Lipoprotein and metabolite associations to breast cancer risk in the HUNT2 study

Julia Debik[1], Hartmut Schaefer[2], Trygve Andreassen[1], Feng Wang[1,3], Fang Fang[2], Claire Cannet[2], Manfred Spraul[2], Tone F. Bathen[1,4], Guro F. Giskeødegård[3,4,5]

[1] Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway

[2] Bruker BioSpin AIC Division, Ettlingen, Germany

[3] Clinic of Surgery, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

[4] Department of Radiology and Nuclear Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

[5] K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway

Corresponding authors:

Julia Debik, Ph.D.
NTNU, Faculty of Medicine and Health Sciences, Department of Circulation and Medical Imaging,
Postboks 8905, 7491 Trondheim, Norway
Phone: +47 4161781 ; E-mail: julia.b.debik@ntnu.no

Guro F. Giskeødegård, Ph.D.
NTNU, Department of Public Health and Nursing, Department of Public Health and Nursing,
Postboks 8905, 7491 Trondheim, Norway
Phone: +47 73598916 ; E-mail: guro.giskeodegard@ntnu.no

Keywords: Breast cancer; Metabolomics; Prediagnostic serum; NMR spectroscopy; Lipoproteins;

Word count: 4699

Abstract

Background: The aim of this study was to gain increased understanding of the etiology of breast cancer, by investigating possible associations between serum lipoprotein subfractions and metabolites and long-term risk of developing the disease.

Methods: From a cohort of 65 200 participants within the Trøndelag Health Study (HUNT study), we identified all women who developed breast cancer within a 22-year follow-up period. Using nuclear magnetic resonance (NMR) spectroscopy, 28 metabolites and 89 lipoprotein subfractions were quantified from prediagnostic serum samples of future breast cancer patients and matching controls (n = 1199 case-control pairs).

Results: Among premenopausal women (554 cases) 14 lipoprotein subfractions were associated with long-term breast cancer risk. In specific, different subfractions of VLDL particles (in particular VLDL-2, VLDL-3 and VLDL-4) were inversely associated with breast cancer. Additionally inverse associations were detected for total serum triglyceride levels and HDL-4 triglycerides. No significant association was found in postmenopausal women.

Conclusions: We identified several associations between lipoprotein subfractions and long-term risk of breast cancer in premenopausal women. Inverse associations between several VLDL subfractions and breast cancer risk were found, revealing an altered metabolism in the endogenous lipid pathway many years prior to a breast cancer diagnosis.

Background

Breast cancer is the most common cancer diagnosis among women, and the incidence rate is following an increasing trend [1]. The overall breast cancer survival rates have increased bringing the 5-year survival up to 90.7% in Norway [2], possibly as a result of tailored treatment strategies and earlier detection due to the establishment of screening programs . However, despite the high overall survival rate,the prognosis greatly depends on the stage of the disease at diagnosis, in addition to treatment efficacy. A better understanding of the etiology of the disease and the biological mechanisms leading to disease could reveal methods for disease prevention and early detection.

Cancer cells have a reprogrammed metabolism for conversion of nutrients to biomass while maintaining a high energy production. This phenomenon is increasingly recognized as a source for biomarkers for early detection. The serum metabolome is affected by the preceding levels of the omics cascade and external factors, providing a detailed snapshot of the current state of the organism [3, 4]. Significant changes in the metabolism of breast cancer patients have been described, both in tumor tissue and biofluids [5-8]. Furthermore, subtle differences in metabolic composition of prediagnostic serum samples have been associated with breast cancer risk [9-16].

Lipids, playing a critical role in cell signaling and membrane formation, have altered levels in many types of cancer [17, 18]. However, the mechanisms governing dysregulated lipid metabolism in cancer development are not fully understood. The two main forms of circulating lipids in the body are triglycerides and cholesterol, which are transported through the bloodstream in lipoproteins. Lipoproteins are particles with an inner core, mainly composed of triglycerides and cholesteryl esters, surrounded by a hydrophilic membrane consisting of free cholesterol, phospholipids and apolipoproteins [19]. There are five main fractions of circulating lipoproteins, ranging from very-low (VLDL) to high-density (HDL) lipoproteins, each with its own characteristic protein and lipid composition. Traditional lipoprotein measurements, however, do not capture the delicate density range within the main fractions. Detailed characterization of the different lipoprotein subfractions and

their content, possible through nuclear magnetic resonance (NMR) spectroscopy, may give important biological information on early breast cancer development.

In this study we aimed to identify associations between lipoprotein subfractions and circulating metabolites in prediagnostic serum samples and breast cancer risk, and to gain insight into the etiology of the disease. Our study aims were accomplished by a case-control study nested in the Trøndelag Health Study (the HUNT study), with samples taken up to 22 years before breast cancer diagnosis.

Methods

Sample Collection and Experimental Design

The Trøndelag Health Study (the HUNT study) is a longitudinal population health study in Norway, including 230 000 participants. It includes a database of questionnaire data, clinical measurements, and biological materials. So far four health surveys have been completed, HUNT1 (1984-86), HUNT2 (1995-97), HUNT3 (2006-08) and HUNT4 (2017-19) [20] in addition to an adolescent part (13 – 19 years). HUNT2 was the first study to include biological material. By matching data from HUNT2 with the Norwegian Cancer Registry, we have identified all participants of HUNT2 that developed breast cancer between data collection and follow-up in 2019 (n = 1208). The mean time from blood collection to a breast cancer diagnosis was 11.7 years (range 0-22 years), while controls had at least 22 years follow-up. For each case, a participant that remained breast cancer free during follow-up was randomly selected as a control, matched for age at inclusion into HUNT2 in intervals of 5 years. To avoid bias towards controls that are healthier than the general population, the cases may have developed another type of cancer, and we did not require the controls to survive the whole follow-up period. This ensured that the controls are representative of true controls. Relevant clinical variables were selected from the HUNT2 databank for both the cases and controls, while cancer-specific variables were retrieved from the Norwegian Cancer Registry. All participants have completed a

written informed consent form, and the study was approved by the Ethics Committee of Central Norway (REK numbers #1995/8395 and #2017/2231).

NMR Experiments

Serum samples were collected in the years 1995-97 and were stored at -80°C until analysis. After thawing at room temperature, 150 µL serum was mixed with 150 µL buffer [D_2O (20% in H_2O) with 0.075 M Na_2HPO_4, 6 mM NaN_3, 4.6 mM 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropanoic acid (TSP-d_4), pH 7.4], and transferred to 3mm NMR tubes. Quality control (QC) samples were prepared from pooled serum samples of 10 healthy donors from the Norwegian blood bank (46 samples in total). One QC sample per 60 HUNT samples was to assess the quality of the NMR acquisitions and identify instrumental drifts. Approximately half of the samples were analyzed locally at the MR Core facility, NTNU in Trondheim, while the second half was shipped to Bruker BioSpin GmbH, Germany, for analyses. The study cohort included 1208 cases and 1208 matched controls, and the selection of samples to be analyzed at the two labs was random and independent of the case-control status. NMR analyses were carried out on a Bruker Avance III HD Ultrashield Plus 600 MHz spectrometer (Bruker BioSpin) equipped with a 5 mm TCI probe in Trondheim, and an Avance-IVDr system, a standardized Bruker Avance 600 MHz spectrometer (Bruker BioSpin) at the lab in Germany. The NMR spectrometers at both labs have been calibrated for use of the same protocol. Sample handling and data acquisition were automatically performed using SampleJet sample changer and the automation software IconNMR on Topspin 3.5 (Bruker BioSpin). NMR spectra were recorded using one-dimensional nuclear Overhauser effect spectroscopy (1D-NOESY) and Carr-Pucell-Meiboom-Gill (CPMG) experiments [21]. Both experiments were performed at 310 K and applied irradiation (25 Hz) on the water resonance during relaxation delay (4 s) and mixing time (10 ms). The 1D-NOESY experiment applied pulse sequence “noesygppr1d” (Bruker nomenclature) using 96k data points and 30 ppm spectral width. 32 scans were recorded, and the free induction decays were Fourier-transformed after zero filling (128k real data points) and 0.3 Hz line broadening. The CPMG experiment (pulse sequence “cpmgpr1d”, Bruker nomenclature) was recorded with 72k data points, 20 ppm spectral width and 32 scans. Data

was zero filled to 128k data points, line broadening (0.3 Hz) before Fourier-transformation. The success of NMR experiments was assessed based on shim quality (linewidth of the alanine doublet at ~1.5 ppm < 1.5Hz including line broadening) and size of residual water signal (its concentration equivalent < 30mmol/L), in accordance with quality requirements of serum NMR experiments described in Bruker protocols (B.I. Quant-PS 2.0^{TM}). After the exclusion of samples for which the NMR acquisition was of poor quality, and their case/control pair, 2398 samples were retained for statistical analysis (1199 complete case and control pairs). The proportion of cases and controls was equivalent across the two labs.

