

Are metabolic signatures mediating the relationship between lifestyle factors and hepatocellular carcinoma risk? Results from a nested case-control study in EPIC

Nada Assi¹, Duncan C Thomas², Michael Leitzmann³, Magdalena Stepien¹, Véronique Chajès¹, Thierry Philip⁴, Paolo Vineis⁵, Christina Bamia^{6,7}, Marie-Christine Boutron-Ruault^{8,9}, Torkjel M Sandanger¹⁰, Amaia Molinuevo^{11,12}, Hendriek C Boshuizen¹³, Anneli Sundkvist¹⁴, Tilman Kühn¹⁵, Ruth C Travis¹⁶, Kim Overvad¹⁷, Elio Riboli⁵, Marc J Gunter¹, Augustin Scalbert¹, Mazda Jenab¹, Pietro Ferrari^{1*}, Vivian Viallon¹⁸

¹ Section of Nutrition and Metabolism, International Agency for Research on Cancer (IARC), 150 cours Albert Thomas, 69008 Lyon, France.

² University of Southern California, Los Angeles, CA 90007, USA.

³ Department of Epidemiology and Preventive Medicine, Regensburg University, Franz-Josef-Strauß-Allee 11, D-93053 Regensburg, Germany.

⁴ Unité Cancer et Environnement, Centre Léon Bérard, 28 rue Laennec, 69373, Lyon 08 Cedex, France.

⁵ Department of Epidemiology and Biostatistics, MRC-HPA Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place W2 1PG London, UK.

⁶ Hellenic Health Foundation, Kaisareias 13 & Alexandroupoleos, GR-115 27, Athens, Greece.

⁷ WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Dept. of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, 75 Mikras Asias street. 115 27 Athens, Greece.

⁸ Université Paris-Saclay, Université Paris-Sud, UVSQ, CESP, INSERM, 16 Avenue Paul Vaillant Couturier, 94800 Villejuif, France.

⁹ Gustave Roussy, F-94805, Villejuif, France.

¹⁰ Department of Community Medicine, UiT the Arctic University of Norway, 9019 Tromsø, Norway.

¹¹ Public Health Division of Gipuzkoa, Regional Government of the Basque Country, Avenida de Navarra, 4, 20013 Donostia-San Sebastián, Spain.

¹² CIBER of Epidemiology and Public Health (CIBERESP), Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0, 28029 Madrid, Spain.

¹³ National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, The Netherlands.

¹⁴ Department of Radiation Sciences Oncology, Umeå University 901 87 Umeå, Sweden.

¹⁵ Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

¹⁶ Cancer Epidemiology Unit, University of Oxford, Oxford OX3 7LF, UK.

¹⁷ The Department of Epidemiology, School of Public Health, Aarhus University, Bartholins Allé 2 - Building 1260, DK-8000 Aarhus, Denmark.

¹⁸ Université de Lyon, Université Claude Bernard Lyon1, Ifsttar, UMRESTTE, UMR T_9405, F- 69373, LYON

Full title: Are metabolic signatures mediating the relationship between lifestyle factors and hepatocellular carcinoma risk? Results from a nested case-control study in EPIC

Running title: Lifestyle-specific metabolic profiles and hepatocellular carcinoma

Keywords: Hepatocellular carcinoma, targeted metabolomics, healthy lifestyle index, mediation analysis, partial least squares.

***Corresponding author:** Pietro Ferrari, International Agency for Research on Cancer, 150 Cours Albert Thomas, 69372 Lyon cedex 08, France. Tel: +33 472 73 8031; Fax: +33 472 73 8361. E-mail: ferrarip@iarc.fr

Word count (before revision): Manuscript 3972 words, Abstract: 249 words

Word count (after revision): Manuscript 4355 words, Abstract: 250 words

Word count (after 2nd revision): Manuscript 4388 words, Abstract: 250 words

Word count (after editorial revision): Manuscript 4410 words, Abstract: 250 words

Total number of tables: 4

Total number of Supplementary tables: 2

Total number of figures: 0

Total number of supplementary figures: 1

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

Abbreviations list:

BMI: Body Mass Index; CVs: Coefficients of Variation; DQs: Dietary Questionnaires; EPIC: European Prospective Investigation into Cancer and nutrition; HBV: Hepatitis B Virus; HCC: Hepatocellular Carcinoma; HCV: Hepatitis C Virus; HLI: Healthy Lifestyle Index; ME: Mediator Effect; MITM: Meeting-in-the-middle; NDE: Natural Direct Effect; NIE: Natural

Indirect Effect; PCA: Principal Component Analysis; PC-PR2: Principal Component Partial R2; PLS: Partial Least Squares; TE: Total Effect.

Grant support:

N. Assi was financially supported by the Université Claude Bernard Lyon I through a doctoral fellowship awarded by the EDISS (Ecole Doctorale InterDisciplinaire Sciences Santé) doctoral school to complete her PhD work. The data on the EPIC-Hepatobiliary dataset was generated through support from the French National Cancer Institute (L'Institut National du Cancer; INCA) (grant number 2009-139; PI: M. Jenab). The work undertaken by D. C. Thomas reported in this publication was supported by the National Institutes of Health under award number P01 CA196559. R. C. Travis is a co-principal investigator of the EPIC-Oxford cohort whose work is supported by Cancer Research UK under grant number C8221/A19170.

Abstract

Background: The “meeting-in-the-middle” (MITM) is a principle to identify exposure biomarkers that are also predictors of disease. The MITM statistical framework was applied in a nested case-control study of hepatocellular carcinoma (HCC) within EPIC where healthy lifestyle index (HLI) variables were related to targeted serum metabolites.

Methods: Lifestyle and targeted metabolomic data were available from 147 incident HCC cases and 147 matched controls. Partial Least Squares analysis related 7 lifestyle variables from a modified HLI to a set of 132 serum-measured metabolites and a liver function score. Mediation analysis evaluated whether metabolic profiles mediated the relationship between each lifestyle exposure and HCC risk.

Results: Exposure-related metabolic signatures were identified. Particularly, the BMI-associated metabolic component was positively related to glutamic acid, tyrosine, PC aaC38:3 and liver function score and negatively to lysoPC aC17:0 and aC18:2. The lifetime alcohol-specific signature had negative loadings on sphingomyelins (SM C16:1, C18:1, SM(OH) C14:1, C16:1 and C22:2). Both exposures were associated with increased HCC with Total Effects (TE)=1.23(95%CI = 0.93,1.62) and 1.40(1.14,1.72), respectively for BMI and alcohol consumption. Both metabolic signatures mediated the association between BMI and lifetime alcohol consumption and HCC with Natural Indirect Effects respectively equal to 1.56(1.24,1.96) and 1.09(1.03,1.15), accounting for a proportion mediated of 100% and 24%.

Conclusions: In a refined MITM framework, relevant metabolic signatures were identified as mediators in the relationship between lifestyle exposures and HCC risk.

Impact: The understanding of the biological basis for the relationship between modifiable exposures and cancer would pave avenues for clinical and public health interventions on metabolic mediators.

Introduction

Hepatocellular carcinoma (HCC) is the most common form of liver cancer(1) and is the second most frequent cause of cancer death worldwide(2). HCC mortality trends have increased in western populations and unfavourable trends have been projected to 2020(3). Similarly, HCC incidence rates have risen dramatically in Europe in the recent decades(4). HCC is a multi-factorial disease strongly associated with lifestyle factors, most of which are modifiable(5). While hepatitis infection, including both HBV and HCV, remains its primary risk factor, other exposures such as obesity, alcohol drinking, diabetes, smoking, physical activity and some dietary items have also been related to HCC risk(6–8). Understanding the underlying etiology of HCC is key to help develop disease prevention strategies, including through the identification of metabolic signatures.

Metabolic profiling is gaining traction in epidemiological investigations of cardiovascular disease and cancer(9). Metabolomics measures a wide array of endogenous and exogenous features providing a comprehensive snapshot of metabolic status at a given time point(10). In this sense, the metabolome is a tool encompassing thousands of detectable molecules involved in various physiological processes that can potentially elucidate underlying molecular and mechanistic pathways that may lead to pathological outcomes(11–14).

In this scenario, the “meeting-in-the-middle” (MITM) principle(15,16), can be used as a research strategy to identify biomarkers that are mechanistic mediators of the relationship between specific exposures and disease outcome. The statistical framework of the MITM was previously extended to multivariate modelling in a Partial Least Squares (PLS) analysis to integrate a set of 21 lifestyle variables and 285 metabolic variables from ^1H NMR spectra in relation to HCC risk(17). In this current work, a more focused methodology was further

developed building on a similar analytical structure. To this end, modifiable lifestyle exposures from a modified healthy lifestyle index (HLI)(18,19), including dietary, anthropometric (BMI), smoking, alcohol consumption and physical activity, and in addition diabetes and hepatitis infection, were investigated with respect to specific metabolic signatures.

