and http://youngfinnsstudy.utu.fi (YFS). Regarding the YFS data the Ethics committee has concluded that under applicable law, the data from this study cannot be stored in public repositories or otherwise made publicly available. The data controller may permit access on case-by-case basis for scientific research, not however to individual participant level data, but aggregated statistical data, which cannot be traced back to the individual participants' data.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

References

1. Borén J, Taskinen M-R, Björnson E, Packard CJ. Metabolism of triglyceride-rich lipoproteins in health and dyslipidaemia. *Nat Rev Cardiol* 2022;**19**:577–592.

2. Eckardstein A von, Nordestgaard BG, Remaley AT, Catapano AL. High-density lipoprotein revisited: biological functions and clinical relevance. *Eur Heart J* 2023;**44**:1394–1407.

3. Kumpula LS, Kumpula JM, Taskinen M-R, Jauhiainen M, Kaski K, Ala-Korpela M. Reconsideration of hydrophobic lipid distributions in lipoprotein particles. *Chem Phys Lipids* 2008;**155**:57–62.

4. Hevonoja T, Pentikäinen MO, Hyvönen MT, Kovanen PT, Ala-Korpela M. Structure of low density lipoprotein (LDL) particles: basis for understanding molecular changes in modified LDL. *Biochim Biophys Acta* 2000;**1488**:189–210.

5. Ala-Korpela M, Zhao S, Järvelin M-R, Mäkinen V-P, Ohukainen P. Apt interpretation of comprehensive lipoprotein data in large-scale epidemiology: disclosure of fundamental structural and metabolic relationships. *Int J Epidemiol* 2022;**51**:996–1011.

6. Cham BE. Importance of apolipoproteins in lipid metabolism. *Chem Biol Interact* 1978;**20**:263–277.

7. Wiesner P, Leidl K, Boettcher A, Schmitz G, Liebisch G. Lipid profiling of FPLC-separated lipoprotein fractions by electrospray ionization tandem mass spectrometry. *J Lipid Res* 2009;**50**:574–585.

8. Kontush A, Chapman MJ. Lipidomics as a Tool for the Study of Lipoprotein Metabolism. *Curr Atheroscler Rep* 2010;**12**:194–201.

9. Goldstein JL, Brown MS. A Century of Cholesterol and Coronaries: From Plaques to Genes to Statins. *Cell* 2015;**161**:161–172.

10. Ala-Korpela M. The culprit is the carrier, not the loads: cholesterol, triglycerides and apolipoprotein B in atherosclerosis and coronary heart disease. *Int J Epidemiol* 2022;**51**:996–1011.

11. Schmidt AF, Joshi R, Gordillo-Marañón M, et al. Biomedical consequences of elevated cholesterol-containing lipoproteins and apolipoproteins on cardiovascular and non-cardiovascular outcomes. *Commun Med* 2023;**3**:9.

12. Würtz P, Kangas AJ, Soininen P, et al. Lipoprotein subclass profiling reveals pleiotropy in the genetic variants of lipid risk factors for coronary heart disease: a note on Mendelian randomization studies. *J Am Coll Cardiol* 2013;**62**:1906–1908.

13. Holmes MV, Millwood IY, Kartsonaki C, et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *J Am Coll Cardiol* 2018;**71**:620–632.

14. Salinas CAA, Chapman MJ. Remnant lipoproteins: are they equal to or more atherogenic than LDL? *Curr Opin Lipidol* 2020;**31**:132–139.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

15. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017;**14**:577–590.

16. Richardson TG, Sanderson E, Palmer TM, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLOS Med* 2020;**17**:e1003062.

17. Karjalainen MK, Holmes MV, Wang Q, et al. Apolipoprotein A-I concentrations and risk of coronary artery disease: A Mendelian randomization study. *Atherosclerosis* 2020;**299**:56–63.

18. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* 2011;**40**:387–426.

19. Alshehry ZH, Barlow CK, Weir JM, Zhou Y, McConville MJ, Meikle PJ. An Efficient Single Phase Method for the Extraction of Plasma Lipids. *Metabolites* 2015;**5**:389–403.

20. Julkunen H, Cichońska A, Tiainen M, et al. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat Commun* 2023;**14**:604.

21. Huynh K, Barlow CK, Jayawardana KS, et al. High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. *Cell Chem Biol* 2019;**26**:71-84.e4.

22. Beyene HB, Olshansky G, Giles C, et al. Lipidomic Signatures of Changes in Adiposity: A Large Prospective Study of 5849 Adults from the Australian Diabetes, Obesity and Lifestyle Study. *Metabolites* 2021;**11**:646.

23. Beyene HB, Olshansky G, T Smith AA, et al. High-coverage plasma lipidomics reveals novel sex-specific lipidomic fingerprints of age and BMI: Evidence from two large population cohort studies. *PLoS Biol* 2020;**18**:e3000870.