Metabolite quantification

CPMG spectral data were transferred to Matlab R2020a for preprocessing. Spectra of samples run at NTNU and at Bruker BioSpin were preprocessed separately due to differences in peak positions of individual metabolites. The left peak of the alanine doublet at 1.47 ppm was used as a chemical shift reference for initial alignment of the spectra, followed by a more thorough peak alignment using the icoshift function, where the mean spectra was used as the reference [22]. The spectral region 0.2 to 9.2 ppm was chosen as the region of interest. Spectral peaks were aligned to metabolites using the human metabolome database (HMDB), published literature, and an in-house overview over previously assigned spectral peaks in serum based on 2D HSQC acquisitions and the STOCSY algorithm [23]. Areas under the spectral peaks were integrated, and corrected for the number of protons giving rise to the signals. Thereafter, peaks were adjusted for T2 relaxation times (Table S1, Supplementary Methods). Peaks arising from the same metabolite were averaged, thus absolute metabolite concentrations were obtained for a total of 28 metabolites. The coefficients of variation (CVs) of the metabolites were below 15% and below 20% for 23 and 26 of the metabolites, respectively (Table S1).

Lipoprotein Parameter Analysis

Lipoprotein subfractions were automatically quantified using Bruker IVDr Lipoprotein Subclass Analysis (B.I.LISA^{TM}) software, from Bruker BioSpin. This method utilizes the broad lipid signals from

the methyl (-CH_3) groups at 0.85 ppm and methylene (-CH_2-) groups at 1.27 ppm of the NMR spectrum, to provide a detailed picture of circulating lipoproteins [24]. The concentrations of lipids [cholesterol (CH), free cholesterol (FC), triglycerides (TG), and phospholipids (PL)] in total serum and in four main lipoprotein classes: very low-, intermediate-, low-, and high-density lipoproteins (VLDL, IDL, LDL, and HDL) and 15 subclasses (VLDL 1−5, LDL 1−6, and HDL 1−4) are provided by the software. In addition, serum levels of apolipoproteins (Apo-A1, Apo-A2, and Apo-B) in the lipoproteins, 12 calculated parameters (ratios of LDL-CH/HDL-CH and Apo- B/Apo-A1), and particle numbers of total serum, VLDL, IDL, LDL, and LDL 1−6 are provided, giving a total of 112 lipoprotein subfractions. However, due to the presence of a contamination in the serum samples (neopentyl glycol, most likely originating from the original collection tubes), which interfered with the broad lipid peak at 0.85 ppm on the ^{1}H spectra, some of the lipoprotein subfractions were excluded from further analysis (Supplementary Methods, Table S2, Figure S1), mostly from LDL-2 and LDL-4 particles. Calculated parameters and particle numbers were also excluded from the statistical analysis, resulting in 89 lipoprotein subfractions. CVs for the lipoprotein subfractions included were below 15% and 20% for 65 and 85 of the variables, respectively (Table S2).

Statistical analysis

Baseline characteristics of the cases and controls were described using mean and standard deviation (SD), and statistical significance was assessed by Student t-tests for continuous variables and the Fisher's exact test for categorical variables.

NMR derived variables were standardized to unit variance prior to statistical analysis. Correlation between the NMR-derived variables was tested using Pearson correlation analysis. Odds ratios (OR) and 95% Wald confidence intervals (CI) were calculated for a one SD increase in the concentration of each variable using unconditional logistic regression. The logistic regression models were fitted separately for pre- and postmenopausal women. P-values were corrected for multiple testing using the Benjamini-Hochberg approach, and significance was considered for $P_{adj} \leq .05$. The baseline model

was adjusted for age (matching factor) and lab at which the NMR measurement took place, due to the presence of a batch effect between labs. In the adjusted model, additional factors were included to correct for confounding. Possible confounding factors were selected a priori based on literature and through the use of a directed acyclic graph (DAG). Confounding factors are factors that may have a direct on the baseline blood sample *and* on the outcome (incidence of breast cancer). Therefore, the following addition factors were included in the full model: number of full-term pregnancies, age at menarche, body mass index (BMI), alcohol consumption (frequency per month) and smoking status (current smoker or non-smoker). Menopausal status was defined by the question: "If no longer menstruate, how old were you at last menstruation?", thus missing values will include women that have not reached menopause and women that did not answer this question in the questionnaire. Menopausal age was available for only 235 of the cases, and these were defined as post-menopausal at baseline. For all remaining cases, age at participation in HUNT2 was used to define menopausal status, and cases aged 51 or higher at baseline were classified as postmenopausal, while the remaining cases as pre-menopausal. This cut-off value of 51 years was chosen at the basis of a large population study in Norway, including more than 300 000 individuals [25]. Because late menopause is a well-established risk factor for breast cancer [26-29] menopausal status of the controls was defined by the menopausal status of the cases. This ensured that the population of controls was not older than the population of cases, and thus an age-bias. Clinical variables with less than 10% missing values for the cohort were included in the logistic regression model, where the missing values were imputed with the median values of the full cohort. The use of hormone replacement therapy (HRT) at baseline (systemic, local, previous or never use) was missing for 20.8% of the study cohort, thus this variable was not included in the adjusted model. To evaluate the influence of HRT usage on the studied associations, we concluded stratified analyses excluding women for whom use of HRT was reported or the information was missing. Stratified analyses were also performed for estrogen receptor positive (ER+) and negative (ER-) breast cancers.

Multivariate predictive models were fit using partial-least squares discriminant analysis (PLS-DA) for discriminating between lipoprotein profiles of pre- and postmenopausal women, including only the controls, and between cases and controls in pre- and postmenopausal women separately. The number of latent variables (LVs) giving the minimum cross validated test error (inner loop) was chosen. The models were validated using double 10-fold cross-validation with 30% of the samples included in the test sets of the inner and outer loops, and their significance ($P_{perm} \leq .05$) was assessed by permutation testing (1000 permutations). Stratified analysis, based on the number of years between sample collection and breast cancer diagnosis were also performed.

Results

Population characteristics

The baseline characteristics of the clinical and lifestyle variables for the participants are summarized in Table 1. Considering traditional breast cancer risk factors, there were significant differences between the groups for age at first pregnancy, the number of full-term pregnancies, height and alcohol intake ($p < .05$). The controls had their first full-term pregnancy at a younger age compared to the cases, and a higher number of full-term pregnancies. The frequency of alcohol intake was significantly higher for the cases. Women who developed breast cancer were also significantly taller than controls, while there was no difference in BMI and waist-hip ratio (WHR). Cases had more often diabetes and use of systemic HRT, however the difference between the cases and controls did not reach statistical significance.

For breast cancer specific variables (Table 2) approximately 84% of the cancer cases were ER+, 69% were PgR+ and 20% were HER2-, however, this information was missing for 36-44% of the participants, depending on the variable. The majority of the cancers (52%) were stage I tumors, whilst below 4% were advanced (with a distant metastasis). The mean lapse of time between sample collection and the breast cancer diagnosis was 11.7 years (range 0-22 years) and the mean age at diagnosis was 64.4 years.

High correlations were observed between the variables, especially among the lipoprotein subfractions (Figure S2). In general, lipoprotein subfractions from the same lipoprotein main fractions were highly correlated with each other, while weak correlations were observed between some of the lipoprotein subfractions and metabolites.

Lipoprotein subfractions associated with breast cancer risk in premenopausal women

From the full study cohort, 554 cases were classified as premenopausal and 645 as postmenopausal at inclusion into HUNT2. Postmenopausal women had significantly different lipid profiles compared to premenopausal women, with elevated levels of most lipoprotein subfraction, except for HDL-3 and HDL-4 cholesterol and phospholipids in postmenopausal women (Figure 1, Table S3).

For premenopausal women, 38 out of the 89 lipoprotein subfractions had a significant inverse association with long-term breast cancer risk using the baseline model, of which 17 remained significant after correcting for multiple testing (Table 3). In the adjusted model, 14 of the lipoprotein subfractions showed a significant inverse association with long-term breast cancer risk (Figure 2, Table 3). All of the lipoprotein subfractions with a significant association in the adjusted model are VLDL subfractions, except for HDL-4 triglycerides. All associations were in the inverse direction, with odds ratios from 0.77-0.83. Excluding all cases who had reported current use of HRT, or for whom the information about HRT usage was missing, and their matched controls, resulted in similar associations (significant ORs from 0.71-0.80; Table S4). Stratified analysis including ER+ cases or ER- cases only (384 and 72 cases, respectively) showed that the associations were not dependent on the ER status (Table S4). For the postmenopausal women, no significant associations were found between lipoprotein subfractions and breast cancer risk, neither in the full cohort nor in stratified analysis (Results not shown).

Circulating metabolites

Analyses performed on premenopausal women showed a significant positive association between acetate and breast cancer risk in the baseline model (P-value = .037). In the adjusted model, acetate

and phenylalanine showed a significant positive association with breast cancer risk (P-values <.05; Table S5). However, none of the associations remained significant after correcting for multiple testing. Similarly as for lipoprotein subfractions, no significant associations were found between circulating metabolites and breast cancer risk for postmenopausal women.