In this study, our objective was to apply a MITM approach to extract metabolic profiles of lifestyle exposures and explore their mediating role in the HCC causal pathway using data from a case-control study nested within the European Prospective Investigation into Cancer and nutrition (EPIC) using targeted metabolomic data.

Material and Methods

The Study population

EPIC is a multicentre prospective study designed to investigate the link between diet, lifestyle and environmental factors with cancer incidence and other chronic disease outcomes. Over 520,000 healthy men and women aged 25-85 were enrolled between 1992 and 2000 across 23 EPIC administrative centres in 10 European countries including Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden, and the United Kingdom(20). Extensive details of the study design, rationale and recruitment methods, dietary assessment, blood collection protocols and follow-up procedures have been previously published(20,21). During the enrolment period (1992-1998), participants gave informed consent and completed questionnaires on diet, lifestyle, anthropometric measures and medical history. Validated country-specific dietary questionnaires (DQs) ensuring high compliance were used to assess habitual diet over the previous 12 months (20). The ethical review boards of the participating institutions and the International Agency for Research on Cancer (IARC) approved this study. Biological samples were collected at recruitment prior to disease onset, and were available for approximately 80% of the cohort. Serum samples were stored at IARC, Lyon, France in -196°C liquid nitrogen for all countries, with the exception of samples originating from Sweden (-80°C freezers) and Denmark (-150°C nitrogen vapour). For information on how to submit an application for gaining access to EPIC data and/or biospecimens, please follow the instructions at <http://epic.iarc.fr/access/index.php>

Nested case-control study design

This study focused on a nested case-control study of HCC in EPIC (22,23) with available biological samples for the case sets identified in the period between subjects' recruitment into the cohort (1993-1998) and 2010(23,24). Cases of HCC (n=147) originated

from all participating EPIC centres except for Norway and France which were not included in this study. For each HCC case, one control (n=147) was selected by incidence density sampling(25) from all cohort members alive and free of cancer (except for non-melanoma skin cancer), and matched by age at blood collection (± 1 year), sex, study centre, time of the day at blood collection (± 3 hours), fasting status at blood collection (<3, 3-6, and >6 hours); among women, the pair was additionally matched by menopausal status (pre-, peri-, and postmenopausal), and hormone replacement therapy use at time of blood collection (yes/no). All subjects were cancer-free at the time of blood collection. Information on follow-up and case ascertainment can be read elsewhere(23). In brief, HCC cases were defined as first primary invasive tumours and identified through the 10th Revision of International Statistical Classification of Diseases, Injury and Causes of Death (ICD10) as C22.0 with morphology codes ICD-O-2 “8170/3” and “8180/3”. Metastatic cases and other types of primary liver cancer were excluded.

The lifestyle variables

The examined predictor variables were restricted to lifestyle factors that have been previously associated with HCC (5–7,26–33) and that were used in the pre-existing HLI(18,19). These included BMI (continuous, kg/m²), average lifetime alcohol intake (continuous, g/day), the diet score (continuous), physical activity (continuous, metabolic equivalent of task MET-h/week), smoking (never, ex-smokers quit>10 years, ex-smokers quit ≤ 10 y, current smokers ≤ 15 cig/day, current smokers > 15 cig/day). Self-reported diabetes at baseline (yes/no) and hepatitis infection (yes/no) were also included as both were already established liver cancer risk factors (26,27,34). Hepatitis infection was assessed from biomarker measures of hepatitis B and hepatitis C viruses’ (HBV, HCV) seropositivity [ARCHITECT HBsAg and anti-HCV chemiluminescent microparticle immunoassays; Abbott Diagnostics, France])(26). The previously mentioned diet score is an *a priori* score

proposed within EPIC based on dietary components that have been posited to affect risk of cancer(18,19) and was detailed elsewhere(18). Briefly, the diet score combined six dietary items including cereal fiber, red and processed meats, ratio of polyunsaturated to saturated fatty acids, margarine (used as a surrogate marker for trans-fat from industrial sources), glycaemic load, and fruits and vegetables. Lifetime alcohol intake was used instead of alcohol at recruitment to dispel with reverse causality.

These seven lifestyle variables will be herein referred to as the X-set. Missing values in the X-set were imputed through a simple EM algorithm using the covariance matrix of the data(35).

Metabolomic data

We measured concentrations of a set of known metabolites using targeted metabolite profiling in serum samples. The samples were analyzed by ultra-performance liquid chromatography (LC; 1290 Series HPLC; Agilent, Les Ulis, France) coupled to a tandem mass spectrometer (MS/MS; QTrap 5500; AB Sciex, Les Ulis, France) and performed at IARC, Lyon, France, using the BIOCRATES AbsoluteIDQ p180 Kit (Biocrates, Innsbruck, Austria). This kit is used to measure over 150 common endogenous metabolites (amino acids, biogenic amines, hexoses, acylcarnitines, lipids) including various metabolites from the intermediate metabolism, all of which are important indicators of individual metabolic status(36,37). Details of the sample preparation and mass spectrometry analyses are provided elsewhere(23,38). ICCs were above 0.50 in 73% and 52% of the metabolites in fasting and non-fasting samples, respectively (38). Metabolites with coefficient of variation (CVs)>20% for analytical replicates were excluded by the IARC laboratory leading to 145 metabolites being detected. Of these, metabolites with >40% of missing values (*i.e.* below the limit of detection/quantification or above the highest calibration standards) were excluded, resulting in a total of 132 metabolites included in this study. Measurements that were below the limit

of detection were set to half that value and those below limit of quantification were set to half that limit. Additionally, measurements that were above the highest concentration calibration standards were set to the highest values. Metabolite nomenclature has been previously described(39). Briefly, fatty acid side chains are labelled “Cx:y”, where x and y are the numbers of carbon atoms and double bonds, respectively. There were 12 acylcarnitines (abbreviated according to the fatty acid side chain), 21 amino acids and 6 biogenic amines (labelled with their full name), 78 phosphatidylcholines (PC) of which there were 11 “LysoPC a” (PCs having one fatty acid side chain with an acyl bound), 34 “PC aa” and 33 “PC ae” (PCs having respectively two acyl side chains [diacyl] and one acyl and one alkyl side chains), a total of 14 sphingomyelins “SM” of which 5 had a hydroxyl group “SM(OH)” (additionally labelled according to the fatty acid side chain) and finally 1 sum of hexoses (including glucose, fructose and galactose). PCs were separated by type of bond and number of fatty acid side chains.

Liver function score

The liver function score is a composite score summarizing the number of abnormal values for six circulating liver blood biomarkers. It reflects possible underlying liver function impairment (17,22,23) and includes the following tests: alanine aminotransferase >55 U/L, aspartate aminotransferase >34 U/L, gamma-glutamyltransferase: men>64 U/L and women>36 U/L, alkaline phosphatase >150 U/L, albumin<35 g/L, total bilirubin > 20.5 µmol/L. These biomarkers were acquired at the same time as the metabolites from the pre-diagnostic blood samples collected at recruitment and the cut-points were provided by the clinical biochemistry laboratory that conducted the analyses (Centre de Biologie République, Lyon, France) based on assay specifications as previously described(23,24).

The set comprising the 132 metabolites and the liver function score will be herein referred to as the M-set of metabolomics data.

Statistical analyses

The flowchart in **Supplementary Figure 1A** recaps the different steps of the statistical analyses. Prior to PLS analysis, the sources of systematic variability within the exposure and the metabolomic data were identified and quantified through PC-PR2 analysis (40). Consequently residuals on country for the lifestyle exposures and on country and batch for the M-set were carried in a series of univariate linear regression models and used in the following analytical steps.

Partial Least Squares (PLS) analyses

Each exposure variable, in turn, was related to metabolomic data through PLS analysis. PLS is a multivariate method that generalizes features of PCA with those of multiple linear regression(41,42). Mathematical and computational details of the PLS method and its applicability within the MITM framework have been thoroughly described previously(17). In brief, PLS extracted linear combinations, referred to as PLS factors, of predictors (in this case the X-variable) and responses (the M-set), allowing a simultaneous decomposition of both sets with the aim of maximizing their covariance(41,42). A series of individual PLS analyses was applied using each HLI variable separately as the predictor to yield exposure-specific metabolic signatures and one PLS factor was retained for each analysis. PLS scores mirroring the metabolic components of each PLS factor, the M-scores, were computed. The PLS factor loadings quantified how much each metabolite contributed to the PLS metabolic signature and variables with loading values lower than the 5th and larger than the 95th percentiles were reported to ease interpretation. In an attempt to yield even more specific metabolic signatures, sensitivity PLS analyses were computed using mutually adjusted lifestyle residuals, where

each lifestyle exposure was regressed on the others, as well as country for each of the exposure variables. Country and batch residuals were used in the M-set.

Mediation analyses

Mediation analysis assessed whether the metabolic profiles mediated the relation between individual lifestyle factors and HCC risk (**Supplementary Figure 1B**). For each individual PLS analysis, mediating effects were computed for each extracted pair of lifestyle variable and M-score, adapting the formulae from VanderWeele and Vansteelandt(43) to accommodate continuous exposures and mediators and conditional logistic regression for our matched setting. For each examined lifestyle variable, estimates of the natural direct effect (NDE), the natural indirect effect (NIE), and the total effect (TE) were obtained, along with the effect of the corresponding M-score adjusted for its counterpart lifestyle exposure and for confounding variables and referred to as the mediator effect (ME).

