Novel biomarkers of habitual alcohol intake and associations with risk of pancreatic and liver cancers and liver disease mortality

Erikka Loftfield, PhD [1]*; Magdalena Stepien, PhD [2]*; Vivian Viallon, PhD [3]; Laura Trijsburg, PhD [3]; Joseph Rothwell, PhD [2,5,6]; Nivonirina Robinot, MSc [4]; Carine Biessy, BSC [3]; Ingvar A. Bergdahl, PhD [7]; Stina Bodén, MSC [8]; Matthias B. Schulze, PhD [9,10]; Manuela Bergman, PhD [9,10]; Elisabete Weiderpass, MD, MSc, PhD [11]; Julie A. Schmidt, PhD [12]; Raul Zamora-Ros, PhD [13]; Therese H. Nøst, PhD [14]; Torkjel M Sandanger, PhD [14]; Emily Sonestedt, PhD [15]; Bodil Ohlsson, PhD [15]; Verena Katzke, PhD [16]; Rudolf Kaaks, PhD [16]; Fulvio Ricceri, PhD [17]; Anne Tjønneland, PhD [18]; Christina C. Dahm, PhD [19]; Maria-Jose Sánchez, PhD [20,21,22]; Antonia Trichopoulou, PhD [23]; Rosario Tumino, MD, MSc, DLSHTM[24]; María-Dolores Chirlaque, PhD [25,26]; Giovanna Masala, PhD [27]; Eva Ardanaz, PhD [28.29.30]; Roel Vermeulen, PhD [31]; Paul Brennan, PhD [32]; Demetrius Albanes, MD[1]; Stephanie J. Weinstein, PhD[1]; Augustin Scalbert, PhD [4]; Neal D. Freedman, PhD [1]; Marc J. Gunter, PhD [2]; Mazda Jenab, PhD [2]; Rashmi Sinha, PhD [1]; Pekka Keski-Rahkonen, PhD [4] ‡; Pietro Ferrari, PhD [3] ‡†

* these first authors contributed equally
‡ these senior authors contributed equally
† corresponding author

[1] Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[2] Nutritional Epidemiology Group, International Agency for Research on Cancer (IARC-WHO), Lyon, France.

[3] Nutritional Epidemiology and Biostatistics Group, International Agency for Research on Cancer (IARC-WHO), Lyon, France.

[4] Biomarkers Group, International Agency for Research on Cancer (IARC-WHO), Lyon, France.

[5] Centre for Epidemiology and Population Health (U1018), Generations and Health team, Faculté de Médecine, Université Paris-Saclay, UVSQ, INSERM, Villejuif, France.

[6] Gustave Roussy, F-94805, Villejuif, France.

[7] Biobank Research Unit, Umeå University, Sweden.

[8] Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden.

[9] Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany.

[10] Institute of Nutritional Science, University of Potsdam, Nuthetal, Germany.

[11] International Agency for Research on Cancer, World Health Organization.

[12] Cancer Epidemiology Unit, Nuffield Deaprtment of Population Health, University of Oxford, Oxford, United Kingdom.

[13] Unit of Nutrition and Cancer, Epidemiology Research Program, Catalan Institute of Oncology, Bellvitge Biomedical Research Institute (IDIBELL), Hospitalet de Llobregat (Barcelona), Spain.

[14] Department of Community Medicine, UiT- The Arctic University of Norway, Tromsø, Norway.

[15] Department of Clinical Sciences in Malmö, Lund University, Malmö, Sweden.

[16] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

[17] Department of Clinical and Biological Sciences, University of Turin, Italy; Unit of Epidemiology, Regional Health Service ASL TO3, Grugliasco (TO), Italy.

[18] Danish Cancer Society Research Center; University of Copenhagen, Department of Public Health

[19] Department of Public Health, Aarhus University, Denmark.

[20] Escuela Andaluza de Salud Pública (EASP), Granada, Spain; Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain.

[21] Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain.

[22] Department of Preventive Medicine and Public Health, University of Granada, Granada, Spain.

[23] Hellenic Health Foundation, Athens, Greece.

[24] Cancer Registry and Histopathology Department, Provincial Health Authority (ASP 7) Ragusa , Italy.

[25] Department of Epidemiology, Regional Health Council, IMIB-Arrixaca, Murcia University, Murcia, Spain.

[26] CIBER in Epidemiology and Public Health (CIBERESP), Madrid, Spain.

[27] Cancer Risk Factors and Life-Style Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network - ISPRO, Florence, Italy.

[28] Navarra Public Health Institute, Pamplona, Spain.

[29] IdiSNA, Navarra Institute for Health Research, Pamplona, Spain.

[30] CIBER Epidemiology and Public Health CIBERESP, Madrid, Spain.

[31] Institute for Risk Assessment Sciences, Division of Environmental Epidemiology, Utrecht University, Utrecht, The Netherlands.

[32] Genetic Epidemiology Group, International Agency for Research on Cancer (IARC-WHO), Lyon, France.

Abstract

Background: Alcohol is an established risk factor for several cancers, but modest alcohol-cancer associations may be missed due to measurement error in self-reported assessments. The identification of biomarkers of habitual alcohol intake may enhance evidence on the role of alcohol in cancer onset.

Methods: Untargeted metabolomics was used to identify metabolites correlated with habitual alcohol intake in a discovery dataset from the European Prospective Investigation into Cancer and Nutrition (EPIC; n=454). Significant correlations were replicated in independent datasets of controls from case-control studies nested within EPIC (n=281) and the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC; n=438) study. Conditional logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals for associations of alcohol-associated metabolites and self-reported alcohol intake with risk of pancreatic cancer, hepatocellular carcinoma (HCC), liver cancer, and liver disease mortality in the contributing studies.

Results: Two metabolites displayed a dose-response association with alcohol intake: 2-hydroxy-3-methylbutyric acid and an unidentified compound ($m/z(+)$:231.0839). A 1-SD increase in $\log_2$-transformed levels of 2-hydroxy-3-methylbutyric acid was associated with risk of HCC (OR=2.14; 1.39-3.31) and pancreatic cancer (OR=1.65; 1.17-2.32) in EPIC and liver cancer (OR=2.00; 1.44-2.77) and liver disease mortality (OR=2.16; 1.63-2.86) in ATBC. Conversely, a 1-SD increase in $\log_2$-transformed questionnaire-derived alcohol intake was not associated with HCC or pancreatic cancer in EPIC or liver cancer in ATBC but was associated with liver disease mortality (OR=2.19; 1.60-2.98) in ATBC.

Conclusions: 2-Hydroxy-3-methylbutyric acid is a candidate biomarker of habitual alcohol intake that may advance the study of alcohol and cancer risk in population-based studies.

Keyword: alcohol intake, untargeted metabolomics, 2-hydroxy-3-methylbutyric acid, biomarkers, EPIC, ATBC

In 2016, an estimated 2.8 million deaths, corresponding to 6.8% and 2.2% of age-standardized deaths in men and women, respectively, were attributed to alcohol use worldwide [1]. Excessive alcohol consumption is an established risk factor for many acute and chronic health conditions [2], including cancers of the upper aerodigestive tract, female breast, liver, colon and rectum [3]. However, the relationship of alcohol, particularly light-to-moderate alcohol consumption, with other cancer sites remains controversial [4].

