Associations between circulating metabolites and incidence of breast, prostate, lung and colorectal cancers in the HUNT Study and UK Biobank

Marion Denos[1*], Abhibhav Sharma[1*], Bjørn Olav Åsvold[1,2], Alya G.A. Arham[1], George Davey Smith[3], Mari Løset[1,4], Ben Michael Brumpton[1,5,6,#], Julia Debik[1,#], Guro F. Giskeødegård[1,7,#]

[1] HUNT Center for Molecular and Clinical Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway.

[2] Department of Endocrinology, Clinic of Medicine, St. Olavs Hospital, Trondheim, Norway.

[3] MRC Integrative Epidemiology Unit, University of Bristol, Bristol Medical School, University of Bristol

[4] Department of Dermatology, Clinic of Orthopedy, Rheumatology and Dermatology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

[5] HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Levanger, Norway.

[6] Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

[7] Department of Surgery, St. Olavs University Hospital, Trondheim, Norway

[*] Contributed equally # Shared last authorship

Corresponding author: Marion Denos; Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway; marion.denos@ntnu.no

ORCID identifier: 0000-0003-4134-9156

Word count

Abstract: 281 words

Main text: 4140 words

Abstract

Background: Breast, prostate, lung and colorectal cancers are the most common cancers in both women and men globally. The role of circulating metabolites, including lipoprotein subfractions, in the development of these cancers remains unclear.

Objective: This prospective cohort study, including two large population-based cohorts, aimed to explore the associations between circulating metabolites and the incidence of breast, prostate, lung and colorectal cancers.

Methods: From the Trøndelag Health Study (HUNT) and the UK Biobank (UKBB), 132 lipoprotein subfractions and metabolites, quantified using nuclear magnetic resonance spectroscopy, were available for participants (n=16,877 in HUNT3, n=260,805 in UKBB). Cox regression was used to investigate the associations with incident cancers.

Results: During a mean follow-up time of 14 years, 549 women and 741 men developed breast (n= 283), prostate (n=452), lung (n=175) or colorectal (n=380) cancer in HUNT3, while 6679 women and 8248 men developed breast (n=4516), prostate (n=5576), lung (n=1985) or colorectal (n=2850) cancer in UKBB. In the UKBB, higher levels of several subfractions of lipoproteins including VLDL and LDL were associated with a slight decreased incidence of prostate and lung cancers (cholesterol in large LDL: HR per 1-SD increment 0.90, 95% CI 0.86-0.94 for lung cancer). Similar associations were observed between albumin, some fatty acids, and lung cancer. However, no clear associations were observed in HUNT3.

Conclusion: Several lipoprotein subfractions and circulating metabolites were associated with a reduced incidence of lung cancer, predominantly in UKBB. Although no conclusive evidence was found for a specific lipid or metabolite profile associated with the incidence of the most common cancers, the observed associations were generally consistent across cohorts. Our

results highlight promising directions for future research into metabolic biomarkers and their potential role in cancer prevention.

Introduction

Globally around 20 million new cancer cases were reported in 2022, and predictions estimate this number to rise to 35 million by 2050 (1). Breast cancer (BC) is the most common cancer in women, while prostate cancer (PCa) ranks second in men (1). Lung cancer (LC) and colorectal cancers (CRC) are also among the most frequently diagnosed cancers in both sexes (1). Although several modifiable risk factors have been associated with cancer incidence, they only partially account for the observed rates. In BC, factors related to reproductive history and lifestyle have been identified, yet they only explain a small proportion of cases in high-income countries (2-7). For PCa, few lifestyle and environmental factors have been established; age, ethnicity, family history and genetic susceptibility are recognized contributors, while associations with smoking and body weight remain inconclusive (1, 3, 8-10). LC is primarily attributed to tobacco smoking, and is often diagnosed at an advanced stage, resulting in poor survival rates. This underscores the need for biomarkers that enable earlier detection (1, 11). CRC risk is associated with factors such as age, smoking, alcohol consumption and diet (12). However, its typically slow progression and lack of early symptoms pose challenges for timely diagnosis. While prostate-specific antigen is used as a diagnostic tool for PCa, it does not predict cancer risk. To date, no clinically established circulating risk markers exist for these four major cancer types.

Metabolomics captures the dynamic interaction between an individual and the surrounding environment, revealing insights closely related to the phenotype (13, 14). The metabolite profile thus reflects the biological condition, providing a snapshot of the organism's current state. Metabolites found in human biofluids may serve as biomarkers of current health and future disease risk, and previous studies have shown the potential of metabolomic biomarkers for disease risk (15-17). Altered cellular metabolism is a well-established hallmark of cancer (18). In particular, changes in the levels of specific lipids and small-molecular have been shown to

support anaerobic glycolysis, thereby meeting the elevated energy and biosynthetic demands of tumor growth (19, 20). However, the role of circulating metabolites and lipids in cancer risk remains inconclusive. Epidemiological studies have found a strong link between elevated blood lipid levels and a higher incidence and recurrence of PCa, while inhibiting lipid biosynthesis with statins has been shown to reduce mortality of BC, LC and colon cancer (21-23). Mass spectrometry (MS) and nuclear magnetic resonance (NMR) based case-control studies have reported varying results on associations between metabolic profiles of prediagnostic blood samples and incidence of BC, PCa, LC and CRC (24-37). Our previous case-control study, based on the second Trøndelag Health Study (HUNT2), found that high levels of very low-density lipoprotein (VLDL) subfractions in blood were associated with a reduced risk of BC in premenopausal women and revealed alterations in lipid metabolism many years before BC diagnosis (29). Recently, studies by Buergel et al. and Julkunen et al. utilized the initial release of UK Biobank metabolomics data from ~118,000 participants (38, 39). They reported no or weak associations between free circulating metabolite levels and incidence of BC, PCa or CRC while they identified several biomarkers for LC incidence (38, 39). Both studies did not account for reproductive factors as confounders in women, which might influence the associations. In contrast, other case-control studies with a prospective design have identified metabolite risk factors for PCa (32, 40, 41). Thus, studies exploring the association between metabolic biomarkers and risk of BC, PCa, LC and CRC have shown inconsistent results and require further investigation.

In this prospective cohort study, we aimed to investigate the associations between circulating lipoprotein subfractions and small-molecular metabolites and incidence of BC, PCa, LC and CRC using data from two large population-based cohorts for independent validation: The Trøndelag Health Study (HUNT) in Norway and the UK Biobank including Phase 1 and Phase 2 metabolomic data releases.

Methods

Study populations

HUNT3

The Trøndelag Health Study (HUNT) is a longitudinal population-based cohort study in Norway, including data from questionnaires, clinical measurements, and biological materials (42). The study enrolled participants aged ≥20 years in 4 health surveys: HUNT1 (1984-1986), HUNT2 (1995-1997), HUNT3 (2006-2008), and HUNT4 (2017-2019) (43). Of the 50,800 participants in HUNT3 (participation rate of 54.1%), a subset of 17,210 participants had serum metabolomics data quantified by NMR spectroscopy, generated by Nightingale Health (44). Each participant was followed up from the date of participation in HUNT3 until the date of first diagnosis of BC, PCa, LC or CRC, the date of death or emigration from Norway, or the end of the follow-up period in December 2023, whichever came first. Diagnoses of BC, PCa, LC and CRC were obtained from the Cancer Registry of Norway. Information on death was collected from the Norwegian Cause of Death Registry. Among the 17,210 participants, we excluded those with a prior diagnosis of the cancer under study (breast, prostate, lung, or colorectal) before inclusion into HUNT3, leaving on average 16,877 participants in the study population.

UK Biobank (UKBB)

UKBB is a large population-based prospective study of adults aged 40-69 years recruited in 2006-2010 (45). Among the 9.2 million individuals invited to the study, around 500,000 (5.5%) participated. A large range of health data, including questionnaires and blood samples, was collected in 22 assessment centers throughout the United Kingdom (UK) (45). Metabolic profiling of all participants' plasma samples is underway using the Nightingale Health NMR

122 platform, and here we used the second release of metabolomics data of 274,352 participants
123 (46). Participants were followed up from the participation date in UKBB to the date of first
124 diagnosis of BC, PCa, LC or CRC, the date of death or lost to follow-up, or the end of the
125 follow-up period in June 2022, whichever came first. UKBB linked data from the National
126 Cancer and Death Registries to its database. Among the 268,047 participants, we excluded those
127 with a prior diagnosis of the cancer under study (breast, prostate, lung, or colorectal) before
128 inclusion into UKBB, leaving on average 260,805 participants in the study population.