24. Hornburg D, Wu S, Moqri M, et al. Dynamic lipidome alterations associated with human health, disease and ageing. *Nat Metab* 2023;**5**:1578–1594.

25. Zwieten A van, Tennant PWG, Kelly-Irving M, Blyth FM, Teixeira-Pinto A, Khalatbari-Soltani S. Avoiding overadjustment bias in social epidemiology through appropriate covariate selection: a primer. *J Clin Epidemiol* 2022;**149**:127–136.

26. Borges MC, Haycock PC, Zheng J, et al. Role of circulating polyunsaturated fatty acids on cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid genetic association data from over 114,000 UK Biobank participants. *BMC Med* 2022;**20**:210.

27. Nordström T, Miettunen J, Auvinen J, et al. Cohort Profile: 46 years of follow-up of the Northern Finland Birth Cohort 1966 (NFBC1966). *Int J Epidemiol* 2022;**50**:1786–1787j.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

28. Raitakari OT, Juonala M, Rönnemaa T, et al. Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol* 2008;**37**:1220–1226.

29. Soininen P, Kangas AJ, Würtz P, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* 2009;**134**:1781–1785.

30. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 2015;**8**:192–206.

31. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol* 2017;**186**:1084–1096.

32. Kettunen J, Demirkan A, Würtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016;**7**:11122.

33. Chambers JC, Zhang W, Sehmi J, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 2011;**43**:1131–1138.

34. Locke AE, Steinberg KM, Chiang CWK, et al. Exome sequencing of Finnish isolates enhances rare-variant association power. *Nature* 2019;**572**:323–328.

35. Inouye M, Kettunen J, Soininen P, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol* 2010;**6**:441.

36. Würtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015;**131**:774–785.

37. Julkunen H, Cichońska A, Slagboom PE, Würtz P, Nightingale Health UK Biobank Initiative. Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *eLife* 2021;**10**:e63033.

38. Bell JA, Richardson TG, Wang Q, et al. Effects of general and central adiposity on circulating lipoprotein, lipid, and metabolite levels in UK Biobank: A multivariable Mendelian randomization study. *Lancet Reg Health Eur* 2022;**21**:100457.

39. Karjalainen MK, Karthikeyan S, Oliver-Williams C, et al. Genome-wide characterization of circulating metabolic biomarkers. *Nature* 2024;**628**:130–138.

40. Lounila J, Ala-Korpela M, Jokisaari J, Savolainen M, Kesäniemi A. Effects of orientational order and particle size on the NMR line positions of lipoproteins. *Phys Rev Lett* 1994;**72**:4049–4052.

41. Tukiainen T, Kettunen J, Kangas AJ, et al. Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. *Hum Mol Genet* 2012;**21**:1444–1455.

42. Ala-Korpela M. 1H NMR spectroscopy of human blood plasma. *Prog Nucl Magn Reson Spectrosc* 1995;**27**:475–554.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

43. Kettunen J, Holmes MV, Allara E, et al. Lipoprotein signatures of cholesteryl ester transfer protein and HMG-CoA reductase inhibition. *PLOS Biol* 2019;**17**:e3000572.

44. Phillips AN, Smith GD. How independent are “independent” effects? relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol* 1991;**44**:1223–1231.

45. Davey Smith G, Phillips AN. Correlation without a cause: an epidemiological odyssey. *Int J Epidemiol* 2020;**49**:4–14.

46. Quehenberger O, Dennis EA. The Human Plasma Lipidome. *N Engl J Med* 2011;**365**:1812–1823.

47. Tham YK, Bernardo BC, Huynh K, et al. Lipidomic Profiles of the Heart and Circulation in Response to Exercise versus Cardiac Pathology: A Resource of Potential Biomarkers and Drug Targets. *Cell Rep* 2018;**24**:2757–2772.

48. Laaksonen R, Ekroos K, Sysi-Aho M, et al. Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. *Eur Heart J* 2016;**37**:1967–1976.

49. Hilvo M, Meikle PJ, Pedersen ER, et al. Development and validation of a ceramide- and phospholipid-based cardiovascular risk estimation score for coronary artery disease patients. *Eur Heart J* 2020;**41**:371–380.

50. Mundra PA, Barlow CK, Nestel PJ, et al. Large-scale plasma lipidomic profiling identifies lipids that predict cardiovascular events in secondary prevention. *JCI Insight* 2018;**3**(17).

51. Smith GD, Phillips A. Declaring independence: why we should be cautious. *J Epidemiol Community Health* 1990;**44**:257–258.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Tables

Table 1. Clinical characteristics and lipoprotein measures in NFBC66 and YFS.