The NDE and NIE were produced through two main models: a linear mediator model and a conditional logistic outcome model. HCC being a rare outcome, direct and indirect effects were estimated taking into account the nested case-control design. This is done by running the mediator regression only for the controls(44). After testing, there was no exposure-mediator interaction, thus the models were written as follows:

Let x be the exposure, m the mediator, c a set of different confounders, y the HCC outcome and j the pair number ranging among the set $\{1, \dots, n=147\}$:

$$E[M|x, c] = \beta_0 + \beta_1 x + \beta_2' c$$

$$\text{logit}[P(Y = 1|x, m, c, j)] = \theta_{0,j} + \theta_1 x + \theta_2 m + \theta_3' c$$

Thus, NDE and NIE are given as follows for a one standard deviation increase in x and m :

$$NDE_{x|c} \approx \exp(\theta_1 \text{sd}(x))$$

$$NIE_{x|c} = \exp(\theta_2 \beta_1 \text{sd}(x))$$

95% CI for NDE and NIE were computed through the following formulae:

$$95\%CI(NDE_{x|c}) = \exp\left(\log(\widehat{NDE}_{x|c}) \pm 1.96 * \widehat{sd}(x) * \sqrt{\widehat{\sigma}_{11}^{\theta}}\right)$$

$$95\%CI(NIE_{x|c}) = \exp\left(\log(\widehat{NIE}_{x|c}) \pm 1.96 * \widehat{sd}(x) * \sqrt{\widehat{\theta}_2^2 \widehat{\sigma}_{11}^{\beta} + \widehat{\beta}_1^2 \widehat{\sigma}_{22}^{\theta}}\right),$$

where $\widehat{\sigma}_{11}^{\theta}, \widehat{\sigma}_{22}^{\theta}$ and $\widehat{\sigma}_{11}^{\beta}$ are the estimated variances of the coefficients $\widehat{\theta}_1, \widehat{\theta}_2$ and $\widehat{\beta}_1$, respectively.

The total effect of X (TE) was computed from the following conditional logistic regression:

$$\text{logit}[P(Y = 1|x, c, j)] = \gamma_{0,j} + \gamma_1 x + \gamma_2' c$$

with TE given by:

$$TE_{x|c} = \exp(\gamma_1 \widehat{sd}(x))$$

Usually TE can be written as the product of NDE and NIE. However, in our setting employing conditional logistic regression, this is no longer the case because discordant pairs in the model adjusted for the mediator are not the same as the model not including the mediator (TE).

The mediator effect (ME), corresponding to the “independent” effect of the M-score adjusted for its counterpart lifestyle exposure and for confounding variables was given by:

$$ME_{m|x,c} = \exp(\theta_2 \widehat{sd}(m))$$

To control for potential confounding, mediation analyses models were adjusted for the modified HLI variables except the one under scrutiny. *P*-values for NDE and NIE were inferred from the 95%CI, whereas for the ME and TE, *p*-values associated with Wald's test for continuous exposure compared with a chi-square distribution with 1 degree of freedom were produced. The false discovery rate (FDR) correction(45) was applied to mediation results stemming from the multiple PLS analyses to control for multiple testing and associated *q*-values were presented.

For each mediation analysis the estimates for the NDE, NIE, TE and ME were reported for an increase in the exposure as follows: an increase of 1-SD for smoking, an increase of 1-SD among the controls for BMI, physical activity and the diet score, an increase of 1 unit (0 to 1) for diabetes and hepatitis infection, and finally an increase of 12 g/day (corresponding to one alcohol unit) for lifetime alcohol.

Since $TE = NDE * NIE$ does not hold in our setting, the mediated proportion was computed using the following formula:

$$\% mediated = \min \left(\max \left(0, \left(\frac{\log(NIE)}{\log(NDE) + \log(NIE)} \right) * 100 \right), 100 \right)$$

Indeed, the proportion mediated makes real sense only when NDE and NIE have the same direction of association and is bounded between 0% and 100%. In this case our formula reduces to:

$$\% mediated = \frac{\log(NIE)}{\log(NDE) + \log(NIE)}.$$

When NDE and NIE have opposite directions, the mediated proportion is not well-defined. For example, if $NDE = 0.5$ and $NIE = 2$ so that $TE = 1$, it is not clear what the mediated proportion would be. In our results, NDE and NIE always had the same direction when they were both statistically significant. For example, in our analyses for diabetes (or equivalently for BMI), the NIE is significantly associated with an increased risk of HCC and the NDE was not significant and had the opposite direction of association. This suggested that $TE = NIE$ and using our first formula above we get the appropriate value of 100%.

All statistical tests were two-sided and after multiple testing correction q -values below 0.05 were considered statistically significant. Statistical analyses were performed using

PROC PLS in SAS(46) for PLS analyses and the R Software(47) for linear and conditional logistic regressions and mediation analyses.

Results

Study population characteristics by case-control status are presented in **Table 1**. Individual PLS analyses yielded metabolic signatures for each component of the modified HLI and the top metabolites contributing to each signature are presented in **Table 2**. For lifetime alcohol, the signature was negatively related to SM C16:1, SM C18:1, SM(OH) C14:1, SM(OH) C16:1 and SM(OH) C22:2 and positively related to glutamic acid and PC aaC32:1. Metabolites associated with smoking included SM C16:1 and C18:1, SM(OH) C14:1 and C22:2, LysoPC aC28:1 and PC aeC30:2 with negative loadings and hexoses with positive loadings. In the sensitivity analysis, smoking was negatively associated with serine, lysine and biogenic taurine and positively with PC aaC36:1 and aaC40:3 (**Supplementary Table 1**). Different diacyl and acyl-alkyl-phosphatidylcholines characterized the metabolic signature related to diet. The metabolic profile of BMI included glutamic acid, tyrosine, PC aaC38:3, the liver function score with positive loadings and glutamine, LysoPC aC17:0 and LysoPC aC18:2 with negative values. Hexoses and amino acids valine, isoleucine and phenylalanine were positively associated with diabetes status.

The association between each of the different lifestyle factors with HCC risk was strongly mediated by the PLS metabolic profiles, with the exception of physical activity and hepatitis infection (**Table 3**). In particular, for both diabetes and BMI, a positive association for the NIE, equal to 5.11 (95%CI=1.99 to 13.14, q -value=2.45E-03) and 1.56 (1.24 to 1.96, q -value=1.20E-03), respectively, was observed, together with a lack of association for the NDE, thus suggesting that the relationship between these two variables and HCC risk was fully mediated by the corresponding metabolic signatures. As for smoking, diet and lifetime alcohol, the mediated proportions were 56%, 38% and 24%, respectively, with NIE equal to 1.22 (1.04 to 1.44, q -value=2.52E-02), 0.85 (0.74 to 0.97, q -value=2.52E-02) and 1.09 (1.03 to 1.15, q -value=4.67E-03), respectively (**Table 3**). Noteworthy, the NIE estimate for

smoking in the sensitivity analysis was 1.98 (1.34 to 2.92, q -value=1.32E-04), and the relation between smoking and HCC was fully mediated by the M-score (**Supplementary Table 2**).

The metabolic profile associated with the diet score showed a strong inverse association with HCC risk, with ME (Mediator Effect) equal to 0.61 (0.41 to 0.89, q -value=1.54E-02). A positive association with HCC was observed for the lifestyle-specific metabolic signatures related to BMI, lifetime alcohol, smoking and diabetes with ME=4.04 (2.22 to 7.36, q -value=3.17E-05), 2.50 (1.57 to 3.97, q -value=2.48E-04), 3.33 (1.96 to 5.66, q -value=3.17E-05) and 2.75 (1.59 to 4.78, q -value=5.55E-04), respectively.

The TE estimates showed strong associations for lifetime alcohol (1.40, 95%CI=1.14 to 1.72, q -value=3.50E-03), diet score (0.66, 0.47 to 0.92, q -value=3.27E-02) and hepatitis infection (16.70, 4.82 to 57.84, q -value=6.24E-05) with HCC risk (**Table 3**). The NDE were not statistically significantly associated with HCC for most of the lifestyle exposures, except for lifetime alcohol intake and hepatitis infection. With the exception of smoking and, to a lesser extent, lifetime alcohol, the PLS metabolic profiles and estimated associations were virtually unchanged in the sensitivity analysis (**Supplementary Tables 1 and 2**).

Discussion

In this nested case-control study, we evaluated the relationships of modifiable lifestyle exposures and other HCC risk factors in models that integrated information on metabolites according to the MITM principle(17). Metabolic signatures mediated varying proportions of the associations of BMI, lifetime alcohol use, diet score, smoking and diabetes status with HCC risk. Our findings provide mechanistic insights into the etiology of HCC with respect to modifiable lifestyle and other HCC risk factors.