129 **Metabolomics data**

130 **HUNT3**

131 Metabolomics analysis of serum samples from 17,210 participants in HUNT3 was generated
132 by NMR spectroscopy using the Nightingale Health NMR biomarker platform (Supplementary
133 Table S1). The serum samples were non-fasting, with an average of 3 hours since the last meal.
134 The samples were analyzed in two batches. Batch 1 (n = 6201) was analyzed using a 600 MHz
135 Bruker AVANCE spectrometer, while batch 2 (n = 11009) was analyzed using one 500 MHz
136 spectrometer and one 600 MHz spectrometer. Batch 2 was adjusted for instrument effects using
137 repeated samples from 96 individuals that were measured at both instruments, by multiplying
138 samples analyzed on 500 MHz by a correction factor per variable, resulting in equal median
139 values for the repeated samples. Finally, batch 2 was adjusted towards batch 1 by multiplying
140 with a correction factor per variable that resulted in equal means in the two batches. This study
141 includes 132 metabolic variables quantified in absolute levels. The variables indicating ratios,
142 percentages and size based on diameters were excluded from the analysis. The metabolic
143 variables include lipoprotein subclasses, fatty acids, and small-molecular metabolites, such as
144 amino acids, ketone bodies, glycolysis metabolites, aromatic and branched chain amino acids.
145 Lipid concentrations and composition were measured for 14 lipoprotein subclasses, including

146 triglycerides, phospholipids, total cholesterol, cholesteryl esters, and free cholesterol, as well
147 as total lipid concentration for each subclass. The average biomarker detection rate was >99%
148 in the serum samples.

149 **UKBB**

150 Nightingale Health has performed metabolic profiling of plasma samples for all UKBB
151 participants. The plasma samples were non-fasting, with an average of 4 hours since the last
152 meal. Measurements took place in Finland, using 500 MHz NMR Bruker AVANCE
153 spectrometers in three different phases, as described elsewhere (39, 46). The first data release
154 covered metabolic variables from a random selection of 118,000, the second release, which we
155 used in the current study, added 157,000 samples, and the remaining were analyzed in the third
156 phase, which is not yet publicly available. Any observations with more than 10% missing
157 metabolite values or flagged with "Low protein" (indicating more severe sample dilution) were
158 removed before imputation. We conducted kNN imputation for data missing at random, and
159 quantile regression imputation of left-censored data (QRILC) for data missing not at random.
160 The same metabolic variables as described for HUNT3 were quantified in UKBB, of which 132
161 were included in this study.

162 **Covariates**

163 Weight and height were measured by health professionals during the clinical examination. Body
164 mass index (BMI) was calculated as weight divided by the squared value of height (kg/m^2) and
165 analyzed as a continuous variable. Information on sociodemographic (age, sex, marital status),
166 lifestyle (smoking status, alcohol consumption) and reproductive factors (age at menarche, age
167 at menopause, number of live births) was collected in questionnaires or through interviews in
168 both cohorts. Covariates were categorized as follows: smoking status (never, former, current),
169 alcohol consumption (never, 1-3, ≥4 times/month), and marital status (married, unmarried).

170 Women with unknown menopausal status were defined as premenopausal if they were below
171 51 years old at recruitment and as postmenopausal if aged 51 or above. The cut-off value of 51
172 years was chosen based on a large study of over 300,000 Norwegian women (47). The variable
173 IDs from both biobanks are provided in Supplementary Table S2.

174 **Cancer ascertainment**

175 Participants in HUNT3 was linked to the Cancer Registry of Norway using the 11-digit personal
176 identification number. The Cancer Registry of Norway provides accurate, close-to-complete
177 and timely data (48). In UKBB, participants were linked to the national cancer registry which
178 consolidates information from regional cancer centres and whose data are considered the gold
179 standard for ascertaining cancer outcomes in the UK (49, 50). Cancer diagnoses were classified
180 according to the Tenth Revision of the International Classification of Diseases (ICD10) as
181 follows: BC (C50), PCa (C61), LC (C33-C34) and CRC (C18-C20) (51).

182 **Statistical analysis**

183 Baseline characteristics of women and men in each cohort were described using mean and
184 standard deviation (SD) for continuous variables, and number of participants (%) for categorical
185 variables. Metabolomics data from UKBB were corrected for the NMR spectrometer used for
186 the measurements, following the procedure by Julkunen et al. (39). We fit a linear regression
187 model with log-transformed concentration as the outcome and spectrometer as the predictor.
188 Scaled residuals from this regression were used as predictors in the analyses. For each
189 metabolite, concentrations outside four interquartile ranges from the median were considered
190 outliers and excluded from the analyses.

191 Cox proportional hazard (PH) models were used to explore the potential associations of
192 metabolomics data with incidence of BC (in women), PCa (in men), LC and CRC, separately.

193 Analyses were sex specific for BC and PCa. Crude and adjusted hazard ratios (HR) with 95%
194 confidence intervals (CIs) were calculated with age as the time scale. The HRs were reported
195 per SD increment in the log-transformed metabolite concentrations. The models were adjusted
196 for potential confounders, including BMI, assessment center (for UKBB), smoking status, and
197 alcohol consumption. Additionally, analyses for BC were adjusted for age at menarche and
198 number of live births, while analyses for PCa were adjusted for marital status, and analyses for
199 LC and CRC were adjusted for sex. We also conducted stratified analyses based on women's
200 menopausal status for BC. P-values were corrected for multiple testing using the Benjamini-
201 Hochberg approach. To assess the robustness of the findings, we performed sensitivity analyses:
202 (1) we restricted the age at participation in HUNT3 from 40 to 69 years old to match the UKBB
203 age frame; (2) we excluded all participants with any type of cancer before baseline.

204 Partial least squares discriminant analysis (PLS-DA) was utilized to evaluate the comparability
205 of NMR data between the UKBB and HUNT3. PLS-DA was applied to metabolomic data from
206 each cohort to discriminate premenopausal and postmenopausal women. Five-fold cross-
207 validation was applied for model training. By examining the loadings of the most explanatory
208 latent variable, the study assessed the consistency of metabolic differences across the two
209 population cohorts. All statistical analyses were performed with R statistical software version
210 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria). Cox-PH analysis was
211 performed using the *survival* package in R (52). PLS-DA was performed using the *mixOmix*
212 package in R (53).

213

214 # Results

215 **Population characteristics**

Table 1 outlines the characteristics of participants in the two cohorts separated by sex. The mean age of participants was around 52 years in HUNT3 and 56 years in UKBB. HUNT3 had a higher number of smokers, while UKBB participants had on average a more frequent alcohol consumption. Moreover, women in HUNT3 had an earlier age of menopause and a higher number of live births.

Among 16,877 participants in HUNT3, 283 women developed BC, 452 men developed PCa, 175 participants developed LC, and 380 participants developed CRC, during a follow-up time of 14 years. Among 260,805 participants in the UKBB, 4516 women developed BC, 5576 men developed PCa, 1985 participants developed LC, and 2850 participants developed CRC, over a 14-year follow-up period.

Lipoprotein and small molecule metabolite associations with cancer incidence

Hazard ratio plots from fully adjusted Cox regression models show the associations between metabolic variables and the incidences of BC (Figure 1), PCa (Figure 2), LC (Figure 3) and CRC (Figure 4).

No lipoprotein subfractions or circulating metabolites were associated with BC in UKBB or HUNT3 after adjusting for multiple testing (Figure 1). Hazard ratios and corresponding 95% CIs for age-adjusted and fully adjusted Cox models for BC are provided in Supplementary Data 1-2.

Several fatty acids and lipoprotein subfractions were associated with PCa in UKBB (Figure 2). Higher levels of total triglycerides in HDL were associated with a decreased PCa incidence in the UKBB cohort (HR per 1-SD increment 0.95, 95% CI 0.92-0.98). A similar trend was observed for LDL and VLDL subfractions. Fatty acids species, including monounsaturated fatty acids (MUFA) (HR 0.96, 95% CI 0.93-0.98) and total fatty acids (HR 0.96, 95% CI 0.93-0.98), were also associated with a decreased risk of PCa, while unsaturated fatty acids were associated

240 with an increased risk of PCa (HR = 1.05, 95% CI 1.02-1.08). There was no association between
241 metabolic variables and PCa in HUNT3. Notably, the point estimates were in the opposite
242 direction for some metabolic variables in HUNT3 compared to those observed in the UKBB.
243 However, the 95% CIs were wide in HUNT3 due to its smaller sample size (Supplementary
244 Data 1-2).