Clinical characteristics	**NFBC66**	**YFS**
Number of participants	5,657	2,036
Number of females (%)	3,168 (56.0 %)	1,109 (54.5 %)
Age (year)	46.6 [46.2 - 47.1]	43.0 [37.0 - 46.0]
BMI (kg/m2)	26.1 [23.5 - 29.3]	25.8 [23.1 - 29.1]
Lipoprotein lipids (mmol/L)		
Total triglycerides	1.13 [0.83 - 1.61]	1.09 [0.79 - 1.58]
Total cholesterol	5.62 [4.97 - 6.35]	5.25 [4.60 - 5.98]
VLDL cholesterol	0.80 [0.63 - 1.02]	0.70 [0.55 - 0.92]
IDL cholesterol	0.92 [0.78 - 1.08]	0.84 [0.71 - 0.99]
LDL cholesterol	2.18 [1.81 - 2.63]	2.00 [1.63 - 2.41]
Remnant cholesterol	1.74 [1.45 - 2.09]	1.57 [1.29 - 1.89]
HDL cholesterol	1.63 [1.36 - 1.92]	1.61 [1.35 - 1.92]
Apolipoproteins (g/L)		
Apolipoprotein B	1.03 [0.88 - 1.21]	1.00 [0.82 - 1.13]
Apolipoprotein A-I	1.73 [1.59 - 1.89]	1.67 [1.54 - 1.84]

The values are median [Q1 – Q3]. BMI, body mass index; VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein. Remnant cholesterol refers to (non-LDL, non-HDL) cholesterol.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Table 2. LC-MS/MS lipidomics measures for the total concentrations of the 24 most abundant circulating lipid classes (µmol/L) in NFBC66 and YFS.

	NFBC66	**YFS**
Cluster no. 1 (mol %)	30.4 %	30.7 %
Oxidised phosphatidylcholine/ lysophosphatidylcholines [Ox-PC/LPC]	0.12 [0.10 – 0.14]	0.13 [0.11 – 0.16]
Oxidised cholesteryl esters [Ox-CE]	5.45 [4.27 – 7.26]	5.62 [4.48 – 7.02]
Phosphatidylcholines [PC]	1901.3 [1667.4 - 2133.3]	2062.7 [1853.7 - 2286.8]
Free cholesterol [Free C]	933.8 [828.4 - 1044.6]	906.9 [793.5 - 1034.9]
Ceramides [Cer]	9.95 [8.21 - 12.0]	10.4 [8.53 - 12.5]
Cluster no. 2 (mol %)	1.1 %	1.2 %
Phosphatidylinositols [PI]	73.1 [61.1 - 87.3]	76.7 [64.8 - 90.7]
Phosphatidylethanolamines [PE]	26.1 [19.7 - 35.2]	28.2 [20.6 - 39.6]
Alkyldiacylglycerols [TG(O)]	3.12 [2.69 - 3.65]	3.14 [2.69 - 3.69]
Cluster no. 3 (mol %)	15.1 %	15.6 %
Diacylglycerols [DG]	88.7 [66.7 - 124.3]	92.9 [65.9 - 136.6]
Triacylglycerols [TG]	1233 [938.6 - 1624]	1337 [1076.2 – 1664.9]
Cluster no. 4 (mol %)	1.8 %	1.9 %
Alkenylphosphatidylethanolamines [PE(P)]	39.4 [32.2 - 47.8]	42.6 [34.2 - 52.8]
Alkylphosphatidylethanolamines [PE(O)]	3.59 [2.86 - 4.56]	3.83 [2.95 - 4.92]
Alkylphosphatidylcholines [PC(O)]	63.9 [54.1 - 75.2]	70.9 [59.5 - 83.8]
GM3 gangliosides [GM3]	2.58 [2.16 - 3.05]	2.66 [2.26 - 3.05]
Alkenylphosphatidylcholines [PC(P)]	40.5 [33.3 – 49.0]	46.0 [37.3 - 54.9]
Dehydrocholesterol esters [DE]	11.0 [9.18 - 13.1]	12.6 [9.47 - 16.1]
Cluster no. 5 (mol %)	1.9 %	2.0 %
Lysophosphatidylcholines [LPC]	166.4 [136.9 - 200.6]	189.6 [157.2 - 222.2]
Lysophosphatidylethanolamines [LPE]	4.53 [3.76 - 5.48]	5.71 [4.63 - 6.87]
Cluster no. 6 (mol %)	0.1 %	0.1 %
Dihexosylceramides [Hex2Cer]	3.54 [2.88 - 4.35]	4.04 [3.31 - 4.91]
Monohexosylceramides [HexCer]	4.91 [4.06 - 5.92]	5.03 [4.22 - 5.97]
Lysoalkylphosphatidylcholines [LPC(O)]	2.51 [2.1 - 2.96]	2.78 [2.30 - 3.28]
Cluster no. 7 (mol %)	4.4 %	4.1 %
Free fatty acids [FFA]	389 [316.5 - 479.8]	372.9 [300.4 - 473.5]
Cluster no. 8 (mol %)	45.2 %	44.4 %
Sphingomyelins [SM]	372.5 [317.9 - 428.5]	382.6 [330.5 - 432.5]
Cholesteryl esters [CE]	3861 [3424 - 4330]	3932 [3479 - 4393]

The clusters refer to **Figure 2**. The values are median [Q1 – Q3].