Carcinogenesis leading to HCC involves various modifications of a number of molecular pathways and its etiology is complex. It involves different risk factors that can overlap (*e.g.* hepatitis infection, cirrhosis, diabetes, obesity) that in turn affect a broad range of mechanisms including activation of pro-inflammatory pathways and induction of the oxidative stress response(48).

PLS analysis was used to relate each of the modified HLI variables, in turn, to the set of metabolites. The exposure-specific metabolic signatures were found to mediate the relation with HCC risk for BMI, lifetime alcohol consumption, smoking, diabetes and diet, with a proportion mediated of 100, 24, 56, 100 and 38%, respectively. These findings suggested that varying proportion of the total effect on HCC is exerted via the metabolic signatures, possibly through specific underlying mechanisms by which the exposure is acting.

Specifically, a recent IARC handbook evaluation on body fatness and obesity reported a positive relationship between BMI and risk of liver cancer(49). Our study suggests that the increase in HCC risk is entirely mediated by a BMI-specific metabolic signature characterized by phosphatidylcholines (LysoPC aC18:2, LysoPC aC17:0 and PC aeC36:2) and tyrosine. PCs are required for lipoprotein assembly and secretion; in particular acyl-alkyl-PCs were correlated with high-density cholesterol(50,51). Tyrosine level's imbalance

has been previously related to insulin resistance and type 2 diabetes(52–55). Correlation studies conducted in the EPIC-Potsdam cohort exploring the association between lifestyle factors and blood metabolite levels, acquired with the same targeted technology showed similar findings, with serum acyl-alkyl-phosphatidylcholines (PC ae), LysoPC aC17:0, aC18:2 and PC aeC36:2 negatively associated with obesity and BMI whereas tyrosine was positively related to BMI(37,36,56). Tyrosine was also related to BMI in the Framingham Offspring cohort(57) and tyrosine and glutamate were positively associated with BMI in a study including participants from the PLCO, Navy and Shanghai studies(58). Glutamate, a neurotransmitter, has been previously implicated in appetite regulation(59). Hypotheses suggested high levels of amino acids may be due to excess protein breakdown in skeletal muscle due to insulin resistance concurrent with obesity, or to higher protein turnover in high-BMI individuals(60–62).

The metabolic signature fully mediated the association between diabetes, a well-established HCC risk factor(5), and HCC. The contributing metabolites were hexoses, phenylalanine and LysoPCs, consistent with previous studies based on targeted(54) and untargeted(63) approaches. These metabolites were further linked with insulin resistance and involved in glycolysis and gluconeogenesis, and their metabolic alterations were associated with an increased diabetes risk(54).

The metabolic signature of lifetime alcohol intake was negatively associated with sphingomyelins and positively associated with phosphatidylcholines. In the CARLA study, in both men and women, PC aaC32:1 was positively associated with alcohol consumption while PC aeC30:2, SM(OH) C14:1, SM(OH) C16:1 and SM(OH) C22:2 were inversely associated with high intakes of alcohol(64). Similar metabolite patterns were observed in a study that focused on alcohol-dependent patients(65). As ethanol has been hypothesised to induce lipogenesis in the liver tissues(66), alcohol can lead to hepatic injuries causing a disruption of

the metabolism of fatty acids including phospholipids(67). A study in EPIC found that alcohol intake was positively associated with plasma phospholipids(68). In particular, alcohol is associated with higher levels of short PCs and lower levels of long-chain PC(67). Alcohol is also responsible of the activation of the enzyme acid sphingomyelinase (ASM) resulting in lower SM levels that could lead to catabolism of SMs into ceramide and PCs(65,69). which, in turn, leads to hepatotoxicity and eventually HCC (70,71).

The identification of specific metabolic signatures for alcohol and smoking was particularly challenging in our study, as these two factors are strongly correlated(72–74). An overlap between the smoking and alcohol-specific metabolic signatures was observed in the preliminary analysis, where four common sphingomyelins, *i.e.* SM C16:1, SM C18:1, SM(OH) C14:1 and SM(OH) C22:2, were identified. In the sensitivity analysis, the different lifestyle exposures were mutually adjusted for prior to PLS, thus leading to a new list of metabolites associated with smoking which included serine, SM(OH) C22:2 and PC aaC36:1, consistently to what was reported in the KORA study(75). As a result, the estimated proportion of mediation increased from 57 to 100 %, resulting in a metabolic signature capturing smoking-related metabolic features that is more predictive of HCC. In the CARLA study, that has also explored lifestyle exposures' relation with serum metabolite levels, a sex-specific positive association was found between PC aaC32:1 and smoking in men and with SM C16:1 and BMI in women(64). Smoking disrupts PC levels, with low acyl-alkyl PCs and high diacyl PCs likely reflecting less lipid remodeling in membranes resulting in inflammation(76) which may in turn lead to liver injury(77). PCs play a major role in essential fatty acids and are a substrate for cell signaling(78). Cigarette smoking contains free radicals(79,80) which promote lipid peroxidation(81) that can lead to liver diseases(82).

One of the limitations of the study was the small sample size, as HCC is a rare malignancy. Consequently both cases and controls were included in the PLS analysis for

identification of the exposure-specific metabolic signatures. However, PLS was done agnostically with respect to case-control status. In addition, mediation analysis assumed a temporal sequence of lifestyle exposures, metabolites and outcome (83,84), for the NDE and NIE to have a causal interpretation. Despite the prospective design of the study, another potential limitation is reverse causality, as the liver is an organ playing a key role in many metabolic pathways(85). In our study lifestyle exposures were assessed at baseline, at the same time of the collection of biological samples that provided metabolomics data. It is worth noting that lifestyle and metabolite concentrations reflect different exposure windows, with metabolites likely reflecting exogenous and endogenous exposures in a limited timeframe (38,39,86). The diet score was derived from questionnaires that covered the dietary habits of participants over the past 12 months prior to baseline(20,21). While lifetime alcohol reflected the history of exposure across adult life, other exposures such as BMI, smoking, physical activity, hepatitis infection and diabetes status were assumed to be relatively stable over time in the middle-age study populations in EPIC. Another key aspect of mediation analysis is what is referred to as the ‘cross-world assumption’, whereby NDE and NIE cannot be identified in the presence of a mediator-outcome confounder that is affected by the exposure(87). In our study the composite liver function score, an index compiled from measures of circulating biomarkers of hepatic function indicating underlying liver impairment(22) was likely affected by lifestyle exposure, and was, in turn, likely influencing metabolite levels and HCC risk. The use of weighting-based estimation methods to look at joint mediators to compute randomized interventional effects has been proposed as a solution in the presence of mediator-outcome confounding(87). In this study the liver function was added to the list of mediators. In this way, the metabolic signatures comprised of relevant information on the liver function, and the link with relevant lifestyle factors was evaluated. Last, we had a substantial number of missing values for hepatitis infection (~25%) which was

imputed by an EM algorithm, and given the small sample size it was not possible to conduct a sensitivity analysis by hepatitis type.

Findings from this comprehensive approach suggested that certain exposure-specific metabolic profiles are intermediate biomarkers on the causal pathway towards hepatocellular carcinogenesis, but replication of these findings in an independent setting and in a larger sample is warranted. Future applications include applying the same methodology to other cancer sites and their associated lifestyle risk factors in order to identify relevant metabolic signatures, possibly relying on Mendelian Randomization analysis to assess causality. Application to untargeted data is also planned.

This study further refined an endeavor for high-throughput data to integrate metabolomics, lifestyle exposures together with disease indicators. Metabolomics lends itself as a promising tool to identify metabolites bridging the link between exposure(s) and disease, as advocated by the MITM principle(15,16). The framework we developed allows the identification of informative metabolic signatures, which are useful to elucidate the underlying biological mechanisms in the relationship between lifestyle exposure to cancer risk(88).

Acknowledgements:

The authors wish to thank Dr Joshua Sampson from the Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA, for useful discussions and insightful comments on this work. The authors would like to extend their thanks to Mr Bertrand Hémon and Ms Carine Biessy from the International Agency for Research on Cancer for their kind help with issues related to data management. The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Nordic Centre of Excellence programme on Food, Nutrition and Health. (Norway); Health Research Fund (FIS), PI13/00061 to Granada), Regional Governments of Andalucía, Asturias, Basque Country, Murcia (no. 6236) and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Scientific Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk and MR/M012190/1 to EPIC-Oxford) (United Kingdom).