245 As presented in Figure 3, total cholesterol, total free cholesterol, and total phospholipids were
246 inversely associated with LC incidence in the UKBB (total cholesterol: HR per 1-SD increment
247 0.93, 95% CI 0.88-0.97). These inverse associations were specifically evident in small and
248 medium VLDL, all LDL and IDL particles. Apolipoprotein B (ApoB) was associated with a
249 reduced LC risk (HR per 1-SD increment 0.92, 95% CI 0.88-0.96). Additionally, higher levels
250 of fatty acids, including docosahexaenoic acid, linoleic acid, omega-3 and -6, and
251 polyunsaturated fatty acids (PUFAs), were associated with a reduced risk of LC. Similarly,
252 albumin level was inversely associated with LC as well as circulating levels of cellular
253 membrane-based lipids, such as choline, phosphatidylcholine, and sphingomyelins. Conversely,
254 glycolysis-related metabolites such as lactate and free circulating ketone bodies (acetone, beta-
255 hydroxybutyrate) were positively associated with LC incidence, while acetate was negatively
256 associated with LC. Notably, the inflammation marker glycoprotein acetylation (GlycA) was
257 positively associated with LC incidence (HR per 1-SD increment 1.23, 95% CI 1.15-1.30). The
258 results from HUNT3 were consistent with UKBB, although not statistically significant after
259 multiple testing adjustment, likely due to the smaller sample size. Hazard ratios and CIs for
260 age-adjusted and fully adjusted Cox models for LC are provided in Supplementary Data 1-2.

261 Higher level of albumin was associated with a reduced risk of CRC in the UKBB (HR per 1-
262 SD increment 0.93, 95% CI 0.89-0.96) (Figure 4). Glycolysis-related metabolites such as
263 glucose and free circulating ketone bodies (acetoacetate, acetone, beta-hydroxybutyrate) were
264 positively associated with CRC incidence, although these associations were not statistically

265 significant after multiple testing correction. The inflammation marker GlycA was positively
266 associated with CRC incidence (HR per 1-SD increment 1.04, 95% CI 1.00-1.09), similarly to
267 its association with LC incidence. No lipoprotein subfractions or other metabolites were
268 associated with CRC in UKBB or HUNT3. Hazard ratios and CIs for age-adjusted and fully
269 adjusted Cox models for CRC are provided in Supplementary Data 1-2.

270 **Sensitivity analysis**

271 Given the previously reported differences in serum metabolomic profiles between
272 premenopausal and postmenopausal women (29), which were confirmed in our study
273 (Supplementary Figure S1), we conducted stratified analyses to investigate the associations of
274 metabolic variables with BC incidence in pre-diagnostic blood samples of premenopausal and
275 postmenopausal women separately. Several metabolic variables were inversely associated with
276 BCa in premenopausal women, but not in postmenopausal women, in UKBB. These included
277 total triglyceride levels, triglycerides and cholesterols in large VLDLs, cholesterol and
278 cholesteryl esters in LDL subfractions, all lipids in small HDL, in addition to several fatty acid
279 (Supplementary Figure S2, Supplementary Data 1-2). No clear associations were observed in
280 HUNT3.

281 Results remained similar when the analysis was performed in the age restricted HUNT3 cohort
282 (in both raw and adjusted models) (Supplementary Figures S3–6, Supplementary Data 1-2).
283 Moreover, the associations remained unchanged when participants with previous history of
284 cancer at baseline were excluded (Supplementary Data 3-4).

285

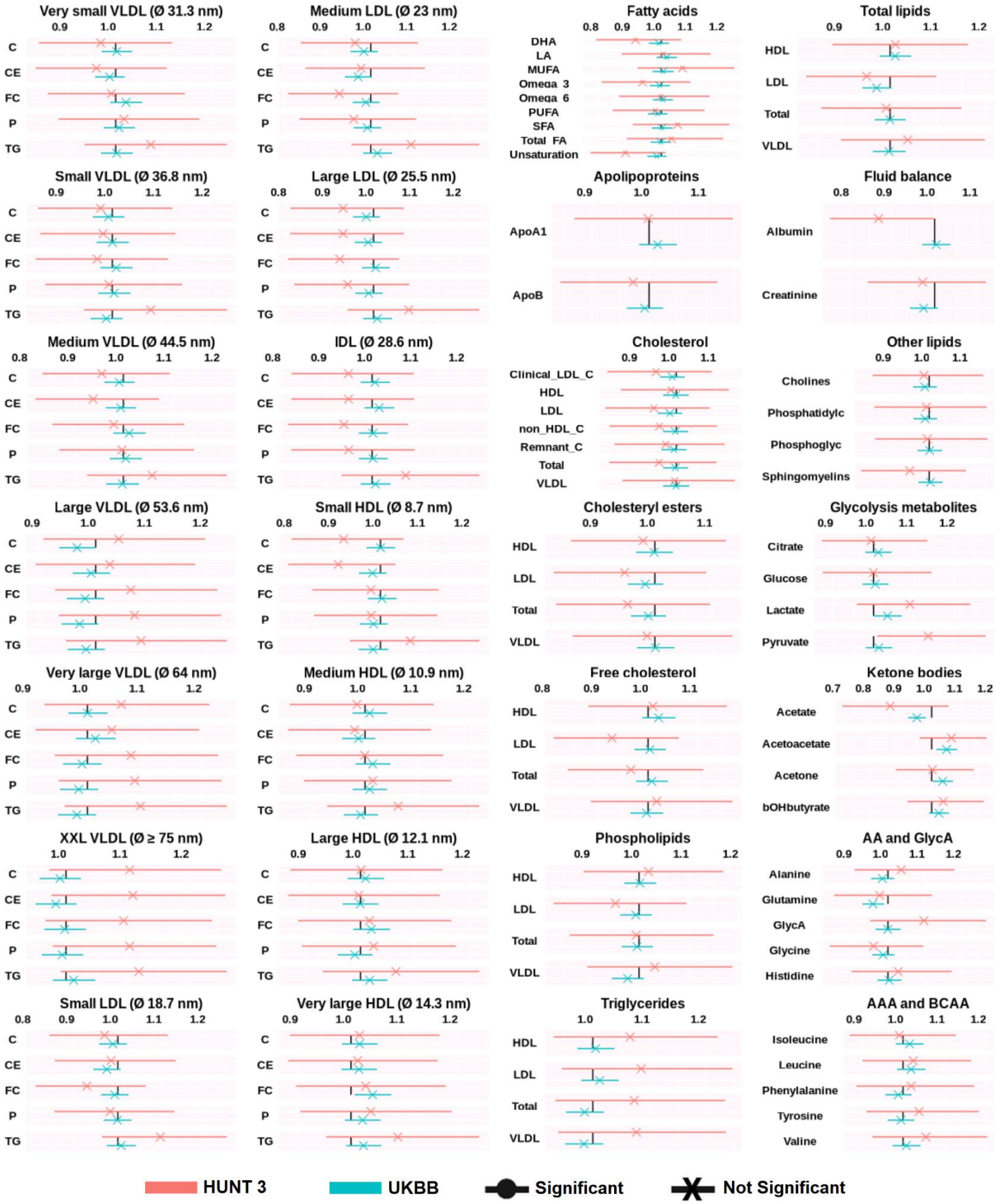
286 **Table 1.** Baseline characteristics of participants in HUNT3 and UK Biobank

	HUNT3		**UK Biobank**	
	Women	Men	Women	Men
Number of participants	9129	7748	139,684	121,121
Female breast cancer cases	283 (3.1)		4516 (3.2)	

Prostate cancer cases		452 (5.8)		5576 (4.6)
Lung cancer cases	91 (1.0)	84 (1.1)	950 (0.7)	1035 (0.8)
Colorectal cancer cases	175 (1.9)	205 (2.6)	1213 (0.8)	1637 (1.3)
Age at cancer diagnosis (yrs)	68.2 ± 12.2	71.8 ± 8.9	64.6 ± 8.2	67.9 ± 6.4
Age at participation (yrs)	51.8 ± 16.0	52.4 ± 15.4	56.3 ± 8.0	56.6 ± 8.2
Fasting time at blood collection (hours)	2.7 ± 2.1	2.9 ± 2.1	3.7 ± 2.3	3.9 ± 2.6
Body mass index (kg/m²)	26.8 ± 4.8	27.4 ± 3.7	27.1 ± 5.2	27.8 ± 4.2
Smoking status				
never	4006 (43.9)	3195 (41.2)	83,066 (59.5)	59,072 (48.8)
former	2566 (28.1)	2617 (33.8)	43,484 (31.1)	46,429 (38.3)
current	2293 (25.1)	1769 (22.8)	12,469 (8.9)	15,021 (12.4)
missing	264 (2.9)	167 (2.2)	545 (0.4)	489 (0.5)
Alcohol consumption (times/month)				
never	478 (5.2)	190 (2.5)	12,940 (9.3)	7605 (6.3)
1 - 3	5774 (63.2)	4027 (52.0)	39,050 (28.0)	19,556 (16.2)
≥4	2600 (28.5)	3413 (44.1)	87,439 (62.6)	93,692 (77.4)
missing	277 (3.0)	118 (1.5)	134 (0.1)	157 (0.1)
Age at menarche (yrs)	13.2 ± 1.4		12.5 ± 2.9	
Age at menopause (yrs)	48.6 ± 5.8		46.4 ± 13.4	
Number of live births (n)	2.5 ± 1.1		1.8 ± 1.2	
Marital status				
married	5233 (57.3)	4776 (61,7)	98,522 (70.5)	93,912 (77.5)
unmarried	3882 (42.5)	2956 (38.1)	41,162 (29.4)	27,209 (22.4)