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Figures and figure captions

Figure 1. An overall illustration on the fundamental factors inherently linking the serum samples to lipoprotein and lipidomics data. **(A)** Blood serum samples are analysed as such in NMR spectroscopy **(B)** but the LC-MS/MS lipidomics analyses **(K)** call for a lipid extraction process **(I)** before experimentation. **(B–D)** In intact serum lipid molecules reside within a complex mixture of lipoprotein particles (and free fatty acids, some lysophosphatidylcholines, and some other lyso-type of lipid molecules with albumin). **(E–H)** Lipoprotein particles can be characterised by different subfractions, each of which has a

ORIGINAL UNEDITED MANUSCRIPT

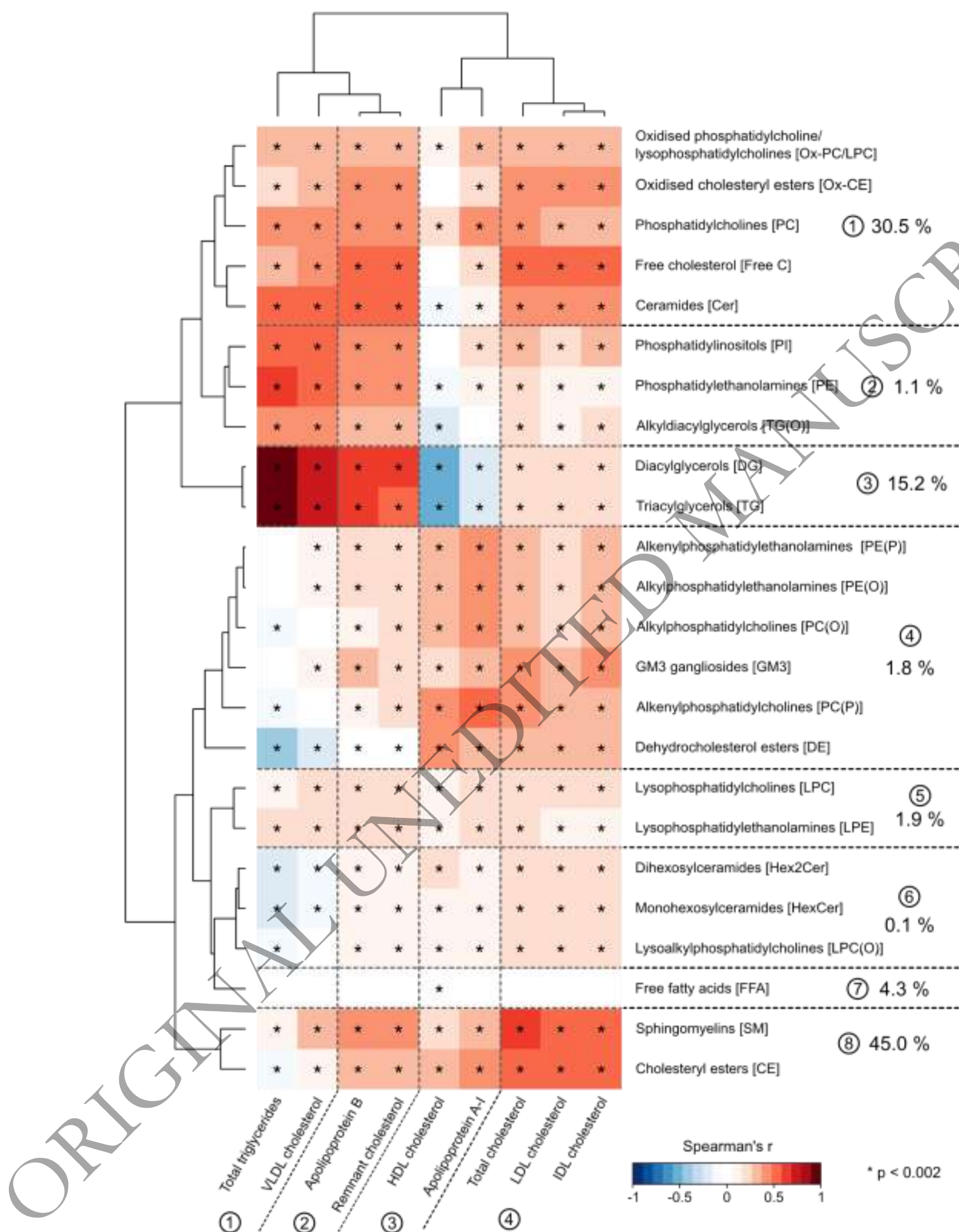
Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