References

1. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer*. 2013;132:1133–45.
2. Ferlay J, Soerjomataram I, Ervik M, Dikshit I, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, Fr. Int. Agency Res. Cancer. 2013.
3. Bertuccio P, Turati F, Carioli G, Rodriguez T, Vecchia C La, Malvezzi M, et al. Global trends and predictions in hepatocellular carcinoma mortality. *J Hepatol* [Internet]. 2017 [cited 2017 Jun 15]; Available from: http://ac.els-cdn.com/S0168827817301459/1-s2.0-S0168827817301459-main.pdf?_tid=23633bda-51c9-11e7-85b5-00000aacb35e&acdnat=1497531178_c667b3166d292f9092c951ea7ed6781f
4. Mittal S, El-Serag HB. Epidemiology of HCC: Consider the Population. *J Clin Gastroenterol* [Internet]. 2013 [cited 2017 Jun 15];47:S2–6. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3683119/pdf/nihms-444898.pdf>
5. Gomaa A-I. Hepatocellular carcinoma: Epidemiology, risk factors and pathogenesis. *World J Gastroenterol*. 2008;14:4300–8.
6. Chiang C-H, Lu C-W, Han H-C, Hung S-H, Lee Y-H, Yang K-C, et al. The Relationship of Diabetes and Smoking Status to Hepatocellular Carcinoma Mortality. *Medicine (Baltimore)* [Internet]. 2016;95:e2699. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005792-201602090-00046>
7. Turati F, Trichopoulos D, Polesel J, Bravi F, Rossi M, Talamini R, et al. Mediterranean diet and hepatocellular carcinoma. *J Hepatol* [Internet]. European Association for the Study of the Liver; 2014;60:606–11. Available from: <http://dx.doi.org/10.1016/j.jhep.2013.10.034>
8. Aleksandrova K, Bamia C, Drogan D, Lagiou P, Trichopoulou A, Jenab M, et al. The association of coffee intake with liver cancer risk is mediated by biomarkers of inflammation and hepatocellular injury: Data from the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr*. 2015;102.

9. Tzoulaki I, Ebbels TMD, Valdes A, Elliott P, Ioannidis JPA. Design and Analysis of Metabolomics Studies in Epidemiologic Research : A Primer on -Omic Technologies. *Am J Epidemiol*. 2014;180:129–39.
10. Abu Bakar MH, Sarmidi MR, Cheng K-K, Ali Khan A, Suan CL, Zaman Huri H, et al. Metabolomics - the complementary field in systems biology: a review on obesity and type 2 diabetes. *Mol Biosyst* [Internet]. 2015 [cited 2016 Jul 28];11:1742–74. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25919044>
11. Baker M. The 'Omics Puzzle. *Nature*. 2013;494:416–9.
12. Nicholson JK, Holmes E, Elliott P. The metabolome-wide association study: a new look at human disease risk factors. *J Proteome Res*. 2008;7:3637–8.
13. Wild CP, Scalbert A, Herceg Z. Measuring the Exposome: A Powerful Basis for Evaluating Environmental Exposures and Cancer Risk. *Environ Mol Mutagen*. 2013;54:480:499.
14. Wild CP. Complementing the Genome with an “‘ Exposome ’”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol biomarkers Prev*. 2005;14:1847–51.
15. Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol Biomarkers Prev*. 2007;16:1954–65.
16. Vineis P, van Veldhoven K, Chadeau-Hyam M, Athersuch TJ. Advancing the application of omics-based biomarkers in environmental epidemiology. *Environ Mol Mutagen* [Internet]. 2013 [cited 2016 Jul 28];54:461–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23519765>
17. Assi N, Fages A, Vineis P, Chadeau-hyam M, Stepien M, Duarte-salles T, et al. A statistical framework to model the meeting-in-the-middle principle using metabolomic data : application to hepatocellular carcinoma in the EPIC study. *Mutagenesis*. 2015;30:743–53.
18. McKenzie F, Biessy C, Ferrari P, Freisling H, Rinaldi S, Chajès V, et al. Healthy Lifestyle and Risk of Cancer in the European Prospective Investigation Into Cancer

- and Nutrition Cohort Study. *Medicine (Baltimore)* [Internet]. 2016 [cited 2016 Jul 28];95:e2850. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27100409>
19. McKenzie F, Ferrari P, Freisling H, Chajès V, Rinaldi S, de Batlle J, et al. Healthy lifestyle and risk of breast cancer among postmenopausal women in the European Prospective Investigation into Cancer and Nutrition cohort study. *Int J cancer* [Internet]. 2015 [cited 2016 Jul 28];136:2640–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25379993>
 20. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* 2002;5:1113–24.
 21. Kaaks R, Slimani N, Riboli E. Pilot Phase Studies on the Accuracy of Dietary Intake Measurements in the EPIC Project : Overall Evaluation of Results. 1997;26:26–36.
 22. Fedirko V, Trichopolou A, Bamia C, Duarte-Salles T, Trepo E, Aleksandrova K, et al. Consumption of fish and meats and risk of hepatocellular carcinoma: the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol.* 2013;24:2166–73.
 23. Stepien M, Duarte-Salles T, Fedirko V, Floegel A, Barupal DK, Rinaldi S, et al. Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study. *Int J Cancer.* 2016;138:348–60.
 24. Trichopoulos D, Bamia C, Lagiou P, Fedirko V, Trepo E, Jenab M, et al. Hepatocellular carcinoma risk factors and disease burden in a European cohort: a nested case-control study. *J Natl Cancer Inst.* 2011;103:1686–95.
 25. Pearce N. Incidence density matching with a simple SAS computer program. *Int J Epidemiol* [Internet]. 1989 [cited 2016 Jul 28];18:981–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2621036>
 26. Akuta N, Suzuki F, Kobayashi M, Hara T, Sezaki H, Suzuki Y, et al. Correlation Between Hepatitis B Virus Surface antigen Level and Alpha-Fetoprotein in Patients Free of Hepatocellular Carcinoma or Severe Hepatitis. *J Med Virol.* 2014;86:131–8.
 27. Yang W-S, Va P, Bray F, Gao S, Gao J, Li H-L, et al. The role of pre-existing diabetes

- mellitus on hepatocellular carcinoma occurrence and prognosis: a meta-analysis of prospective cohort studies. *PLoS One*. 2011;6:e27326.
28. Berzigotti A, Saran U, Dufour J-F. Physical Activity and Liver Diseases. *Hepatology*. 2016;63:1026–40.
 29. Liu X, Xu J. Body Mass Index and Waistline are Predictors of Survival for Hepatocellular Carcinoma After Hepatectomy. *Med Sci Monit* [Internet]. 2015;21:2203–9. Available from: <http://www.medscimonit.com/abstract/index/idArt/894202>
 30. Niu J, Lin Y, Guo Z, Niu M, Su C. The Epidemiological Investigation on the Risk Factors of Hepatocellular Carcinoma: A Case-Control Study in Southeast China. *Medicine (Baltimore)* [Internet]. 2016;95:e2758. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26871825>
 31. Raffetti E, Portolani N, Molfino S, Baiocchi GL, Limina RM, Caccamo G, et al. Role of aetiology, diabetes, tobacco smoking and hypertension in hepatocellular carcinoma survival. *Dig Liver Dis* [Internet]. Editrice Gastroenterologica Italiana; 2015;47:950–6. Available from: <http://dx.doi.org/10.1016/j.dld.2015.07.010>
 32. Rong X, Wei F, Geng Q, Ruan J, Shen H, Li A, et al. The Association Between Body Mass Index and the Prognosis and Postoperative Complications of Hepatocellular Carcinoma: A Meta-Analysis. *Medicine (Baltimore)* [Internet]. 2015;94:e1269. Available from: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=ovftq&AN=00005792-201508010-00026>
 33. Testino G, Leone S, Borro P. Alcohol and hepatocellular carcinoma: A review and a point of view. *World J Gastroenterol*. 2014;20:15943–54.
 34. Trichopoulos D, Bamia C, Lagiou P, Fedirko V, Trepo E, Jenab M, et al. Hepatocellular carcinoma risk factors and disease burden in a European cohort: a nested case-control study. *J Natl Cancer Inst*. 2011;103:1686–95.
 35. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J R Stat Soc Ser B*. 1977;39:1–38.

36. Floegel A, Wientzek A, Bachlechner U, Jacobs S, Drogan D, Prehn C, et al. Linking diet, physical activity, cardiorespiratory fitness and obesity to serum metabolite networks: findings from a population-based study. *Int J Obes (Lond)*. Nature Publishing Group; 2014;1–9.
37. Bachlechner U, Floegel A, Steffen A, Prehn C, Adamski J, Pischon T, et al. Associations of anthropometric markers with serum metabolites using a targeted metabolomics approach: results of the EPIC-potsdam study. *Nutr Diabetes* [Internet]. Nature Publishing Group; 2016 [cited 2016 Sep 23];6:e215. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27348203>
38. Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, et al. Reliability of serum metabolites over a two-year period: A targeted metabolomic approach in fasting and non-fasting samples from EPIC. *PLoS One* [Internet]. 2015;10:1–10. Available from: <http://dx.doi.org/10.1371/journal.pone.0135437>
39. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, et al. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS One*. 2011;6.
40. Fages A, Ferrari P, Monni S, Dossus L, Floegel A, Mode N, et al. Investigating sources of variability in metabolomic data in the EPIC study: the Principal Component Partial R-square (PC-PR2) method. *Metabolomics*. 2014;10:1074–83.
41. Abdi H. Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdiscip Rev Comput Stat*. 2010;2:97–106.
42. Tenenhaus M. *La régression PLS*. Technip. Paris; 1998.
43. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172:1339–48.
44. VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation Analysis With Matched Case-Control Study Designs. *Am J Epidemiol* [Internet]. 2016;183:1–2. Available from: <http://aje.oxfordjournals.org/content/early/2016/04/12/aje.kww038.short>
45. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57:289–300.