Data are given as mean ± standard deviation or as number of individuals (%)

Breast Cancer



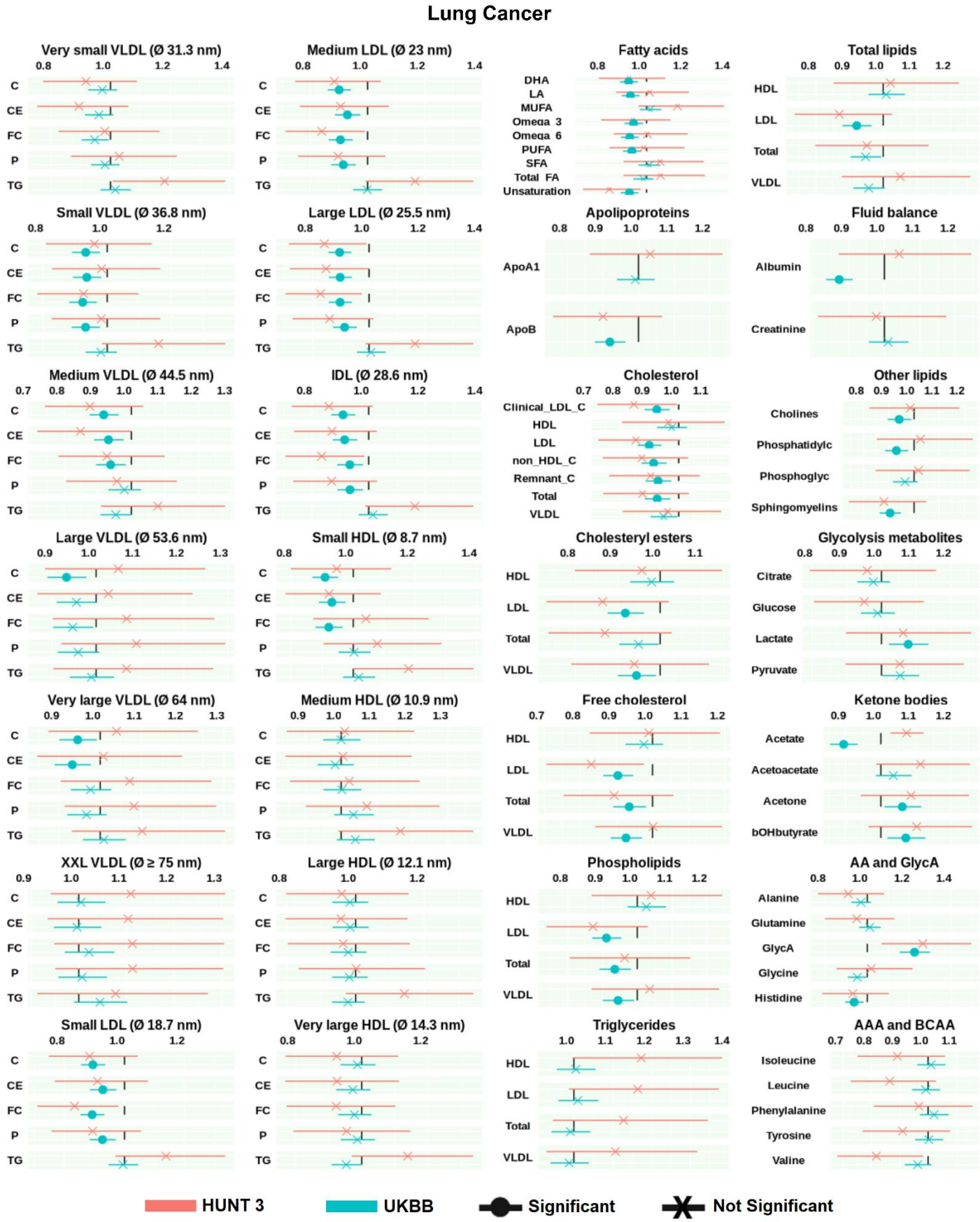
289

Figure 1. Hazard ratios (HR) of circulating metabolites in relation to breast cancer incidence were assessed in HUNT3 (events = 283, n = 9129) and the UKBB (events = 4516, n = 139,684). The data are presented as HRs per 1-SD increment with 95% confidence intervals (CIs), based on SD-scaled concentrations. Cox proportional hazards regression models were adjusted for UKBB assessment center (for UKBB), BMI, alcohol consumption, smoking status, menarche status, and number of live births, with age used as the timescale. AA=Amino acids, AAA = Aromatic amino acids, ApoA1 = Apolipoprotein A1, ApoB = Apolipoprotein B, bOHbutyrate = Beta-hydroxybutyrate, BCAA = Branched-chain amino acids, CE = Cholesteryl esters, C = Cholesterol, DHA = Docosahexaenoic acid, FC = Free cholesterol, GlycA = Glycoprotein Acetylation, HDL = High-density lipoprotein, LA = Linoleic acid, LDL = Low-density lipoprotein, MUFA = Monounsaturated fatty acids, Phosphatidylc = Phosphatidylcholine, Phosphoglyc = Phosphoglycerides, P = Phospholipids, PUFA = Polyunsaturated fatty acids, SFA = Saturated fatty acids, TG = Triglycerides, Total FA = Total fatty acids, VLDL = Very low-density lipoprotein, XXL = Extremely large, Ø = Diameter.

Prostate Cancer
Very small VLDL (Ø 31.3 nm)
Medium LDL (Ø 23 nm)
Fatty acids
Total lipids
Small VLDL (Ø 36.8 nm)
Large LDL (Ø 25.5 nm)
Apolipoproteins
Fluid balance
Medium VLDL (Ø 44.5 nm)
IDL (Ø 28.6 nm)
Cholesterol
Other lipids
Large VLDL (Ø 53.6 nm)
Small HDL (Ø 8.7 nm)
Cholesteryl esters
Glycolysis metabolites
Very large VLDL (Ø 64 nm)
Medium HDL (Ø 10.9 nm)
Free cholesterol
Ketone bodies
XXL VLDL (Ø ≥ 75 nm)
Large HDL (Ø 12.1 nm)
Phospholipids
AA and GlycA
Small LDL (Ø 18.7 nm)
Very large HDL (Ø 14.3 nm)
Triglycerides
AAA and BCAA
C
CE
FC
P
TG
DHA
LA
MUFA
Omega 3
Omega 6
PUFA
SFA
Total FA
Unsaturation
HDL
LDL
Total
VLDL
ApoA1
ApoB
Albumin
Creatinine
Clinical_LDL_C
non_HDL_C
Remnant_C
Cholines
Phosphatidylc
Phosphoglyc
Sphingomyelins
Citrate
Glucose
Lactate
Pyruvate
Acetate
Acetoacetate
Acetone
bOHbutyrate
Alanine
Glutamine
GlycA
Glycine
Histidine
Isoleucine
Leucine
Phenylalanine
Tyrosine
Valine
HUNT 3
UKBB
Significant
Not Significant

303

Figure 2. Hazard ratios (HR) of circulating metabolites in relation to prostate cancer incidence were assessed in HUNT3 (events = 452, n = 7,748) and the UKBB (events = 5,576, n = 121,121). The data are presented as HRs per 1-SD increment with 95% confidence intervals (CIs), based on SD-scaled concentrations. Cox proportional hazards regression models were adjusted for UKBB assessment center (for UKBB), BMI, alcohol consumption, smoking status, marital status, with age used as the timescale. AA=Amino acids, AAA = Aromatic amino acids, ApoA1 = Apolipoprotein A1, ApoB = Apolipoprotein B, bOHbutyrate = Beta-hydroxybutyrate, BCAA = Branched-chain amino acids, CE = Cholesteryl esters, C = Cholesterol, Clinical_LDL_C, DHA = Docosahexaenoic acid, FC = Free cholesterol, GlycA = Glycoprotein Acetylation, HDL = High-density lipoprotein, LA = Linoleic acid, LDL = Low-density lipoprotein, MUFA = Monounsaturated fatty acids, Phosphatidylc = Phosphatidylcholine, Phosphoglyc = Phosphoglycerides, P = Phospholipids, PUFA = Polyunsaturated fatty acids, SFA = Saturated fatty acids, TG = Triglycerides, Total FA = Total fatty acids, VLDL = Very low-density lipoprotein, XXL = Extremely large, Ø = Diameter.