Prediction of a future cancer from prediagnostic serum metabolic profiles

A weak but significant discrimination between breast cancer cases and controls was obtained for premenopausal women (Accuracy = 53%; P-value = .027; Table S6) from lipoprotein subfractions. Adding information on established breast cancer risk factors did not increase the prediction accuracy (Accuracy = 53%; P-value = .021; Table S6). No significant discrimination was obtained for postmenopausal women.

Discussion

The discovery of novel biomarkers for early breast cancer development has several clinical applications, such as insight into metabolic pathways that may represent new therapeutic targets, and early identification of individuals eligible for primary prevention. In this study we analyzed the association of circulating lipoprotein subfractions and metabolites with breast cancer incidence, and assessed the predictive value of serum metabolic profiles of healthy females. We found significant associations between multiple circulating lipoprotein subfractions and breast cancer risk 0-22 years after blood collection. This study is the first to report associations between lipoprotein subfractions and long-term breast cancer risk.

Our results reveal alterations in the lipid metabolism of premenopausal women (at baseline) many years before they develop breast cancer. We found that high levels of circulating cholesterol, free cholesterol, phospholipids, and triglycerides in VLDL subfractions have a protective effect from developing breast cancer, even when adjusting for clinical risk factors including lifestyle factors, however only for premenopausal women. More specifically, several VLDL2-4 subfractions were inversely associated with breast cancer risk. VLDLs are large particles consisting mainly of triglycerides.

They are produced by the liver and take part in the endogenous path, where they are transported through the bloodstream, and deliver free fatty acids to the peripheral tissues and muscles (Figure 3). During this process VLDLs are reduced to IDLs and finally to LDLs as fatty acids are cleaved off. These are taken up by the liver, however a residual fraction of LDLs may turn into foam cells and form atherosclerotic plaque [19]. Their size depends on the content of triglycerides while the rate of synthesis depends on the availability of triglycerides (Figure 3) [30].

Other studies, utilizing traditional clinical chemistry methods to assess serum lipids have suggested that levels of triglycerides may be inversely associated with breast cancer risk. A study by Jobard et al. has shown weak inverse associations between fatty acids, mainly from LDLs and VLDLs and breast cancer risk in premenopausal women, which is in accordance with our findings [31]. No study to date has looked at triglyceride levels or VLDL subfractions and premenopausal breast cancer risk, however a recent study by Bendinelli et al. investigated associations between VLDL subfractions and high mammographic breast density [32], which is a strong risk factor for breast cancer. They found free cholesterol, triglycerides, cholesterol and Apo-B levels in VLDL main fraction, and the subfractions VLDL-1 and VLDL-5 to be inversely associated with high mammographic breast density, supporting our finding of inverse associations between VLDL parameters and breast cancer risk. Similarly, a study on NMR metabolic profiles found an inverse association between lipids and lipoproteins and increased breast cancer risk, however, this study did not include subfraction analysis or stratified analysis based on the menopausal status [11].

Although we have stratified analysis based on the menopausal status, it is important to keep in mind that we are referring to menopausal status at baseline. Given a mean age at diagnosis of 64.4 years, most of the cancer incidences have occurred post menopause. Studies have shown that circulating estrogens are positively associated with breast cancer risk [33], especially in postmenopausal women [34-36]. The ovarian function ceases during menopause, leading to a fall in the estrogen levels. The synthesis of estrogen in postmenopausal women occurs mainly in peripheral tissues, such as adipose

tissue, and estrogen levels correlate with BMI for these women [37]. Estrogen levels play an important role in the regulation of lipid metabolism, and are shown to be negatively correlated with triglycerides and VLDLs [38, 39]. Impaired estrogen signaling is associated with the development of metabolic disorders, and an estrogen deficiency will lead to insulin resistance, which in turn will cause increased lipogenesis, triglyceride accumulation and increased VLDL production in the liver [40]. In our study we found that in addition to VLDL subfractions, total triglycerides, IDL, LDL and HDL 2-4 triglycerides are inversely associated with premenopausal breast cancer risk. The protective effect of VLDL subfractions and triglycerides for premenopausal breast cancer risk observed in this study might thus reflect hormonal activity. Recently, the influence of menopausal hormone therapy (HRT) on breast cancer incidence and mortality was assessed through two placebo-controlled randomized clinical trials, involving in total 27 347 postmenopausal women, with a long-term follow-up. The study showed that the use of estrogen alone, compared with placebo, was significantly associated with lower breast cancer incidence and mortality, while the use of estrogen in combination with progestin was associated with a higher breast cancer incidence but not mortality [41]. Other studies have shown that use of HRT is associated with an increased risk of developing breast cancer, and that this risk is dependent on the type [42], timing and duration of HRT, and BMI [28]. Sensitivity analysis showed that our found associations were not dependent on the ER status of the tumor or the use of HRT. Detailed information on the type and duration of HRT is not available for our cohort.

Several LDL 5-6 subfractions were inversely associated with breast cancer (Apo-B in LDL 5-6, phospholipids and cholesterol in LDL-5 and triglycerides in LDL-6), although these associations did not reach statistical significance when correcting for multiple testing. Interestingly, no association was found between LDL-1 or LDL-3 and breast cancer risk. As illustrated in Figure 3, LDL is synthesized in the same lipid pathway as VLDL, and in general smaller LDLs are the most atherogenic [43], thus these findings are surprising. However, due to the presence of a contamination in the serum samples (Figure S1, Supplementary methods), most parameters from LDL-2 (except triglycerides) and all parameters from LDL-4 subfractions were removed, and the associations of these parameters and breast cancer

risk could not be assessed. Collectively, our results suggest that early breast cancer formation in premenopausal women is likely driven by hormonal activity rather than an unhealthy lifestyle. This should thus be further elucidated.

We found no significant associations between lipoprotein subfractions and long-term breast cancer risk for postmenopausal women. This is in accordance with a recent study by Jobard et al. showing that a perturbed metabolism is associated with increased breast cancer risk in premenopausal women only [31]. Furthermore, our finding of different lipid profiles between pre- and postmenopausal women is in accordance with previous studies which have shown that the lipid profile is highly dependent on the menopausal status [44-46], reflected in significantly higher cholesterol and total, LDL and VLDL (but not HDL) triglyceride levels in postmenopausal women. Other studies have shown that a weight gain or high BMI is associated with increased breast cancer risk in postmenopausal women, while the opposite is true for premenopausal women [47, 48], thus the lipid metabolism is clearly affected by hormone activities.

A few studies have employed mendelian randomization to evaluate the relationship between circulating lipids and breast cancer risk [49-51]. The studies by Johnson et al. and Nowak et al. showed that genetically elevated plasma HDL-cholesterol and LDL-cholesterol levels are associated with increased breast cancer risk [49, 50]. These studies did not include analyses stratified on menopausal status. In the study by Beeghly-Fadiel et al., they found that HDL cholesterol was associated with an increased risk of breast cancer for both pre- and post-menopausal women, while triglycerides were inversely associated with breast cancer risk in postmenopausal women only [51]. Taken together, our findings do not reflect genetical predisposition of an increased breast cancer risk.

Although several metabolites significantly associated with breast cancer risk have been reported in other studies [10-15, 31, 52-54], there is a heterogeneity in the analytical platforms used and type of biological medium. Two recent studies have reported serum metabolic alterations, measured by NMR, associated with overall breast cancer risk [11] or premenopausal breast cancer risk [15, 31], with

metabolic panels that overlap with ours. We found weak associations between acetate and phenylalanine, and premenopausal breast cancer risk, however no longer significant after multiple testing correction. A similar association between phenylalanine and increased breast cancer risk has been reported previously, detected by liquid chromatography tandem mass spectrometry (LC-MS) in prediagnostic plasma, however this study did not include stratification on menopausal status [16]. Concerning the remaining metabolites from our panel, recent studies have reported positive associations between creatine [15], ethanol, leucine, ornithine, glutamine, glutamate, pyruvate and histidine [31], and premenopausal breast cancer risk, while no such associations were found in our study. The former study included 1057 case-control pairs, of which approx. 80% were premenopausal, had a mean follow-up of 8 years, and metabolites were quantified using LC-MS [15]. In the later study, the analytical platform for metabolite detection was ^{1}H NMR, however metabolite concentrations were not absolute, the median follow-up period was 4.75 years, and the number of pre-menopausal case-control pairs was limited to 180 [31]. The study by Lécuyer et al. with a patient cohort including 206 cases and 396 matched controls, of which approximately 60 % were premenopausal, several amino acids were significantly associated with breast cancer risk [11], which were not significant in our study.

There are limitations of our study that need to be addressed. Firstly, the HUNT2 questionnaire data are self-reported, thus a self-reported bias and a recall bias will in general often be present. Because the lipid profiles undergo substantial changes in the menopausal transition, analyses were stratified based on the menopausal status. Only 20% of the study participants had reported menopausal age, while the mean age of the cohort was 52.6 years, thus most of the participants were likely postmenopausal, and thus this variable was imputed for most of the participants. In addition, the perimenopausal period is on average 4 years, and thus perimenopausal women may be included in both the analyses on pre- and postmenopausal women. Another limitation in this study, is related to our NMR experiments. As described in the Supplementary material, our biobank samples contained a contamination by neopentylglycol, and some lipoprotein variables had to be removed due to this. The

NMR experiments were also performed at two different labs, and although the same protocol was followed, a slight batch-effect in the lipid profiles was observed, which was taken into account in statistical analysis. Importantly, samples were run in a completely randomized order, with an equivalent proportion of cases and controls across the two labs. Another limitation of the study is the lack of an external validation cohort. However, no such cohort was available for this study, and our findings should thus be validated in future studies.