Self-reported alcohol intake is, like other dietary factors, prone to underreporting [5]. Although the extent and distribution of exposure misclassification is unknown [6], it is likely that observed associations between alcohol use and disease risk in prospective studies are attenuated and that estimates of alcohol-attributable death and disease are underestimated. Biomarkers of liver function and oxidative stress are used to study alcohol-related liver injury and alcoholic liver disease (ALD) [7, 8], but most alcohol consumers, particularly light-to-moderate consumers, will never manifest ALD. There are also biomarkers of recent (e.g., ethyl glucuronide) and heavy alcohol use (e.g., carbohydrate deficient transferrin and phosphatidylethanol (PEth)) [9-11]. However, biomarkers of habitual alcohol use, including light-to-moderate drinking, are needed to better assess alcohol exposure in epidemiological studies and to improve risk estimates for diseases including cancer where modest associations may exist.

Metabolomics is a powerful tool for discovering dietary biomarkers. When used in an untargeted mode, it can detect a wide range of compounds in biological samples including metabolites formed during digestion, metabolism and microbial fermentation [12, 13], making it well-suited for discovering novel biomarkers of exposure or response to habitual alcohol consumption. Herein we applied a multistep design, using untargeted metabolomics and independent discovery and replication datasets, to identify serum metabolites associated with habitual alcohol consumption among free-living individuals with a wide range of intake. We then estimated the associations of these candidate alcohol biomarkers with risk of pancreatic cancer, liver cancers, and liver disease mortality in the European Prospective Investigation into Cancer and Nutrition (EPIC) study and the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC).

METHODS

Study design

EPIC Study

EPIC recruitment and study procedures, including dietary assessment methods and blood collection are described extensively elsewhere [14]. Briefly, EPIC is a large cohort study of over half a million men and women recruited between 1992 and 2000 in 23 European centers. Diet, including average daily alcohol intake, over the 12 months before enrolment was assessed by validated country-specific food frequency questionnaires (FFQ) designed to capture local dietary habits with high compliance. Country-specific alcohol intake was calculated based on the estimated average glass volume and ethanol content for wine, beer, cider, sweet liquor, distilled spirits, or fortified wines, using information collected in standardized 24-hr dietary recalls from a subset of the cohort [15]. The correlation between alcohol intake estimated by FFQ and 24-hour dietary recall was 0.79 [16]. Blood samples were collected and stored at -196°C under liquid nitrogen at the International Agency for Research on Cancer (IARC) for all countries except Sweden (-80°C freezers), and Denmark (-150°C, nitrogen vapor).

Our study included a discovery and two replication datasets (**Figure 1**). The discovery set (n=454) was nested in the EPIC cross-sectional study [17, 18]. The first replication set included control subjects from two EPIC nested case-control studies of hepatocellular carcinoma (HCC; n=129) and pancreatic cancer (n=152) with untargeted metabolomics data [19-21]. Non-metastatic incident HCC (n=129) and pancreatic cancer (n=152) cases, were matched 1:1 with cancer-free controls on study center, sex, age at blood collection (± 1 year), date (± 6 months) and time of the day (± 2 h) of blood collection, fasting status, and, for women, exogenous hormone use. Follow-up was based on a combination of methods, including health insurance records, registries and active follow-up [14]. Approval for the EPIC study was obtained from the IARC ethics review board (Lyon, France) and local review bodies of participating institutions.

ATBC Study

The second replication set included two nested case-control studies in the ATBC cohort of male Finnish smokers [22]. In ATBC, participants reported on demographics, lifestyle, and medical history via questionnaires and donated a fasting serum sample at baseline, which was stored at -70°C. Participants were passively followed during the post-intervention period via linkage with the Finnish Cancer Registry and death registry. Liver cancer (n=229) and liver disease mortality (n=248) cases were individually matched 1:1 with controls, selected by incidence density sampling, on baseline age (+/-5 years) and serum draw date (+/-30 days) [23]. For this study, we excluded cases and controls with missing data on alcohol intake (n=72) and those with samples that failed laboratory analysis (n=14) resulting in an analytic sample of n=438 controls for cross-sectional analysis and 192 and 199 complete case-control sets for prospective analyses of liver cancer incidence and liver disease mortality, respectively. Approval for the ATBC study was obtained from the Institutional Review Boards of National Cancer Institute (Bethesda, Maryland), and the National Public Health Institute of Finland. EPIC and ATBC studies were conducted according to the guidelines of the Declaration of Helsinki; all participants provided written informed consent.

Metabolomics analyses

Sample analysis, data pre-processing, matching of features across datasets, and compound identification are desribed in detail in the **Supplementary Methods**. Briefly, all samples were analyzed by the same laboratory at IARC with a UHPLC-QTOF-MS system (1290 Binary LC system, 6550 QTOF mass spectrometer; Agilent Technologies, Santa Clara, CA) using reversed phase chromatography and electrospray ionization. Raw data were processed using Agilent MassHunter Qualitative analysis B.06.00, ProFinder B.08.00, and Mass Profiler Professional B.12.1 software with Agilent's recursive feature finding procedure. The *m/z* values of the features of interest were searched against the Human Metabolome Database (HMDB) [24] and METLIN [25].Compound identity was confirmed by comparison of chemical standards and representative samples.

Statistical analyses

We used an integrated workflow for metabolomics data analysis [26]. Features detected in <50% of the discovery set samples abd ackground features, (i.e., feature intensities present in all blanks with ratio of geometric mean intensities of non-blank:blank samples <5) were excluded. Feature intensities were $\log_2$-transformed. Study participants with >50% missing features and those identified as outliers by a PCA-based approach were excluded [27]. Missing values were imputed within each plate by a K-nearest neighbours method, with K=10 [28]. Last, feature intensities measured across plates within any single batch were normalised by applying a random forest-based approach to correct for unwanted variation [29]. In the EPIC discovery set and replication sets, these steps were applied on feature matrices acquired in positive and negative modes separately. In ATBC, these steps were applied on each batch.

In the discovery and replication sets, alcohol intake (g/day) was adjusted for age, sex, country (in EPIC only), body mass index (BMI, kg/m^2), smoking status and intensity, coffee consumption (g/day, log-transformed) via the residual method in linear regression models. Coffee drinking and coffee-associated metabolites have been strongly associated with risk of liver cancer and liver disease mortality in ATBC [30, 31]; for consistency, coffee drinking was considered a potential confounder across discovery and replication sets. Residuals for feature intensities were also adjusted for well plate number within the analytical batch, position within the plate (row and column indexes), and the study (EPIC replication) or batch indicator (ATBC replication) as random effects. We used the principal component partial-R^2 (PC-PR2) method [32] to quantify the contribution of alcohol and potential confounders to the variability of the 67 features intensities that were statistically significantly associated with alcohol intake in the discovery set [33].