Lung Cancer
Very small VLDL (Ø 31.3 nm)
Medium LDL (Ø 23 nm)
Fatty acids
Total lipids
Small VLDL (Ø 36.8 nm)
Large LDL (Ø 25.5 nm)
Apolipoproteins
Fluid balance
Medium VLDL (Ø 44.5 nm)
IDL (Ø 28.6 nm)
Cholesterol
Other lipids
Large VLDL (Ø 53.6 nm)
Small HDL (Ø 8.7 nm)
Cholesteryl esters
Glycolysis metabolites
Very large VLDL (Ø 64 nm)
Medium HDL (Ø 10.9 nm)
Free cholesterol
Ketone bodies
XXL VLDL (Ø ≥ 75 nm)
Large HDL (Ø 12.1 nm)
Phospholipids
AA and GlycA
Small LDL (Ø 18.7 nm)
Very large HDL (Ø 14.3 nm)
Triglycerides
AAA and BCAA
C
CE
FC
P
TG
DHA
LA
MUFA
Omega 3
Omega 6
PUFA
SFA
Total FA
Unsaturation
HDL
LDL
Total
VLDL
ApoA1
ApoB
Albumin
Creatinine
Clinical_LDL_C
non_HDL_C
Remnant_C
Cholines
Phosphatidylc
Phosphoglyc
Sphingomyelins
Citrate
Glucose
Lactate
Pyruvate
Acetate
Acetoacetate
Acetone
bOHbutyrate
Alanine
Glutamine
GlycA
Glycine
Histidine
Isoleucine
Leucine
Phenylalanine
Tyrosine
Valine
HUNT 3
UKBB
Significant
Not Significant

317

Figure 3. Hazard ratios (HR) of circulating metabolites in relation to lung cancer incidence were assessed in HUNT3 (events = 175, n = 16,877) and the UKBB (events =1985, n = 260,805). The data are presented as HRs per 1-SD increment with 95% confidence intervals (CIs), based on SD-scaled concentrations. Cox proportional hazards regression models were adjusted for sex, UKBB assessment center (for UKBB), BMI, alcohol consumption, smoking status, with age used as the timescale. AA=Amino acids, AAA = Aromatic amino acids, ApoA1 = Apolipoprotein A1, ApoB = Apolipoprotein B, bOHbutyrate = Beta-hydroxybutyrate, BCAA = Branched-chain amino acids, CE = Cholesteryl esters, C = Cholesterol, Clinical_LDL_C, DHA = Docosahexaenoic acid, FC = Free cholesterol, GlycA = Glycoprotein Acetylation, HDL = High-density lipoprotein, LA = Linoleic acid, LDL = Low-density lipoprotein, MUFA = Monounsaturated fatty acids, Phosphatidylc = Phosphatidylcholine, Phosphoglyc = Phosphoglycerides, P = Phospholipids, PUFA = Polyunsaturated fatty acids, SFA = Saturated fatty acids, TG = Triglycerides, Total FA = Total fatty acids, VLDL = Very low-density lipoprotein, XXL = Extremely large, Ø = Diameter.

Colorectal Cancer

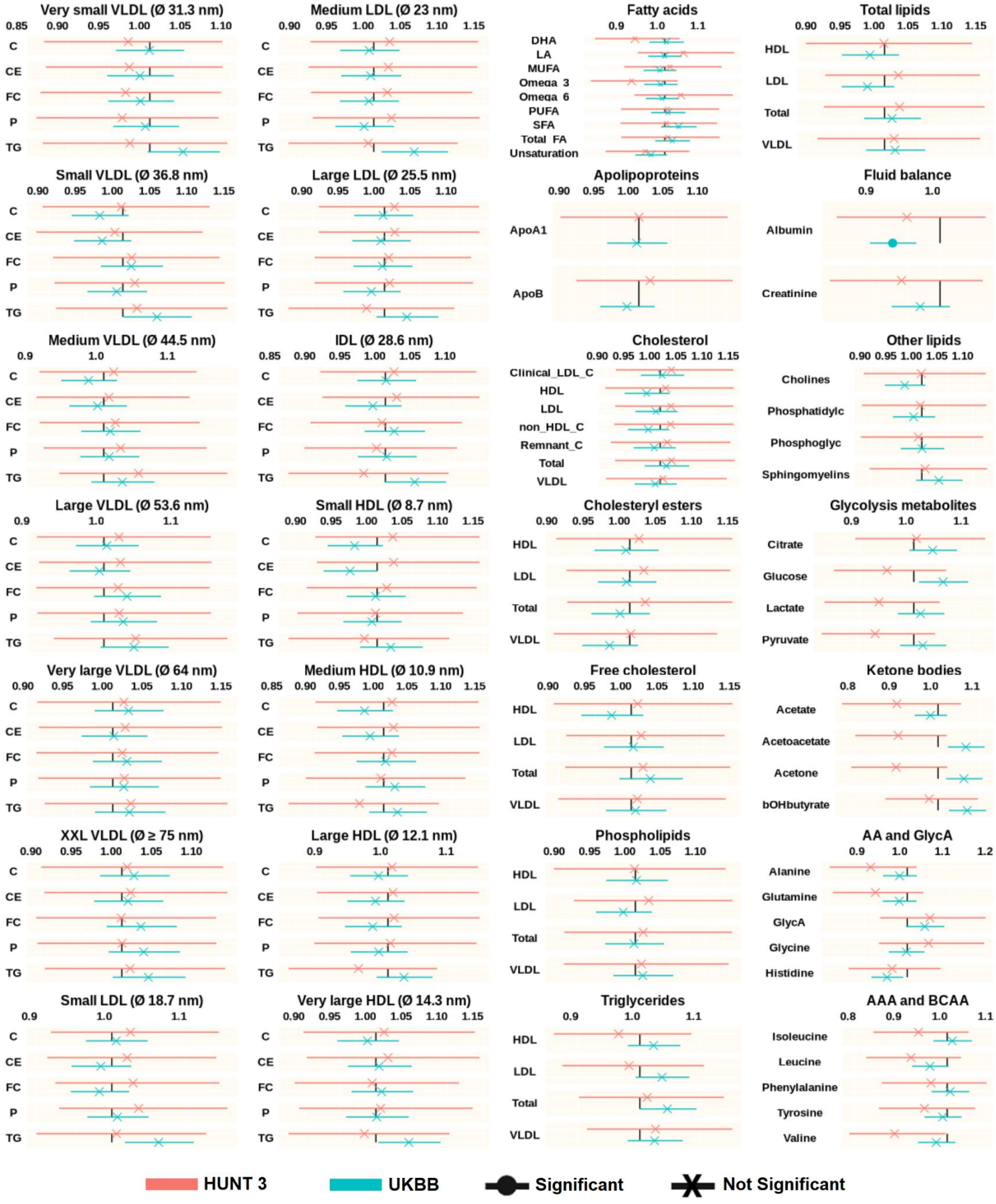


Figure 4. Hazard ratios (HR) of circulating metabolites in relation to colorectal cancer incidence were assessed in HUNT3 (events = 380, n = 16,877) and the UKBB (events = 2850, n = 260,805). The data are presented as HRs per 1-SD increment with 95% confidence intervals (CIs), based on SD-scaled concentrations. Cox proportional hazards regression models were adjusted for sex, UKBB assessment center (for UKBB), BMI, alcohol consumption, smoking status, with age used as the timescale. AA=Amino acids, AAA = Aromatic amino acids, ApoA1 = Apolipoprotein A1, ApoB = Apolipoprotein B, bOHbutyrate = Beta-hydroxybutyrate, BCAA = Branched-chain amino acids, CE = Cholesteryl esters, C = Cholesterol, Clinical_LDL_C, DHA = Docosahexaenoic acid, FC = Free cholesterol, GlycA = Glycoprotein Acetylation, HDL = High-density lipoprotein, LA = Linoleic acid, LDL = Low-density lipoprotein, MUFA = Monounsaturated fatty acids, Phosphatidylc = Phosphatidylcholine, Phosphoglyc = Phosphoglycerides, P = Phospholipids, PUFA = Polyunsaturated fatty acids, SFA = Saturated fatty acids, TG = Triglycerides, Total FA = Total fatty acids, VLDL = Very low-density lipoprotein, XXL = Extremely large, Ø = Diameter.