This study presents one of the largest prospective analysis of serum metabolic profiles and breast cancer risk to date. The large study cohort, long follow-up period and availability of numerous lifestyle factors allowed for evaluating the behavior of significant associations when adjusting for established breast cancer risk factors. Associations between metabolic factors and breast cancer risk are in general modest, in terms of their odds ratios, as compared to other diseases such as diabetes, where metabolomics research findings have been replicated several times [55, 56]. This implies a lack of accurate predictive value of serum metabolic profiles of healthy females and breast cancer, which we have observed in this study.

In conclusion, we identified several associations between lipoprotein subfractions and long-term risk of breast cancer in premenopausal women. In particular, we found inverse associations in several VLDL subfractions and breast cancer, revealing an altered metabolism in the endogenous lipid pathway many years prior to a breast cancer diagnosis.

Additional information

Acknowledgements: The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology NTNU), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The NMR analyses were performed at the MR Core Facility, Norwegian University of Science and Technology (NTNU), funded by the Faculty of Medicine at NTNU and Central Norway Regional Health Authority, and at Bruker BioSpin GmbH, Germany.

Authors' contributions: Conceptualization: JD, TFB, GFG. Data curation: JD, TFB, GFG. Formal analysis: JD, TA, TFB, GFG. Funding acquisition: TFB, GFG. Investigation: JD, HS, TA, FW, FF, CC, GFG. Methodology: JD, HS, TA, MS, TFB, GFG. Software: HS, MS. Supervision: TFB, GFG. Visualization: JD. Writing – original draft: JD. Writing – review & editing: JD, HS, TA, FW, FF, CC, MS, TFB, GFG.

Ethical approval and consent to participate: All participants have completed a written informed consent form, and the study was approved by the Ethics Committee of Central Norway (REK numbers #1995/8395 and #2017/2231).

Consent for publication: Not applicable.

Data availability: The Trøndelag Health Study (HUNT) has invited persons aged 13 - 100 years to four surveys between 1984 and 2019. Comprehensive data from more than 140,000 persons having participated at least once and biological material from 78,000 persons are collected. The data are stored in HUNT databank and biological material in HUNT biobank. HUNT Research Centre has permission from the Norwegian Data Inspectorate to store and handle these data. The key identification in the data base is the personal identification number given to all Norwegians at birth or immigration, whilst de-identified data are sent to researchers upon approval of a research protocol by the Regional Ethical Committee and HUNT Research Centre. To protect participants' privacy, HUNT Research Centre aims to limit storage of data outside HUNT databank, and cannot deposit data in open repositories. HUNT databank has precise information on all data exported to different projects

and are able to reproduce these on request. There are no restrictions regarding data export given approval of applications to HUNT Research Centre. For more information see: http://www.ntnu.edu/hunt/data.

Competing interests: The authors declare that they have no competing interests.

Funding information: This work has been supported by the Norwegian Financial Mechanism (2014-2021, JD, TFB, Project 2019/34/H/NZ7/00503) the Norwegian Cancer Society (GFG, 6834362 and 202021); Stiftelsen DAM (FW, 2020/FO298770); The K.G. Jebsen Foundation, the Liaison Committee for education, research and innovation in Central Norway (FW, 2020/3806-4), and the Joint Research Committee between St. Olavs hospital and the Faculty of Medicine and Health Sciences, NTNU (GFG, 28328 and TFB, 28346).

Authors' Information: HS, FF, CC and MS are employed at Bruker BioSpin. Bruker BioSpin has funded release of the samples from the HUNT biobank.

References

1. Ferlay, J., I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, et al., *Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012.* International Journal of Cancer, 2015. **136**(5): p. E359-E386.
2. Cancer Registry of Norway, *Cancer in Norway 2018 - Cancer incidence, mortality, survival and prevalence in Norway*. 2019.
3. van Roekel, E.H., L. Trijsburg, N. Assi, M. Carayol, D. Achaintre, N. Murphy, et al., *Circulating Metabolites Associated with Alcohol Intake in the European Prospective Investigation into Cancer and Nutrition Cohort.* Nutrients, 2018. **10**(5).
4. Carayol, M., M.F. Leitzmann, P. Ferrari, R. Zamora-Ros, D. Achaintre, M. Stepien, et al., *Blood Metabolic Signatures of Body Mass Index: A Targeted Metabolomics Study in the EPIC Cohort.* J Proteome Res, 2017. **16**(9): p. 3137-3146.
5. Bathen, T.F., B. Geurts, B. Sitter, H.E. Fjosne, S. Lundgren, L.M. Buydens, et al., *Feasibility of MR metabolomics for immediate analysis of resection margins during breast cancer surgery.* PLoS One, 2013. **8**(4): p. e61578.
6. Gu, H., Z. Pan, B. Xi, V. Asiago, B. Musselman, and D. Raftery, *Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer.* Anal Chim Acta, 2011. **686**(1-2): p. 57-63.
7. Slupsky, C.M., H. Steed, T.H. Wells, K. Dabbs, A. Schepansky, V. Capstick, et al., *Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers.* Clin Cancer Res, 2010. **16**(23): p. 5835-41.
8. Furberg, A.S., M.B. Veierod, T. Wilsgaard, L. Bernstein, and I. Thune, *Serum high-density lipoprotein cholesterol, metabolic profile, and breast cancer risk.* J Natl Cancer Inst, 2004. **96**(15): p. 1152-60.
9. Bro, R., M.H. Kamstrup-Nielsen, S.B. Engelsen, F. Savorani, M.A. Rasmussen, L. Hansen, et al., *Forecasting individual breast cancer risk using plasma metabolomics and biocontours.* Metabolomics, 2015. **11**(5): p. 1376-1380.
10. His, M., V. Viallon, L. Dossus, A. Gicquiau, D. Achaintre, A. Scalbert, et al., *Prospective analysis of circulating metabolites and breast cancer in EPIC.* BMC Med, 2019. **17**(1): p. 178.
11. Lecuyer, L., A. Victor Bala, M. Deschasaux, N. Bouchemal, M. Nawfal Triba, M.P. Vasson, et al., *NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer.* Int J Epidemiol, 2018. **47**(2): p. 484-494.
12. Kuhn, T., A. Floegel, D. Sookthai, T. Johnson, U. Rolle-Kampczyk, W. Otto, et al., *Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study.* BMC Med, 2016. **14**: p. 13.
13. Moore, S.C., M.C. Playdon, J.N. Sampson, R.N. Hoover, B. Trabert, C.E. Matthews, et al., *A Metabolomics Analysis of Body Mass Index and Postmenopausal Breast Cancer Risk.* J Natl Cancer Inst, 2018. **110**(6): p. 588-597.
14. Playdon, M.C., R.G. Ziegler, J.N. Sampson, R. Stolzenberg-Solomon, H.J. Thompson, M.L. Irwin, et al., *Nutritional metabolomics and breast cancer risk in a prospective study.* Am J Clin Nutr, 2017. **106**(2): p. 637-649.
15. Zeleznik, O.A., R. Balasubramanian, Y. Zhao, L. Frueh, S. Jeanfavre, J. Avila-Pacheco, et al., *Circulating amino acids and amino acid-related metabolites and risk of breast cancer among predominantly premenopausal women.* Npj Breast Cancer, 2021. **7**(1).
16. Lecuyer, L., C. Dalle, B. Lyan, A. Demidem, A. Rossary, M.P. Vasson, et al., *Plasma Metabolomic Signatures Associated with Long-term Breast Cancer Risk in the SU.VI.MAX Prospective Cohort.* Cancer Epidemiology Biomarkers & Prevention, 2019. **28**(8): p. 1300-1307.