We calculated Pearson correlation coefficients using the residuals for alcohol intake and for feature intensities; correlations with a false discovery rate (FDR)-corrected p-value<0.05 were considered statistically significant, and these f_1 features were carried forward for replication. In the first (EPIC) replication step, f_1 residual-adjusted correlation coefficients were computed and corrected by the more conservative Bonferroni method. The f_2 correlations with a p-value $<0.05/f_1$ were considered statistically significant and carried forward to the second replication step in ATBC, again using the residuals for alcohol intake and for feature intensities. The linearity of the association

between standardized residuals of 2-hydroxy-3-methylbutyric acid and of alcohol intake was evaluated with cubic regression splines with 5 knots [34], by comparing the log-likelihood of models with and without the non-linear terms to a chi-distribution with 2 degrees of freedom.

The EPIC replication set was used to define high- (quartile 4: alcohol intake >33.1 and >12.3 g/day in men and women, respectively)) and low-consumers (quartile 1: alcohol intake >0.1 but <4.9 and <1.1 g/day in men and women, respectively) . We used logistic regression to estimate the area under receiver operating characteristics (AUROC) curves [35] and evaluated the predictive accuracy of the residuals of each main feature (i.e., candidate biomarkers) to discriminate high-consumers from low-consumers for metabolites that replicated across studies. We used the leave-one-out cross validation scheme, to mitigate issues related to over fitting [36].

We estimated odds ratios (OR) and 95% confidence intervals (95% CI) for candidate features and HCC and pancreatic cancer in EPIC and liver cancer and fatal liver disease in ATBC using conditional logistic regression models. In crude (conditioned on the matching criteria only) and multivariable models, adjusting for potential confounders, feature intensities were $\log_2$-transformed, centered and scaled (i.e., mean=0, standard deviation=1) to ensure comparability of OR across different endpoints.

All statistical analyses were performed using the Statistical Analysis Software, release 9.4 (SAS Institute Inc., Cary, NC, USA) and R version 3.6.0 [37].

RESULTS

Population characteristics

Baseline participant characteristics are presented in **Table 1**. In the EPIC discovery set, most participants were women (57.5%) and never (52.2%) or former (26.4%) smokers. In the EPIC replication set, there was a higher percentage of men (52.7%) and a lower percentage of never smokers (46.2%). In the ATBC replication set, all participants were Finnish men and current smokers. Median alcohol intake was 10.0 g/day, 6.6 g/day, and 11.5 g/day in the EPIC discovery, EPIC and ATBC replication sets, respectively.

Biomarker discovery analysis

After excluding participant samples identified as outliers or as having too many missing values, the final discovery set comprised 451 and 452 study participants in positive and negative ionization mode datasets, respectively. The final EPIC replication set comprised 271 and 277 study participants in positive and negative ionization datasets, respectively. Residuals of 205 features in the discovery set were significantly correlated with residuals of alcohol intake (163 features in positive and 42 features in negative ionization mode; **Figure 1)**, with correlation coefficients ranging from -0.29 to 0.50 in log-log plots (**Table S1**).

Of the 205 features in the discovery set, 51 features in positive and 16 features in negative ionization mode (f_1=67) matched by mass and retention time with equivalent features in the EPIC replication set, and PC-PR2 analyses showed that alcohol intake explained >7% of variability in the feature intensities (f_1=67; **Figure 2**). Residuals of f_2=10 features were statistically significantly correlated with residuals of alcohol intake (**Table 2**). The first two features corresponded to a compound that could not be unequivocally identified, but had an identical mass, isotope pattern, ion formation (mostly $[M+Na]^+$ and $[M+HCOOH-H]^-$) and retention time to ethyl glucoside (HMDB0029968) [38]. However, chromatograms (**Supplementary Methods**) indicated a lack of specificity, and although fragmentation of the $[M+Na]^+$ ion could not be induced, our results suggest the unknown is a combination of ethyl-α-D-glucoside, ethyl-β-D-glucoside, and an additional structural isomer. The remaining eight features corresponded to a single compound, which was confirmed by comparison with an authentic standard as 2-hydroxy-3-methylbutyric acid (HMDB0000407).

For subsequent analyses, the feature with the greatest chromatographic intenstity (i.e., main feature) for each metabolite was used (**Table 2**). The discriminatory accuracy for high versus low alcohol consumption in cross-validated models that included both 2-hydroxy-3-methylbutyric acid and the unknown compound was 75% (95% CI: 69-86%) (**Figure 3**). The test for non-linearity using restricted regression spline was borderline significant (p=0.06; **Figure S1**). All seven positive ionization mode features selected in the EPIC replication set were confirmed in the ATBC replication set (**Table 2**). In the ATBC replication set, the discriminatory accuracy for alcohol consumption (high

vs. low consumers as defined in men in EPIC) of 2-hydroxy-3-methylbutyric acid and the unknown compound was 76% (95% CI: 68-84%), only 2% higher than the discriminatory accuracy observed for 2-hydroxy-3-methylbutyric acid alone (74%, 95% CI: 66-82%).

Disease risk associations

In multivariable models (**Table 3**), 2-hydroxy-3-methylbutyric acid was associated with increased odds of HCC (OR_{1-SD}=2.14: 1.39, 3.31) and pancreatic cancer (OR_{1-SD}=1.65: 1.17, 2.32) in EPIC, as well as liver cancer (OR_{1-SD}=2.00; 1.44, 2.77) and fatal liver disease (OR_{1-SD}=2.16; 1.63, 2.86) in ATBC; the unknown candidate biomarker was associated with increased odds of liver cancer (OR_{1-SD}=1.70; 95% CI: 1.29, 2.25) and liver disease mortality (OR=1.85; 95% CI: 1.39-2.46) in ATBC, but not with HCC or pancreatic cancer in EPIC. Alcohol intake was not associated with HCC (OR_{1-SD}=0.78; 95% CI: 0.56, 1.09) or pancreatic cancer risk (OR_{1-SD}=1.03: 0.77, 1.39) in EPIC, but was strongly associated with liver disease mortality (OR_{1-SD}=2.19: 95% CI, 1.60, 2.98) in ATBC. The alcohol findings are in line with previously published EPIC and ATBC analyses [39-41].

DISCUSSION

Using untargeted metabolomics data from a discovery and two independent replication sets, we found two serum metabolites that were highly correlated with self-reported habitual alcohol intake. One compound was identified as 2-hydroxy-3-methylbutyric acid; the other remains unknown but is likely a combination of isomers of ethyl glucoside. Of note, ethyl-α-D-glucoside is a known constituent of some alcoholic beverages [38]. Both compounds could discriminate high alcohol consumers from low consumers in EPIC and ATBC despite marked differences in the study designs. Notably, 2-hydroxy-3-methylbutyric acid was strongly associated with HCC and pancreatic cancer risks in EPIC, and with liver cancer and fatal liver disease in ATBC. In contrast, self-reported alcohol intake was only consistently associated with risk of liver disease mortality in ATBC. Further research is needed to elucidate the potential metabolic cascade from alcohol drinking to 2-hydroxy-3-methylbutyric acid to disease. Additionally, studies measuring circulating concentrations of 2-hydroxy-3-methylbutyric acid rather than relative levels are now needed to compare across studies and

improve risk estimation; this is especially important for diseases such as pancreatic cancer, for which the literature is suggestive [42] yet inconsistent [43].