typical lipid content and associations with other subfractions and lipid measures. **(I)** Lipid extraction is a key process to make serum samples analysable by LC-MS/MS lipidomics **(K)**. This results in a sample that consists of all the individual lipid molecules in the original serum sample as a homogeneous lipid-soluble mixture with all the lipoprotein particles broken and free fatty acids as well as the lyso-type of lipid molecules dissolved from albumin **(J)**; all the information on the specific lipoprotein origin of the individual lipid molecules is lost **(K)**. However, LC-MS/MS lipidomics leads to a very rich and detailed analyses on serum lipids. In **(L–N)** unique results are illustrated for the 24 most abundant molecular lipid classes in two independent large population studies, namely the Northern Finland Birth Cohort 1966 (NFBC66) with 5,657 participants and the Cardiovascular Risk in Young Finns Study (YFS) with 2,036 participants. The lipid class data are shown as the percentage of total lipids (mol%) in serum. The box plots represent medians with interquartile ranges and the minimum (Q1 - 1.5*IQR) and maximum (Q3 + 1.5*IQR) values with potential outliers. The absolute concentration data for the lipoprotein measures are given in **Table 1** and for the lipidomics measures in **Table 2**. Please note the different size-scale in **(B)** and **(D & K)**.

VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein; ALB, albumin; NMR, nuclear magnetic resonance; LC-MS/MS, Liquid chromatography–mass spectrometry; ApoA-I, apolipoprotein A-I; ApoB, apolipoprotein B; TG, triglycerides; C, cholesterol; FC, free cholesterol; CE, cholesteryl esters; PL, phospholipids; PC, phosphatidylcholines; SM, sphingomyelins; LPC, lysophosphatidyl-cholines; PI, phosphatidylinositols; PC(O), alkylphosphatidylcholines; PC(P), alkenylphos-phatidylcholines; PE(P), alkenylphosphatidylethanolamines; PE, phosphatidylethanolamines; Cer(d), ceramides; HexCer, monohexosylceramides; LPE, lysophosphatidylethanolamines; PE(O), alkylphosphatidylethanolamines; Hex2Cer, dihexosylceramides; GM3, GM3 gangliosides; LPC(O), lysoalkylphosphatidylcholines; Ox-PC/LPC, oxidised phosphatidyl-cholines/lysophosphatidylcholines; DG, diacylglycerols; DE, dehydrocholesterol esters; Ox-CE, oxidised cholesteryl esters; TG(O), alkyldiacylglycerols; FFA, free fatty acids.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025



ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

Figure 2. Associations of the nine key lipoprotein measures with the 24 most abundant lipidomics lipid classes as indicated by partial Spearman's rank correlations. The data are from two independent large population studies, NFBC66 with 5,657 participants and YFS

with 2,036 participants. All correlations were adjusted for sex in both cohorts and age in YFS. Both cohorts were analysed separately and then meta-analysed. The two-dimensional hierarchical clustering is based on the meta-analysed results for both cohorts, and the resulting ordering is preserved in all the following heat maps in **Figure 3**. The general association characteristics are summarised via four and eight metabolic clusters for the lipoprotein and lipidomics data, respectively. The percentages shown for the lipidomics clusters depict the contribution of each cluster to the circulating total lipid concentration. P-value <0.002 is marked with an asterisk in the map to indicate a multiple testing corrected association. The abbreviations are as explained in the caption for **Figure 1**.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