17. Baenke, F., B. Peck, H. Miess, and A. Schulze, *Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development.* Dis Model Mech, 2013. **6**(6): p. 1353-63.
18. Santos, C.R. and A. Schulze, *Lipid metabolism in cancer.* FEBS J, 2012. **279**(15): p. 2610-23.
19. Feingold, K.R. and C. Grunfeld, *Introduction to Lipids and Lipoproteins*, in *Endotext*, K.R. Feingold, et al., Editors. 2000: South Dartmouth (MA).
20. Krokstad, S., A. Langhammer, K. Hveem, T.L. Holmen, K. Midthjell, T.R. Stene, et al., *Cohort Profile: the HUNT Study, Norway.* Int J Epidemiol, 2013. **42**(4): p. 968-77.
21. Dona, A.C., B. Jimenez, H. Schafer, E. Humpfer, M. Spraul, M.R. Lewis, et al., *Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping.* Anal Chem, 2014. **86**(19): p. 9887-94.
22. Tomasi, G., F. Savorani, and S.B. Engelsen, *icoshift: An effective tool for the alignment of chromatographic data.* J Chromatogr A, 2011. **1218**(43): p. 7832-40.
23. Cloarec, O., M.E. Dumas, A. Craig, R.H. Barton, J. Trygg, J. Hudson, et al., *Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets.* Anal Chem, 2005. **77**(5): p. 1282-9.
24. Jimenez, B., E. Holmes, C. Heude, R.F. Tolson, N. Harvey, S.L. Lodge, et al., *Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by (1)H NMR Spectroscopy in a Multilaboratory Trial.* Anal Chem, 2018. **90**(20): p. 11962-11971.
25. Bjelland, E.K., S. Hofvind, L. Byberg, and A. Eskild, *The relation of age at menarche with age at natural menopause: a population study of 336 788 women in Norway.* Human Reproduction, 2018. **33**(6): p. 1149-1157.
26. Clavel-Chapelon, F. and E.N. Group, *Cumulative number of menstrual cycles and breast cancer risk: results from the E3N cohort study of French women.* Cancer causes & control : CCC, 2002. **13**(9): p. 831-838.
27. Collaborative Group on Hormonal Factors in Breast, C., *Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. .* Lancet, 1997. **350**(9084): p. 1047-59.
28. Collaborative Group on Hormonal Factors in Breast, C., *Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence.* Lancet, 2019. **394**(10204): p. 1159-1168.
29. Lacey, J.V., Jr., A.R. Kreimer, S.S. Buys, P.M. Marcus, S.C. Chang, M.F. Leitzmann, et al., *Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort.* BMC Cancer, 2009. **9**: p. 84.
30. Morrisett, J.D., R.L. Jackson, and A.M. Gotto Jr, *Lipoproteins: structure and function.* Annual review of biochemistry, 1975. **44**(1): p. 183-207.
31. Jobard, E., L. Dossus, L. Baglietto, M. Fornili, L. Lecuyer, F.R. Mancini, et al., *Investigation of circulating metabolites associated with breast cancer risk by untargeted metabolomics: a case-control study nested within the French E3N cohort.* British Journal of Cancer, 2021. **124**(10): p. 1734-1743.
32. Bendinelli, B., A. Vignoli, D. Palli, M. Assedi, D. Ambrogetti, C. Luchinat, et al., *Prediagnostic circulating metabolites in female breast cancer cases with low and high mammographic breast density.* Sci Rep, 2021. **11**(1): p. 13025.
33. Key, T.J., P.N. Appleby, G.K. Reeves, R.C. Travis, A.J. Alberg, A. Barricarte, et al., *Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies.* Lancet Oncology, 2013. **14**(10): p. 1009-1019.

34. Key, T.J., P.N. Appleby, G.K. Reeves, A.W. Roddam, K.J. Helzlsouer, A.J. Alberg, et al., *Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies.* British Journal of Cancer, 2011. **105**(5): p. 709-722.
35. Kaaks, R., S. Rinaldi, T.J. Key, F. Berrino, P.H.M. Peeters, C. Biessy, et al., *Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition.* Endocrine-Related Cancer, 2005. **12**(4): p. 1071-1082.
36. Zhang, X.H., S.S. Tworoger, A.H. Eliassen, and S.E. Hankinson, *Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up.* Breast Cancer Research and Treatment, 2013. **137**(3): p. 883-892.
37. Key, T.J., P.N. Appleby, G.K. Reeves, A. Roddam, J.F. Dorgan, C. Longcope, et al., *Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women.* J Natl Cancer Inst, 2003. **95**(16): p. 1218-26.
38. Mesalic, L., E. Tupkovic, S. Kendic, and D. Balic, *Correlation between hormonal and lipid status in women in menopause.* Bosnian Journal of Basic Medical Sciences, 2008. **8**(2): p. 188-192.
39. Palmisano, B.T., L. Zhu, and J.M. Stafford, *Role of Estrogens in the Regulation of Liver Lipid Metabolism.* Adv Exp Med Biol, 2017. **1043**: p. 227-256.
40. Faulds, M.H., C.Y. Zhao, K. Dahlman-Wright, and J.A. Gustafsson, *The diversity of sex steroid action: regulation of metabolism by estrogen signaling.* Journal of Endocrinology, 2012. **212**(1): p. 3-12.
41. Chlebowski, R.T., G.L. Anderson, A.K. Aragaki, J.E. Manson, M.L. Stefanick, K. Pan, et al., *Association of Menopausal Hormone Therapy With Breast Cancer Incidence and Mortality During Long-term Follow-up of the Women's Health Initiative Randomized Clinical Trials.* JAMA, 2020. **324**(4): p. 369-380.
42. Roman, M., S. Sakshaug, S. Graff-Iversen, S. Vangen, E. Weiderpass, G. Ursin, et al., *Postmenopausal hormone therapy and the risk of breast cancer in Norway.* Int J Cancer, 2016. **138**(3): p. 584-93.
43. Ivanova, E.A., V.A. Myasoedova, A.A. Melnichenko, A.V. Grechko, and A.N. Orekhov, *Small Dense Low-Density Lipoprotein as Biomarker for Atherosclerotic Diseases.* Oxidative Medicine and Cellular Longevity, 2017. **2017**.
44. Pardhe, B.D., S. Ghimire, J. Shakya, S. Pathak, S. Shakya, A. Bhetwal, et al., *Elevated Cardiovascular Risks among Postmenopausal Women: A Community Based Case Control Study from Nepal.* Biochemistry Research International, 2017. **2017**: p. 3824903.
45. Shenoy, R. and P. Vernekar. *Fasting Lipid Profile in Pre- and Post-Menopausal Women: A Prospective Study*. 2015.
46. Carr, M.C., *The emergence of the metabolic syndrome with menopause.* J Clin Endocrinol Metab, 2003. **88**(6): p. 2404-11.
47. Alsaker, M.D., I. Janszky, S. Opdahl, L.J. Vatten, and P.R. Romundstad, *Weight change in adulthood and risk of postmenopausal breast cancer: the HUNT study of Norway.* Br J Cancer, 2013. **109**(5): p. 1310-7.
48. Liu, K., W.N. Zhang, Z.M. Dai, M. Wang, T. Tian, X.H. Liu, et al., *Association between body mass index and breast cancer risk: evidence based on a dose-response meta-analysis.* Cancer Management and Research, 2018. **10**: p. 143-150.
49. Johnson, K.E., K.M. Siewert, D. Klarin, S.M. Damrauer, K.M. Chang, P.S. Tsao, et al., *The relationship between circulating lipids and breast cancer risk: A Mendelian randomization study.* PLoS Med, 2020. **17**(9): p. e1003302.
50. Nowak, C. and J. Ärnlöv, *A Mendelian randomization study of the effects of blood lipids on breast cancer risk.* Nature Communications, 2018. **9**(1): p. 3957.
51. Beeghly-Fadiel, A., N.K. Khankari, R.J. Delahanty, X.-O. Shu, Y. Lu, M.K. Schmidt, et al., *A Mendelian randomization analysis of circulating lipid traits and breast cancer risk.* International Journal of Epidemiology, 2019. **49**(4): p. 1117-1131.

52. Zeleznik, O.A., R. Balasubramanian, Y. Zhao, L. Frueh, S. Jeanfavre, J. Avila-Pacheco, et al., *Circulating amino acids and amino acid-related metabolites and risk of breast cancer among predominantly premenopausal women.* npj Breast Cancer, 2021. **7**(1): p. 54.
53. Lecuyer, L., C. Dalle, S. Lefevre-Arbogast, P. Micheau, B. Lyan, A. Rossary, et al., *Diet-Related Metabolomic Signature of Long-Term Breast Cancer Risk Using Penalized Regression: An Exploratory Study in the SU.VI.MAX Cohort.* Cancer Epidemiol Biomarkers Prev, 2020. **29**(2): p. 396-405.
54. Jobard, E., L. Dossus, L. Baglietto, M. Fornili, L. Lécuyer, F.R. Mancini, et al., *Investigation of circulating metabolites associated with breast cancer risk by untargeted metabolomics: a case–control study nested within the French E3N cohort.* British Journal of Cancer, 2021.
55. Craig, A., O. Cloarec, E. Holmes, J.K. Nicholson, and J.C. Lindon, *Scaling and normalization effects in NMR spectroscopic metabonomic data sets.* Anal Chem, 2006. **78**(7): p. 2262-7.
56. Wang, T.J., M.G. Larson, R.S. Vasan, S. Cheng, E.P. Rhee, E. McCabe, et al., *Metabolite profiles and the risk of developing diabetes.* Nat Med, 2011. **17**(4): p. 448-53.

Figure captions

Figure 1. Scores and loading plots from PLS-DA for discrimination of pre- or postmenopausal women from their lipid profiles (analysis has been restricted to the controls). The model has been orthogonalized. Prediction accuracy = 69.0% (P-value < .001) with 4 latent variables included. LV1: first latent variable; LV2: second latent variable; TP: total plasma; VLDL: very-low density lipoprotein; IDL: intermediate-density lipoprotein; LDL: low-density lipoprotein; HDL: high-density lipoprotein; CH: cholesterol; FC: free cholesterol; PL: phospholipids; TG: triglycerides. AB: apolipoprotein-B; A1: apolipoprotein-A1; A2: apolipoprotein-A2.