Prior population-based studies have used a targeted or semi-targeted metabolomics approach to identify alcohol-specific metabolomic profiles of alcohol intake. Three studies, including one in EPIC, used targeted metabolomics, measuring 123 to 163 metabolites, to gain insight into metabolic pathways linking alcohol drinking to human health [44-46]; ten alcohol-metabolite associations were common to all three studies and included phosphatidylcholines (PCs), LysoPCs, acylcarnitines and sphingomyelins. Of note, PCs contribute to the formation of PEth in human tissues [47], which is a known biomarker of recent and heavy alcohol consumption used to diagnose alcohol abuse [48, 49]. A fourth targeted study used nuclear magnetic resonance to evaluate cross-sectional associations of 76 lipids, fatty acids, amino acids, ketone bodies and gluconeogenesis-related metabolites with alcohol consumption [50]. The endogenous metabolites identified by these targeted platforms did not overlap with the compounds most highly correlated with alcohol intake in our untargeted study, underscoring the breadth of the metabolome and discovery potential of untargeted metabolomics methods.

Metabolomics analyses that limit biomarker discovery to previously annotated compounds have also identified several alcohol-related biomarkers. For example, using prediagnostic serum samples from a nested breast cancer case-control study within a U.S. cohort, alcohol intake was associated with 16 of the 617 annotated metabolites, including 2-hydroxy-3-methylbutyric acid, 2,3-dihydroxyisovaleric acid (i.e., 2,3-hydroxy-3-methylbutyric acid), ethyl glucuronide and several endogenous metabolites related to androgen metabolism [51]. Other cross-sectional analyses, measuring hundreds of metabolites, also found associations of 2-hydroxy-3-methylbutyric acid, 2,3-dihydroxyisovaleric acid (i.e., 2,3-hydroxy-2-methylbutyric acid) and ethyl glucuronide with alcohol intake using prediagnostic serum [52, 53]. However, these studies lacked separate discovery and replication steps, and estimated correlations in cases and controls combined rather than in controls only. One study, which included discovery and replication sets, evaluated associations between alcohol intake and 356 known metabolites among African Americans [54] and found that alcohol was associated with five 2-hydroxybutyrate-related metabolites including 2-hydroxy-3-methylbutyric acid. A Japanese study of 107 metabolites in men identified positive associations between 2-hydroxybutyric

acid and alcohol intake in a discovery and a replication set [55]. The production of 2-hydroxy-3-methylbutyric acid and other hydroxybutyric acid-related metabolites is linked to the rate of hepatic glutathione synthesis, which can increase considerably in response to oxidative stress or detoxification of xenobiotics in the liver [56]. A targeted metabolomics investigation in EPIC found evidence suggesting that glutathione metabolism is involved in the development of HCC [20]. Additionally, 2-hydroxy-3-methylbutyric acid is a product of branched-chain amino acid metabolism, which has been linked to alcohol drinking [55, 57].

To our knowledge, this study is unique in its untargeted metabolomics approach without preselected metabolites and its use of a discovery and two independent replication sets. By considering nearly 7,000 features, many of which are correlated, we greatly increased the number of potential candidates, but we also incurred a stronger penalisation for multiple testing. Consequently, our approach may have missed features that did not meet stringent thresholds for statistical significance. A strength of our approach was the use of large independent discovery and replications sets; although matching features by mass and retention time across sets may have resulted in the loss of relevant information. Other potential limitations relate to generalizability and measurement error. Circulating metabolite levels reflect environmental exposures as well as host and microbial metabolism [58-60], and identification of candidate biomarkers that are sufficiently specific to ethanol and generalizable to diverse populations is challenging. Measurement error, both systematic and random, is inherent to self-reported assessments [61-63], including alcohol intake, and likely biases association estimates in not only aetiological studies, but also in biomarker discovery studies. Despite our use of cutting-edge untargeted metabolomics methods, a robust study design, and an aetiological component to evaluate the associations of our candidate biomarkers with disease outcomes, we cannot dismiss the possibility that our findings were impacted by measurement error in self-reported alcohol intake.

In summary, we observed robust correlations between self-reported habitual alcohol intake and 2-hydroxy-3-methylbutyric acid and an unidentified compound in a discovery set and two independent replication sets of cancer-free participants. Associations for 2-hydroxy-3-methylbutyric acid with risk of HCC and pancreatic cancer in the EPIC study and with liver cancer in ATBC were stronger than those for either self-reported alcohol intake or the unidentified compound. In conclusion,

2-hydroxy-3-methylbutyric acid is a promising candidate biomarker for studying the relationship

between habitual alcohol intake and health [51-54], but further research, preferably in the context a

randomized-controlled trial, is needed to better characterize the relationship between 2-hydroxy-3-

methylbutyric acid and alcohol at varying levels of intake.

References

1.   Collaborators GBDA. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018;392(10152):1015-1035.
2.   World Health Organization. *Alcohol fact sheet*. http://www.who.int/mediacentre/factsheets/fs349/en/.
3.   Humans IWGotEoCRt. Personal habits and indoor combustions. Volume 100 E. A review of human carcinogens. IARC Monogr Eval Carcinog Risks Hum 2012;100(Pt E):1-538.
4.   World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. In. *Alcoholic drinks and the risk of cancer.*
5.   Klatsky AL, Udaltsova N, Li Y, *et al.* Moderate alcohol intake and cancer: the role of underreporting. Cancer Causes Control 2014;25(6):693-9.
6.   Kroke A, Klipstein-Grobusch K, Hoffmann K, *et al.* Comparison of self-reported alcohol intake with the urinary excretion of 5-hydroxytryptophol:5-hydroxyindole-3-acetic acid, a biomarker of recent alcohol intake. Br J Nutr 2001;85(5):621-7.
7.   Das SK, Nayak P, Vasudevan DM. Biochemical markers for alcohol consumption. Indian J Clin Biochem 2003;18(2):111-8.
8.   Das SK, Vasudevan DM. Biochemical diagnosis of alcoholism. Indian J Clin Biochem 2005;20(1):35-42.
9.   Peterson K. Biomarkers for alcohol use and abuse - A summary. Alcohol Research & Health 2004;28(1):30-37.
10.   Torrente MP, Freeman WM, Vrana KE. Protein biomarkers of alcohol abuse. Expert Rev Proteomics 2012;9(4):425-36.
11.   Helander A, Bottcher M, Dahmen N, *et al.* Elimination Characteristics of the Alcohol Biomarker Phosphatidylethanol (PEth) in Blood during Alcohol Detoxification. Alcohol and Alcoholism 2019;54(3):251-257.
12.   Scalbert A, Brennan L, Manach C, *et al.* The food metabolome: A window over dietary exposure. Am J Clin Nutr 2014;99(6):1286-1308.
13.   Edmands WMB, Ferrari P, Rothwell JA, *et al.* Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. Am J Clin Nutr 2015;102(4):905-913.
14.   Riboli E, Hunt KJ, Slimani N, *et al.* European prospective investigation into cancer and nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5.
15.   Slimani N, Ferrari P, Ocke M, *et al.* Standardization of the 24-hour diet recall calibration method used in the european prospective investigation into cancer and nutrition (EPIC): general concepts and preliminary results. Eur J Clin Nutr 2000;54(12):900-17.
16.   Kaaks R, Slimani N, Riboli E. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S26-36.
17.   Slimani N, Bingham S, Runswick S, *et al.* Group level validation of protein intakes estimated by 24-hour diet recall and dietary questionnaires against 24-hour urinary nitrogen in the European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study. Cancer Epidemiol Biomarkers Prev 2003;12(8):784-95.
18.   Rothwell JA, Keski-Rahkonen P, Robinot N, *et al.* A Metabolomic Study of Biomarkers of Habitual Coffee Intake in Four European Countries. Mol Nutr Food Res 2019;63(22):e1900659.