46. SAS Institute Inc., Cary N. Base SAS® 9.4 Procedures Guide. 2012.
47. R Foundation for Statistical Computing, R Core Team. R: A language and environment for statistical computing. Vienna, Austria.; 2013.
48. Sanyal AJ, Yoon SK, Lencioni R. The etiology of hepatocellular carcinoma and consequences for treatment. *Oncologist* [Internet]. AlphaMed Press; 2010 [cited 2017 Sep 18];15 Suppl 4:14–22. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21115577>
49. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer — Viewpoint of the IARC Working Group. *N Engl J Med* [Internet]. 2016 [cited 2016 Sep 28];375:794–8. Available from: <http://www.nejm.org/doi/10.1056/NEJMSr1606602>
50. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost H-G, et al. Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. *Diabetes* [Internet]. 2013 [cited 2016 Sep 23];62:639–48. Available from: <http://diabetes.diabetesjournals.org/cgi/doi/10.2337/db12-0495>
51. Cole LK, Vance JE, Vance DE. Phosphatidylcholine biosynthesis and lipoprotein metabolism. *Biochim Biophys Acta - Mol Cell Biol Lipids* [Internet]. 2012 [cited 2016 Nov 16];1821:754–61. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S138819811100179X>
52. Kawanaka M, Nishino K, Oka T, Urata N, Nakamura J, Suehiro M, et al. Tyrosine levels are associated with insulin resistance in patients with nonalcoholic fatty liver disease. *Hepatic Med Evid Res*. 2015;7:29–35.
53. Hellmuth C, Kirchberg FF, Lass N, Harder U, Peissner W, Koletzko B, et al. Tyrosine Is Associated with Insulin Resistance in Longitudinal Metabolomic Profiling of Obese Children. *J Diabetes Res* [Internet]. Hindawi Publishing Corporation; 2016 [cited 2016 Nov 16];2016:1–10. Available from: <http://www.hindawi.com/journals/jdr/2016/2108909/>
54. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes*. 2013;62:639–48.

55. Rauschert S, Uhl O, Koletzko B, Hellmuth C. Metabolomic Biomarkers for Obesity in Humans : A Short Review. *Ann Nutr Metab*. 2014;64:314–24.
56. Kim JY, Park JY, Kim OY, Ham BM, Kim H-J, Kwon DY, et al. Metabolic Profiling of Plasma in Overweight/Obese and Lean Men using Ultra Performance Liquid Chromatography and Q-TOF Mass Spectrometry (UPLC–Q-TOF MS). *J Proteome Res* [Internet]. American Chemical Society; 2010 [cited 2016 Nov 16];9:4368–75. Available from: <http://pubs.acs.org/doi/abs/10.1021/pr100101p>
57. Ho JE, Larson MG, Ghorbani A, Cheng S, Chen H, Keyes M, et al. Metabolomic Profiles of Body Mass Index in the Framingham Heart Study Reveal Distinct Cardiometabolic Phenotypes. *PLoS One* [Internet]. 2016;11:e0148361. Available from: <http://dx.doi.org/10.1371/journal.pone.0148361>
58. Moore SC, Matthews CE, Sampson JN, Stolzenberg-Solomon RZ, Zheng W, Cai A, et al. Human metabolic correlates of body mass index. *Metabolomics*. 2014;10:259–69.
59. Wu Q, Clark MS, Palmiter RD. Deciphering a neuronal circuit that mediates appetite. *Nature* [Internet]. 2012 [cited 2017 Sep 18];483:594–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22419158>
60. Felig P, Marliss E, Cahill GF. Plasma Amino Acid Levels and Insulin Secretion in Obesity. *N Engl J Med* [Internet]. 1969 [cited 2017 Sep 18];281:811–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/5809519>
61. Forlani G, Vannini P, Marchesini G, Zoli M, Ciavarella A, Pisi E. Insulin-dependent metabolism of branched-chain amino acids in obesity. *Metabolism* [Internet]. 1984 [cited 2017 Sep 18];33:147–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6363875>
62. Tremblay F, Brule S, Hee Um S, Li Y, Masuda K, Roden M, et al. Identification of IRS-1 Ser-1101 as a target of S6K1 in nutrient- and obesity-induced insulin resistance. *Proc Natl Acad Sci* [Internet]. 2007 [cited 2017 Sep 18];104:14056–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17709744>
63. Drogan D, Dunn WB, Lin W, Buijsse B, Schulze MB, Langenberg C, et al. Untargeted Metabolic Profiling Identifies Altered Serum Metabolites of Type 2 Diabetes Mellitus in a Prospective, Nested Case Control Study. *Clin Chem* [Internet]. 2015 [cited 2016

- Sep 23];61:487–97. Available from:
<http://www.clinchem.org/cgi/doi/10.1373/clinchem.2014.228965>
64. Lacruz ME, Kluttig A, Tiller D, Medenwald D, Giegling I, Rujescu 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* [Internet]. 2016 [cited 2017 Apr 25];9:487–94. Available from: <http://circgenetics.ahajournals.org/content/9/6/487.long#sec-4>
 65. Reichel M, Hönig S, Liebisch G, Lüth A, Kleuser B, Gulbins E, et al. Alterations of plasma glycerophospholipid and sphingolipid species in male alcohol-dependent patients. *Biochim Biophys Acta - Mol Cell Biol Lipids*. 2015;1851:1501–10.
 66. You M, Fischer M, Deeg MA, Crabb DW. Ethanol Induces Fatty Acid Synthesis Pathways by Activation of Sterol Regulatory Element-binding Protein (SREBP). *J Biol Chem*. 2002;277:29342–7.
 67. Glen I, Skinner F, Glen E, Mbch B, Macdonell L, Al GET. The Role of Essential Fatty Acids in Alcohol Dependence and Tissue Damage. *Alcohol Clin Exp Res*. 1987;11:37–41.
 68. Saadatian-Elahi M, Slimani N, Chajès V, Jenab M, Goudable J, Biessy C, et al. Plasma phospholipid fatty acid profiles and their association with food intakes : results from a cross-sectional study within the European Prospective Investigation into Cancer and Nutrition 1 – 3. *Am J Clin Nutr*. 2009;89:331–46.
 69. Liu JJ, Wang JY, Hertervig E, Cheng Y, Nilsson A, Duan RD. Activation of neutral sphingomyelinase participates in ethanol-induced apoptosis in Hep G2 cells. *Alcohol Alcohol* [Internet]. 2000 [cited 2017 Sep 15];35:569–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11093963>
 70. Saito M, Saito M. Involvement of sphingolipids in ethanol neurotoxicity in the developing brain. *Brain Sci* [Internet]. Multidisciplinary Digital Publishing Institute (MDPI); 2013 [cited 2017 Sep 15];3:670–703. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24961420>
 71. Deaciuc I V, Nikolova-Karakashian M, Fortunato F, Lee EY, Hill DB, McClain CJ. Apoptosis and dysregulated ceramide metabolism in a murine model of alcohol-

- enhanced lipopolysaccharide hepatotoxicity. *Alcohol Clin Exp Res* [Internet]. 2000 [cited 2017 Sep 15];24:1557–65. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11045865>
72. Gubner NR, Delucchi KL, Ramo DE. Associations between binge drinking frequency and tobacco use among young adults. *Addict Behav* [Internet]. Elsevier Ltd; 2016;60:191–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27156220>
 73. Barrett SP, Tichauer M, Leyton M, Pihl RO. Nicotine increases alcohol self-administration in non-dependent male smokers. *Drug Alcohol Depend.* 2006;81:197–204.
 74. Kuper H, Tzonou A, Kaklamani E, Hsieh C-C, Lagiou P, Adami H-O, et al. Tobacco Smoking , Alcohol Consumption and Their Interaction in the Causation of Hepatocellular Carcinoma. *Int J Cancer.* 2000;502:498–502.
 75. Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, et al. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med* [Internet]. 2013 [cited 2016 Sep 28];11:60. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23497222>
 76. González-Reimers E, Santolaria-Fernández F, Martín-González MC, Fernández-Rodríguez CM, Quintero-Platt G. Alcoholism: a systemic proinflammatory condition. *World J Gastroenterol* [Internet]. Baishideng Publishing Group Inc; 2014 [cited 2017 Sep 17];20:14660–71. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25356029>
 77. Nagata K, Suzuki H, Sakaguchi S. Common pathogenic mechanism in development progression of liver injury caused by non-alcoholic or alcoholic steatohepatitis. *J Toxicol Sci.* 2007;32:453–68.
 78. Yamaji-Hasegawa A, Tsujimoto M. Asymmetric Distribution of Phospholipids in Biomembranes. *Biol Pharm Bull* [Internet]. The Pharmaceutical Society of Japan; 2006 [cited 2017 Sep 18];29:1547–53. Available from: https://www.jstage.jst.go.jp/article/bpb/29/8/29_8_1547/_article
 79. Church DF, Pryor WA. Free-radical chemistry of cigarette smoke and its toxicological implications. *Environ Health Perspect* [Internet]. National Institute of Environmental