Discussion

This prospective study explored the association between circulating metabolites, including lipoprotein subfractions and lipids, and the incidence of BC, PCa, LC or CRC, using data from two large cohorts: HUNT3 and the UKBB. Several metabolites were associated with cancer incidence in UKBB, particularly for LC. Higher levels of triglycerides were associated with a lower risk of PCa. In contrast, higher levels of most lipid concentrations in LDL and IDL subfractions, except triglycerides, were associated with lower risk of LC. No clear associations were observed for BC before analyses were stratified based on women's menopausal status, where several lipoprotein subfractions showed a protective association in premenopausal women. Higher levels of albumin were inversely associated with both LC and CRC, while the inflammation marker GlycA showed a positive association with both cancers. In addition, free circulating ketone bodies (acetone, beta-hydroxybutyrate) were positively associated with incidence of LC and CRC. Except for PCa, results from HUNT3 were generally consistent with UKBB, but with wider confidence intervals, likely due to smaller sample size.

Lipids, including fats, oils, phospholipids, sterols, and triglycerides, are essential for energy storage, membrane structure, and cellular signaling (26, 28, 54-58). Lipoproteins, transporters of lipids in the blood, are classified into five types—VLDL, IDL, LDL, HDL, and chylomicrons. While LDL carries triglycerides and cholesterol from the liver to tissues in the

364 endogenous pathway, HDL collects excess cholesterol from tissues and blood vessels and
365 transports it back to the liver for removal (59). So far, few studies have investigated the role of
366 lipoprotein subfractions in cancer risk. We found the level of several lipoprotein subfractions
367 in the endogenous pathway to be inversely associated with LC. Inverse associations between
368 LDL-cholesterol and lung cancer risk have been seen in other large prospective cohorts (60),
369 and a Mendelian randomization study showed evidence for a causal effect of high levels of
370 LDL-cholesterol on a reduced risk of LC (61). However, results are conflicting as a recent
371 Mendelian randomization study reported no association between lipid subfractions and cancer
372 risk (62). Furthermore, several prospective studies based on NMR reported inverse associations
373 of free circulating lipids and fatty acids with long-term PCa risk (26, 55-58). In line with these
374 studies, we observed negative associations between the levels of total lipids or total fatty acids
375 and PCa risk in UKBB. Associations with fatty acid metabolites, which were also observed for
376 LC incidence, may reflect dietary risk factors, or reflect metabolic shifts years before diagnosis,
377 including increased β-oxidation (16, 63-65). We have previously reported inverse associations
378 of lipid fractions in VLDL with BC incidence in premenopausal women (29), and this finding
379 was replicated in premenopausal women in UKBB in the current study.

380 Positive associations of ketone bodies—acetoacetate, acetone, and β-hydroxybutyrate—were
381 observed for all cancers in our study. Ketone bodies result mainly from fatty acid oxidation in
382 the liver, and changes in ketone bodies may reflect metabolic reprogramming that both may
383 promote or protect against cancer (66, 67).

384 Chronic low-grade inflammation is a well-known risk factor for cancer development, and
385 tumor-promoting inflammation is one of the hallmarks of cancer (18). Blood levels of GlycA
386 products are emerging as biomarkers for systemic inflammation (68-70). Our analyses indicated
387 that elevated GlycA levels were associated with increased incidences of both LC and CRC.
388 Similarly, low blood concentration of albumin may reflect inflammation (71), and albumin was

389 found to be inversely associated with incidence of LC, CRC and PCa. Surprisingly, neither
390 GlycA nor albumin was associated with incidence of BC. Our results are not in agreement with
391 a previous prospective case-cohort study in the EPIC cohort, which reported an inverse
392 association between albumin levels and BC incidence, but not with PCa, LC, or CRC incidence
393 (72). Thus, further studies are needed to determine the association between albumin and cancer
394 risk.

395 Dietary factors may influence several circulating metabolite levels, including glycolysis
396 metabolites, lipids, fatty acids, and glycerol-derived compounds which are all affected by
397 dietary fat and carbohydrate intake (73-76). High-fat diets tend to elevate circulating lipid
398 metabolites, whereas low-fat or calorie-restricted diets may reduce their availability. In this
399 study, we did not account for dietary intake for two main reasons —a) due to the differences in
400 the variables and questionnaires used to capture diet and lifestyle features between HUNT3 and
401 UKBB and b) dietary questionnaires are predominantly self-reported, which may introduce a
402 response bias. However, future investigations incorporating detailed dietary information are
403 warranted.

404 As metabolite levels were measured in blood samples collected years before diagnosis in our
405 study, the observed associations between metabolite concentrations and cancer incidence may
406 reflect broader metabolic alterations that promote the development and growth of cancer cells.
407 Comparing results from metabolomics studies poses certain challenges due to variations in
408 population characteristics and analytical techniques (77). Metabolites in both cohorts were
409 quantified by Nightingale Health; however, different blood sample types were used: EDTA-
410 plasma in UKBB and serum in HUNT3, which may influence metabolite measurements.
411 Nonetheless, replicating potential associations—both for individual metabolites and more
412 broadly across metabolite classes—is crucial for reliable interpretation of findings. While
413 metabolic associations in UKBB and HUNT3 were mainly showing the same trends for LC and

414 CRC, striking discrepancies were observed for PCa, suggesting potential cohort-specific
415 differences in population characteristics, data collection methods, or underlying biological
416 mechanisms.

417 The key strengths of this study include: 1) the utilization of two distinct and large population-
418 based biobanks, 2) the use of NMR metabolomics, a highly robust method for quantifying
419 metabolites and lipoproteins, and 3) a prospective design with extended follow-up.
420 Additionally, we conducted a sensitivity analysis by excluding participants with any form of
421 cancer prior to or at recruitment. The associations remained consistent, indicating no
422 confounding influence from other malignancies on the observed relationships. The metabolic
423 profiles for both UKBB and HUNT3 were measured in a non-fasting state, with only a one-
424 hour difference in the average time since last meal, which likely had minimal impact,
425 maintaining cohort comparability. However, differences in dietary habits between UKBB and
426 HUNT3 could still pose a comparability issue. Previous studies have demonstrated that lipid
427 levels exhibit minimal changes in response to typical food intake in the general population (73-
428 76, 78), and that non-fasting lipid profiles could reliably predict an increased risk of
429 cardiovascular events (79). However, some small-molecular metabolites are very sensitive to
430 diet and may be affected by the time since last meal (80). Although HUNT3 mirrored the
431 associations observed in UKBB for LC incidence, the lower number of cases in HUNT3 may
432 have contributed to less precise results. The incidence of PCa in HUNT3 (5.9%) was higher
433 than in UKBB (4.6%). This difference may explain the discrepancy in the direction of
434 associations in the two cohorts, potentially due to varying baseline risk profiles or differing
435 environmental and genetic factors influencing disease risk. Moreover, UKBB represents an
436 older population, while HUNT3 encompasses a broader age range (18–100 years vs. 40–69
437 years). Additionally, the low participation rate in UKBB (5.5%) may have introduced selection
438 bias, as participants tend to be healthier than non-participants (81). In contrast, HUNT3, with a

higher participation rate (54.1%), is considered more representative of the general population (82).

Conclusion

This prospective metabolomic study examined the associations between blood metabolites and the incidence of breast, prostate, lung and colorectal cancers in two large population biobanks. The study highlights inverse associations between most lipids in LDL and IDL subfractions and lung cancer incidence, predominantly observed in the UK Biobank. Similar associations were observed for albumin and some fatty acids. Ketone bodies, including acetoacetate, acetone, and β-hydroxybutyrate, as well as the inflammation-related metabolite glycoprotein acetylation, showed positive associations with incidence of lung cancer. Although the direction of these associations was generally consistent in the HUNT Study, none reached statistical significance, likely due to the smaller sample size.

Supplementary material

The Supplementary data is available on GitHub.

Competing interests

The authors declare no conflict of interest.

Ethics declarations

The UKBB study was approved by the Northwest Multi-Centre Research Ethics Committee (reference number 11/NW/0382). The use of UKBB data in the current study was approved by the Regional Committees for Medical and Health Research Ethics (REK 427916). All participants signed an informed consent form. The UKBB database was made in accordance with the Declaration of Helsinki. The HUNT Study was approved by the Regional

Committee for Ethics in Medical Research (REK; reference number 468380). All participants signed informed written consent on participation in HUNT, with linkage to previous HUNT surveys and specific registries in accordance with the Declaration of Helsinki.