19.	Stepien M, Keski-Rahkonen P, Kiss A, *et al.* Metabolic perturbations prior to hepatocellular carcinoma diagnosis - Findings from a prospective observational cohort study [published online ahead of print July 31, 2020]. Int J Cancer. 2020; doi.org/10.1002/ijc.33236.

20.	Stepien M, Duarte-Salles T, Fedirko V, *et al.* Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study. Int J Cancer 2016;138(2):348-60.

21.	Gasull M, Pumarega J, Kiviranta H, *et al.* Methodological issues in a prospective study on plasma concentrations of persistent organic pollutants and pancreatic cancer risk within the EPIC cohort. Environ Res 2019;169:417-433.

22.	The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. Ann Epidemiol 1994;4(1):1-10.

23.	Loftfield E, Rothwell JA, Sinha R, *et al.* Prospective investigation of serum metabolites, coffee drinking, liver cancer incidence, and liver disease mortality. J Natl Cancer Inst 2019; 10.1093/jnci/djz122.

24.	Wishart DS, Feunang YD, Marcu A, *et al.* HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 2018;46(D1):D608-D617.

25.	Smith CA, O'Maille G, Want EJ, *et al.* METLIN: a metabolite mass spectral database. Ther Drug Monit 2005;27(6):747-51.

26.	Kirpich AS, Ibarra M, Moskalenko O, *et al.* SECIMTools: a suite of metabolomics data analysis tools. BMC Bioinformatics 2018;19(1):151.

27.	Edmands WM, Barupal DK, Scalbert A. MetMSLine: an automated and fully integrated pipeline for rapid processing of high-resolution LC-MS metabolomic datasets. Bioinformatics 2015;31(5):788-90.

28.	Do KT, Wahl S, Raffler J, *et al.* Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. Metabolomics 2018;14(10):128.

29.	Fan S, Kind T, Cajka T, *et al.* Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. Anal Chem 2019;91(5):3590-3596.

30.	Lai GY, Weinstein SJ, Albanes D, *et al.* The association of coffee intake with liver cancer incidence and chronic liver disease mortality in male smokers. Br J Cancer 2013;109(5):1344-51.

31.	Loftfield E, Rothwell JA, Sinha R, *et al.* Prospective Investigation of Serum Metabolites, Coffee Drinking, Liver Cancer Incidence, and Liver Disease Mortality. J Natl Cancer Inst 2020;112(3):286-294.

32.	Fages A, Ferrari P, Monni S, *et al.* Investigating sources of variability in metabolomic data in the EPIC study: the Principal Component Partial R-square (PC-PR2) method. Metabolomics 2014;10(6):1074-1083.

33.	Perrier F, Novoloaca A, Ambatipudi S, *et al.* Identifying and correcting epigenetics measurements for systematic sources of variation. Clin Epigenetics 2018;10:38.

34.	Chambers J, Hastie T, Pregibon D. Statistical Models in S: Chapter 7. Generalized additive models. *Heidelberg*, *1990*, p. 317-321. Physica-Verlag HD.

35.	Wentzensen N, Wacholder S. From differences in means between cases and controls to risk stratification: a business plan for biomarker development. Cancer Discov 2013;3(2):148-57.

36.	Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag; 2001.

37.	R Core Team. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing; 2013.

38.	Mishima T, Harino S, Sugita J, *et al.* Plasma kinetics and urine profile of ethyl glucosides after oral administration in the rat. Biosci Biotechnol Biochem 2008;72(2):393-7.

39.	Trichopoulos D, Bamia C, Lagiou P, *et al.* Hepatocellular carcinoma risk factors and disease burden in a European cohort: a nested case-control study. J Natl Cancer Inst 2011;103(22):1686-95.

40.	Rohrmann S, Linseisen J, Vrieling A, *et al.* Ethanol intake and the risk of pancreatic cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). Cancer Causes Control 2009;20(5):785-94.

41.	Schwartz LM, Persson EC, Weinstein SJ, *et al.* Alcohol consumption, one-carbon metabolites, liver cancer and liver disease mortality. PLoS One 2013;8(10):e78156.

42.   Naudin S, Li K, Jaouen T, *et al.* Lifetime and baseline alcohol intakes and risk of pancreatic cancer in the European Prospective Investigation into Cancer and Nutrition study. Int J Cancer 2018;143(4):801-812.

43.   World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and pancreatic cancer. In; 2018.

44.   Jaremek M, Yu Z, Mangino M, *et al.* Alcohol-induced metabolomic differences in humans. Transl Psychiatry 2013;3:e276.

45.   van Roekel EH, Trijsburg L, Assi N, *et al.* Circulating Metabolites Associated with Alcohol Intake in the European Prospective Investigation into Cancer and Nutrition Cohort. Nutrients 2018;10(5).

46.   Lacruz ME, Kluttig A, Tiller D, *et al.* Cardiovascular Risk Factors Associated With Blood Metabolite Concentrations and Their Alterations During a 4-Year Period in a Population-Based Cohort. Circ Cardiovasc Genet 2016;9(6):487-494.

47.   Brühl A, Faldum A, Löffelholz K. Degradation of phosphatidylethanol counteracts the apparent phospholipase D-mediated formation in heart and other organs. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 2003;1633(2):84-89.

48.   Walther L, de Bejczy A, Lof E, *et al.* Phosphatidylethanol is superior to carbohydrate-deficient transferrin and gamma-glutamyltransferase as an alcohol marker and is a reliable estimate of alcohol consumption level. Alcohol Clin Exp Res 2015;39(11):2200-8.

49.   Zheng Y, Beck O, Helander A. Method development for routine liquid chromatography-mass spectrometry measurement of the alcohol biomarker phosphatidylethanol (PEth) in blood. Clin Chim Acta 2011;412(15-16):1428-35.

50.   Wurtz P, Cook S, Wang Q, *et al.* Metabolic profiling of alcohol consumption in 9778 young adults. Int J Epidemiol 2016;45(5):1493-1506.