Figure 2. Odds ratio per SD for lipoprotein subfractions associated with long-term breast cancer risk in premenopausal women participating in the HUNT2 study. Colored subfractions indicate subfractions significantly associated with breast cancer risk in the fully adjusted model, after correction for multiple testing. Each color represents a property of the subfraction. TP: total plasma; VLDL: very-low density lipoprotein; IDL: intermediate-density lipoprotein; LDL: low-density lipoprotein; HDL: high-density lipoprotein; CH: cholesterol; FC: free cholesterol; PL: phospholipids; TG: triglycerides. AB: apolipoprotein-B; A1: apolipoprotein-A1; A2: apolipoprotein-A2.

Figure 3. Lipoprotein metabolism. Exogeneous pathway: The exogeneous pathway starts in the intestine, where dietary lipids are hydrolyzed. The lipids are then assembled to chylomicrons, which are transported through the bloodstream to the liver. In this process free fatty acids are delivered to adipose tissue and muscles through the enzyme lipoprotein lipase (LPL), for energy and storage. Endogeneous pathway: VLDL is synthesized by the liver and transported through the bloodstream. There free fatty acids are delivered to adipose tissue and muscles through LPL, and VLDLs are reduced to IDLs and finally to LDLs as fatty acids are cleaved off. These are taken up by the liver, however a residual fraction of LDLs turn into foam cells and may form atherosclerotic plaque [19]. The reverse transport pathway: HDL is synthesized in the liver and enters the bloodstream, where it removes cholesterol from peripheral tissue, through the action of Apo-A which acts as an acceptor, and

transfers it back to the liver. Lipoprotein subfractions are numbered according to increasing density, as illustrated for VLDL in the figure.

Table captions

Table 1. Baseline characteristics for the study cohort.

Table 2. Characteristics of breast cancer specific variables for the cases.

Table 3. Odds ratios and 95% confidence intervals (CI) per 1 SD increase for lipoprotein subfractions significantly associated with risk of overall breast cancer (p-value < 0.05 and q-value < 0.05) in premenopausal women of the HUNT2 study.

Tables

Table 1. Baseline characteristics for the study cohort.

Variable	Breast cancer cases (n = 1199)	Controls (n = 1199)	P-value[a]
Age at blood collection (years)	52.7 [14.1]	52.7 [14.2]	.93
Age at first menstrual period (years)	13.3 [1.4]	13.4 [1.4]	.45
Missing	48 (4.0%)	66 (5.5%)	
Number of full-term pregnancies (n)			<.001 *
0	119 (9.9%)	86 (7.2%)	
1	136 (11.3%)	130 (10.8%)	
2	419 (34.7%)	401 (33.4%)	
3	334 (27.9%)	336 (28.0%)	
≥ 4	188 (15.7%)	236 (19.7%)	
Missing	3 (0.3%)	10 (0.8%)	
Age at first full-term pregnancy (years)	23.8 [4.6]	23.4 [4.1]	.03*
Missing	127 (10.6%)	100 (8.3%)	
Age at last full-term pregnancy (years)	30.5 [5.2]	30.3 [5.1]	.46
Missing	256 (21.3%)	226 (18.8%)	
Family history of cancer (mother) (n)	173 (14.4%)	144 (12.0%)	.86
Hormone replacement therapy[b] (n)			.07
Systemic use	153 (12.8%)	115 (9.6%)	
Local use	38 (3.2%)	41 (3.4%)	
Previous use	61 (5.1%)	74 (6.2%)	
Never use	697 (58.1%)	720 (60.0%)	
Missing	250 (20.9%)	249 (20.8%)	
Menopausal age[c] (years)	48.0 [5.0]	47.4 [5.2]	.18
Missing	964 (80.4%)	951 (79.3%)	
Removal of part of or one ovary (n)			.60
Yes	49 (4.1%)	43 (3.6%)	
No	226 (18.8%)	218 (18.2%)	
Unknown / Missing	924 (77.1%)	938 (78.2%)	
Removal of both ovaries (n)			.30
Yes	32 (2.7%)	32 (2.7%)	
No	357 (29.8%)	354 (29.5%)	
Unknown / Missing	810 (67.6%)	813 (67.8%)	
Fasting status at blood collection (time since last meal)			.92
< 3 h	830 (69.2%)	826 (68.9%)	
3-6 h	308 (25.7%)	307 (25.6%)	
> 6 h	55 (4.6%)	62 (5.2%)	
Unknown	6 (0.5%)	4 (0.3%)	
Alcohol consumption at recruitment (freq./month)	1.82	1.50	.002*
Unknown	112 (9.3%)	114 (9.5%)	
Height (cm)	164.5 [6.3]	163.9 [6.3]	.02*
Missing	2 (0.2%)	8 (0.7%)	
Weight (kg)	71.9 [12.4]	71.0 [12.7]	.08
Missing	4 (0.3%)	9 (0.8%)	
BMI (kg/m^2)	26.6 [4.6]	26.5 [4.6]	.41
Missing	4 (0.3%)	9 (0.8%)	

Waist circumference (cm)	82.3 [11.5]	82.2 [11.3]	.97
Missing	6 (0.5%)	12 (1.0%)	
Hip circumference (cm)	102.6 [9.4]	102.5 [9.4]	.68
Missing	6 (0.5%)	12 (1.0%)	
WHR	0.80 [0.06]	0.80 [0.06]	.56
Missing	6 (0.5%)	12 (1.0%)	
Current smoker (n)			.09
Yes	367 (30.6%)	333 (27.8%)	
No	762 (63.6%)	807 (67.3%)	
Missing	70 (5.8%)	59 (4.9%)	
Light exercise duration last 12 months in hours/week (n)			.99
0	86 (7.4%)	87 (7.4%)	
< 1	185 (15.4%)	178 (14.8%)	
1 – 2	403 (33.6%)	404 (33.7%)	
> 3	335 (27.9%)	329 (27.4%)	
Missing	190 (15.8%)	201 (16.8%)	
Hard exercise duration last 12 months in hours/week (n)			.71
0	354 (29.5%)	369 (30.8%)	
< 1	191 (15.9%)	182 (15.2%)	
1 - 2	150 (12.5%)	143 (11.9%)	
> 3	57 (4.8%)	48 (4.0%)	
Missing	447 (37.3%)	457 (38.1%)	

Baseline characteristics for the HUNT2 breast cancer study cohort. SD: Standard deviation; BMI: Body mass index; WHR: Waist-to-hip ratio. Values are reported as mean [SD]. [a] P-value for the comparison between breast cancer cases and controls using Student t-test for continuous variables or Pearson's Chi-squared test for categorical variables; [b] Current use of systemic estrogen in the form of tablet or patches; [c] Missing values include women that have not reached menopause and women that did not give an answer; * Implies significance (P-value < .05)

Table 2. Characteristics of breast cancer specific variables for the cases.

Variable	Mean [SD] or Counts
Length of follow-up from blood collection (years)	11.7 [6.1]
Age at diagnosis (years)	64.4 [13.3]
ER status[a]	
Positive	639 (53.3%)
Negative	123 (10.3%)
Missing	437 (36.4%)
PgR status	
Positive	527 (44.0%)
Negative	233 (19.4%)
Missing	439 (36.3%)
HER2 status	
Positive	132 (11.0%)
Negative	538 (44.9%)
Missing	529 (45.0%)
Local metastasis[b]	362 (30.2%)
Distant metastasis	35 (2.9%)
Tumor size (mm)	17 (0.9-120.0)*
Stage (TNM classification)	
I	499 (41.6%)
II	331 (27.6%)
III	90 (7.5%)
IV	35 (2.9%)
Unknown	119 (9.9%)
Detection method	
At screening	372 (31.0%)
Interval cancer	87 (7.3%)
Outside the screening program	740 (61.7%)

Baseline characteristics of breast cancer specific variables for the cases in the HUNT2 breast cancer study cohort. SD: Standard deviation; ER: Estrogen receptor; PgR: Progesterone receptor; HER2: Human epidermal growth factor receptor 2; TNM: Tumor, nodes and metastases. [a]The cut-off value for defining a breast tumor ER- was changed in 2010 from <10% to <1% in Norway; [b]Metastasis in reginal lymph nodes, or a local infiltration in skin and/or in the wall of the breast; *Median and range.