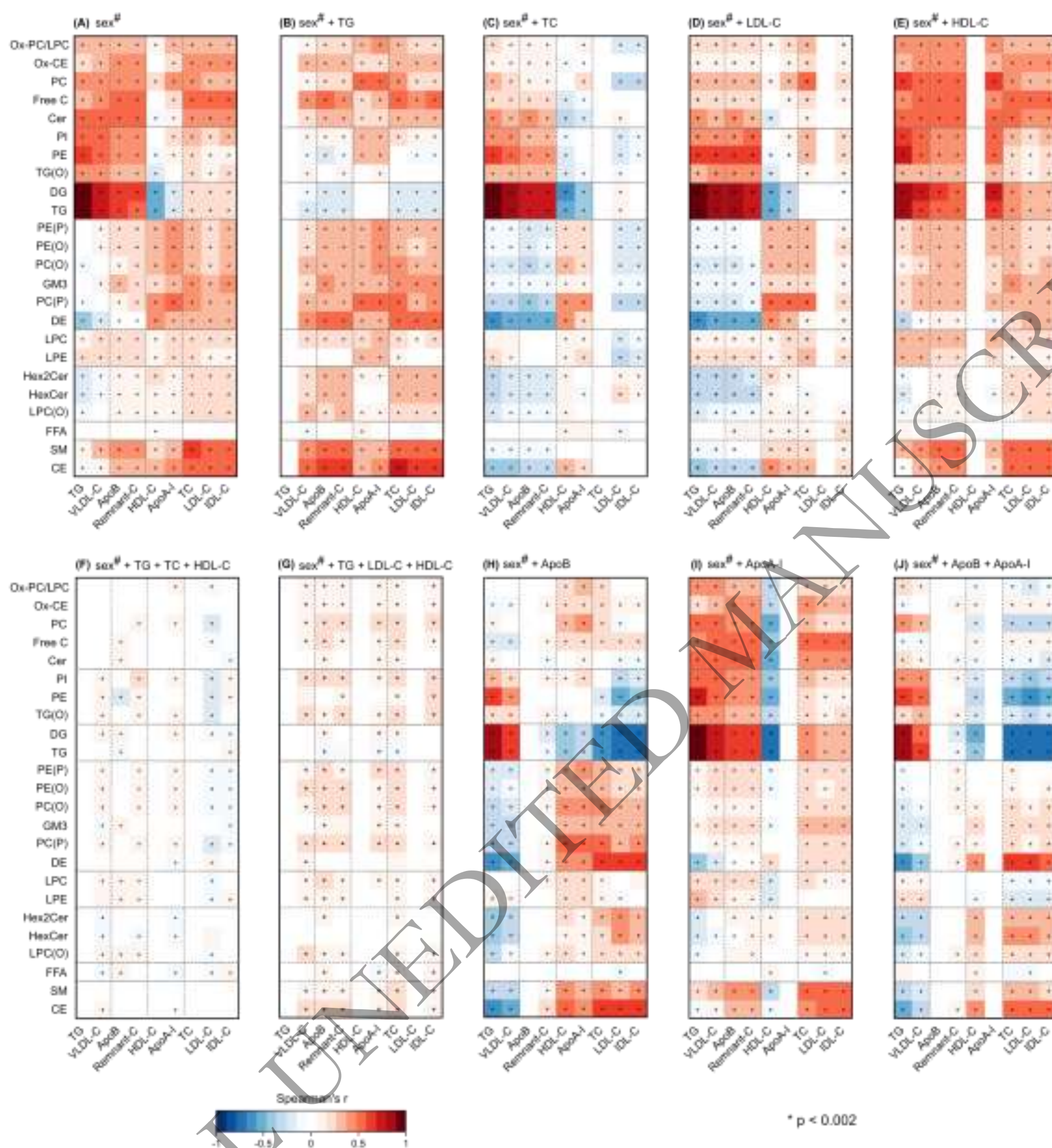


Figure 3. Associations of the nine key lipoprotein measures with the 24 most abundant lipidomics lipid classes adjusted for different lipoprotein measures and their combinations. The data are from two independent large population studies, NFBC66 with 5,657 participants and YFS with 2,036 participants. The lipoprotein adjustments in **(B–J)** are in addition to the adjustment for sex in both cohorts and age in YFS (marked as sex#) **(A)**. Both cohorts were analysed separately and then meta-analysed. All correlations are partial Spearman's rank correlations, and the organisation of the heat maps is from **Figure 2**. P-value <0.002 is marked with an asterisk in the maps to indicate a multiple testing corrected association. The abbreviations are as explained in the caption for **Figure 1**.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