51.   Playdon MC, Ziegler RG, Sampson JN, *et al.* Nutritional metabolomics and breast cancer risk in a prospective study. Am J Clin Nutr 2017;106(2):637-649.

52.   Guertin KA, Moore SC, Sampson JN, *et al.* Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr 2014;100(1):208-17.

53.   Playdon MC, Sampson JN, Cross AJ, *et al.* Comparing metabolite profiles of habitual diet in serum and urine. Am J Clin Nutr 2016;104(3):776-89.

54.   Zheng Y, Yu B, Alexander D, *et al.* Metabolomic patterns and alcohol consumption in African Americans in the Atherosclerosis Risk in Communities Study. Am J Clin Nutr 2014;99(6):1470-8.

55.   Harada S, Takebayashi T, Kurihara A, *et al.* Metabolomic profiling reveals novel biomarkers of alcohol intake and alcohol-induced liver injury in community-dwelling men. Environ Health Prev Med 2016;21(1):18-26.

56.   Lord RS, Bralley JA. Clinical applications of urinary organic acids. Part I: Detoxification markers. Altern Med Rev 2008;13(3):205-15.

57.   Pallister T, Jennings A, Mohney RP, *et al.* Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. PLoS One 2016;11(6):e0158568.

58.   Vipperla K, O'Keefe SJ. Intestinal microbes, diet, and colorectal cancer. Current Colorectal Cancer Reports 2013;9(1):95-105.

59.   Putignani L, Dallapiccola B. Foodomics as part of the host-microbiota-exposome interplay. Journal of Proteomics 2016;147:3-20.

60.   Shin SY, Fauman EB, Petersen AK, *et al.* An atlas of genetic influences on human blood metabolites. Nature Genetics 2014;46(6):543-550.

61.   Kipnis V, Subar AF, Midthune D, *et al.* Structure of dietary measurement error: Results of the OPEN biomarker study. Am J Epidemiol 2003;158(1):14-21.

62.   Prentice RL, Mossavar-Rahmani Y, Huang Y, *et al.* Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. Am J Epidemiol 2011;174(5):591-603.

63.   Willett W. *Nutritional epidemiology*. Oxford: Oxford University Press; 2013.

64.   Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Statistical Methodology 1995;57(1):289-300.

Figure legends

Figure 1. Flowchart of the study, displaying features and samples size of the discovery and Replication I (EPIC) and II (ATBC) sets (blue box), as well as of the aetiological components in nested-case-control studies (red box).

Figure 2. PC-PR2 (Principal Component Partial R^2) analysis to quantify the contribution of potential confounder variables to the variability of the set of f_1=67 feature intensities that were statistically significantly associated to alcohol intake in the discovery set.

Figure 3: ROC curves with AUC for high alcohol intake, in the range 33.1- 138.9 g/day in men (EPIC n=36; ATBC n=70) and 12.3-93.5 g/day in women (EPIC n=26), *vs.* low alcohol intake, in the range of >0.1-4.9 g/day in men (EPIC n=35; ATBC n=90) and >0.1-1.0 g/day in women (EPIC n=25), in **(A)** EPIC replication set and **(B)** ATBC replication set for the two identified biomarker candidates after cross-validation.

Table 1. Descriptive statistics of the discovery and replication sets.

	EPIC Discovery [1] (n=454)	EPIC Replication [2] (n=281)	ATBC Replication [3] (n=438)
Men (%)	42.5	52.7	100
BMI (median kg/m^2; 10-90th %)	25.8 (20.9-31.6)	26.6 (20.7-34.1)	26.2 (22.5-31.3)
Age (median years; 10-90th %)	55.2 (42.5-63.9)	59.4 (49.0-68.6)	56.0 (51.0-63.0)
Smoking status (%)			
Current	18.5	19.2	100
Former	26.4	33.5	
Never	52.2	46.2	
Unknown	2.9	1.1	
Smoking intensity (median cig/day; 10-90th %)	11.5 (2-26)	15 (4-30)	20 (10-30)
Country (%)			
France	14.5	0.4	
Italy	34.8	18.5	
Spain	-	10.0	
United Kingdom	-	17.1	
The Netherlands	-	10.3	
Greece	12.3	10.7	
Germany	38.3	24.9	
Denmark	-	8.2	
Finland	-	-	100
Alcohol non-drinkers (%) [4]	8	14	9
Alcohol intake (median g/day; 10th-90th %)			
Men	21.4 (1.3-50.4)	14.9 (1.0-51.7)	11.5 (0.2-42.1)
Women	5.2 (0.02-24.9)	2.0 (0.01-23.3)	--
Coffee intake (median g/day; 10th-90th %)	146.3 (21.4, 580.2)	190 (3, 857)	550 (220-1,100)

[1]EPIC cross-sectional sample;

[2]Controls from both liver and pancreatic cancer EPIC nested case-control studies;

[3]Controls from liver cancer and liver disease mortality ATBC nested case-control studies excluding those with missing data on alcohol intake;

[4]Alcohol non-drinkers are considered as those with alcohol intake ≤0.1g/day.

Table 2. Feature-specific intensity and reproducibility (coefficient of variation=CV) in quality control (QC) samples, and adjusted Pearson correlation coefficients (r) with between alcohol intake in the discovery and replication sets.

m/z [4]	RT [5] (min)	Method	Associated metabolite	QC samples [1] (n=38) Mean intensity	CV (%)	EPIC Discovery (n=454) [2] r	p-value	q-value [6]	EPIC Replication (n=281) [3] r	p-value [7]	ATBC Replication (n=438) r	p-value [8]
231.0839 [9]	**0.89**	RP+	Unknown	58378	18.5	0.41	1.2×10^{-19}	4.4×10^{-16}	0.38	7.0×10^{-11}	0.40	6.3×10^{-18}
253.0925	0.93	RP-	Unknown	11140	13.2	0.39	2.6×10^{-18}	4.6×10^{-15}	0.32	3.2×10^{-8}	-[10]	-
203.0227 [9]	**2.78**	RP+	2-hydroxy-3-methylbutyric acid	204079	14.8	0.26	1.9×10^{-8}	2.0×10^{-6}	0.24	5.3×10^{-5}	0.40	1.1×10^{-18}
217.9895	2.78	RP+	2-hydroxy-3-methylbutyric acid	36539	11.7	0.30	9.0×10^{-11}	2.1×10^{-8}	0.25	2.3×10^{-5}	0.38	2.4×10^{-16}
250.0134	2.78	RP+	2-hydroxy-3-methylbutyric acid	122838	12.5	0.28	9.0×10^{-10}	1.6×10^{-7}	0.27	8.2×10^{-6}	0.40	3.5×10^{-18}
221.0605	2.78	RP+	2-hydroxy-3-methylbutyric acid	56192	11.2	0.28	2.6×10^{-9}	3.2×10^{-7}	0.25	2.1×10^{-5}	0.39	1.9×10^{-17}
218.9958	2.78	RP+	2-hydroxy-3-methylbutyric acid	115590	11.7	0.28	1.3×10^{-9}	2.1×10^{-7}	0.26	1.8×10^{-5}	0.40	1.7×10^{-18}
235.0479	2.78	RP+	2-hydroxy-3-methylbutyric acid	34447	15.5	0.20	2.3×10^{-5}	1.0×10^{-3}	0.26	2.1×10^{-5}	0.38	4.2×10^{-16}
117.0559	2.78	RP-	2-hydroxy-3-methylbutyric acid	211842	12.1	0.28	1.3×10^{-9}	2.2×10^{-7}	0.28	2.0×10^{-6}	-[10]	-
261.9788	2.78	RP-	2-hydroxy-3-methylbutyric acid	15985	11.9	0.27	7.2×10^{-9}	8.3×10^{-7}	0.28	2.7×10^{-6}	-[10]	-