Acknowledgement

This research has been conducted using data from UK Biobank, a major biomedical database (https://www.ukbiobank.ac.uk/), under application number 40135. The Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Center (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health.

Funding

The study was funded by the Norwegian Cancer Society (project number 221998-2021). Open access funding is provided by the Norwegian University of Science and Technology - NTNU (including St. Olavs Hospital - Trondheim University Hospital).

Data availability

The UK Biobank data are available to researchers, subject to successful registration and application process via their website (https://www.ukbiobank.ac.uk/). Data from the HUNT Study used in research projects will, when reasonably requested by others, be made available on request to the HUNT Data Access Committee. The HUNT data access information describes the policy regarding data availability (https://www.ntnu.edu/hunt/data).

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63.

2. Global cancer burden growing, amidst mounting need for services Lyon, France; Geneva, Switzerland: World Health Organization (WHO); 2024 [Available from:

https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services#:~:text=In%202022%2C%20there%20were%20an,women%20die%20from%20the%20disease.

3. Cancer Registry of Norway. Norwegian Institute of Public Health. Cancer in Norway 2023 - Cancer incidence, mortality, survival and prevalence in Norway. 2024.

4. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft: Helsedirektoratet; 2024 [updated 12 Sep 2024; cited 2024. Available from: https://www.helsebiblioteket.no/innhold/nasjonal-faglig-retningslinje/brystkreft-handlingsprogram.

5. Howell A, Anderson AS, Clarke RB, Duffy SW, Evans DG, Garcia-Closas M, et al. Risk determination and prevention of breast cancer. Breast Cancer Res. 2014;16(5):446.

6. Fabian CJ, Kimler BF, Hursting SD. Omega-3 fatty acids for breast cancer prevention and survivorship. Breast Cancer Res. 2015;17(1):62.

7. Huang J, Chan PS, Lok V, Chen X, Ding H, Jin Y, et al. Global incidence and mortality of breast cancer: a trend analysis. Aging (Albany NY). 2021;13(4):5748–803.

8. Rawla P. Epidemiology of Prostate Cancer. World J Oncol. 2019;10(2):63–89.

9. Odedina FT, Akinremi TO, Chinegwundoh F, Roberts R, Yu D, Reams RR, et al. Prostate cancer disparities in Black men of African descent: a comparative literature review of prostate cancer burden among Black men in the United States, Caribbean, United Kingdom, and West Africa. Infect Agent Cancer. 2009;4 Suppl 1(Suppl 1):S2.

10. Kimura T. East meets West: ethnic differences in prostate cancer epidemiology between East Asians and Caucasians. Chin J Cancer. 2012;31(9):421–9.

11. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet. 2018;391(10125):1023–75.

12. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet. 2014;383(9927):1490–502.

13. Baharum SN, Azizan KA. Metabolomics in Systems Biology. In: Aizat WM, Goh H-H, Baharum SN, editors. Omics Applications for Systems Biology. Cham: Springer International Publishing; 2018. p. 51–68.

14. Barrett JC, Esko T, Fischer K, Jostins-Dean L, Jousilahti P, Julkunen H, et al. Metabolomic and genomic prediction of common diseases in 700,217 participants in three national biobanks. Nature Communications. 2024;15(1):10092.

15. Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: Current trends and future perspectives. Journal of Pharmaceutical and Biomedical Analysis. 2014;87:1–11.

16. Schmidt DR, Patel R, Kirsch DG, Lewis CA, Vander Heiden MG, Locasale JW. Metabolomics in cancer research and emerging applications in clinical oncology. CA: A Cancer Journal for Clinicians. 2021;71(4):333–58.

17. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of Toxicology. 2011;85(1):5–17.

18. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31–46.

19. Koundouros N, Poulogiannis G. Reprogramming of fatty acid metabolism in cancer. British Journal of Cancer. 2020;122(1):4–22.

20. Giallourou N, Urbaniak C, Puebla-Barragan S, Vorkas PA, Swann JR, Reid G. Characterizing the breast cancer lipidome and its interaction with the tissue microbiota. Communications Biology. 2021;4(1):1229.

21. Pelton K, Freeman MR, Solomon KR. Cholesterol and prostate cancer. Curr Opin Pharmacol. 2012;12(6):751–9.

22. Shafique K, McLoone P, Qureshi K, Leung H, Hart C, Morrison DS. Cholesterol and the risk of grade-specific prostate cancer incidence: evidence from two large prospective cohort studies with up to 37 years' follow up. BMC Cancer. 2012;12:25.

23. Nielsen SF, Nordestgaard BG, Bojesen SE. Statin use and reduced cancer-related mortality. N Engl J Med. 2012;367(19):1792–802.

24. Zeleznik OA, Balasubramanian R, Zhao Y, Frueh L, Jeanfavre S, Avila-Pacheco J, 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):54.

25. Lécuyer L, Dalle C, Lyan B, Demidem A, Rossary A, Vasson MP, et al. Plasma Metabolomic Signatures Associated with Long-term Breast Cancer Risk in the SU.VI.MAX Prospective Cohort. Cancer Epidemiol Biomarkers Prev. 2019;28(8):1300–7.

26. Kühn T, Floegel A, Sookthai D, Johnson T, Rolle-Kampczyk U, Otto W, et al. Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. BMC Medicine. 2016;14(1):13.

27. His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Medicine. 2019;17(1):178.

28. Lécuyer L, Victor Bala A, Deschasaux M, Bouchemal N, Nawfal Triba M, Vasson MP, et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. Int J Epidemiol. 2018;47(2):484–94.

29. Debik J, Schäfer H, Andreassen T, Wang F, Fang F, Cannet C, et al. Lipoprotein and metabolite associations to breast cancer risk in the HUNT2 study. Br J Cancer. 2022;127(8):1515–24.

30. Schmidt JA, Fensom GK, Rinaldi S, Scalbert A, Appleby PN, Achaintre D, et al. Patterns in metabolite profile are associated with risk of more aggressive prostate cancer: A prospective study of 3,057 matched case-control sets from EPIC. Int J Cancer. 2020;146(3):720–30.

31. Lécuyer L, Victor Bala A, Demidem A, Rossary A, Bouchemal N, Triba MN, et al. NMR metabolomic profiles associated with long-term risk of prostate cancer. Metabolomics. 2021;17(3):32.

32. Schmidt JA, Fensom GK, Rinaldi S, Scalbert A, Appleby PN, Achaintre D, et al. Pre-diagnostic metabolite concentrations and prostate cancer risk in 1077 cases and 1077 matched controls in the European Prospective Investigation into Cancer and Nutrition. BMC Med. 2017;15(1):122.

33. His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med. 2019;17(1):178.

34. Jobard E, Dossus L, Baglietto L, Fornili M, Lécuyer L, Mancini FR, et al. Investigation of circulating metabolites associated with breast cancer risk by untargeted metabolomics: a case-control study nested within the French E3N cohort. Br J Cancer. 2021;124(10):1734–43.

35. Cross AJ, Moore SC, Boca S, Huang WY, Xiong X, Stolzenberg-Solomon R, et al. A prospective study of serum metabolites and colorectal cancer risk. Cancer. 2014;120(19):3049–57.

36. Vidman L, Zheng R, Bodén S, Ribbenstedt A, Gunter MJ, Palmqvist R, et al. Untargeted plasma metabolomics and risk of colorectal cancer-an analysis nested within a large-scale prospective cohort. Cancer Metab. 2023;11(1):17.

37. Wikoff WR, Hanash S, DeFelice B, Miyamoto S, Barnett M, Zhao Y, et al. Diacetylspermine Is a Novel Prediagnostic Serum Biomarker for Non-Small-Cell Lung Cancer and Has Additive Performance With Pro-Surfactant Protein B. J Clin Oncol. 2015;33(33):3880–6.

38. Buergel T, Steinfeldt J, Ruyoga G, Pietzner M, Bizzarri D, Vojinovic D, et al. Metabolomic profiles predict individual multidisease outcomes. Nat Med. 2022;28(11):2309–20.

39. Julkunen H, Cichońska A, Tiainen M, Koskela H, Nybo K, Mäkelä V, et al. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. Nat Commun. 2023;14(1):604.

40. Huang J, Mondul AM, Weinstein SJ, Koutros S, Derkach A, Karoly E, et al. Serum metabolomic profiling of prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial. Br J Cancer. 2016;115(9):1087–95.

41. Kelly RS, Vander Heiden MG, Giovannucci E, Mucci LA. Metabolomic Biomarkers of Prostate Cancer: Prediction, Diagnosis, Progression, Prognosis, and Recurrence. Cancer Epidemiol Biomarkers Prev. 2016;25(6):887–906.

42. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort Profile: the HUNT Study, Norway. Int J Epidemiol. 2013;42(4):968–77.