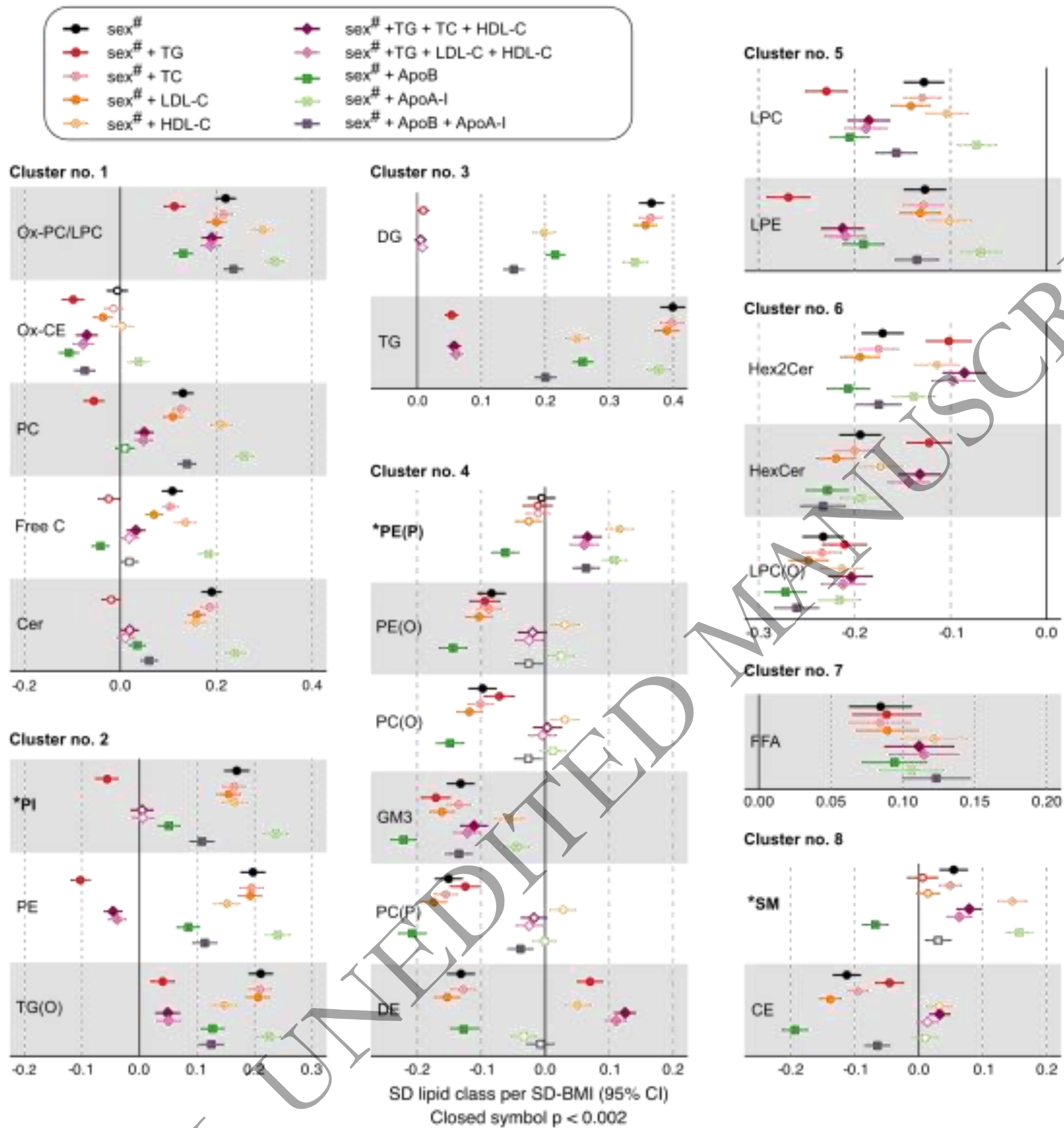
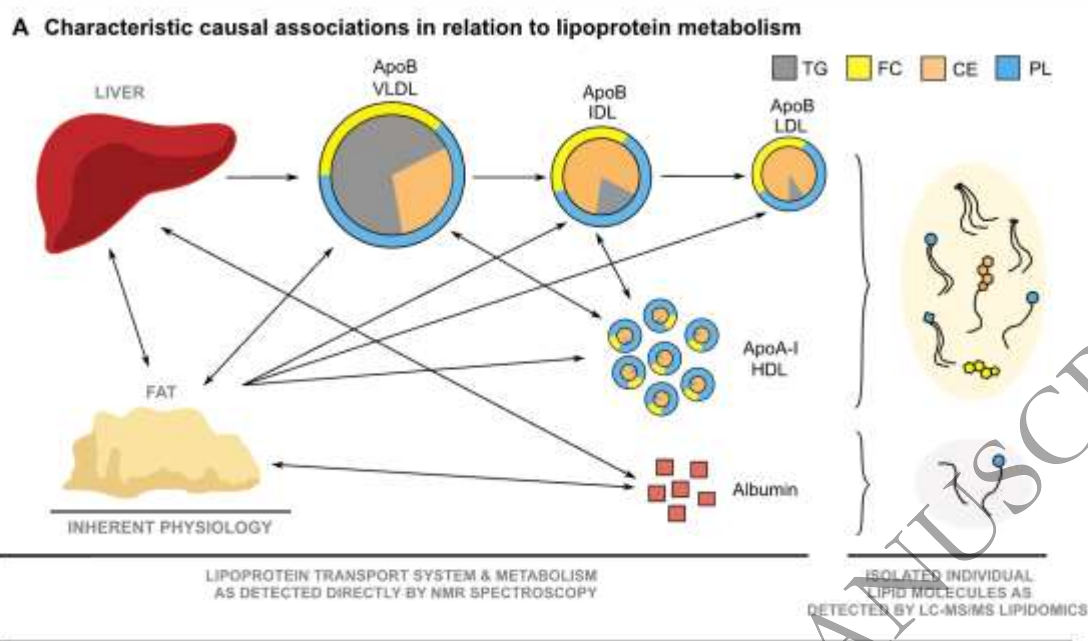
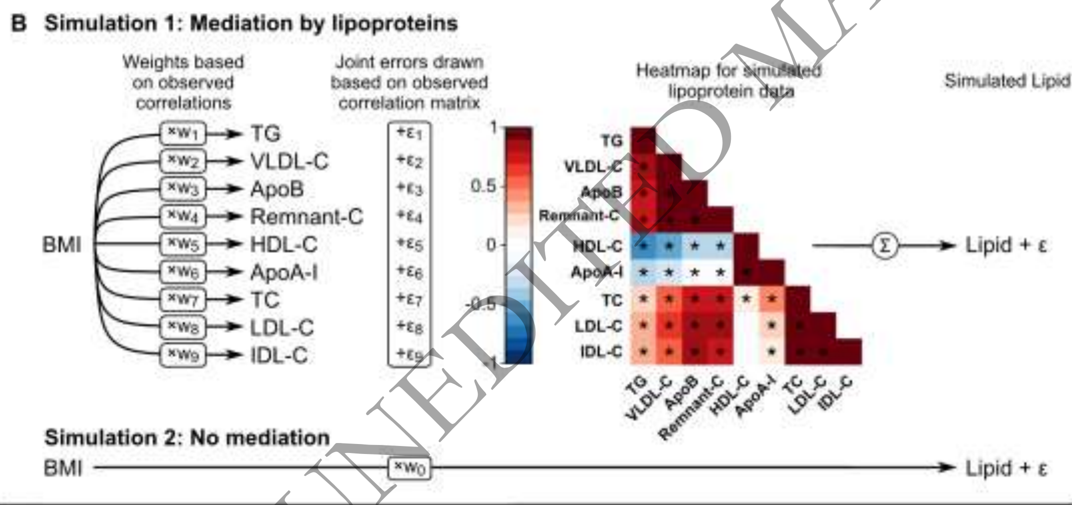