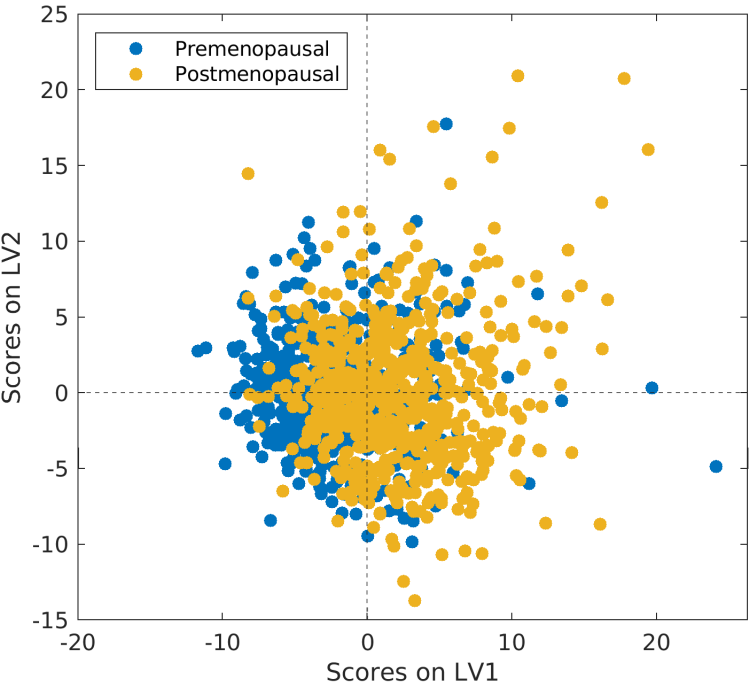
Table 3. Odds ratios and 95% confidence intervals (CI) per 1 SD increase for lipoprotein subfractions significantly associated with risk of overall breast cancer (p-value < 0.05 and q-value < 0.05) in premenopausal women of the HUNT2 study.

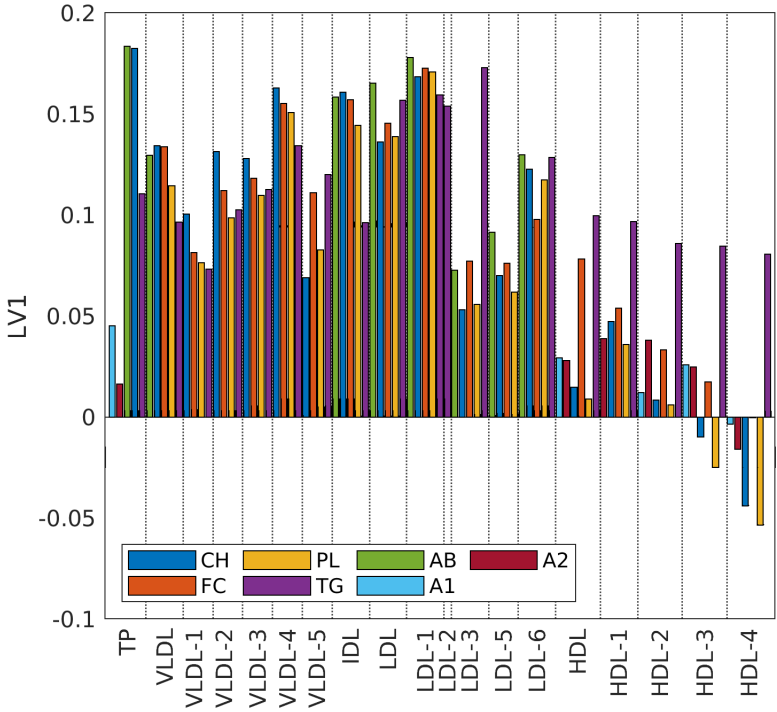
		Baseline model*			Adjusted model**		
		OR (95% CI)	P	P_{adj}	OR (95% CI)	P	P_{adj}
Total plasma	TPA2	1.16 (1.01,1.33)	.04	.11	1.12 (0.97,1.29)	.11	.209
	TPTG	**0.81 (0.69,0.95)**	**.009**	**.046**	0.81 (0.69,0.96)	.01	.07
VLDL	**VLAB**	**0.78 (0.67,0.91)**	**.002**	**.04**	**0.77 (0.65,0.90)**	**.002**	**.049**
	VLCH	**0.80 (0.69,0.93)**	**.005**	**.046**	**0.80 (0.67,0.94)**	**.007**	**.049**
	VLFC	**0.79 (0.67,0.92)**	**.002**	**.04**	**0.78 (0.66,0.92)**	**.003**	**.049**
	VLPL	**0.79 (0.68,0.92)**	**.002**	**.04**	**0.78 (0.66,0.91)**	**.002**	**.049**
	VLTG	0.82 (0.70,0.95)	.009	.046	0.81 (0.69,0.96)	.01	.07
VLDL-1	V1CH	0.84 (0.72,0.98)	.02	.08	0.84 (0.72,0.99)	.04	.12
	V1FC	0.83 (0.72,0.96)	.02	.06	0.83 (0.71,0.98)	.03	.08
	V1PL	0.85 (0.73,0.98)	.03	.08	0.85 (0.72,0.99)	.04	.12
	V1TG	0.86 (0.74,1.00)	.05	.12	0.87 (0.74,1.01)	.08	.16
VLDL-2	**V2CH**	**0.79 (0.68,0.92)**	**.003**	**.04**	**0.79 (0.67,0.93)**	**.005**	**.049**
	V2FC	0.81 (0.69,0.94)	.007	.046	0.81 (0.69,0.95)	.01	.06
	V2PL	**0.80 (0.69,0.93)**	**.003**	**.04**	**0.80 (0.68,0.93)**	**.005**	**.049**
	V2TG	**0.80 (0.69,0.93)**	**.004**	**.04**	**0.80 (0.69,0.94)**	**.006**	**.049**
VLDL-3	V3CH	0.81 (0.69,0.94)	.007	.046	0.80 (0.68,0.95)	.01	.06
	V3FC	**0.80 (0.69,0.94)**	**.005**	**.046**	**0.80 (0.67,0.94)**	**.007**	**.049**
	V3PL	**0.79 (0.68,0.92)**	**.003**	**.04**	**0.78 (0.67,0.92)**	**.003**	**.049**
	V3TG	**0.80 (0.69,0.93)**	**.004**	**.04**	**0.79 (0.68,0.93)**	**.004**	**.049**
VLDL-4	V4CH	0.82 (0.70,0.96)	.01	.06	0.82 (0.70,0.96)	.02	.07
	V4FC	0.83 (0.71,0.97)	.02	.07	0.82 (0.69,0.97)	.02	.07
	V4PL	**0.81 (0.69,0.94)**	**.006**	**.046**	**0.80 (0.68,0.93)**	**.005**	**.049**
	V4TG	**0.81 (0.69,0.94)**	**.006**	**.046**	**0.80 (0.68,0.93)**	**.004**	**.049**
VLDL-5	V5CH	0.86 (0.75,0.99)	.04	.10	0.84 (0.73,0.97)	.02	.07
	V5PL	**0.84 (0.73,0.96)**	**.01**	**.06**	**0.82 (0.71,0.94)**	**.006**	**.049**
IDL	IDPL	0.82 (0.70,0.97)	.02	.07	0.81 (0.68,0.97)	.02	.07
	IDTG	0.83 (0.71,0.96)	.02	.06	0.83 (0.71,0.98)	.03	.08
LDL	LDTG	0.83 (0.70,0.96)	.02	.06	0.82 (0.69,0.97)	.03	.07
LDL-5	L5AB	0.86 (0.74,0.99)	.04	.10	0.83 (0.72,0.97)	.03	.07
	L5CH	0.88 (0.76,1.01)	.07	.13	0.85 (0.73,0.98)	.03	.09
	L5PL	0.87 (0.75,1.01)	.06	.12	0.84 (0.72,0.97)	.02	.07
LDL-6	L6AB	0.84 (0.72,0.98)	.03	.08	0.85 (0.72,1.00)	.05	.12
	L6TG	0.84 (0.72,0.98)	.03	.10	0.86 (0.73,1.00)	.06	.146
HDL	HDTG	0.85 (0.75,0.96)	.009	.046	0.85 (0.74,0.96)	.01	.06
HDL-1	H1TG	0.86 (0.76,0.98)	.02	.07	0.87 (0.76,0.98)	.03	.09
HDL-2	H2TG	0.88 (0.78,1.00)	.04	.11	0.89 (0.78,1.01)	.06	.15
HDL-3	H3CH	1.15 (1.01,1.30)	.04	.10	1.11 (0.98,1.27)	.11	.21
	H3FC	1.16 (1.01,1.32)	.04	.10	1.13 (0.98,1.29)	.10	.19
HDL-4	**H4TG**	**0.85 (0.75,0.96)**	**.01**	**.06**	**0.83 (0.73,0.95)**	**.008**	**.049**