43. Åsvold BO, Langhammer A, Rehn TA, Kjelvik G, Grøntvedt TV, Sørgjerd EP, et al. Cohort Profile Update: The HUNT Study, Norway. Int J Epidemiol. 2023;52(1):e80–e91.

44. Næss M, Kvaløy K, Sørgjerd EP, Sætermo KS, Norøy L, Røstad AH, et al. Data Resource Profile: The HUNT Biobank. International Journal of Epidemiology. 2024;53(3).

45. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.

46. Nightingale Health. NMR metabolomics companion document (Phase 2 update) [21.11.2024]. Available from: https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=130.

47. Bjelland EK, Hofvind S, Byberg L, Eskild A. The relation of age at menarche with age at natural menopause: a population study of 336 788 women in Norway. Hum Reprod. 2018;33(6):1149–57.

48. Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer. 2009;45(7):1218–31.

49. Conroy MC, Lacey B, Bešević J, Omiyale W, Feng Q, Effingham M, et al. UK Biobank: a globally important resource for cancer research. Br J Cancer. 2023;128(4):519–27.

50. Møller H, Richards S, Hanchett N, Riaz SP, Lüchtenborg M, Holmberg L, et al. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. Br J Cancer. 2011;105(1):170–6.

51. World Health Organization. International classification of diseases for oncology (ICD-O). 3rd ed: World Health Organization; 2013.

52. Therneau TM. A Package for Survival Analysis in R. 2024.

53. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Computational Biology. 2017;13(11):e1005752.

54. Bro R, Kamstrup-Nielsen MH, Engelsen SB, Savorani F, Rasmussen MA, Hansen L, et al. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. Metabolomics. 2015;11(5):1376–80.

55. Schmidt JA, Fensom GK, Rinaldi S, Scalbert A, Appleby PN, Achaintre D, et al. Patterns in metabolite profile are associated with risk of more aggressive prostate cancer: A prospective study of 3,057 matched case–control sets from EPIC. International Journal of Cancer. 2020;146(3):720–30.

56. Schmidt JA, Fensom GK, Rinaldi S, Scalbert A, Appleby PN, Achaintre D, et al. Pre-diagnostic metabolite concentrations and prostate cancer risk in 1077 cases and 1077 matched controls in the European Prospective Investigation into Cancer and Nutrition. BMC Medicine. 2017;15(1):122.

57. Mondul AM, Moore SC, Weinstein SJ, Karoly ED, Sampson JN, Albanes D. Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocopherol, beta-carotene cancer prevention (ATBC) study. International Journal of Cancer. 2015;137(9):2124–32.

58. Mondul AM, Moore SC, Weinstein SJ, Männistö S, Sampson JN, Albanes D. 1-stearoylglycerol is associated with risk of prostate cancer: results from serum metabolomic profiling. Metabolomics. 2014;10(5):1036–41.

59. Feingold KR. Introduction to lipids and lipoproteins. Endotext [internet]. 2024.

60. Lyu Z, Li N, Wang G, Feng X, Chen S, Su K, et al. Independent and joint associations of blood lipids and lipoproteins with lung cancer risk in Chinese males: A prospective cohort study. Int J Cancer. 2019;144(12):2972–84.

61. Carreras-Torres R, Johansson M, Haycock PC, Wade KH, Relton CL, Martin RM, et al. Obesity, metabolic factors and risk of different histological types of lung cancer: A Mendelian randomization study. PLoS One. 2017;12(6):e0177875.

62. Carter P, Vithayathil M, Kar S, Potluri R, Mason AM, Larsson SC, et al. Predicting the effect of statins on cancer risk using genetic variants from a Mendelian randomization study in the UK Biobank. Elife. 2020;9.

63. His M, Zelek L, Deschasaux M, Pouchieu C, Kesse-Guyot E, Hercberg S, et al. Prospective associations between serum biomarkers of lipid metabolism and overall, breast and prostate cancer risk. Eur J Epidemiol. 2014;29(2):119–32.

64. Röhrig F, Schulze A. The multifaceted roles of fatty acid synthesis in cancer. Nat Rev Cancer. 2016;16(11):732–49.

65. Günther UL. Metabolomics Biomarkers for Breast Cancer. Pathobiology. 2015;82(3-4):153–65.

66. Puchalska P, Crawford PA. Multi-dimensional Roles of Ketone Bodies in Fuel Metabolism, Signaling, and Therapeutics. Cell Metab. 2017;25(2):262–84.

655 67. Suresh VV, Sivaprakasam S, Bhutia YD, Prasad PD, Thangaraju M, Ganapathy V. Not Just an
656 Alternative Energy Source: Diverse Biological Functions of Ketone Bodies and Relevance of
657 HMGCS2 to Health and Disease. Biomolecules. 2025;15(4).
658 68. Gruppen EG, Connelly MA, Vart P, Otvos JD, Bakker SJ, Dullaart RP. GlycA, a novel
659 proinflammatory glycoprotein biomarker, and high-sensitivity C-reactive protein are inversely
660 associated with sodium intake after controlling for adiposity: the Prevention of Renal and Vascular
661 End-Stage Disease study. Am J Clin Nutr. 2016;104(2):415–22.
662 69. Chung CP, Ormseth MJ, Connelly MA, Oeser A, Solus JF, Otvos JD, et al. GlycA, a novel
663 marker of inflammation, is elevated in systemic lupus erythematosus. Lupus. 2016;25(3):296–300.
664 70. Lawler PR, Akinkuolie AO, Chandler PD, Moorthy MV, Vandenburgh MJ, Schaumberg DA,
665 et al. Circulating N-Linked Glycoprotein Acetyls and Longitudinal Mortality Risk. Circulation
666 Research. 2016;118(7):1106–15.
667 71. Sheinenzon A, Shehadeh M, Michelis R, Shaoul E, Ronen O. Serum albumin levels and
668 inflammation. International Journal of Biological Macromolecules. 2021;184:857–62.
669 72. Kühn T, Sookthai D, Graf ME, Schübel R, Freisling H, Johnson T, et al. Albumin, bilirubin,
670 uric acid and cancer risk: results from a prospective population-based study. Br J Cancer.
671 2017;117(10):1572–9.
672 73. Mensink RP, Katan MB. Effect of dietary fatty acids on serum lipids and lipoproteins. A meta-
673 analysis of 27 trials. Arteriosclerosis and Thrombosis: A Journal of Vascular Biology. 1992;12(8):911–
674 9.
675 74. Ahrens E, Insull W, Blomstrand R, Hirsch J, Tsaltas T, Peterson M. THE INFLUENCE OF
676 DIETARY FATS ON SERUM-LIPID LEVELS IN MAN. The Lancet. 1957;269(6976):943–53.
677 75. Rasmussen OW, Thomsen C, Hansen KW, Vesterlund M, Winther E, Hermansen K. Effects on
678 Blood Pressure, Glucose, and Lipid Levels of High-Monounsaturated Fat Diet Compared With a High-
679 Carbohydrate Diet in NIDDM Subjects. Diabetes Care. 1993;16(12):1565–71.
680 76. Kinsell LW, Partridge J, Boling L, Margen S, Michaels G. Dietary modification of serum
681 cholesterol and phospholipid levels. J Clin Endocrinol Metab. 1952;12(7):909–13.

682 77. Yu Z, Kastenmüller G, He Y, Belcredi P, Möller G, Prehn C, et al. Differences between Human
683 Plasma and Serum Metabolite Profiles. PLOS ONE. 2011;6(7):e21230.
684 78. Schwingshackl L, Hoffmann G. Comparison of effects of long-term low-fat vs high-fat diets
685 on blood lipid levels in overweight or obese patients: a systematic review and meta-analysis. J Acad
686 Nutr Diet. 2013;113(12):1640–61.
687 79. Langsted A, Freiberg JJ, Nordestgaard BG. Fasting and nonfasting lipid levels: influence of
688 normal food intake on lipids, lipoproteins, apolipoproteins, and cardiovascular risk prediction.
689 Circulation. 2008;118(20):2047–56.
690 80. Giskeødegård GF, Andreassen T, Bertilsson H, Tessem MB, Bathen TF. The effect of sampling
691 procedures and day-to-day variations in metabolomics studies of biofluids. Anal Chim Acta.
692 2019;1081:93–102.
693 81. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of
694 Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the
695 General Population. Am J Epidemiol. 2017;186(9):1026–34.
696 82. Langhammer A, Krokstad S, Romundstad P, Heggland J, Holmen J. The HUNT study:
697 participation is associated with survival and depends on socioeconomic status, diseases and symptoms.
698 BMC Med Res Methodol. 2012;12:143.

699