[1] Quality control samples within the discovery set;

[2] The analyses of features acquired in positive and negative modes used data from 451 and 452 participants, respectively, after the exclusion of outliers and samples with too many missing values;

[3] The analyses of features acquired in positive and negative modes used data from 271 and 277 participants, respectively, after the exclusion of outliers and samples with too many missing values;

[4] m/z= monoisotopic mass divided by the charge state values, as observed in the discovery set;

[5] Retention time;

[6] Q-values associated to False Discovery Rate (FDR) procedure to correct for multiple testing [64], alpha=0.05;

[7] Threshold for statistical significance corrected with Bonferroni method for multiple testing, equal to 0.0007463 (0.05/f_1, with f_1=67).

[8] Threshold for statistical significance corrected with Bonferroni method for multiple testing, equal to 0.007 (0.05/f_3, with f_3=7);

[9] Feature chosen for analysis of disease see Table 3;

[10] Feature not available in ATBC.

Table 3. Crude and adjusted [1] odds ratios (OR, 95% CI) of self-reported alcohol intake (12 g/day) and the main features of the unknown compound and 2-hydroxy-3-methylbutyric acid (per 1-SD) with hepatocellular carcinoma (HCC; 129 case-control sets) and pancreatic cancer (152 case-control sets) in EPIC, and with liver cancer (194 case-control sets) and liver disease mortality (201 case-control sets) in ATBC

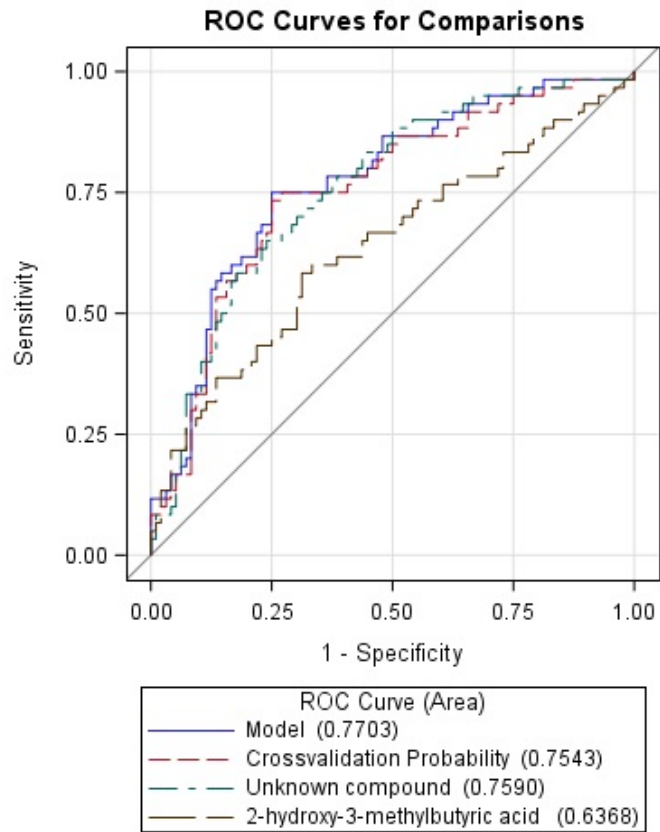|	**Crude**	**models**		**Adjusted**	**models**[1]	
	OR	(95% CI)	p-value	OR	(95% CI)	p-value
HCC, EPIC (128 case-control sets)						
Alcohol intake (12g/day)	1.13	(1.00, 1.27)	0.05	1.04	(0.89, 1.20)	0.65
Alcohol intake (log$_2$-transformed, 1-SD)	0.93	(0.73, 1.20)	0.59	0.78	(0.56, 1.09)	0.14
Unknown compound (log$_2$-transformed, 1-SD) [2]	1.17	(0.90, 1.52)	0.25	1.01	(0.73, 1.40)	0.94
2-hydroxy-3-methylbutyric acid (log$_2$-transformed, 1-SD) [3]	1.75	(1.29, 2.39)	3.8×10^{-4}	2.14	(1.39, 3.31)	5.5×10^{-4}
Pancreatic cancer, EPIC (139 case-control sets)						
Alcohol intake (12g/day)	1.07	(0.92, 1.25)	0.36	1.04	(0.88, 1.24)	0.65
Alcohol intake (log$_2$-transformed, 1-SD)	1.08	(0.83, 1.40)	0.58	1.03	(0.77, 1.39)	0.83
Unknown compound (log$_2$-transformed, 1-SD) [2]	1.20	(0.95, 1.50)	0.13	1.17	(0.91, 1.51)	0.22
2-hydroxy-3-methylbutyric acid (log$_2$-transformed, 1-SD) [3]	1.52	(1.13, 2.04)	5.2×10^{-3}	1.65	(1.17, 2.32)	3.9×10^{-3}
Liver cancer, ATBC (192 case-control sets)						
Alcohol intake (12g/day)	1.25	(1.09, 1.43)	1.2×10^{-3}	1.17	(1.01, 1.36)	0.03
Alcohol intake (log$_2$-transformed, 1-SD)	1.33	(1.05, 1.67)	0.016	1.23	(0.94, 1.60)	0.13
Unknown compound (log$_2$-transformed, 1-SD) [2]	1.34	(1.07, 1.68)	0.01	1.70	(1.29, 2.25)	2.0×10^{-4}
2-hydroxy-3-methylbutyric acid (log$_2$-transformed, 1-SD) [3]	2.08	(1.53, 2.82)	2.6×10^{-6}	2.00	(1.44, 2.77)	3.4×10^{-5}
Liver disease mortality, ATBC (199 case-control sets)						
Alcohol intake (12g/day)	1.38	(1.22, 1.54)	1.1×10^{-7}	1.32	(1.16, 1.50)	1.6×10^{-5}
Alcohol intake (log$_2$-transformed, 1-SD)	2.37	(1.78, 3.14)	2.8×10^{-8}	2.19	(1.60, 2.98)	8.4×10^{-7}
Unknown compound (log$_2$-transformed, 1-SD) [2]	1.95	(1.49, 2.54)	9.1×10^{-7}	1.85	(1.39, 2.46)	2.6×10^{-5}
2-hydroxy-3-methylbutyric acid (log$_2$-transformed, 1-SD) [3]	2.26	(1.73, 2.95)	2.1×10^{-9}	2.16	(1.63, 2.86)	9.6×10^{-8}