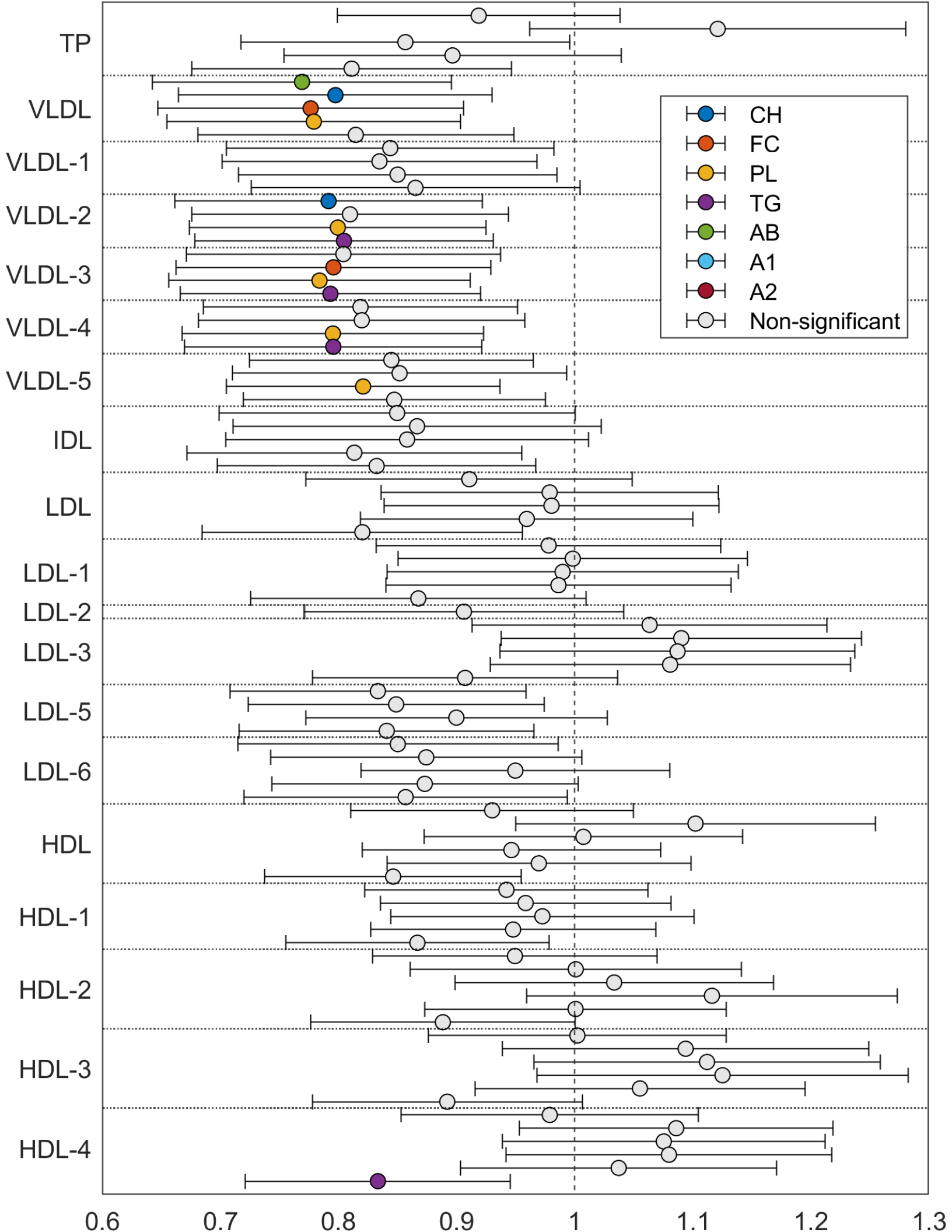
TP: total plasma; VLDL: very-low density lipoprotein; IDL: intermediate-density lipoprotein; LDL: low-density lipoprotein; HDL: high-density lipoprotein; CH: cholesterol; FC: free cholesterol; PL: phospholipids; TG: triglycerides. AB: apolipoprotein-B; A2: apolipoprotein-2.

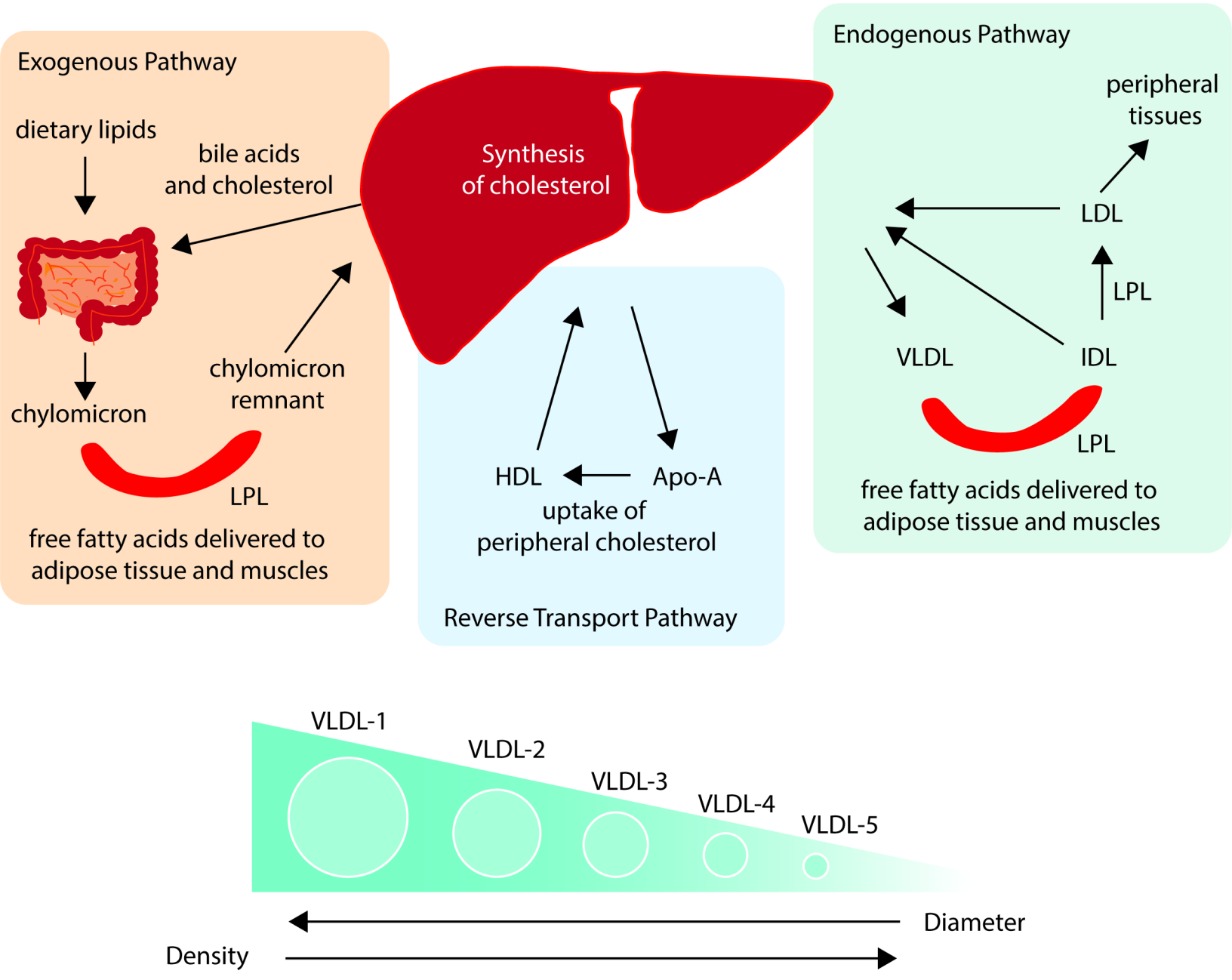
*Baseline model: adjusted for lab for NMR analyses and matching variable (age at participation in the HUNT2 study)

**Adjusted model: in addition to lab and age, this model is adjusted for no. of full-term pregnancies, age at menarche, BMI, alcohol consumption (frequency per month) and smoking status. Bold font indicates variables significant in the adjusted model after multiple testing correction.


Premenopausal
Postmenopausal
Scores on LV2
Scores on LV1
25
20
15
10
5
0
-5
-10
-15
-20
-10
0
10
20


LV1
0.2
0.15
0.1
0.05
0
-0.05
-0.1
CH
PL
AB
A2
FC
TG
A1
TP
VLDL
VLDL-1
VLDL-2
VLDL-3
VLDL-4
VLDL-5
IDL
LDL
LDL-1
LDL-2
LDL-3
LDL-5
LDL-6
HDL
HDL-1
HDL-2
HDL-3
HDL-4


TP
VLDL
VLDL-1
VLDL-2
VLDL-3
VLDL-4
VLDL-5
IDL
LDL
LDL-1
LDL-2
LDL-3
LDL-5
LDL-6
HDL
HDL-1
HDL-2
HDL-3
HDL-4
CH
FC
PL
TG
AB
A1
A2
Non-significant
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3


Exogenous Pathway
dietary lipids
bile acids
and cholesterol
chylomicron
chylomicron
remnant
LPL
free fatty acids delivered to
adipose tissue and muscles
Synthesis
of cholesterol
HDL
Apo-A
uptake of
peripheral cholesterol
Reverse Transport Pathway
Endogenous Pathway
peripheral
tissues
LDL
LPL
VLDL
IDL
LPL
free fatty acids delivered to
adipose tissue and muscles

VLDL-1
VLDL-2
VLDL-3
VLDL-4
VLDL-5
Diameter
Density