- Health Science; 1985 [cited 2017 Sep 18];64:111–26. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/3007083>
80. Girotti AW. Mechanisms of lipid peroxidation. *J Free Radic Biol Med* [Internet]. 1985 [cited 2017 Sep 18];1:87–95. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/3915303>
 81. Ionescu N, de Freitas C, Bueno AA. Perturbations in blood phosphatidylcholine and sphingomyelin Fatty Acid composition in a sample population of cigarette smokers. *Indian J Clin Biochem* [Internet]. Springer; 2013 [cited 2017 Sep 15];28:361–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24426238>
 82. Li S, Tan H-Y, Wang N, Zhang Z-J, Lao L, Wong C-W, et al. The Role of Oxidative Stress and Antioxidants in Liver Diseases. *Int J Mol Sci* [Internet]. Multidisciplinary Digital Publishing Institute (MDPI); 2015 [cited 2017 Sep 18];16:26087–124. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26540040>
 83. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18:137–50.
 84. Gelfand LA, Mensinger JL, Tenhave T. Mediation Analysis: A Retrospective Snapshot of Practice and More Recent Directions. *J Gen Psychol*. 2009;136:153–76.
 85. Mitra V, Metcalf J. Metabolic functions of the liver. *Anaesth Intensive Care Med*. Elsevier Ltd.; 2009;10:334–5.
 86. Sampson JN, Boca SM, Shu XO, Stolzenberg-Solomon RZ, Matthews CE, Hsing AW, et al. Metabolomics in Epidemiology: Sources of Variability in Metabolite Measurements and Implications. *Cancer Epidemiol Biomarkers Prev* [Internet]. American Association for Cancer Research; 2013 [cited 2016 Sep 12];22:631–40. Available from: <http://cebp.aacrjournals.org/cgi/doi/10.1158/1055-9965.EPI-12-1109>
 87. Vanderweele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*. 2014;25:300–6.
 88. 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 [Internet]. Springer US; 2015;11:1376–80. Available from:
<http://dx.doi.org/10.1007/s11306-015-0793-8>

Tables and Figures

Table 1: Baseline characteristics of the study population of the EPIC nested case-control study on hepatocellular carcinoma.

Characteristics ¹	Cases	Controls
N	147	147
Sex (N)		
Male	102	102
Female	45	45
Age at blood collection (years)	60.1 (50.7-68.8)	60.1 (50.9-68.9)
Height (cm)	167.7 (152.3-180.7)	169.3 (156.1-181.0)
Weight (kg)	79.8 (59.0-102.2)	78.3 (60.6-93.0)
BMI (kg/m²)	28.2 (23.0-34.9)	27.3 (22.0-32.5)
Total energy intake (kcal/day)	2260.8 (1381.4-3169.3)	2276.6 (1495.1-3140.6)
Alcohol intake at recruitment (g/day)		
among consumers	29.6 (1.08-80.76)	16.8 (1.27-41.27)
% of never drinkers (<0.1g/day)	13.6	6.1
Physical activity (Mets-hour/week)	77.9 (18.9-150.5)	83.3 (21.6-157.2)
Lifetime alcohol intake (g/day)²		
among consumers	33.6 (2.13-74.42)	19.7 (2.17-41.32)
% of never drinkers (<0.1g/day)	3.1	4.1
Dietscore²	25.7 (16.0-33.0)	27.4 (20.6-30.0)
Hepatitis infection^{2,3}		
Yes	41	5
No	106	142
Diabetes at baseline²		
Yes	19	10
No	128	137
Smoking status² (N)		
Current smokers, > 15 cigarettes/day	25	23
Current smokers, ≤ 15 cigarettes/day	34	10
Former smokers, quit ≤10years	17	25
Former smokers, quit >10years	29	29
Never	42	60
Fasting status (N)		
Unknown	17	16
In between (3-6h)	31	29
No (<3h)	59	61
Yes (>6h)	40	41

¹ Values are presented as means and 10th and 90th percentiles in parentheses for continuous variables and as frequencies for categorical variables. ²There were respectively 42, 12, 76, 29 and 7 missing values for lifetime alcohol consumption, dietscore, hepatitis, diabetes and smoking; they were imputed with an EM algorithm using the covariance matrix of the data. ³Prior to EM, there were 41 hepatitis infections (3 in controls and 38 in cases). After imputation there were 46 hepatitis infections (5 in controls and 41 in cases). Among the initial distribution of the 41 hepatitis patients, there were 26 with hepatitis C, 15 with hepatitis B with 3 subjects having both HBV and HCV.

Table 2: Metabolites contributing to the PLS factor of each HLI component (N=294, X-set=1, M-set=133)*. Results from multiple PLS models performed using residuals based on country (X- and M-sets) and batch residuals (M-set only).

Metabolites	Loadings	Metabolites	Loadings	Metabolites	Loadings
BMI		Lifetime alcohol		Diet score	
Glutamine	-0.186	Glutamic Acid	0.17	PC aaC36:1	-0.178
Glutamic Acid	0.230	SM C16:1	-0.171	PC aaC38:0	0.195
Tyrosine	0.243	SM C18:1	-0.167	PC aaC38:6	0.23
LysoPC aC17:0	-0.218	SM(OH) C14:1	-0.18	PC aaC40:6	0.204
LysoPC aC18:2	-0.236	SM(OH) C16:1	-0.184	PC aaC42:2	0.263
PC aeC36:2	-0.203	SM(OH) C22:2	-0.211	PC aeC34:1	-0.195
Liver function score	0.191	PC aaC32:1	0.211	PC aeC40:6	0.167
Physical activity		Smoking		Hepatitis infection	
Biogenic Creatinine	-0.199	Hexoses	0.136	SM C20:2	-0.179
Biogenic Taurine	-0.181	SM C16:1	-0.238	SM(OH) C16:1	-0.178
Glutamic Acid	-0.212	SM C18:1	-0.194	PC aaC32:2	0.188
PC aaC34:2	-0.188	SM(OH) C14:1	-0.214	PC aaC34:1	0.184
PC aeC34:2	0.209	SM(OH) C22:2	-0.182	PC aaC34:3	0.18
PC aeC34:3	0.176	LysoPC aC28:1	-0.204	PC aaC34:4	0.197
PC aeC36:3	0.193	PC aeC30:2	-0.264	PC aaC36:5	0.189
Diabetes status					
Biogenic Alpha AAA	0.236				
Isoleucine	0.168				
Phenylalanine	0.158				
Valine	0.211				
Hexoses	0.551				
Lyso PC aC16:1	-0.145				
Liver function score	0.226				

* Metabolite variables contributing to each PLS factor were selected based on extreme loading values i.e. below or above the 2.5th and 97.5th percentiles.

Table 3: Results from the mediation analyses, with ORs and their associated 95% confidence intervals for the natural direct effect (NDE), natural indirect effect (NIE) and total effects (TE) using residuals based on country (X- and M-sets) and batch (M-set only).

Models*	NDE	<i>q</i> -value	NIE	<i>q</i> -value	Total Effect	<i>q</i> -value	% mediated
BMI	0.85 (0.60,1.20)	4.81E-01	1.56 (1.24,1.96)	1.20E-03	1.23 (0.93,1.62)	1.74E-01	100
Lifetime Alcohol	1.31 (1.06,1.61)	4.20E-02	1.09 (1.03,1.15)	4.67E-03	1.40 (1.14,1.72)	3.50E-03	24
Diet score	0.77 (0.54,1.11)	3.92E-01	0.85 (0.74,0.97)	2.52E-02	0.66 (0.47,0.92)	3.27E-02	38
Physical activity	0.98 (0.72,1.35)	9.18E-01	0.97 (0.87,1.09)	6.17E-01	0.98 (0.71,1.34)	8.84E-01	**
Smoking	1.17 (0.77,1.77)	5.36E-01	1.22 (1.04,1.44)	2.52E-02	1.42 (0.99,2.03)	1.02E-01	57
Hepatitis Infection	17.99 (5.15,62.80)	4.11E-05	0.94 (0.83, 1.06)	3.48E-01	16.70 (4.82,57.84)	6.24E-05	0
Diabetes	0.46 (0.11,1.93)	4.82E-01	5.11 (1.99,13.14)	2.45E-03	2.45 (0.84,7.18)	1.41E-01	100

* Models were mutually adjusted for all HLI variables. Cases and controls were matched on age at blood collection (± 1 year), sex, study centre, date (± 2 months) and time of day at blood collection (± 3 h), fasting status at blood collection ($<3/3-6/>6$ h); women were additionally matched on menopausal status (pre/peri/postmenopausal) and hormone replacement therapy. The mediator models were linear. The outcome models were computed through conditional logistic regressions. In the mediation analysis, the exposure was the original modified HLI lifestyle factor, the mediator was the associated M-score (metabolic profile) and the outcome was HCC. ** As the associations were null for direct and indirect effects, the proportion mediated was not computed. NDE and NIE and their 95%CI computed from formulae detailed in Material and Methods.