[1] Models for hepatocellular carcinoma (HCC) were adjusted for body mass index(BMI, kg/m^2), waist circumference (cm), recreational and household physical activity (Met-hours/week), smoking status, level of educational attainment, and coffee intake (grams/day, log$_2$-transformed); models for pancreatic cancer were adjusted for BMI (kg/m^2), sex-specific physical activity categories and smoking; ATBC liver cancer and fatal liver disease models were adjusted for

BMI (kg/m^2), leisure time physical activity, smoking intensity (cigarettes/day), level of educational attainment, and coffee intake (grams/day, $\log_2$-transformed);
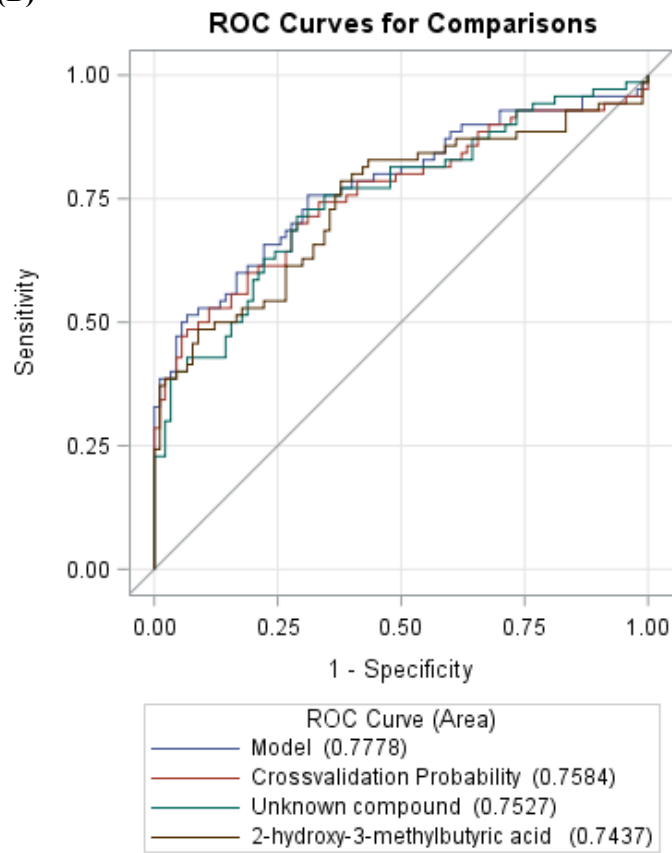
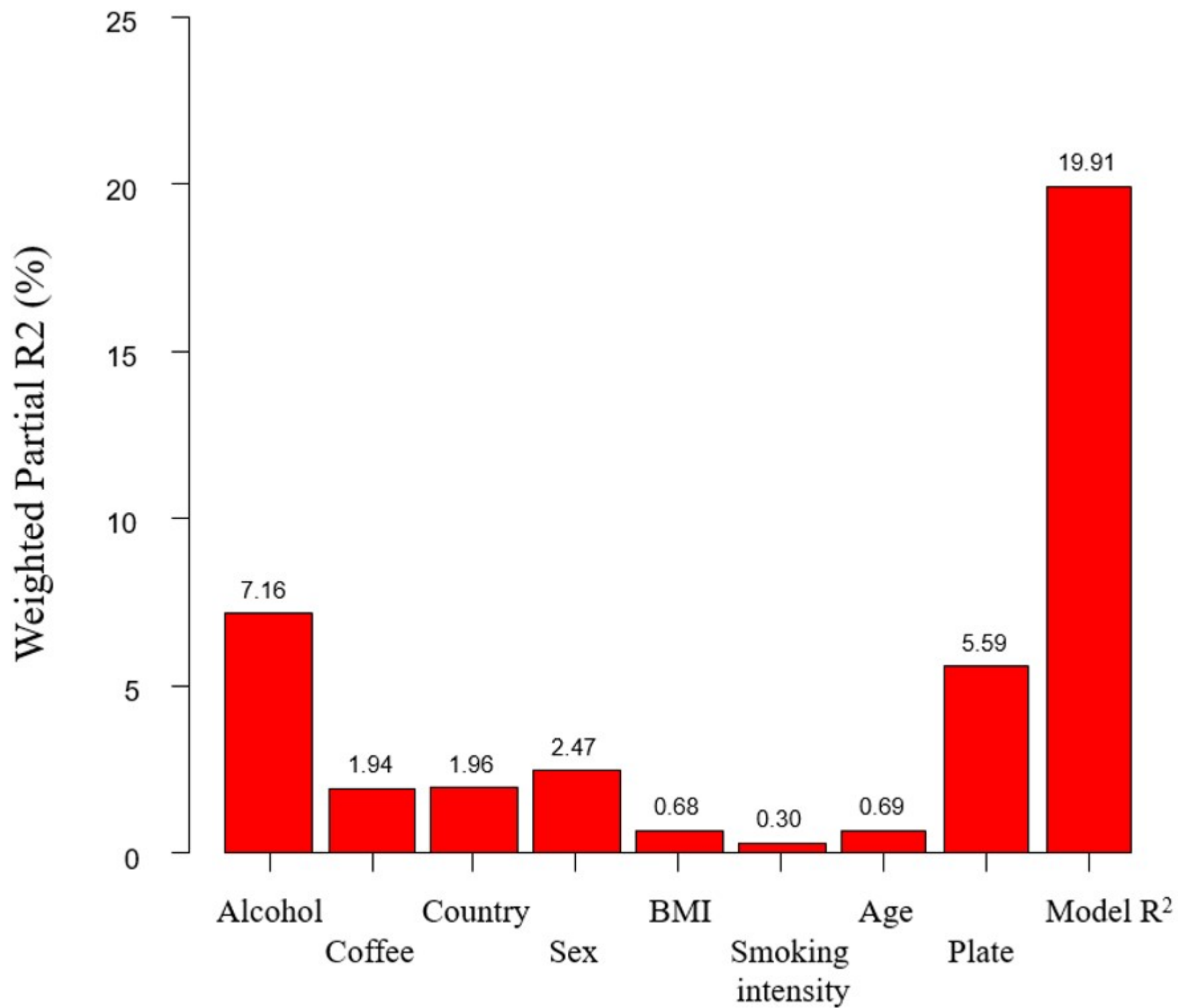[2] Unknown compound (*m/z*=231.0839);

[3] 2-hydroxy-3-methylbutyric acid (*m/z*=203.0227).

(A)

ROC Curves for Comparisons

ROC Curve (Area)
— Model (0.7703)
– – – Crossvalidation Probability (0.7543)
– · – Unknown compound (0.7590)
— — 2-hydroxy-3-methylbutyric acid (0.6368)

(B)

ROC Curves for Comparisons

ROC Curve (Area)
— Model (0.7778)
— Crossvalidation Probability (0.7584)
— Unknown compound (0.7527)
— 2-hydroxy-3-methylbutyric acid (0.7437)