Figure 4. Illustration of the effects of various lipoprotein adjustments on the associations of the 24 most abundant lipidomics lipid classes with BMI via linear regression analyses. The data are from two independent large population studies, NFBC66 with 5,657 participants and YFS with 2,036 participants. Both cohorts were analysed separately and then meta-analysed. Association magnitudes are reported in standard deviation (SD) units throughout to ease the comparison across multiple measures with markedly varying absolute concentrations (**Table 2**). Sex# refers to adjusting for sex in both cohorts and age in YFS. P-value <0.002 is marked with an asterisk in the maps to indicate a multiple testing corrected association. The results for the emphasised lipid classes, namely **PI**, **PE(P)**, and **SM** are discussed in more detail in the text. The abbreviations are as explained in the caption for **Figure 1**.

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

ORIGINAL UNEDITED MANUSCRIPT





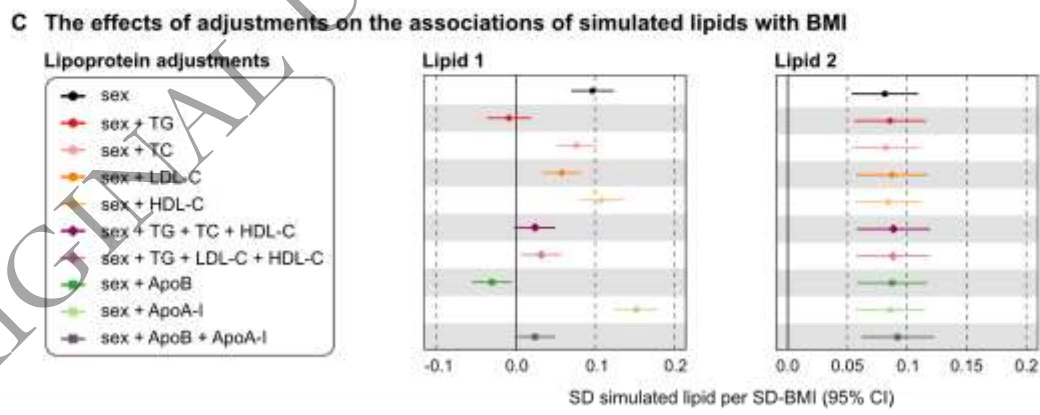


Figure 5. A. Illustration of causal associations between various tissues and lipoprotein metabolism as well as causal associations between various lipoprotein particles. The arrowheads represent a causal direction. This figure is a schematic representation and

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025

ORIGINAL UNEDITED MANUSCRIPT

should not be interpreted as a formal directed acyclic graph. **B.** The basis for simulated lipid data representing pure lipoprotein mediation (Simulation 1 with no simulated direct association between lipid 1 and BMI) and no mediation by lipoprotein particles (Simulation 2 with simulated direct positive association between lipid 2 and BMI). **C.** The effects of various lipoprotein adjustments on the associations of the two simulated lipids with BMI via linear regression analyses.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/aje/advance-article/doi/10.1093/aje/kwae445/7912636 by Turku University user on 20 January 2025