Table 4: Results from the mediation analyses, with ORs and their associated 95% confidence intervals for the mediator effects (ME) using residuals based on country (X- and M-sets) and batch (M-set only).

Models*	ME	<i>q</i> -value
BMI	4.04 (2.22,7.36)	3.17E-05
Lifetime Alcohol	2.50 (1.57,3.97)	2.48E-04
Diet score	0.61 (0.41,0.89)	1.54E-02
Physical activity	0.90 (0.60,1.35)	6.15E-01
Smoking	3.33 (1.96,5.66)	3.17E-05
Hepatitis Infection	1.22 (0.88,1.69)	2.60E-01
Diabetes	2.75 (1.59,4.78)	5.55E-04

* Models were mutually adjusted for all HLI variables. Cases and controls were matched on age at blood collection (± 1 year), sex, study centre, date (± 2 months) and time of day at blood collection (± 3 h), fasting status at blood collection ($<3/3-6/>6$ h); women were additionally matched on menopausal status (pre/peri/postmenopausal) and hormone replacement therapy. The mediator models were linear. The outcome models were computed through conditional logistic regressions. In the mediation analysis, the exposure was the original modified HLI lifestyle factor, the mediator was the associated M-score (metabolic profile) and the outcome was HCC.

Supplementary Tables and Figures

Supplementary Table 1: Metabolites contributing* to two selected modified HLI variable-specific PLS factors: smoking and lifetime alcohol (N=294, X-set=1, M-set=133) – Results reported and from the sensitivity analysis, using mutually adjusted lifestyle residuals as well as residuals for country and batch (the latter only in the M-set).

Sensitivity Analysis			
Lifetime Alcohol		Smoking	
Metabolites	Loadings	Metabolites	Loadings
SM C18:1	-0.161	Biogenic Taurine	-0.201
SM(OH) C16:1	-0.168	Lysine	-0.211
SM(OH) C22:1	-0.168	Serine	-0.189
SM(OH) C22:2	-0.203	SM(OH) C14:1	-0.195
LysoPC aC16:1	0.162	PC aaC36:1	0.23
PC aaC32:1	0.234	PC aaC40:3	0.202
Acylcarnitine C2	0.152	PC aeC30:2	-0.206

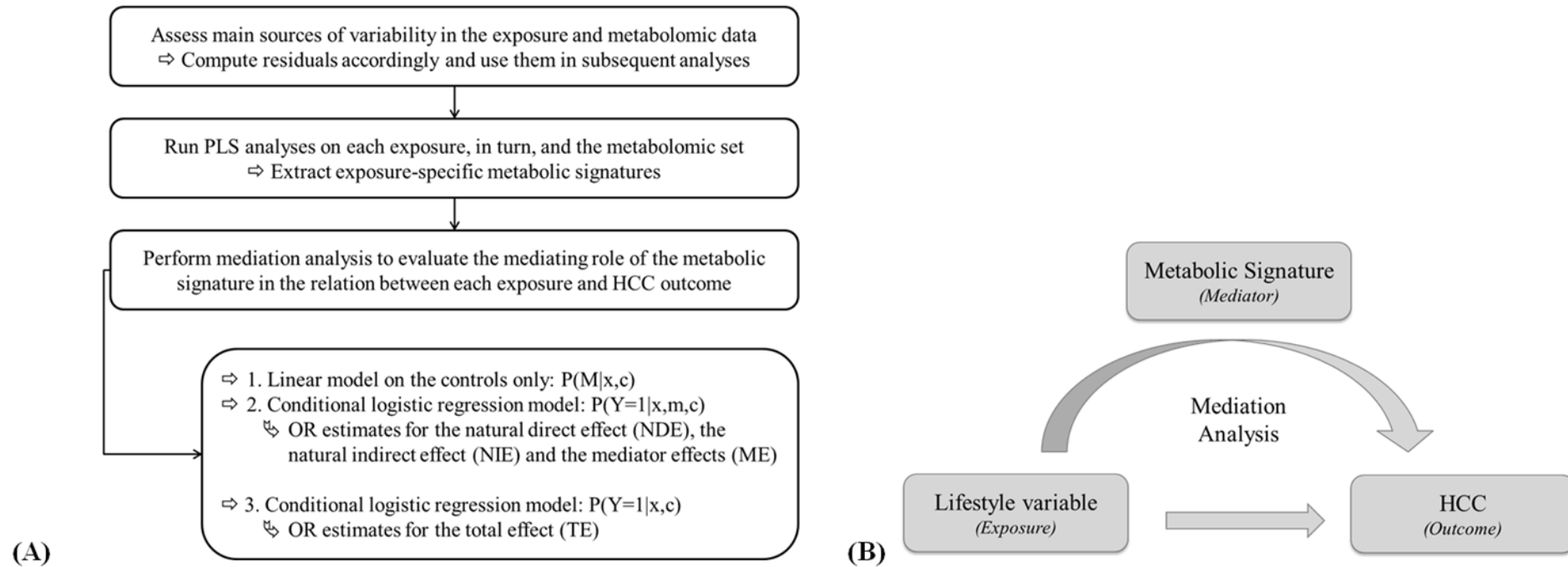
* Metabolite variables contributing to each PLS factor were selected based on extreme loading values, i.e. below or above the 2.5th and 97.5th percentiles.

Supplementary Table 2: Results from the mediation analyses, with ORs and their associated 95% confidence intervals for the natural direct effect (NDE), natural indirect effect (NIE), total effects (TE) and mediator effects (ME). Results reported for the sensitivity analysis, using mutually adjusted lifestyle residuals as well as residuals for country and batch (the latter only in the M-set).

Models*	NDE	q-value	NIE	q-value	TE	q-value	ME	q-value	% mediated
BMI	0.84 (0.58,1.22)	4.12E-01	1.52 (1.22,1.89)	1.20E-03	1.23 (0.93,1.62)	1.74E-01	3.22 (1.90,5.46)	4.73E-05	100
Lifetime Alcohol	1.30 (1.05,1.60)	5.25E-02	1.10 (1.04,1.17)	2.94E-03	1.40 (1.14,1.72)	5.00E-03	2.60 (1.62,4.18)	1.93E-04	27
Diet score	0.79 (0.54,1.15)	4.12E-01	0.85 (0.75,0.97)	1.96E-02	0.66 (0.47,0.92)	3.24E-02	0.57 (0.38,0.85)	8.04E-03	41
Physical activity	0.94 (0.68,1.32)	7.30E-01	0.85 (0.74, 0.98)	2.89E-02	0.98 (0.71,1.34)	8.84E-01	0.57 (0.36,0.88)	1.41E-02	100
Smoking	0.71 (0.36,1.39)	4.12E-01	1.98 (1.34,2.92)	1.32E-04	1.42 (0.99,2.03)	1.02E-01	11.63 (4.33,31.28)	8.12E-06	100
Hepatitis Infection	15.56 (4.44,54.54)	1.25E-04	1.07 (0.91,1.26)	4.33E-01	16.70 (4.82, 57.84)	6.24E-05	0.88 (0.64,1.21)	4.17E-01	2
Diabetes	0.43 (0.10,1.87)	4.12E-01	5.39 (2.07,14.0)	1.32E-04	2.45 (1.50,3.88)	1.41E-01	2.83 (1.63,4.94)	4.18E-04	100

* Models were mutually adjusted for all HLI variables. Cases and controls were matched on age at blood collection (± 1 year), sex, study centre, date (± 2 months) and time of day at blood collection (± 3 h), fasting status at blood collection ($<3/3-6/>6$ h); women were additionally matched on menopausal status (pre/peri/postmenopausal) and hormone replacement therapy. The mediator models were linear. The outcome models were computed through conditional logistic regressions. In the mediation analysis, the exposure was in turn the original HLI lifestyle factor, the mediator was the associated M-score (metabolic profile) and the outcome was HCC. NDE and NIE and their 95%CI computed from formulae detailed in Material and Methods.

Supplementary Figure 1: (A) Flowchart recapping the different steps of the analysis and **(B)** Directed Acyclic Graph (DAG) for the mediation analysis*.



* The TE is estimated from model 3. (panel A), arrow from Exposure to Outcome (panel B), unadjusted for the Mediator. The NDE is estimated from model 2. (panel A), arrow from Exposure to Outcome (panel B), adjusted for the Mediator. The ME is estimated from model 2. (panel A), arrow from Mediator to Outcome (panel B) adjusted for the Exposure. The NIE is estimated from models 1. and 2. (panel A) by coefficient multiplication, arrow from Exposure to Mediator to Outcome (panel B).